

Méthodes de recherche en psychologie 1re édition

Méthodes de recherche en psychologie

1re édition

1re édition

*RAJIV S. JHANGIANI; I-CHANT A. CHIANG; CARRIE
CUTTLER; DANA C. LEIGHTON; MOLLY A. METZ; JOSÉE
TURCOTTE; AND LAURA AUCHINLECK*

LAURENTIAN UNIVERSITY
SUDBURY, ON



Méthodes de recherche en psychologie 1re édition Copyright © 2026 by Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler, Dana C. Leighton, Josée Turcotte & Laura Auchinleck is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/), except where otherwise noted.

Il s'agit d'une traduction et adaptation de la

[quatrième édition du manuel de Rajiv S. Jhangiani \(Kwantlen Polytechnic University\), I-Chant A. Chiang \(Quest University Canada\), Carrie Cuttler \(Washington State University\) et Dana C. Leighton \(Texas A&M University–Texarkana\)](#)

. Cette édition s'appuie sur [la deuxième édition canadienne de Rajiv S. Jhangiani et I-Chant A. Chiang](#), la [deuxième édition américaine de Dana C. Leighton](#), ainsi que sur la [troisième édition américaine de Carrie Cuttler](#), et tient compte des commentaires de plusieurs évaluateurs et évaluatrices externes coordonnés par la [communauté Rebus](#).

Cette édition est publiée sous la

[licence internationale Creative Commons Attribution – Pas d'utilisation commerciale – Partage dans les mêmes conditions 4.0 \(CC BY-NC-SA 4.0\)](#)

Contents

Remerciements	ix
À propos de ce livre	xii
À propos de l'équipe	xxv
Préface	xxvi

Chapitre 1-La science de la psychologie

1. Méthodes de connaissance	3
2. Comprendre la science	6
3. Objectifs de la science	11
4. Science et bon sens	13
5. Psychologues expérimentaux et cliniques	16
6. Points clés et exercices	20
7. Références	23

Chapitre 2-Une vue d'ensemble de la méthode scientifique

11. Élaborer une hypothèse	27
14. Tirer des conclusions et présenter les résultats	33
16. Références	36
13. Analyser les données	37
15. Points clés et exercices	40
10. Générer de bonnes questions de recherche	44
12. Concevoir une étude de recherche	49
9. Trouver un sujet de recherche	54
8. Un modèle de recherche scientifique en psychologie	62

Chapitre 3-L'éthique de la recherche

17. Fondements moraux de la recherche éthique	69
18. Des principes moraux aux codes éthiques	76
19. Mettre l'éthique en pratique	86

20. Points clés et exercices	92
21. Références	94

Chapitre 4-La mesure psychologique

22. Comprendre les mesures psychologiques	101
23. Fidélité et validité des mesures	112
24. Stratégies pratiques pour la mesure psychologique	119
25. Points clés et exercices	125
26. Bibliographie	127

Chapitre 5-La recherche expérimentale

27. Expérience : notions de base	133
28. Plan expérimental	141
29. Expérimentation et validité	149
30. Considérations pratiques	154
31. Points clés et exercices	163
32. Références	166

Chapitre 6-La recherche non expérimentale

33. Aperçu de la recherche non expérimentale	171
34. Recherche corrélacionnelle	176
35. Corrélation complexe	185
36. Recherche qualitative	191
37. Recherche par observation	197
38. Points clés et exercices	208
39. Références	211

Chapitre 7-La recherche par enquête

40. Aperçu de la recherche par enquête	217
41. Construire des enquêtes	220
42. Réaliser des enquêtes	230
43. Points clés et exercices	237
44. Références	240

Chapitre 8-La recherche quasi-expérimentale

45. Devis à un groupe	245
46. Devis à groupes non équivalents	251
47. Points clés et exercices	256
48. Références	258

Chapitre 9-Les plans factoriels

49. Mettre en place une expérience factorielle	263
50. Interpréter les résultats d'une expérience factorielle	269
51. Points clés et exercices	279
52. Références	281

Chapitre 10-La recherche à cas unique

53. Aperçu de la recherche à cas unique	285
54. Protocole de recherche à cas unique	288
55. Le « débat » entre cas unique et groupe	299
56. Points clés et exercices	304
57. Références	307

Chapitre 11-Présenter votre recherche

58. Le style APA (Association américaine de psychologie)	311
59. Rédiger un article de recherche dans le style de l'Association américaine de psychologie (APA)	325
60. Autres formats de présentation	340
61. Points clés et exercices	346
62. Références	348

Chapitre 12-Les statistiques descriptives

63. Décrire les variables uniques	353
64. Décrire les relations statistiques	366
65. Décrire vos résultats	378
66. Faire vos analyses	390
67. Points clés et exercices	395
68. Références	399

Chapitre 13-Les statistiques inférentielles

69. Comprendre le test de l'hypothèse nulle	403
70. Quelques tests de base de l'hypothèse nulle	410
71. Considérations supplémentaires	425
72. De la « crise de répliquabilité » aux pratiques de science ouverte	433
73. Points clés et exercices	439
74. Références	442
Glossaire	445
Références	480
Ressources pour le corps enseignant	497

Remerciements

Édition originale

Ce manuel est le fruit d'un travail d'amour et d'un engagement profond envers la population étudiante. Chacun et chacune d'entre nous avaient déjà travaillé à l'adaptation, à la mise à jour et à l'amélioration des éditions successives de ce manuel depuis sa publication initiale.

En nous réunissant pour produire cette quatrième édition en collaboration, nous avons pu nous appuyer sur notre propre expertise et notre expérience en classe ainsi que sur les commentaires judicieux de plusieurs pairs impliqués dans l'évaluation.

Nous tenons à remercier la communauté Rebus, en particulier Zoe Wake Hyde et Apurva Ashok, de nous avoir guidés et soutenus tout au long du processus d'évaluation par les pairs et pour avoir créé une communauté d'auteurs et d'éducateurs ouverts qui nous apporte un soutien intellectuel et nous encourage.

Nous sommes infiniment reconnaissants pour l'évaluation faite par Judy Grissett (Georgia Southwestern State University), Amy Nusbaum (Washington State University) et une autre personne anonyme, qui ont donné de leur temps et de leur énergie pour fournir des suggestions et des commentaires précieux qui ont permis d'améliorer la qualité et la cohérence de la 4^e édition de cet ouvrage.

Enfin, nous remercions Lana Radomsky pour son aide dans la mise en forme et la compilation du glossaire et des références.

Rajiv, Carrie et Dana (mai 2019)



Rajiv S. Jhangiani, Carrie Cuttler, & Dana C. Leighton

Traduction adaptée

Nous souhaitons tout d'abord remercier les auteurs de l'édition originale, de partager le fruit de leur labeur avec la communauté universitaire en tant que ressource éducationnelle libre sous licence ouverte. Nous voulons leur exprimer ma gratitude pour leur accueil chaleureux, leur enthousiasme et tout le soutien offert dès le début de cette entreprise.

La traduction adaptée de ce livre de méthode de recherche en psychologie a été rendue possible grâce au travail collaboratif de membres du corps étudiant, de bibliothécaires, d'une chargée de cours et de professeures bilingues, tous et toutes unis par un lieu du savoir : l'Université Laurentienne.

Nous voulons remercier les étudiantes et l'étudiant qui ont fait la traduction appuyée par les logiciels DeepL, puis Antidote pendant les belles journées d'été : Grace Raddon, Joël Pitre, Amelia Faubert et Monica Jean-Baptiste. Un merci à Émilie Ouellet pour la rédaction de sections en avant-propos.

S'ajoute à l'équipe des collègues horspair. Nous voulons exprimer notre plus sincère gratitude à Dan Scott, bibliothécaire de l'Université Laurentienne, qui nous a guidés tout au long du processus d'adaptation d'un

livre existant et qui a travaillé à la publication de la nouvelle édition par l'intermédiaire de la bibliothèque de l'Université Laurentienne. Il nous a également aidés à mettre en place notre adaptation sur Pressbooks. Bruce Oddson, professeur et spécialiste du contenu, a eu la générosité de faire la révision de la matière dans tous les chapitres. Nous ne pouvons passer sous silence la contribution exceptionnelle de Leïla Saadaoui, bibliothécaire responsable des ressources en français à l'université. Elle a été d'une patience infinie et d'une très grande générosité en agissant comme consultante sur divers sujets et en étant l'avant-dernière relectrice de tous les chapitres.

Enfin, un merci tout spécial à Murielle Marois qui a été la dernière relectrice du livre en son entier.

Nous avons eu la chance de travailler avec une équipe formidable.

Josée Turcotte et Laura Auchinleck (avril 2026)

Ce projet de traduction adaptée a été rendu possible grâce au soutien financier du ministère des Collèges et Universités et du gouvernement du Canada dans le cadre de l'Entente Canada-Ontario relative à l'enseignement dans la langue de la minorité et à l'enseignement de la seconde langue officielle.

À propos de ce livre

Ce manuel est une adaptation d'un manuel écrit par [auteur original anonyme] et adapté par la Fondation Saylor sous une licence Creative Commons Attribution-NonCommercial-ShareAlike 3.0 sans attribution telle que demandée par le créateur original de l'œuvre ou le détenteur de la licence. Le texte original est disponible ici : <http://www.saylor.org/site/textbooks/>

La **première édition canadienne** (publiée en 2013) a été rédigée par Rajiv S. Jhangiani (Kwantlen Polytechnic University) et placée sous licence Creative Commons Attribution-NonCommercial-ShareAlike 3.0. Les révisions comprenaient l'ajout d'une table des matières, des changements au chapitre 3 (Éthique de la recherche) pour inclure un exemple contemporain de manquement à l'éthique et pour refléter les lignes directrices canadiennes en matière d'éthique et les lois sur la protection de la vie privée, des informations supplémentaires concernant la collecte de données en ligne au chapitre 9 (Recherche par sondage), des corrections d'erreurs dans le texte et les formules, des changements orthographiques des conventions américaines aux conventions canadiennes, l'ajout d'une page de couverture et d'autres ajustements nécessaires à la mise en forme.

La **deuxième édition canadienne** (publiée en 2015) a été coécrite par Rajiv S. Jhangiani (Kwantlen Polytechnic University) et I-Chant A. Chiang (Quest University Canada) et placée sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Révisions incluses : (tout au long) révision linguistique, orthographe et formatage, liens vidéo et liens vers des sites web supplémentaires, visualisations interactives, figures, tableaux et exemples ; (chapitre 1) Many Labs Replication Project ; (chapitre 2) examen par les pairs en double aveugle, bases de données de la littérature contemporaine, comment lire des articles universitaires ; (chapitre 3) éthique canadienne ; (chapitre 4) lois, effets, théorie ; (chapitre 5) description plus complète du MMPI, suppression de l'IAT, descriptions de la validité ; (Chapitre 6) descriptions de la validité et du réalisme, plan en carré latin ; (Chapitre 7) études mixtes, débat qualitatif-quantitatif ; (Chapitre 8) exercice factoriel 2×2 ; (Chapitre 9) études électorales canadiennes, ordre et questions ouvertes ; (Chapitre 13) courbe p et annonce de la BASP concernant l'interdiction des valeurs p ; "crise de la reproductibilité" en psychologie ; (Glossaire) ajout de termes clés.

La **deuxième édition américaine** (publiée en 2017) a été rédigée par Dana C. Leighton (Southern Arkansas University) et placée sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Les révisions comprennent la réversion de l'orthographe de l'anglais canadien à l'anglais américain et l'ajout d'une photo de couverture : "Great Wave off Kanagawa" d'après Katsushika Hokusai (葛飾北斎) est du domaine public.

La **troisième édition américaine** (publiée en 2017) a été rédigée par Carrie Cuttler (Washington State University) et placée sous une licence internationale Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Les révisions ont porté sur la réorganisation générale, la révision de la langue, l'orthographe, le formatage, les liens vidéo supplémentaires et les exemples. Plus précisément, la section sur le modèle global a été déplacée du chapitre 1 au chapitre 2, de nouvelles sections ont été ajoutées au chapitre 1 sur les méthodes de connaissance et les objectifs de la science, et un lien sur la crise de la réplication en psychologie a été ajouté au chapitre 1. Le chapitre 2 a également été réorganisé en déplaçant la section sur l'analyse de la littérature de recherche plus tôt dans le chapitre et en prenant des sections du chapitre 4 (sur les théories et les hypothèses), en les déplaçant dans le chapitre 2 et en coupant le reste du chapitre 4.

Les sections du chapitre 2 sur la corrélation ont également été déplacées vers le chapitre 6. De nouvelles sections sur les caractéristiques des bonnes questions de recherche, une vue d'ensemble de la recherche expérimentale par rapport à la recherche non expérimentale, une description des études sur le terrain par rapport aux études en laboratoire et la formulation de conclusions ont également été ajoutées au chapitre 2. Le chapitre 3 a été enrichi d'une définition de l'anonymat, d'un développement sur le rapport Belmont (les principes de respect de la personne et de bienfaisance ont été ajoutés) et d'un lien vers un clip réfutant le mythe selon lequel les vaccins causent l'autisme. Des sections du chapitre 4 (sur la définition des théories et des hypothèses) ont été déplacées vers le chapitre 2 et le reste du chapitre 4 précédent (sur le phénomène, les théories et les hypothèses) a été supprimé. Le chapitre 5 a été réorganisé en déplaçant les sections sur les quatre types de validité, les contrôles de manipulation et les effets placebo plus loin dans le chapitre. Les descriptions des plans à un facteur à deux niveaux, des plans à un facteur à plusieurs niveaux, des plans à groupes appariés, des effets d'ordre et du contrebalancement aléatoire ont été ajoutées au chapitre 5 et le concept de validité statistique a été développé. Le chapitre 6 a également été réorganisé en déplaçant les sections décrivant les coefficients de corrélation des chapitres 2 et 12 au chapitre 6. La section de l'ouvrage consacrée à la corrélation complexe a également été déplacée vers le chapitre 6 et la section consacrée aux quasi-expériences a été déplacée du chapitre 6 vers son propre chapitre (chapitre 8). Les catégories de recherche non expérimentale décrites au chapitre 6 ont été modifiées pour devenir la recherche transversale, la recherche corrélationnelle et la recherche observationnelle. Le chapitre 6 a été complété pour décrire les études transversales, la corrélation partielle, la régression simple, l'utilisation de la régression pour faire des prédictions, les études de cas, l'observation des participants, l'observation déguisée et non déguisée, et l'observation structurée. Les termes "variable indépendante" et "variable dépendante" utilisés dans le contexte de la régression ont été remplacés respectivement par "variable prédictive" et "variable de résultat/critère". Une distinction entre l'échantillonnage stratifié proportionnel et l'échantillonnage stratifié disproportionné a été ajoutée au chapitre 7. La section sur les modèles quasi-expérimentaux a été déplacée dans son propre chapitre (chapitre 8) et a été développée pour inclure l'instrumentation et le test comme menaces à la validité interne des modèles prétest-post-test à un groupe, et pour inclure des sections décrivant le modèle post-test à un groupe seulement, le modèle prétest-post-test à groupes non équivalents, les séries temporelles interrompues avec le modèle à groupes non équivalents, le modèle prétest-post-test avec réplication par permutation, et la réplication par permutation avec suppression du traitement. La section du chapitre 9 sur les plans factoriels a été divisée en deux sections et le reste du chapitre a été déplacé ou coupé. En outre, des exemples d'interactions quotidiennes ont été ajoutés et une description des effets simples a été ajoutée au chapitre 9. La section sur les études de cas qui figurait au chapitre 10 a été modifiée et déplacée au chapitre 6. En outre, des étiquettes ont été ajoutées aux modèles à bases multiples pour les comportements, les contextes et les participants, et un paragraphe de conclusion sur les preuves convergentes a été ajouté au chapitre 10. Seules des modifications mineures ont été apportées aux autres chapitres (chapitres 11, 12 et 13).

La **quatrième édition** (publiée en 2019) a été coécrite par Rajiv S. Jhangiani (Kwantlen Polytechnic University), Carrie Cuttler (Washington State University) et Dana C. Leighton (Texas A&M University-Texarkana) et est placée sous licence Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. Les révisions apportées à la présente édition comprennent la modification du système de numérotation des chapitres et des sections afin de mieux tenir compte des adaptations qui suppriment ou réorganisent les chapitres ; la poursuite du retour à l'édition canadienne ; des révisions grammaticales générales ; le remplacement de "he/she" par "they" et de "his/her" par "their" ; la suppression ou la mise à

jour des liens morts ; l'intégration de vidéos qui n'étaient pas intégrées ; le déplacement des principaux points à retenir et des exercices de la fin de chaque section de chapitre à la fin de chaque chapitre ; une nouvelle conception de la couverture. En outre, les révisions suivantes ont été apportées à certains chapitres :

Chapitre 1 :

- Mise à jour de la liste des thérapies empiriques

Chapitre 2 :

- Ajout d'une description des recherches de suivi menées par Drews, Pasupathi et Strayer (2004) démontrant que les conversations au téléphone cellulaire en conduisant présentent un risque plus élevé que les conversations avec un passager
- Ajout du terme méta-analyse accompagné de sa définition
- Remplacement des termes "hommes" et "femmes" par "mâles" et "femelles"
- Mise à jour de la description du nombre d'enregistrements obtenus avec différents termes de recherche en une description élargie du nombre relative de résultats (qui ne changera pas au fur et à mesure qu'il y a ajout d'articles à PsychINFO)
- Remplacement du terme "définir opérationnellement" les variables par un énoncé plus général sur la mesure des variables, étant donné que l'expression "définition opérationnelle" n'est définie formellement que plus loin dans le texte
- Ajout d'une citation à propos de la recherche de Zajonc (1965)
- Ajout d'une brève description des facteurs (c'est-à-dire petite taille de l'échantillon, seuil alpha strict) qui augmentent la probabilité d'une erreur de type II

Chapitre 3 :

- Suppression des titres des tableaux dans les renvois aux tableaux
- Ajout d'un énoncé indiquant que de nombreuses personnes, y compris des enfants, sont mortes parce que des personnes ont évité le vaccin RRO
- Ajout d'un énoncé sur l'auto-plagiat en tant que comportement non éthique avec l'exemple de soumettre le même travail dans plusieurs cours
- Explication du principe de respect des personnes
- Révision des niveaux d'examen du CÉR pour correspondre à la terminologie utilisée dans les réglementations fédérales
- Les notes de bas de page pour les références sont devenues de véritables notes de bas de page dans Pressbooks.

Chapitre 4 :

- Suppression des exemples potentiellement offensants ou stigmatisants
- Clarification de la définition des niveaux de mesure
- Ajout de références pour les différentes échelles décrites
- Ajout d'une description plus détaillée de la raison pour laquelle le QI est mesuré sur une échelle d'intervalle

- Ajout de descriptions des indicateurs de tendance centrale qu'il convient de calculer et de rapporter pour chacune des échelles de mesure (nominale, ordinale, intervalle, rapport).
- Ajout d'un paragraphe sur la définition opérationnelle du construit, expliquant le passage d'une définition conceptuelle vers quelque chose qui peut être directement observé et mesuré
- Ajout d'une brève description des PsycTESTS et d'un lien vers ces tests
- Suppression de l'énoncé disant que la famille et les amis peuvent constituer de bons sujets pilotes

Chapitre 5 :

- Clarification de la distinction entre variables indépendantes et dépendantes
- Avancement de la discussion sur la condition de contrôle
- Brève discussion sur l'éthique de la recherche dans le cadre de la description de l'étude de Guéguen & de Gail (2003)
- Définition plus claire d'une analyse de puissance et insistance sur l'importance d'en effectuer une
- Mention des variables confondantes dans le cadre de la discussion sur la validité interne
- Mention que les expériences intrasujets nécessitent moins de participants
- Suppression de la référence en double
- Ajout de références
- Mise à jour de la langue

Chapitre 6 :

- Clarification des situations où les approches non expérimentales sont appropriées
- Ajout d'informations sur l'étude non expérimentale de Milgram sur l'obéissance à l'autorité
- Ajout d'une discussion sur les études transversales, longitudinales et trans-séquentielles
- Révision de l'organisation des approches non expérimentales
- Suppression de la description de la variable indépendante choisie par l'expérimentateur
- Spécification des types de variables pouvant être mesurées dans la recherche corrélationnelle
- Ajout d'un exemple d'étude corrélationnelle utilisant des variables catégorielles
- Ajout d'un tableau d'analyse factorielle
- Plus d'exemples de techniques d'analyse de données non statistiques
- Ajout d'un tableau résumant certaines différences entre la recherche quantitative et la recherche qualitative
- Description de certaines dynamiques de groupe et caractéristiques de personnalité susceptibles d'influencer la participation à des groupes de discussion
- Discussion de la recherche de Festinger sur la dissonance cognitive qui a utilisé l'observation participante déguisée
- Description de l'effet Hawthorne
- Ajout d'un exemple d'étude utilisant l'observation structurée dans un environnement de laboratoire.

Chapitre 7 :

- Clarification du langage concernant les méthodes de collecte de données par rapport aux devis de recherche

- Mention de la randomisation de l'ordre de présentation des questions comme un autre moyen de réduire les effets de l'ordre des réponses
- Explication de la cotation inversée
- Description d'autres types d'échantillonnage non probabiliste
- Réitère l'importance d'effectuer une analyse de puissance
- Ajout de sites de collecte de données en ligne

Chapitre 8 :

- Discussion sur la manière dont l'inclusion d'un groupe contrôle élimine certaines menaces à la validité interne dans les études à un seul groupe.

Chapitre 9 :

- Clarification de la discussion sur les plans factoriels non expérimentaux.

Chapitre 10 : aucun changement substantiel

Chapitre 11 :

- Ajout des conférences régionales de psychologie à la liste des conférences
- Discussion condensée et plus claire sur les documents finaux
- Mise à jour de la discussion sur le partage ouvert des résultats afin de reconnaître certaines revues qui exigent des données ouvertes
- Ajout d'une explication sur le langage à la première personne

Chapitre 12 :

- Correction de recommandations erronées concernant le style APA et ajout de références précises au *Publication Manual*
- Uniformisation de l'utilisation des termes "figure" et "graphique" pour mieux correspondre au style APA.
- Modifications mineures de la discussion concernant le format des communications par affiches
- Déplacement de la liste des conférences à la fin de la discussion afin de ne pas interrompre le contenu.

Chapitre 13 :

- Définition du p-hacking et clarification de la discussion sur le p-hacking
- Définition de la valeur p plus juste d'un point de vue technique

Méthodes de recherche en psychologie (première édition canadienne en français)

Le présent document est une traduction en français de la quatrième édition de la version canadienne

(publiée en 2019). Il a été traduit par le biais de différentes méthodes et avec deux objectifs en tête 1) opter pour une écriture inclusive de manière à inclure tout le monde et 2) respecter le texte original autant que possible. Pour la traduction, on a occasionnellement eu recours à DeepL et Chat GPT.

Plusieurs révisions ont été faites pour s'assurer de la justesse du message et obtenir une écriture inclusive. Des termes neutres ou épicènes ont été utilisés tout au long du livre. Pour nommer quelques exemples, les termes genrés chercheur et chercheuse ont été remplacés par les personnes en recherche, ou les scientifiques; les termes étudiante et étudiant ont été remplacés par les personnes aux études, les personnes étudiantes, le corps étudiant, les personnes inscrites à un cours; les termes participants et participantes ont été remplacés par les personnes participantes. Des noms ont été mis au pluriel pour les rendre épicènes (par exemple, des psychologues). D'autres ont été remplacés dans des mots épicènes ou neutre (par exemple, débutants et débutantes par novices, intéressé et intéressée par enthousiaste). Les pronoms genrés ils et elles ont été la plupart du temps enlevés soit en répétant le nom des personnes concernées, soit en y référant autrement (comme de dire ces personnes). Les phrases ont parfois dû être reformulées pour garder une écriture inclusive. Notons que le pronom elles est parfois utilisé pour référer aux personnes mentionnées avant et les mots sont accordés au féminin. Ceci n'est pas fait dans l'optique de féminiser le lectorat, mais plutôt d'accorder les termes au sujet de la phrase, ce sont les personnes. Que les personnes aient un genre fluide, non binaire, masculin ou féminin, nous souhaitons inclure tout le monde.

Depuis la dernière édition de ce livre, une nouvelle édition du style APA a été publiée. Nous avons adapté le document pour suivre cette dernière édition de l'APA. Pour suivre le style APA, nous avons placé toutes les références dans une section références à la fin de chaque chapitre, plutôt que dans des notes en bas de page. En plus du texte et du glossaire, toutes les illustrations et tableaux ont été traduits. Une nouvelle page couverture ainsi que des photos ont été ajoutées au début de chaque chapitre. Des codes QR ont également été créés pour toutes les ressources existantes. Afin de soutenir le lectorat francophone, de nouveaux liens et ressources ont été ajoutés au texte, tandis que les liens morts ont été supprimés.

Chapitre 1:

- L'image 1 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Antoni Shkraba.
- Ajout d'informations et d'un lien vers Les Septiques du Québec (<https://www.sceptiques.qc.ca/dictionnaire.php>)
- Ajout d'une mise à jour indiquant qu'au moment de la publication initiale, le cannabis était illégal, mais que depuis, cela a changé au Canada.
- Ajout d'une référence en français:
[Lilienfeld, S. O. \(2011\). Les 10 commandements pour aider les étudiants à distinguer la science de la pseudoscience en psychologie. Revue Électronique de Psychologie Sociale, 5, 23-29.](#)

Chapitre 2:

- L'image 2 a été ajoutée à partir du site web Pixabay (<https://pixabay.com/>), téléchargée par le créateur StockSnap.
- La Figure 2.2 a été ajoutée et créée par l'illustratrice Laura Auchinleck.
- Ajout de la ressource Revue québécoise de psychologie.
- Ajout que « la recherche se fait en anglais dans cette base de données » et ajout la bonne terminologie

en anglais trouvée dans le Thesaurus de psychINFO

- Ajout que les fournisseurs américains sont généralement ProQuest ou EBSCOhost.
- Ajout que vous pouvez emprunter des ressources d'une autre bibliothèque grâce au service de prêt entre bibliothèques (PEB).
- Ajout de références au Canada; personnes aux études aux États-Unis, au Mexique et au Canada; personnes adolescentes canadiennes
- Ajouter le fait que le style de rédaction de l'American Psychological Association (style APA) est le plus fréquemment demandé dans les revues américaines ou canadiennes en psychologie

Chapitre 3:

- L'image 3 a été ajoutée à partir du site web Pixabay (<https://pixabay.com/>), téléchargée par le créateur whitesession.
- Renvoi à la dernière version de la politique (EPTC 2 2022) publiée en 2022
- Ajout que certaines recherches sont exemptées de l'évaluation par le CÉR
- Ajout de liens en français :
 - Le Code de Nuremberg [en français] :
https://frq.gouv.qc.ca/app/uploads/2021/03/code_nuremberg_1947.pdf
 - La Déclaration d'Helsinki de L'Association Médicale Mondiale (AMM) [en français] :
https://www.frq.gouv.qc.ca/app/uploads/2021/03/declaration_helsinki_2013.pdf
 - Le Rapport Belmont [en français] :
https://frq.gouv.qc.ca/app/uploads/2021/04/rapport_belmont_1974.pdf
 - Politique fédérale en matière de protection des sujets humains :
<https://www.hhs.gov/ohrp/international/ethical-codes-and-research-standards/index.html>
- Ajout des différences organisationnelles entre les systèmes américain et canadien et ajout de liens aux codes et organisations concernées par l'éthique au Canada.
 - Lignes directrices et politiques du Conseil canadien de protection des animaux :
<https://ccac.ca/fr/lignes-directrices-et-politiques/principes-fondamentaux.html>
 - Code canadien d'éthique pour les psychologues de la Société canadienne de psychologie :
<https://cpa.ca/fr/aproposdelascp/comites/ethics/codeofethics/>
- Donne deux exemples d'autoplagiat, comme la 4^e édition américaine

Chapitre 4:

- L'image 4 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Ichad Windhiagiri.
- Ajout d'un lien à la traduction et validation de l'échelle de l'estime de soi de Rosenberg :
<https://onlinelibrary.wiley.com/doi/epdf/10.1080/00207599008247865>
- Ajout d'un lien au critique de la mesure du stress, Guillet et Hermand, 2006,
https://hal.science/hal-00755690/file/Article_Critique_de_L_A_valuation_du_stress.pdf
- Ajout d'une note pour dire « Daniel Gabriel Fahrenheit (l'inventeur de l'échelle de température du même nom) »

- Ajout d'un passage et d'une référence à la version traduite en français auprès des personnes adolescentes de 10 à 14 ans du "Need for Cognition Scale" (Ginet et Py, 2000)
- Mention qu'on peut trouver des ouvrages de référence dans les collections de la bibliothèque de votre université ou dans la bibliothèque du département de psychologie.

Chapitre 5:

- L'image 5 a été ajoutée à partir du site web Pexels (<https://www.pexels.com/>), prise par le photographe Davi Pimentel.
- La Figure 5.3 a été ajoutée et créée par l'illustratrice Laura Auchinleck.
- Ajout de l'article de Peggy Chekroun portant sur les processus conduisant à l'inaction, les situations pouvant réduire l'effet du témoin ou même l'empêcher, et la description plus exacte de ce qu'on vu et entendu les 38 témoins.
 - Chekroun P. (2008). Pourquoi les individus aident-ils moins autrui lorsqu'ils sont nombreux ?. Revue électronique de Psychologie Sociale, 2, 9-16.
<https://psychologiescientifique.org/wp-content/uploads/2018/02/Chekroun-2008-Pourquoi-les-individus-aident-ils-moins-autrui-lor.pdf>

Chapitre 6:

- L'image 6 a été ajoutée à partir du site web Pexels (<https://www.pexels.com/>), prise par le photographe Andrea Piacquadio.
- Ajout d'information dans le texte pour expliquer plus en détail le processus de sélection et pour expliquer l'énoncé où l'on décrit que les personnes participantes ont été réparties de manière aléatoire entre deux conditions (section 29. Recherche corrélacionnelle)
- Ajout du vrai nom du patient, H.M, dans l'une des études de cas les plus célèbres en psychologie, Henry Molaise
- À la phrase où l'on dit que les personnes participantes sont aléatoirement réparties on a ajouté entre deux conditions.
- Changement d'unité de mesure en mètre au lieu de pied dans l'exemple référant à l'étude de Levine et Noranzayan (1999)
- Ajout d'information dans le texte quant à la position des personnes qui observent dans le couloir
- Ajout de ressources en français:
 - Que se passe-t-il lorsque l'on retire l'hippocampe du cerveau ? – Sam Kean [traduit si vous activez les sous-titres en français]
<https://ed.ted.com/lessons/what-happens-when-you-remove-the-hippocampus-sam-kean>
 - Gros (13 mars 2017). Aux frontières du cerveau. L'image de la semaine: «Henry Molaison, l'homme qui ne pouvait plus se souvenir». Blog à la une. CNRS Le Journal: Donner du sens à la science.
<https://lejournalcnrs.fr/nos-blogs/aux-frontieres-du-cerveau/limage-de-la-semaine-henry-molaison-lhomme-qui-ne-pouvait-plus>
- Ajout de références plus récentes comme,
 - Milgram, S. (2017). Soumission à l'autorité: un point de vue expérimental (E. Molinié, Trad.; Nouvelle

éd. augmentée).

- Geertz, C. (2017). The interpretation of cultures (L'interprétation des cultures) 3e éd. Basic Books.
- Corbin, J. et Strauss, A. (2014). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (Les bases de la recherche qualitative : Techniques et procédures pour développer une théorie ancrée) 4e éd. Sage Publications.

Chapitre 7:

- L'image 7 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), capturée par le photographe Cottonbro-studio
- Ajout de références canadiennes, comme les sections portant sur
 - L'échantillonnage stratifié proportionnel: 15,3 % de la population canadienne est asiatique
 - L'échantillonnage stratifié disproportionné: communautés noires au Canada représentent un pourcentage relativement faible de la population canadienne (environ 2,9 %)
 - L'échantillonnage en grappes: petite ville de Saskatchewan
 - Pourquoi un échantillon d'environ 1 000 personnes est-il considéré comme adéquat : la population canadienne
- Spécification de l'endroit sur la page (voir l'encadré "Créer des enquêtes en ligne")
- Ajout d'outils d'enquête en ligne dans leur version française.
 - SurveyMonkey <https://surveymonkey.com> version française <https://fr.surveymonkey.com/>
 - Qualtrics <https://www.qualtrics.com/> version française <https://www.qualtrics.com/fr/>
 - Lime Survey <https://www.limesurvey.org> en français: <https://www.limesurvey.org/fr>
- Retrait de l'allusion au coût de l'utilisation de MTurk

Chapitre 8:

- L'image 8 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Mikhail Nilov.

Chapitre 9:

- L'image 9 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par la photographe Polina Tankilevitch.
- Ajout d'un renvoi au chapitre 6 pour les explications détaillées des problèmes
- Précision qu'il s'agit d'un test de mémoire verbale dans l'exemple portant sur la consommation de caféine
- Ajout d'une explication de ce qu'est le MCAT et des autres ressources

Chapitre 10:

- L'image 10 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Ron Lach.
- Ajout d'un mot précisant la notion de temps à la description de l'étude de Hall et al. (1968)

- Ajout d'un exemple et d'une ressource portant sur l'efficacité de la thérapie cognitivo-comportementale faite par visioconférence pour l'insomnie
 - Lonergan, M., Lafontaine, M.F., Gosselin, D., Guérin, E., Gosselin, J., Joanisse, M., Greenman, P., Chomienne, M.H. et Grenier, J. (2021). La psychothérapie par visioconférence pour l'insomnie chez les francophones en situation minoritaire: un protocole expérimental à cas unique à mesures répétées. *Revue Science et Comportement*, 31(1) 35 – 47.
https://www.science-et-comportement.ugam.ca/AnciensNumeros/V31/Article_024_2021.pdf
- Ajout d'une explication de ce qu'est la stabilité de la réponse: (c'est-à-dire qu'il y a peu ou pas de variabilité de la réponse pendant le niveau de base)
- Ajout d'un exemple sur diverses utilisations de différents protocoles d'expérience à cas unique
 - Lanovaz, M.J. (2013). L'utilisation de devis expérimentaux à cas unique en psychoéducation. *Revue de psychoéducation*, 42(1), 161-183. DOI:10.7202/1061728ar Lien: <https://www.erudit.org/fr/revues/psyedu/2013-v42-n1-psyedu04724/1061728ar/>
- Ajout d'un exemple illustrant le concept de latence (porte sur l'intervention visant à réduire les comportements perturbateurs d'un enfant)

Chapitre 11:

- L'image 11 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Pavel Danilyuk.
- L'information référant au style APA a été mise à jour compte tenu de la récente édition (7e): nombre de pages, chapitres associés au niveau élevé et bas, emplacement des figures, tableaux et annexe dans le document.
- Ajout d'une phrase décrivant brièvement les trois niveaux de base.
- Paragraphe créé pour le premier niveau. Ajout d'une phrase énumérant ce que comprend le format général de l'article.
- Ajout d'un code QR au format de l'article [en anglais]
- Le deuxième niveau, on mentionne le titre des chapitres pertinents.
- Ajout de la traduction d'une citation directe de l'introduction "Biais free language".
- Ajout de quelques exemples associés à l'âge, aux handicaps, au statut socio-économique, à la participation à la recherche.
- Dans le tableau 11.1, certains exemples ont été changés ou adaptés.
- Retrait de la section concernant l'utilisation du terme sujet.
- Ajout sur la rédaction inclusive selon l'APA 7e édition.
- Ajout d'un encadré sur la rédaction inclusive en français qui inclut plusieurs ressources en ligne.
- Pour le tableau 11.2 présentant les erreurs les plus fréquentes, une note indique que ceci a été remarqué avec l'édition précédente du style APA et que certaines d'entre elles ne s'appliquent pas en français.
- Ajout d'un encadré sur le Style APA de l'anglais au français dans lequel on présente des différences.
- Pour les exemples de références en style APA, nous avons suivi la 7e édition.
- Pour le titre de revue scientifique, ajout d'exemples concrets qui illustrent comment les écrire si l'on réfère au titre d'une revue en anglais ou en français.
- Ajout précisant le type de citation: narrative et entre parenthèse

- Ajout d'un encadré sur la paraphrase avec des ressources interactives.
- Retrait de la section sur les auteurs après remerciement et avant adressés.
- Ajout du logo de l'affiliation comme exemple d'éléments visuels; ajout d'une carte d'affaires comme moyen de fournir ses coordonnées
- Ajout de congrès académiques français et retrait de congrès américains locaux destinés aux étudiants.
- Dans les points clés, ajout d'un langage libre de biais
- Ajout de références et d'activités associées à la dernière édition de l'APA (7e édition).

American Psychological Association. (2020). Publication manual of the American Psychological Association 2020: the official guide to APA style (7th ed.). American Psychological Association.

American Psychological Association. (s.d.) Student Paper Setup Guide. <https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf>

American Psychological Association. (s.d.). APA Style. <https://apastyle.apa.org>

American Psychological Association. (s.d.). APA Style: Paraphrasing and citation activities. Repéré le 23 septembre 2025. <https://apastyle.apa.org/instructional-aids/paraphrasing-citation-activities.pdf>

- Ajout de ressources associées à l'adaptation française de l'APA (7e éd.).
 - Couture, M. (2025, 25 mars). *Adaptation française de l'APA (7e éd.)*, sous licence CC BY-NC 4.0. <http://profmcouture.ca/apa/index.htm>
 - Université de Montréal (s.d.). Citer selon les normes APA, 7e édition. Repéré le 23 septembre 2025. <https://bib.umontreal.ca/citer/styles-bibliographiques/apa>
 - Université d'Ottawa (s.d.). Citer et évaluer les sources. Repéré le 23 septembre 2025. <https://www.uottawa.ca/bibliotheque/ecriture-citati>
- Ajout de ressources sur l'écriture inclusive
 - Gouvernement du Canada. Écriture inclusive — les lignes directrices et ressources. <https://www.noslangues-ourlanguages.gc.ca/fr/cles-de-la-redaction/ecriture-inclusive-lignes-directrices-ressources>
 - Institut national de la recherche scientifique (s.d.) *Inclusivement vôtres! Guide de rédaction inclusive* Repéré le 22 avril 2026. <https://inrs.ca/linrs/equite-diversite-et-inclusion/guide-de-redaction-inclusive/>
 - Office québécois de la langue française. (s.d.) *Féminisation et rédaction épiciène*, Dépannage linguistique. Repéré le 22 avril 2026. <https://vitrinelinguistique.oqlf.gouv.qc.ca/banque-de-depannage-linguistique/la-redaction-et-la-communication/feminisation-et-redaction-epicene>
 - Université Laval. *Outils et formations pertinentes en matière d'équité, diversité et inclusion*. Repéré le 22 avril 2026. <https://nutriss.ulaval.ca/boite-a-outils/equite-diversite-et-inclusion-edi/outils-et-formations-pertinentes>
 - Université de Montréal (s.d.). *Inclusivement: Guide d'écriture pour toutes et tous*. Repéré le 22 avril 2026. <https://français.umontreal.ca/soutiller-en-français/ecriture-inclusive/>
 - Université de Sherbrooke. *Le français langue commune à UdeS, le français en outil: La rédaction*

inclusive. Repéré le 22 avril 2026.

<https://www.usherbrooke.ca/langue/le-francais-en-outils/la-redaction-inclusive>

- Université du Québec (2023). *Guide de communication inclusive: pour des communications qui mobilisent, transforment et ont du style!*
<https://reseau.uquebec.ca/system/files/documents/guide-communication-inclusive-universite-du-quebec-2023.pdf>
- Ajout d'une activité formative sur les paraphrases en français.

Université de Montréal (s.d.). *Quiz interactif sur l'écriture en paraphrases*. Repéré le 22 avril 2026.

<https://integrite.umontreal.ca/quiz-generaux/paraphraser/>

Chapitre 12:

- L'image 12 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Fauxels.
- Ajout du mot scores, dans la phrase parlant de la distribution asymétrique... vers l'extrémité supérieure de son étendue de scores.
- Pour la moyenne, on a retiré la phrase "Il s'agit d'une moyenne" puisqu'il s'agit du même terme en français.
- Dans la 3e édition canadienne du livre en anglais, il y a retrait de la description des figures 12.2 portant sur l'inventaire de dépression de Beck et 12.4 décrivant deux histogrammes. Nous avons fait de même.
- Dans la section "Différences entre les groupes ou les conditions", après la phrase entre parenthèses, il y a ajout de la phrase suivante: La gravité de la peur semble être moindre dans les conditions traitements que dans la condition contrôle.
- Les descriptions associées aux figures 12.6 et 12.7 ont été reformulées pour les rendre plus faciles à comprendre en français et ajout du mot nuage.
- Le style APA de la 7e édition a été suivi dans les lignes directrices portant sur la présentation des figures. Le terme anglais captions a été traduit par titres et figures.
- Dans la description de la figure 12.12, on dit deux figures plutôt que deux types de figures.
- Dans les exercices, l'unité de mesure est en centimètre plutôt qu'en pouce.

Chapitre 13:

- L'image 13 a été ajoutée à partir du site web Pexel (<https://www.pexels.com/>), prise par le photographe Fauxels.
- Ajout de JASP comme outil statistique gratuit en science ouverte, développé à University of Amsterdam, disponible dans différentes langues, dont le français. <https://jasp-stats.org/>
- Dans l'exemple d'un psychologue de la santé, on ne réfère pas à des majeures, mais plutôt à des programmes en nutrition et en psychologie.
- Figure 13.4 image adaptée et traduite de l'originale
- Traduction des nouvelles lignes directrices pour la promotion de la transparence et l'ouverture (TOP 2025) disponible en ligne à partir de la référence suivante:
 - Grant, S., Corker, K. S., Mellor, D., Stewart, S. L. K. Cashin, A. G., Lagisz, M., Mayo-Wilson, E., Moher, D.,

Umpierre, D., Barbour, V., Buck, S., Collins, G. S., Hazlett, H. F., Hrynaskiewicz, I., Lee, C. J., Parker, T. H., Rethlefsen, M. L., Toomey, E. and Nosek, B. A. (2025). *TOP 2025: An Update to the Transparency and Openness Promotion Guidelines*. [Center for Open Science](https://osf.io/preprints/metaarxiv/nmfs6_v2). MetaArXiv Preprints
https://osf.io/preprints/metaarxiv/nmfs6_v2

- Ajout d'une référence portant sur la science ouverte
 - Thériault (2024). L'importance de la science ouverte en recherche en psychologie. Accepté dans *Fragments : Revue de psychologie*. Preprint: <https://doi.org/10.31234/osf.io/758dx>
- Ajout d'une référence associée à la valeur p
 - R. Nuzzo (2018). Mathématiques. La malédiction de la valeur-p. *Pour la science* (98). <https://www.pourlascience.fr/sd/mathematiques/la-malediction-de-la-valeur-p-10029.php>
- Badges numériques en français
 - Thériault, R. (2025). Les badges numériques en français tirés de « Bonnes pratiques et science ouverte : Un document d'accompagnement au PRD », disponible au https://remi-theriault.com/files/science_ouverte.pdf. Reproduite avec la permission de l'auteur.
- Ajout d'une référence portant sur la répliquabilité.
 - Vu Van, B. A. (journaliste), (2025, 16 mars). Les Jeux de la répliquabilité: mettre des études à l'épreuve. Dans Lemyre, É. et Caron, S. (réalisation), *Découverte, Radio-Canada Info*. Repéré le 24 août 2025. https://www.youtube.com/watch?v=QcC2cFG_1so

À propos de l'équipe

Auteurs et autrices: Rajiv S. Jhangiani, I-Chant A. Chiang, Carrie Cuttler, Dana C. Leighton, Molly A. Metz, Josée Turcotte et Laura Auchinleck

Traductrices: Josée Turcotte, Monica Jean-Baptiste, Grace Raddon, Joël pitre, Amelia Faubert

Autrice de nouveau contenu et responsable de l'adaptation: Josée Turcotte

Dessinateur-maquettiste et autrice de toutes les figures, graphiques et de l'infographie: Laura Auchinleck

Relectrices et relecteur: Josée Turcotte, Murielle Marois, Leïla Saadaoui, Monica Jean-Baptiste, Bruce Oddson

Éditrice: Josée Turcotte,

Producteur du livre: Daniel B. Scott

Préface

La psychologie, comme la plupart des autres sciences, dispose de son propre ensemble d'outils pour étudier les questions de recherche importantes dans son domaine. Contrairement à d'autres sciences plus anciennes et plus mûres, la psychologie est un domaine relativement nouveau et, comme un adolescent, elle apprend et évolue rapidement. Les personnes menant des recherches en psychologie apprennent et évoluent en même temps que la science émergente. Ce manuel présente aux personnes étudiantes les principes fondamentaux de ce que c'est que de penser comme une personne faisant des recherches en psychologie dans le monde contemporain de la recherche en psychologie.

Historiquement, la psychologie a développé des pratiques et des méthodes basées sur les sciences physiques établies. Contrairement aux sciences physiques, la psychologie a été confrontée à la variation inhérente à ses sujets : les personnes. Pour mieux en tenir compte, nous avons développé des pratiques et des méthodes statistiques que nous considérions (naïvement) comme infaillibles. Au fil du temps, nous avons établi une base de résultats de recherche que nous considérions comme solide.

Ces dernières années, la psychologie s'est tournée vers l'introspection, se tournant vers l'intérieur et réexaminant les connaissances que nous considérions comme fondamentales. Nous avons commencé à découvrir que certaines de ces fondations inébranlables n'étaient pas aussi solides que nous le pensions ; certaines des découvertes fondamentales de la psychologie ont été remises en question et n'ont pas été confirmées par un examen plus approfondi. Comme beaucoup de conversations introspectives, celle-ci a provoqué une crise de foi.

Les psychologues se demandent maintenant si nous savons vraiment ce que nous pensions savoir ou si nous avons simplement eu de la chance. Nous nous efforçons de comprendre comment ce que nous choisissons de publier et de ne pas publier, ce que nous choisissons de rapporter et de ne pas rapporter, et la manière dont nous formons nos personnes aux études au monde de la recherche ont un effet sur ce que nous appelons la "connaissance" en psychologie. Nous commençons à nous demander si ces connaissances représentent le comportement réel et les processus mentaux des êtres humains, ou si elles représentent simplement les effets de notre choix de méthodes. Cette question a déclenché une tempête de feu parmi les chercheurs en psychologie, mais elle doit être résolue. Dans un ouvrage destiné à des personnes aux études qui débutent en psychologie, il est tentant de passer ces questions sous silence et de proclamer que nos "connaissances" sont la "vérité". Ce serait pourtant un mauvais service à rendre à nos étudiants, qui doivent être capables de remettre en question la recherche de manière critique. Au lieu d'éviter cette controverse, ce manuel invite le lecteur à s'y plonger.

À chaque étape, le processus de recherche en psychologie est jalonné de décisions, de compromis et d'incertitudes. Nous décidons d'étudier une variable et pas une autre ; nous mettons en balance les coûts de la recherche et ses avantages ; nous ne savons pas si nos résultats seront reproduits. Chaque étape est une décision qui nous fait prendre une direction différente et nous rapproche ou nous éloigne de la vérité. La recherche n'est pas un chemin facile à parcourir, mais nous espérons que ce manuel sera une carte de randonnée qui pourra au moins inspirer la direction que les étudiants peuvent prendre, et fournir quelques itinéraires absolus pour commencer à voyager.

Comme nous l'avons écrit au début de cette préface, la psychologie est une science jeune. Comme tout adolescent, la psychologie est aux prises avec son identité en tant que science, apprend à utiliser

de meilleurs outils, comprend l'importance de la transparence et a des conversations plus ouvertes pour améliorer sa compréhension du comportement humain. Nous grandirons et mûrirons ensemble. C'est une période passionnante pour participer à cette croissance, alors que la psychologie devient une science plus mature.

CHAPITRE 1-LA SCIENCE DE LA PSYCHOLOGIE



Image 1 – Trois personnes parlent

Nombreuses sont les personnes qui pensent que les femmes ont tendance à parler plus que les hommes — certaines suggérant même que cette différence a une base biologique. Selon une estimation largement citée, les femmes prononceraient en moyenne 20 000 mots par jour, contre 7 000 pour les hommes. Cette affirmation semble plausible, mais est-elle vraie ? Un groupe de scientifiques en psychologie dirigé par Matthias Mehl a décidé de le vérifier. Ces gens ont cherché si quelqu'un avait déjà essayé de compter le nombre de mots prononcés quotidiennement par les femmes et les hommes. Personne ne l'a fait. L'équipe de recherche a donc mené une étude au cours de laquelle des personnes de sexe féminin et masculin étudiant à l'université (369 au total) ont porté des enregistreurs audio pendant que tout le monde vaquait à leurs occupations. Résultat ? Les femmes prononçaient en moyenne 16 215 mots par jour et les hommes, 15 669 — une différence extrêmement faible qui pourrait facilement s'expliquer par le hasard. Dans un article publié dans la revue *Science*, ces scientifiques ont résumé leurs conclusions comme suit : « Nous concluons donc, sur la base des données empiriques disponibles, que le stéréotype très répandu et très médiatisé sur

la loquacité des femmes n'est pas fondé» (Mehl, Vazire, Ramirez-Exparza, Slatcher et Pennebaker, 2007, p. 82).

La psychologie est généralement définie comme l'étude scientifique du comportement humain et des processus mentaux, et cet exemple illustre les caractéristiques qui la rendent scientifique. Dans ce chapitre, nous examinerons de plus près ces caractéristiques, nous passerons en revue les objectifs de la psychologie et nous répondrons à plusieurs questions fondamentales que les personnes aux études se posent souvent à ce sujet. Qui mène des recherches scientifiques en psychologie ? Pourquoi ? La psychologie scientifique nous apprend-elle quelque chose que le sens commun ne nous apprend pas ? Pourquoi devrais-je me donner la peine d'apprendre l'approche scientifique — surtout si je veux travailler en clinique en tant que psychologue, et non en tant que personne en recherche ? Ce sont de très bonnes questions, et y répondre maintenant vous fournira une base solide pour apprendre le reste de la matière de votre cours.

1. Méthodes de connaissance

Objectifs d'apprentissage

1. Décrire les 5 méthodes d'acquisition des connaissances.
2. Comprendre les avantages et les problèmes de chacune d'entre elles.

Prenez une minute pour réfléchir à ce que vous savez et à la manière dont vous avez acquis ces connaissances. Peut-être que vous savez que vous devriez faire votre lit le matin, car votre mère ou votre père vous a dit ce que vous deviez faire. Peut-être savez-vous que les cygnes sont blancs parce que tous les cygnes que vous avez vus étaient blancs. Ou peut-être que vous savez que votre ami·e vous ment parce qu'il/elle agit bizarrement et ne vous regarde pas dans les yeux. Mais devons-nous faire confiance aux connaissances obtenues de ces manières ? Les méthodes d'acquisition des connaissances peuvent être divisées en cinq catégories, chacune ayant ses forces et ses faiblesses.

Intuition

La première méthode de connaissance est **l'intuition**. Lorsque nous utilisons notre intuition, nous nous fions à nos tripes, à nos émotions et/ou à nos instincts pour nous guider. Plutôt que d'examiner les faits ou d'utiliser la pensée rationnelle, l'intuition consiste à croire ce qui semble être vrai. Le problème, c'est que nos intuitions peuvent être erronées parce qu'elles sont guidées par des biais cognitifs et motivationnels, plutôt que par un raisonnement logique ou des données scientifiques. Bien que le comportement étrange de votre ami·e peut vous amener à penser qu'il/elle vous ment, il se peut qu'il/elle retienne un peu de gaz ou soit préoccupé·e par un autre problème qui n'a rien à voir avec vous. Cependant pour certaines personnes, peser les alternatives et penser à toutes les possibilités peut être paralysant et, parfois, les décisions fondées sur l'intuition sont en fait supérieures à celles basées sur l'analyse (les personnes intéressées par cette idée devraient lire le livre *Blink* de Malcolm Gladwell, 2005).

Autorité

L'une des méthodes les plus courantes d'acquisition de connaissances est celle de **l'autorité**. Cette méthode consiste à accepter de nouvelles idées parce qu'une figure d'autorité affirme qu'elles sont vraies. Ces autorités comprennent les parents, les médias, les médecins, les prêtres et autres autorités religieuses, le

gouvernement, et surtout les membres du corps professoral. Dans un monde idéal, nous devrions pouvoir faire confiance aux figures d'autorité, mais l'histoire nous a appris le contraire. De nombreuses atrocités commises contre l'humanité sont la conséquence de gens suivant aveuglément l'autorité (par exemple, les procès des sorcières de Salem, les crimes de guerre nazis). À un niveau plus anodin, bien que vos parents vous aient peut-être dit de faire votre lit le matin, faire votre lit crée un environnement chaud et humide dans lequel les acariens prospèrent. Garder les draps ouverts rend l'environnement moins hospitalier pour les acariens. Ces exemples illustrent le problème de l'utilisation de l'autorité comme source de connaissance: l'autorité peut se tromper, elle peut simplement utiliser son intuition pour parvenir à ses conclusions, et elle peut avoir ses propres raisons de vous induire en erreur. Néanmoins, une grande partie de l'information que nous acquérons provient de l'autorité, car nous n'avons pas le temps de remettre en question et d'effectuer des recherches indépendantes sur chaque information que nous apprenons grâce à l'autorité. Mais nous pouvons apprendre à évaluer les compétences des figures d'autorité, à évaluer les méthodes qu'elles ont utilisées pour arriver à leurs conclusions, et à déterminer si elles ont des raisons de nous induire en erreur.

Rationalisme

Le **rationalisme** consiste à utiliser la logique et le raisonnement pour acquérir de nouvelles connaissances. Avec cette méthode, on énonce des prémisses puis on suit des règles logiques pour parvenir à des conclusions solides. Par exemple si l'on me donne la prémisse que tous les cygnes sont blancs et la prémisse que ceci est un cygne, je peux arriver à la conclusion rationnelle que ce cygne est blanc sans même avoir vu le cygne. Le problème de cette méthode est que, si les prémisses sont erronées ou s'il y a une erreur dans le raisonnement logique, alors la conclusion ne sera pas valide. Par exemple, la prémisse selon laquelle tous les cygnes sont blancs est incorrecte; il y a des cygnes noirs en Australie. De plus, si l'on n'est pas formé aux règles formelles de la logique, il est facile de commettre une erreur. Néanmoins, si les prémisses sont correctes et si les règles logiques sont suivies correctement, il s'agit d'un bon moyen d'acquérir des connaissances.

Empirisme

L'**empirisme** consiste à acquérir des connaissances par l'observation et l'expérience. Une fois encore, beaucoup d'entre vous ont pu croire que tous les cygnes sont blancs parce que vous n'avez vu que des cygnes blancs. Pendant des siècles, les gens ont cru que la Terre était plate parce qu'elle semblait l'être. Ces exemples et les nombreuses illusions visuelles qui trompent nos sens illustrent les problèmes liés au fait de se fier uniquement à l'empirisme pour obtenir des connaissances. Il y a des limites dans ce que nous pouvons expérimenter et observer, et nos sens peuvent nous tromper. De plus, nos expériences antérieures peuvent modifier la façon dont nous percevons les événements. Néanmoins, l'empirisme est au cœur de la méthode scientifique. La science repose sur des observations. Mais pas n'importe quelles observations, la science s'appuie sur des observations structurées, appelées empirisme systématique.

La méthode scientifique

La **méthode scientifique** est un processus de collecte et d'évaluation systématique de données afin de tester des idées et de répondre à des questions. Bien que les scientifiques puissent utiliser l'intuition, l'autorité, le rationalisme et l'empirisme pour générer de nouvelles idées, ces personnes ne s'arrêtent pas là. Les scientifiques vont plus loin en utilisant l'empirisme systématique pour faire des recherches approfondies. Ces personnes font des observations dans diverses conditions contrôlées afin de tester leurs idées, et utilisent le rationalisme pour parvenir à des conclusions valides. Bien que la méthode scientifique soit la plus susceptible de produire des connaissances valides, comme toutes les méthodes d'acquisition de connaissances, elle présente également des inconvénients. L'un des principaux problèmes est qu'il n'est pas toujours possible d'utiliser la méthode scientifique; cette méthode peut nécessiter beaucoup de temps et de ressources. Un autre problème de la méthode scientifique est qu'elle ne peut pas être utilisée pour répondre à toutes les questions. Comme décrit dans la section suivante, la méthode scientifique ne peut être utilisée que pour répondre à des questions empiriques. Ce livre et votre cours sur les méthodes de recherche sont conçus pour vous fournir un examen approfondi de la manière dont les psychologues utilisent la méthode scientifique pour faire progresser notre compréhension du comportement humain et de l'esprit.

2. Comprendre la science

Objectifs d'apprentissage

1. Définir la science.
2. Décrire les trois caractéristiques fondamentales de la science.
3. Expliquer pourquoi la psychologie est une science.
4. Définir la pseudoscience et en donner quelques exemples.

Qu'est-ce que la science?

Certaines personnes sont surprises d'apprendre que la psychologie est une **science**. Elles admettent généralement que l'astronomie, la biologie et la chimie sont des sciences, mais elles se demandent ce que la psychologie a en commun avec ces autres domaines. Toutefois avant de répondre à cette question, il vaut la peine de réfléchir à ce que l'astronomie, la biologie et la chimie ont en commun entre elles. Il est clair que ce n'est pas leur objet d'étude. Les astronomes étudient les corps célestes, les biologistes étudient les organismes vivants, et les chimistes étudient la matière et ses propriétés. Ce n'est pas non plus l'équipement et les techniques utilisés. Peu de biologistes sauraient quoi faire avec un radiotélescope par exemple, et peu de chimistes sauraient comment suivre une population d'originaux dans la nature. Pour ces raisons et d'autres encore, les philosophes et les scientifiques, qui ont longuement réfléchi à cette question, ont conclu que ce que les sciences ont en commun, c'est une approche générale visant à comprendre le monde naturel. La psychologie est une science parce qu'elle adopte cette même approche générale pour comprendre un aspect du monde naturel : le comportement humain.

Caractéristiques de la science

L'approche scientifique générale possède trois caractéristiques fondamentales (Stanovich, 2010). La première est **l'empirisme systématique**. L'empirisme fait référence à l'apprentissage fondé sur l'observation, et les scientifiques apprennent à connaître le monde naturel de manière systématique, en planifiant, en réalisant, en enregistrant, et en analysant soigneusement des observations. Comme nous le verrons, le raisonnement logique et même la créativité jouent également un rôle important dans la science. Mais les scientifiques sont uniques dans leur insistance à vérifier leurs idées sur le fonctionnement du monde à l'aide de leurs observations systématiques. Remarquez, par exemple, que Mehl et ses collègues ne se fiaient pas aux

stéréotypes d'autres personnes ou même à leurs propres observations informelles. Au contraire, cette équipe de recherche a plutôt enregistré, compté et comparé systématiquement le nombre de mots prononcés par un large échantillon de femmes et d'hommes. De plus lorsque leurs observations systématiques ont contredit les stéréotypes des gens, cette équipe a fait confiance à leurs observations systématiques.

La deuxième caractéristique de l'approche scientifique — qui découle directement de la première — est qu'elle s'intéresse aux **questions empiriques**. Il s'agit de questions portant sur la manière dont le monde est réellement, et par conséquent auxquelles l'observation systématique peut répondre. La question de savoir si les femmes parlent plus que les hommes est empirique en ce sens. Soit les femmes parlent réellement plus que les hommes, soit elles ne le font pas, et cela peut être déterminé en observant systématiquement le nombre de mots prononcés par des femmes et des hommes. Cela dit, il existe de nombreuses questions intéressantes et importantes qui ne peuvent être testées empiriquement et auxquelles la science n'est pas en mesure de répondre. Parmi celles-ci figurent les questions relatives aux valeurs — à savoir si les choses sont bonnes ou mauvaises, justes ou injustes, belles ou laides, et comment le monde devrait être. Ainsi bien que la question de savoir, si un stéréotype est exact ou inexact, soit une question empiriquement testable à laquelle la science peut répondre, la question — ou plutôt le jugement de valeur — de savoir, s'il est mal pour les gens d'avoir des stéréotypes inexacts, ne l'est pas. De même la question de savoir, si le comportement criminel a une base génétique, est une question empirique mais la question de savoir, quelles actions devraient être considérées comme illégales, ne l'est pas. Il est particulièrement important que les personnes menant des recherches en psychologie prennent conscience de cette distinction.

La troisième caractéristique de la science est qu'elle crée un **savoir public**. Après avoir posé leurs questions empiriques, effectué leurs observations systématiques et tiré leurs conclusions, les scientifiques publient leurs travaux. Cela signifie généralement rédiger un article destiné à être publié dans une revue scientifique, dans lequel ces scientifiques situent leurs questions de recherche dans le contexte des recherches antérieures, décrivent de manière détaillée les méthodes utilisées pour y répondre, et présentent clairement leurs résultats et leurs conclusions. De plus en plus, les scientifiques choisissent de publier leurs travaux dans des revues en libre accès, où les articles sont accessibles gratuitement à tout le monde — scientifiques comme non-scientifiques. Ce choix important permet à la recherche financée par des fonds publics de créer un savoir véritablement public.

La publication est une caractéristique essentielle de la science pour deux raisons. La première est que la science est un processus social — une collaboration à grande échelle entre de nombreuses personnes en recherche réparties dans le temps et dans l'espace. Notre savoir scientifique actuel sur la plupart des sujets repose sur de nombreuses études différentes menées par de nombreuses personnes en recherche qui ont partagé leur travail publiquement au fil des années. La seconde est que la publication permet à la science de se corriger. Les scientifiques comprennent que, malgré tous leurs efforts, leurs méthodes peuvent être imparfaites, et leurs conclusions peuvent être erronées. La publication permet aux autres membres de la communauté scientifique de détecter et de corriger ces erreurs, de sorte qu'au fil du temps, le savoir scientifique reflète de plus en plus fidèlement la réalité du monde.

Un bon exemple de la nature autocorrective de la science est le «[Many Labs Replication Project](#)» — un vaste effort coordonné par d'éminentes personnes menant des recherches en psychologie à travers le monde pour tenter de reproduire les résultats de 13 études classiques et contemporaines (Klein et al., 2013). Un des résultats sélectionnés par ces personnes en recherche pour être reproduits est l'effet fascinant, rapporté pour la première fois par Simone Schnall et ses collègues de l'University of Plymouth, selon lequel le fait de se laver les mains amène les gens à juger des transgressions morales — allant du fait de garder

l'argent trouvé dans un portefeuille à l'utilisation d'un chaton à des fins d'excitation sexuelle — comme moins répréhensibles (Schnall, Benton et Harvey, 2008). Si cet effet est **fidèle**, il pourrait contribuer à expliquer pourquoi tant de traditions religieuses associent la propreté physique à la pureté morale. Toutefois, malgré l'utilisation du même matériel et des procédures quasi identiques avec un échantillon beaucoup plus vaste, les personnes en recherche de « Many Labs » n'ont pas réussi à reproduire le résultat initial (Johnson, Cheung et Donnellan, 2013), ce qui suggère que le résultat pourrait provenir de la taille relativement petite de l'échantillon dans l'étude initiale (qui peut mener à des résultats peu fidèles). Pour clarifier, à ce stade, nous ne pouvons toujours pas conclure définitivement que l'effet du lavage des mains n'existe pas; cependant, les efforts déployés pour tester sa fidélité démontrent certainement la nature collaborative et prudente du progrès scientifique.

Pour en savoir plus sur la crise de la réplication en psychologie, voir :
nobaproject.com/modules/the-replication-crisis-in-psychology

La science contre la pseudoscience

Les **pseudosciences** désignent des activités et des croyances qui sont présentées comme scientifiques par leurs adeptes — et qui peuvent sembler scientifiques à première vue — mais qui ne le sont pas. Prenons l'exemple de la théorie des biorythmes (à ne pas confondre avec les cycles du sommeil ou les rythmes circadiens qui, eux, ont une base scientifique). L'idée est que les capacités physiques, intellectuelles et émotionnelles des individus suivent des cycles commençant à la naissance, et se poursuivent jusqu'à leur mort. On affirme que le cycle physique aurait une période de 23 jours, le cycle intellectuel, une période de 33 jours et le cycle émotionnel, une période de 28 jours. Par exemple si vous aviez la possibilité de choisir le moment de votre examen, vous souhaiteriez le programmer à un moment où votre cycle intellectuel est à son apogée. La théorie des biorythmes existe depuis plus de 100 ans, et vous pouvez trouver de nombreux livres, et sites web populaires sur les biorythmes, qui contiennent souvent des termes impressionnants et à consonance scientifique, tels qu'onde sinusoïdale et bioélectricité. Cependant le problème avec les biorythmes est que les données scientifiques indiquent qu'ils n'existent pas (Hines, 1998).

Un ensemble de croyances ou d'activités peut être qualifié de pseudoscientifique si (a) ses adeptes affirment ou laissent entendre qu'il est scientifique, mais (b) qu'il manque une ou plusieurs des trois caractéristiques fondamentales de la science. Par exemple il peut lui manquer l'empirisme systématique. Soit il n'existe pas de recherche scientifique pertinente, soit, comme dans le cas des biorythmes, il existe des recherches scientifiques pertinentes, mais elles sont ignorées. Il peut également manquer de savoir public. Les personnes qui promeuvent ces croyances ou activités peuvent prétendre avoir mené des recherches scientifiques, mais ne les publient jamais de manière à permettre leur évaluation par d'autres.

Un ensemble de croyances et d'activités peut également être pseudoscientifique parce qu'il n'aborde pas de questions empiriques. Le philosophe Karl Popper s'est particulièrement intéressé à cette idée (Popper, 2002). Il a soutenu plus précisément que toute affirmation scientifique doit être formulée de telle sorte qu'il existe des observations qui pourraient — si elles étaient faites — constituer des preuves contre cette affirmation. En d'autres termes, les affirmations scientifiques doivent être **falsifiables**. Par exemple l'affirmation selon laquelle les femmes parlent plus que les hommes est falsifiable parce que des observations systématiques pourraient révéler soit qu'elles parlent plus que les hommes, soit qu'elles ne le font pas.

Comme exemple d'affirmation infalsifiable, considérons que de nombreuses personnes croyant à la perception extrasensorielle (PES) et à d'autres pouvoirs psychiques affirment que ces pouvoirs peuvent disparaître lorsqu'ils sont observés de trop près. Cela fait en sorte qu'aucune observation possible ne peut être considérée comme une preuve contre la PES. Si un test rigoureux d'une personne se proclamant voyante montrait qu'elle prédit l'avenir avec une précision supérieure au hasard, cela serait compatible avec l'affirmation selon laquelle elle possède des pouvoirs psychiques. Mais si elle échouait à prédire l'avenir à un niveau supérieur au hasard, cela serait également compatible avec l'affirmation, puisque ses pouvoirs peuvent soi-disant disparaître lorsqu'ils sont observés de trop près.

Pourquoi devrions-nous nous préoccuper des pseudosciences ? Il y a au moins trois raisons. La première est que l'étude des pseudosciences permet de mieux cerner les caractéristiques fondamentales de la science — et leur importance. La deuxième est que les biorythmes, les pouvoirs psychiques, l'astrologie et bien d'autres croyances pseudoscientifiques sont largement répandus et sont promus sur Internet, à la télévision, ainsi que dans des livres et des magazines. Loin d'être inoffensive, la promotion de ces croyances entraîne souvent de graves conséquences personnelles, par exemple lorsque les adeptes des pseudosciences optent pour des « traitements » tels que l'homéopathie, dans le cas de maladies graves, au lieu de traitements reposant sur des données empiriques. Comprendre, ce qui les rend pseudoscientifiques, peut aider à identifier et à évaluer de telles croyances et pratiques, lorsqu'on y est confronté. La troisième raison est que de nombreuses pseudosciences prétendent expliquer certains aspects du comportement humain et des processus mentaux, notamment les biorythmes, l'astrologie, la graphologie (analyse de l'écriture manuscrite) et la thérapie magnétique pour le soulagement de la douleur. Il est important que les personnes étudiant en psychologie distinguent clairement leur propre discipline de cette « pseudo-psychologie ».

Le dictionnaire des sceptiques

Le *Skeptic's Dictionary* (www.skeptdic.com) est une excellente source d'informations sur les pseudosciences. Ce dictionnaire a été traduit en français par Les Septiques du Québec (www.sceptiques.qc.ca/dictionnaire.php). Parmi les croyances et pratiques pseudoscientifiques que vous pouvez découvrir, citons les suivantes :

- **Cryptozoologie.** L'étude des créatures « cachées » telles que Bigfoot, le monstre du Loch Ness et le Chupacabra.
- **Psychothérapies pseudoscientifiques.** Régression dans la vie antérieure, thérapie de rebirthing et thérapie de bioscream entre autres.
- **Homéopathie.** Traitement des affections médicales à l'aide de substances naturelles qui ont été diluées, parfois au point de ne plus être présentes.
- **Pyramidologie.** Théories étranges sur l'origine et la fonction des pyramides égyptiennes (par exemple, qu'elles ont été construites par des extraterrestres) et l'idée que les pyramides, en général, ont des pouvoirs de guérison et d'autres pouvoirs spéciaux.

Une autre excellente ressource en ligne est *Neurobonkers* (neurobonkers.com), qui publie régulièrement des articles sur des affirmations relevant spécifiquement de la science psychologique.

3. Objectifs de la science

Objectifs d'apprentissage

1. Décrire les trois objectifs de la science et donner un exemple pour chacun.
2. Distinguer la recherche fondamentale de la recherche appliquée.

Les objectifs généraux de la recherche scientifique en psychologie

Les êtres humains ont toujours été curieux du monde naturel, y compris d'eux-mêmes et de leur comportement (en fait, c'est probablement la raison pour laquelle vous étudiez la psychologie). La science est née de cette curiosité naturelle, et est devenue le meilleur moyen d'obtenir un savoir détaillé et exact. Gardons à l'esprit que la plupart des phénomènes et des théories qui figurent dans les manuels de psychologie sont le produit de la recherche scientifique. Dans un manuel typique d'introduction à la psychologie par exemple, on peut en apprendre davantage sur les zones corticales spécifiques au langage et à la perception, sur les principes du conditionnement classique et opérant, sur les biais de raisonnement et de jugement, et sur la tendance surprenante des gens à obéir aux personnes qui sont en position d'autorité. Et la recherche scientifique se poursuit, car ce que nous savons en ce moment ne représente qu'une infime partie de ce qu'il est possible de connaître.

Les trois objectifs de la science

Le premier et le plus fondamental objectif de la science est de **décrire**. Cet objectif est atteint en réalisant des observations minutieuses. À titre d'exemple, je peux m'intéresser à mieux comprendre les conditions médicales pour lesquelles les bénéficiaires utilisent la marijuana à des fins médicales. Dans ce cas, je pourrais essayer d'accéder aux dossiers de plusieurs grands centres ayant une licence pour l'utilisation de la marijuana à des fins médicales afin de voir quelles conditions amènent les gens à être autorisés à utiliser la marijuana à des fins médicales¹. Je pourrais également interroger un large échantillon de bénéficiaires utilisant de la marijuana à des fins médicales, et leur demander d'indiquer quelles conditions médicales ces personnes traitent ou gèrent avec la marijuana. En effet des recherches, basées sur des enquêtes auprès de bénéficiaires utilisant la marijuana à des fins médicales, ont été menées et ont révélé que le principal

symptôme pour lequel ces personnes utilisent la marijuana à des fins médicales est la douleur, suivie de l'anxiété et de la dépression (Sexton, Cuttler, Finnell et Mischley, 2016).

Le deuxième objectif de la science est de **prédire**. Une fois que nous avons observé de façon régulière que deux comportements ou événements sont systématiquement liés l'un à l'autre, nous pouvons utiliser cette information pour prédire si un événement ou un comportement se produira dans une situation donnée. Lorsque je sais que la plupart des bénéficiaires, qui utilisent la marijuana à des fins médicales, l'utilisent pour traiter la douleur, je peux tenir compte de cette information pour prédire qu'une personne qui utilise du cannabis médical ressent probablement de la douleur. Bien sûr, mes prédictions ne seront pas exactes à 100 %, mais si la relation entre l'usage de la marijuana à des fins médicales et la douleur est forte, alors mes prédictions auront une exactitude supérieure au hasard.

Le troisième et ultime objectif de la science est d'**expliquer**. Cet objectif consiste à déterminer les causes d'un comportement. Par exemple, les personnes en recherche pourraient essayer de comprendre les mécanismes par lesquels la marijuana réduit la douleur. Est-ce que la marijuana réduit l'inflammation, ce qui réduit à son tour la douleur? Ou bien est-ce que la marijuana réduit simplement la détresse associée à la douleur plutôt que la douleur elle-même? Comme vous pouvez le constater, ces questions portent sur les mécanismes sous-jacents et les relations causales.

Recherche fondamentale et recherche appliquée

La recherche scientifique est souvent classée en deux catégories: **fondamentale et appliquée**. La **recherche fondamentale** en psychologie est menée principalement dans le but de parvenir à une compréhension plus détaillée et plus précise du comportement humain, sans nécessairement essayer de résoudre un problème pratique particulier. La recherche de Mehl et de ses collègues appartient à cette catégorie. La **recherche appliquée**, quant à elle, est menée principalement pour répondre à un problème pratique. La recherche sur les effets de l'utilisation du téléphone cellulaire au volant, par exemple, a été motivée par des préoccupations en matière de sécurité, et a mené à l'adoption de lois visant à limiter cette pratique. Bien que la distinction entre recherche fondamentale et recherche appliquée soit pratique, elle n'est pas toujours évidente. Par exemple la recherche fondamentale sur les différences entre les sexes en matière de loquacité pourrait éventuellement avoir une incidence sur la pratique de la thérapie conjugale, et la recherche appliquée sur les effets de l'utilisation du téléphone cellulaire au volant pourrait produire de nouvelles connaissances sur les processus fondamentaux de la perception, de l'attention et de l'action.

Notes

1. Le cannabis était illégal au moment où cette question s'est posée.

4. Science et bon sens

Objectifs d'apprentissage

1. Expliquer les limites du bon sens lorsqu'il s'agit d'atteindre une compréhension détaillée et exacte du comportement humain.
2. Donner plusieurs exemples de bon sens ou de psychologie populaire qui sont incorrects.
3. Définir le scepticisme et son rôle dans la psychologie scientifique.

Peut-on compter sur le bon sens?

Certaines personnes se demandent si l'approche scientifique de la psychologie est nécessaire. Ne pourrions-nous pas parvenir aux mêmes conclusions en nous fiant simplement au bon sens ou à l'intuition? Il est certain que tout le monde a des croyances intuitives sur le comportement, les pensées et les émotions des gens — et ces croyances sont collectivement désignées sous le nom de **psychologie populaire**. Bien qu'une grande partie de notre psychologie populaire soit probablement raisonnablement exacte, il est clair qu'une grande partie ne l'est pas. Par exemple la plupart des gens croient que la colère peut être soulagée en « se défoulant » — peut-être en donnant un coup de poing ou en criant très fort. La recherche scientifique a cependant montré que cette approche tend à accroître la colère, et non à la réduire (Bushman, 2002). De même la plupart des gens croient que personne n'avouerait un crime qu'il n'a pas commis, sauf peut-être si cette personne était physiquement torturée. Mais encore une fois, de nombreuses recherches empiriques ont montré que les faux aveux sont étonnamment courants et se produisent pour diverses raisons (Kassin et Gudjonsson, 2004).

Quelques grands mythes

Dans *50 Great Myths of Popular Psychology*, le psychologue Scott Lilienfeld et ses collègues évoquent plusieurs croyances intuitives largement répandues sur le comportement humain, dont la recherche scientifique a montré qu'elles étaient erronées (Lilienfeld, Lynn, Ruscko et Beyerstein, 2010). En voici une courte liste :

- Les gens n'utilisent que 10 % de la capacité de leur cerveau.

- La plupart des gens vivent une crise de la quarantaine ou de la cinquantaine.
- Les personnes aux études apprennent mieux lorsque les styles d'enseignement sont adaptés à leurs styles d'apprentissage.
- Une faible estime de soi est une cause majeure de problèmes psychologiques.
- Les hospitalisations en psychiatrie et les crimes augmentent pendant les pleines lunes.

Comment avons-nous pu nous tromper à ce point?

Comment se fait-il que tant de nos croyances intuitives sur le comportement humain soient erronées? Remarquez qu'il s'agit ici d'une question empirique et il se trouve que les psychologues ont mené des recherches scientifiques à ce sujet, et ont identifié de nombreux facteurs contributifs (Gilovich, 1991). L'un des facteurs est que la formation de croyances détaillées et exactes exige des capacités d'observation, de mémoire et d'analyse à un niveau que nous ne possédons pas naturellement. Il serait pratiquement impossible de compter le nombre de mots prononcés par les femmes et les hommes que nous rencontrons, d'estimer le nombre de mots prononcés par jour, d'en faire une moyenne pour les deux groupes et de les comparer — et ce, entièrement dans notre tête. C'est pourquoi nous avons tendance à utiliser des raccourcis mentaux (ce que les psychologues appellent des **heuristiques**) pour former et entretenir nos croyances. Par exemple, si une croyance est largement partagée — surtout si elle est approuvée par des « spécialistes » — et semble intuitive, nous avons tendance à supposer qu'elle est vraie. Ce phénomène est amplifié par le fait que nous avons ensuite tendance à nous concentrer sur les cas qui confirment nos croyances intuitives et non sur ceux qui les infirment. C'est ce qu'on appelle le **biais de confirmation**. Par exemple lorsque nous commençons à croire que les femmes sont plus loquaces que les hommes, nous avons tendance à remarquer et à nous souvenir des femmes bavardes et des hommes silencieux, mais à ignorer ou à oublier les femmes silencieuses et les hommes bavards. Nous entretenons également des croyances erronées en partie parce qu'il serait agréable qu'elles soient vraies. Par exemple de nombreuses personnes pensent que les régimes hypocaloriques constituent un traitement efficace à long terme contre l'obésité, alors qu'un examen approfondi des données scientifiques a montré que ce n'était pas le cas (Mann et al., 2007). Les gens peuvent continuer à croire en l'efficacité des régimes, en partie parce que cela leur donne l'espoir de perdre du poids si ces personnes sont obèses, ou cela leur procure un sentiment de satisfaction quant à leur propre « maîtrise de soi » si ces personnes ne le sont pas.

Les scientifiques — et particulièrement les psychologues — se savent tout aussi vulnérables que quiconque aux croyances intuitives mais erronées. C'est pourquoi les scientifiques cultivent une attitude de **scepticisme**. Être sceptique ne signifie pas être cynique ou méfiant·e, ni remettre en question chaque croyance ou affirmation que l'on rencontre (ce qui serait de toute façon impossible). Cela signifie plutôt prendre le temps d'envisager des alternatives et de chercher des données — en particulier des données empiriques probantes recueillies systématiquement — lorsque l'enjeu est suffisamment important pour le justifier. Imaginons, par exemple, que vous lisiez un article de magazine affirmant que de donner une allocation hebdomadaire aux enfants est un bon moyen de les aider à développer leur sens des

responsabilités financières. Il s'agit d'une affirmation intéressante et potentiellement importante (surtout si vous avez des enfants). Adopter une attitude sceptique consisterait à se demander pendant un moment si, au contraire, recevoir une allocation apprend simplement aux enfants à dépenser de l'argent — voire à devenir plus matérialistes. Adopter une attitude sceptique consisterait aussi à se demander quels éléments probants soutiennent l'affirmation initiale. L'auteur ou l'autrice est-il/elle une personne travaillant en recherche scientifique? Des données probantes scientifiques sont-elles citées? Si l'enjeu est suffisamment important, cela pourrait impliquer de consulter la littérature de recherche pour voir si d'autres ont étudié la question.

Parce qu'il n'y a souvent pas assez d'éléments probants pour évaluer pleinement une croyance ou une affirmation, les scientifiques cultivent également une **tolérance à l'incertitude**. Les scientifiques acceptent qu'il y ait beaucoup de choses qui ne sont tout simplement pas connues. Par exemple il n'existe aucune donnée probante scientifique démontrant que le fait de recevoir une allocation rend les enfants plus responsables financièrement. Il n'y a pas non plus de preuve scientifique que ça les rend plus matérialistes. Bien que ce type d'incertitude puisse être problématique d'un point de vue pratique — par exemple, il est difficile de décider quoi faire lorsque nos enfants demandent de l'argent de poche —, il est stimulant d'un point de vue scientifique. Si nous ne connaissons pas la réponse à une question intéressante et empiriquement testable, la science et peut-être même vous, en tant que personne en recherche, pourraient être en mesure de la fournir.



Voici 10 conseils pour aider les personnes aux études à distinguer la science de la pseudoscience en psychologie (Lilienfeld, 2011).

<https://psychologiescientifique.org/ressources/pedagogie/10-conseils-pour-aider-vos-etudiants-a-distinguer-la-science-de-la-pseudoscience/>

*Lecture sur version
imprimée? Scannez ce code
QR avec votre téléphone
cellulaire*

5. Psychologues expérimentaux et cliniques

Objectifs d'apprentissage

1. Définir la pratique clinique de la psychologie et la distinguer de la psychologie expérimentale.
2. Expliquer en quoi la science est pertinente pour la pratique clinique.
3. Définir le concept de traitement empirique validé et donner quelques exemples.

Qui mène la recherche scientifique en psychologie?

Psychologues expérimentaux

La recherche scientifique en psychologie est généralement menée par des personnes titulaires d'un doctorat (souvent le doctorat **en philosophie** [Ph.D.]) ou d'une maîtrise en psychologie ou dans des domaines connexes, souvent soutenues par des auxiliaires de recherche possédant un baccalauréat ou une autre formation pertinente. Ces personnes travaillent parfois pour des organismes gouvernementaux (par exemple, pour étudier l'impact des politiques publiques), des associations nationales (par exemple, Société canadienne de psychologie, l'American Psychological Association), des organisations à but non lucratif (par exemple, la National Alliance on Mental Illness) ou dans le secteur privé (par exemple, dans la mise en marché et le développement de produits; en comportement organisationnel). Toutefois la majorité travaille dans des collèges et des universités, et collabore souvent avec leurs étudiant·e·s au premier cycle et aux cycles supérieurs. Bien que certaines personnes en recherche aient la formation et l'autorisation de pratiquer en santé mentale — en particulier les personnes menant des recherches en psychologie clinique — la majorité ne l'a pas. Au contraire, leurs expertises se trouvent dans un ou plusieurs autres sous-domaines de la psychologie: les neurosciences comportementales, la psychologie cognitive, la psychologie du développement, la psychologie de la personnalité, la psychologie sociale, etc. Le travail des personnes en recherche titulaires d'un doctorat peut être de mener des recherches à temps plein ou, comme plusieurs membres du corps professoral de collèges et d'universités, de mener des recherches, en plus de donner des cours, et de servir leur établissement et leur communauté d'autres manières.

Bien entendu, les personnes mènent également des recherches en psychologie pour les défis intellectuels et techniques qu'elles comportent, et pour la satisfaction de contribuer aux connaissances scientifiques

du comportement humain. Vous pourriez vous-même découvrir que vous aimez ce processus. Si c'est le cas, votre établissement d'enseignement supérieur pourrait vous offrir des occasions de participer à des recherches en cours en tant qu'auxiliaire de recherche ou en tant que personne participante. Bien sûr il est aussi possible que vous n'aimiez pas le processus de recherche scientifique en psychologie. Mais au minimum, vous aurez une meilleure compréhension de l'origine des connaissances scientifiques en psychologie, une appréciation de ses forces et de ses limites, et une conscience de la manière dont elles peuvent être appliquées pour résoudre des problèmes pratiques en psychologie, et dans la vie de tous les jours.

Blogs portant sur la psychologie scientifique

Un moyen simple et amusant de suivre l'actualité de la recherche scientifique en psychologie consiste à lire l'un des nombreux et excellents blogs consacrés à résumer et commenter les nouvelles découvertes. En voici quelques-uns (en anglais):

Research Digest, digest.bps.org.uk/
Talk Psych, www.talkpsych.com/
Brain Blogger, brainblogger.com/
Mind Hacks, mindhacks.com/
PsyBlog, www.spring.org.uk

Psychologues cliniciens

La psychologie est l'étude scientifique du comportement et des processus mentaux. Mais c'est aussi l'application de la recherche scientifique pour « aider les gens, les organisations et les communautés à mieux fonctionner » (American Psychological Association, 2011). De loin, l'application la plus courante et la plus connue est la **pratique de la psychologie clinique** — le diagnostic et le traitement des troubles psychologiques et des problèmes connexes. Utilisons le terme de *pratique clinique* au sens large pour désigner les activités des personnes travaillant en psychologie clinique, psychologie scolaire, thérapie conjugale et familiale, travail social clinique, service conseil, et d'autres personnes professionnelles travaillant avec des personnes individuellement ou en petits groupes pour identifier et aider à résoudre leurs problèmes psychologiques. Il est important de considérer la relation entre la recherche scientifique et la pratique clinique, car de nombreuses personnes aux études s'intéressent particulièrement à la pratique clinique, peut-être même comme carrière.

Le point principal est que les troubles psychologiques et autres problèmes comportementaux font partie du monde naturel. Cela signifie que les questions relatives à leur nature, à leurs causes et à leurs conséquences peuvent être testées empiriquement et donc faire l'objet d'une étude scientifique. Comme pour d'autres questions portant sur le comportement humain, nous ne pouvons pas nous fier à notre intuition ou à notre bon sens pour obtenir des réponses détaillées et exactes. Considérons par exemple, que des

dizaines de livres populaires et des milliers de sites Internet affirment que les adultes ayant grandi avec un parent alcoolique ont un profil de personnalité distinct, notamment une faible estime de soi, un sentiment d'impuissance et des difficultés avec l'intimité. Bien que cela semble plausible, la recherche scientifique a démontré que les adultes ayant grandi avec un parent alcoolique ne sont pas plus susceptibles que quiconque d'avoir ces problèmes (Lilienfeld et al., 2010). De même les questions portant sur l'efficacité d'une psychothérapie particulière sont des questions testables empiriquement qui peuvent recevoir une réponse grâce à la recherche scientifique. Si une nouvelle psychothérapie est un traitement efficace pour la dépression, alors l'observation systématique devrait révéler que les personnes déprimées qui reçoivent cette psychothérapie s'améliorent davantage qu'un groupe similaire de personnes déprimées qui ne reçoivent pas cette psychothérapie (ou qui reçoivent un traitement alternatif). Les traitements dont l'efficacité a été démontrée de cette manière sont appelés **traitements validés empiriquement**.

Traitements validés empiriquement

Un traitement validé empiriquement est un traitement qui a été étudié scientifiquement et qui a démontré qu'il entraînait une amélioration plus grande qu'aucun traitement, qu'un placebo ou qu'un traitement alternatif. Ceux-ci comprennent de nombreuses formes de psychothérapie, qui peuvent être aussi efficaces que les traitements médicamenteux standards. Parmi les formes de psychothérapie bénéficiant d'un soutien empirique solide, on peut citer les suivantes :

- **Thérapie d'acceptation et d'engagement** (de l'anglais *Acceptance and Commitment Therapy*, ACT). Pour la dépression, les troubles anxieux mixtes, la psychose, la douleur chronique et les troubles obsessionnels compulsifs.
- **Thérapie de couple comportementale**. Pour les troubles liés à la consommation d'alcool.
- **La thérapie cognitivo-comportementale** (TCC). Pour de nombreux troubles, notamment les troubles alimentaires, la dépression, les troubles anxieux, etc.
- **Thérapie d'exposition**. Pour le syndrome de stress post-traumatique et les phobies.
- **Thérapie d'exposition avec prévention de la réponse**. Pour les troubles obsessionnels compulsifs.
- **Thérapie familiale**. Pour les troubles de l'alimentation.

Pour une liste plus complète, voir le site web suivant, géré par la division 12 de l'American Psychological Association, la Society for Clinical Psychology : www.div12.org/psychological-treatments

Beaucoup, dans la communauté de la psychologie clinique, ont affirmé que leur domaine n'avait pas accordé suffisamment d'attention à la recherche scientifique — par exemple en négligeant d'utiliser les traitements validés empiriquement — et ont suggéré divers changements dans la formation en psychologie clinique, dans l'évaluation et la mise en pratique des traitements. D'autres pensent que ces affirmations sont exagérées et que les changements suggérés sont inutiles (Norcross, Beutler et Levant, 2005). Les deux camps s'accordent cependant sur le fait qu'une approche scientifique de la psychologie clinique est essentielle,

si l'objectif est de diagnostiquer et de traiter les problèmes psychologiques en s'appuyant sur des connaissances détaillées et exactes concernant ces problèmes, et les traitements les plus efficaces pour les résoudre. Il est donc non seulement important que la recherche scientifique en psychologie clinique se poursuive, mais il est aussi important que les personnes en clinique, qui pourraient ne jamais mener d'étude scientifique elles-mêmes, soient scientifiquement lettrées, afin de pouvoir lire et évaluer de nouvelles recherches, et prendre des décisions thérapeutiques fondées sur les meilleures données probantes disponibles. C'est ce qu'on appelle une pratique fondée sur les données probantes.

6. Points clés et exercices

Points clés

- Les connaissances s'acquièrent de différentes manières, notamment par l'intuition, l'autorité, le rationalisme, l'empirisme et la méthode scientifique.
- La science est une manière générale de comprendre le monde naturel. Ses trois caractéristiques fondamentales sont l'empirisme systématique, les questions empiriques et les connaissances publiques.
- La psychologie est une science parce qu'elle adopte l'approche scientifique pour comprendre le comportement humain.
- Les pseudosciences renvoient à des croyances et des activités présentées comme scientifiques, mais qui ne possèdent pas une ou plusieurs des trois caractéristiques de la science. Il est important de distinguer l'approche scientifique visant à comprendre le comportement humain des nombreuses approches pseudoscientifiques.
- Les psychologues mènent des recherches afin de décrire des phénomènes de base, de faire des prédictions sur les comportements futurs et d'expliquer les causes du comportement.
- La recherche fondamentale vise à mieux comprendre le comportement humain en tant que tel, tandis que la recherche appliquée vise à résoudre un problème pratique. Les deux sont utiles et la distinction entre les deux n'est pas toujours évidente.
- Les intuitions des gens au sujet du comportement humain, également appelées psychologie populaire, se révèlent souvent erronées. C'est l'une des principales raisons pour lesquelles la psychologie repose sur la science plutôt que sur le bon sens.
- Les personnes menant des recherches en psychologie cultivent certaines attitudes de la pensée critique. L'une d'entre elles est le scepticisme. Ces personnes cherchent des preuves et considèrent des alternatives avant d'accepter une affirmation sur le comportement humain comme vraie. Une autre est la tolérance à l'incertitude. Ces personnes suspendent leur jugement lorsqu'il n'y a pas suffisamment de données pour déterminer si une affirmation est vraie ou fausse.
- La recherche scientifique en psychologie est principalement menée par des personnes titulaires d'un doctorat en psychologie ou dans un domaine connexe, dont la plupart sont des membres du corps professoral de collèges et d'universités. Ces personnes le font pour des raisons professionnelles et personnelles, ainsi que pour contribuer aux connaissances scientifiques portant sur le comportement humain. La plupart des psychologues sont des psychologues expérimentaux qui mènent des recherches.
- La pratique clinique de la psychologie — le diagnostic et le traitement des problèmes psychologiques — est une application importante de la discipline scientifique qu'est la psychologie.

- La recherche scientifique est essentielle à la pratique clinique, parce qu'elle fournit des connaissances détaillées et exactes sur les problèmes psychologiques, et permet de déterminer si les traitements sont efficaces.

Exercices

- **Pratique** : Considérez trois choses que vous savez et déterminez comment vous avez acquis ces connaissances (autorité, intuition, rationalisme, empirisme, méthode scientifique).
- **Pratique** : Essayez de générer différentes questions de recherche pour décrire, prédire et expliquer un phénomène qui vous intéresse.
- **Pratique** : En utilisant votre propre expérience ou ce que vous avez déjà appris en psychologie, énumérez trois questions de recherche fondamentale et trois questions de recherche appliquée qui vous intéressent.
- **Pratique** : Citez trois questions empiriques portant sur le comportement humain. Citez trois questions non empiriques portant sur le comportement humain.
- **Pratique** : Pour chacune des croyances intuitives suivantes concernant le comportement humain, énoncez trois raisons pour lesquelles elle pourrait être vraie et trois raisons pour lesquelles elle pourrait ne pas l'être :
 - On ne peut pas vraiment aimer une autre personne si on ne s'aime pas soi-même.
 - Les personnes qui bénéficient d'un « conseil de crise » immédiatement après avoir vécu un événement traumatisant sont plus en mesure d'y faire face à long terme.
 - L'étude est plus efficace lorsqu'elle se fait toujours au même endroit.
- Regardez la vidéo suivante, dans laquelle le psychologue Scott Lilienfeld parle du biais de confirmation, de la vision étroite et de l'utilisation de preuves pour évaluer le monde qui nous entoure :
[Scott Lilienfeld-biais de confirmation](#)
- **Discussion** : Considérez l'affirmation psychologique suivante. « Le choix d'un.e conjoint.e est fortement influencé par la perception qu'ont les gens de leurs propres parents. Certains choisissent un.e conjoint.e qui ressemble d'une certaine manière à l'un de leurs parents. D'autres choisissent un.e conjoint.e différent.e de l'un de leurs parents ». Cette affirmation est-elle falsifiable? Pourquoi ou pourquoi pas?
- **Discussion** : Les gens suggèrent parfois que la psychologie ne peut pas être une science parce que (a) le comportement humain ne peut pas être prédit avec une précision parfaite ou (b) une grande partie de son sujet (par exemple, les pensées et les sentiments) ne peut pas être observée directement. Êtes-vous d'accord ou non avec chacune de ces idées? Pourquoi?



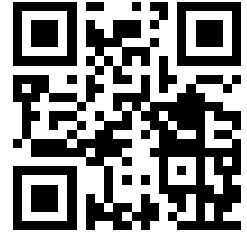
*Lecture sur version imprimée?
Scannez ce code QR avec votre
téléphone cellulaire*

- Visionner la vidéo suivante, réalisée par PHD Comics, qui donne un aperçu de la publication en libre accès et de son importance: [Open Access Explained](#)

- **Discussion :** Des psychologues cliniciens affirment pratiquer une « forme d'art » fondé sur l'intuition et l'expérience personnelle, et que leur pratique ne peut donc pas être évaluée scientifiquement. Rédigez un paragraphe sur le degré de satisfaction que vous éprouveriez à l'égard d'une telle personne clinicienne et expliquez pourquoi, selon chacun des trois points de vue suivants:

- une cliente potentielle du clinicien
- une juge qui doit décider si elle autorise le clinicien à témoigner en tant qu'expert dans une affaire de maltraitance d'enfant
- un représentant de la compagnie d'assurance qui doit décider de rembourser ou non la clinicienne pour ses services

- **Pratique:** Créez une courte liste de questions que le client pourrait poser à une clinicienne pour déterminer si cette dernière accorde suffisamment d'attention à la recherche scientifique.



*Lecture sur version imprimée?
Scannez ce code QR avec votre
téléphone cellulaire*

Attributions des vidéos

- ESC 2017 – Scott Lilienfeld : "Tunnel Vision : Confirmation Bias" © ESC European Skeptics Congress. est placé sous licence Creative Commons (réutilisation autorisée)
- "Open Access Explained !" par Piled Higher and Deeper (PHD Comics). est placé sous licence Creative Commons (réutilisation autorisée)

7. Références

- American Psychology Association (2020). *About APA*. Extrait de <https://www.apa.org/about>
- Bushman, B. J. (2002). L'évacuation de la colère alimente-t-elle ou éteint-elle la flamme? Catharsis, rumination, distraction, colère et réponse agressive. *Personality and Social Psychology Bulletin*, 28, 724-731.
- Gilovich, T. (1991). *How we know what isn't so : La faillibilité de la raison humaine dans la vie quotidienne*. New York, NY: Free Press.
- Gladwell, M. E. (2005). *Blink: The power of thinking without thinking*. (9e éd.). New York: Little, Brown & Co.
- Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports*, 83, 19-64.
- Kassin, S. M. et Gudjonsson, G. H. (2004). The psychology of confession evidence: A review of the literature and Issues. *Psychological Science in the Public Interest*, 5, 33-67. <https://doi.org/10.1111/j.1529-1006.2004.00016.x>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, S., Bernstein, M. J., . . . Nosek, B. A. (2013). Investigating variation in replicability : A "many labs" replication project. *Social Psychology*, 45(3), 142-152. <https://doi.org/10.1027/1864-9335/a000178>
- Lilienfeld, S. O. (2011). Les 10 commandements pour aider les étudiants à distinguer la science de la pseudoscience en psychologie. *Revue Électronique de Psychologie Sociale*, 5, 23-29.
- Lilienfeld, S. O., Lynn, S. J., Ruscio, J. et Beyerstein, B. L. (2010). *50 great myths of popular psychology* (50 grands mythes de la psychologie populaire). Malden, MA : Wiley-Blackwell.
- Mann, T., Tomiyama, A. J., Westling, E., Lew, A., Samuels, B. et Chatman, J. (2007). Medicare's search for effective obesity treatments : Diets are not the answer. *American Psychologist*, 62, 220-233. <https://doi.org/10.1037/0003-066x.62.3.220>
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82. <https://psycnet.apa.org/doi/10.1126/science.1139940>
- Norcross, J. C., Beutler, L. E. et Levant, R. F. (dir.). (2005). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC : American Psychological Association.
- Popper, K. R. (2002). *Conjectures et réfutations : La croissance de la connaissance scientifique*. New York, NY : Routledge.
- Schnall, S., Benton, J. et Harvey, S. (2008). With a clean conscience : La propreté réduit la sévérité des jugements moraux. *Psychological Science*, 19(12), 1219-1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>
- Sexton, M., Cuttler, C., Finnell, J. et Mischley, L (2016). Une enquête transversale sur les utilisateurs de

cannabis médical : Patterns of use and perceived efficacy (Modes d'utilisation et efficacité perçue). *Cannabis and Cannabinoid Research*, 1, 131- 138. <https://doi.org/10.1089/can.2016.0007>

Stanovich, K. E. (2010). *How to think straight about psychology* (9e éd.). Boston, MA : Allyn & Bacon.

CHAPITRE 2-UNE VUE D'ENSEMBLE DE LA MÉTHODE SCIENTIFIQUE



Image 2. Personne écrivant à la main

Voici le résumé et la traduction d'un article paru en 2014 dans la revue *Psychological Science*.

Il est de plus en plus courant de prendre des notes sur un ordinateur portable plutôt qu'à la main. Plusieurs personnes en recherche ont suggéré que la prise de notes sur ordinateur portable est moins efficace que la prise de notes à la main pour l'apprentissage. Les études antérieures se sont principalement concentrées sur la capacité des personnes aux études à effectuer plusieurs tâches à la fois, et à se laisser distraire lors de l'utilisation d'ordinateurs portables. La présente étude suggère que, même lorsque les ordinateurs portables sont utilisés uniquement pour prendre des notes, ils peuvent tout de même nuire à l'apprentissage parce que leur utilisation entraîne un traitement plus superficiel de l'information. Dans trois études, nous avons constaté que les personnes aux études prenant des notes sur des ordinateurs portables obtenaient de moins bons résultats aux questions conceptuelles que celles prenant des notes à la main. Nous montrons que, bien que prendre davantage de notes pourrait être bénéfique, la tendance des personnes prenant des notes sur un

ordinateur portable à transcrire les cours mot pour mot, plutôt qu'à traiter l'information et à la reformuler avec leurs propres mots, nuit à l'apprentissage (Mueler et Oppenheimer, 2014, p. 1159).

Dans ce résumé, cette équipe de recherche a identifié une question de recherche — portant sur l'effet de la prise de notes avec un ordinateur portable sur l'apprentissage — et a spécifié pourquoi elle mérite d'être étudiée — parce que cette pratique est omniprésente et qu'elle peut nuire à l'apprentissage. Dans ce chapitre, nous vous donnons un aperçu des différentes étapes du processus de recherche. Il s'agit notamment de trouver un sujet de recherche, d'analyser la littérature de recherche, d'affiner votre question de recherche et de formuler une hypothèse, de concevoir et de mener une étude, d'analyser les données, de tirer des conclusions et de présenter les résultats.

11. Élaborer une hypothèse

Objectifs d'apprentissage

1. Faire la distinction entre une théorie et une hypothèse.
2. Découvrir comment les théories sont utilisées pour générer des hypothèses et comment les résultats des études peuvent être utilisés pour éclairer davantage les théories.
3. Comprendre les caractéristiques d'une bonne hypothèse.

Théories et hypothèses

Avant de décrire comment élaborer une hypothèse, il est important de faire la distinction entre une **théorie** et une hypothèse. Une théorie est une explication ou une interprétation cohérente d'un ou de plusieurs phénomènes. Bien que les théories puissent prendre des formes diverses, elles ont en commun d'aller au-delà des phénomènes qu'elles expliquent en incluant des variables, des structures, des processus, des fonctions ou des principes d'organisation qui n'ont pas été observés directement. Prenons l'exemple de la théorie de la facilitation et de l'inhibition sociales de Zajonc (1965). Il a proposé que le fait d'être observé par d'autres personnes pendant l'exécution d'une tâche crée un état général d'excitation physiologique, qui augmente la probabilité de la réponse dominante (la plus probable). Ainsi pour les tâches très pratiquées, le fait d'être observé augmente la tendance à donner des réponses correctes, mais pour les tâches relativement peu pratiquées, le fait d'être observé augmente la tendance à donner des réponses incorrectes. Remarquons que cette théorie — que l'on appelle désormais *la théorie de la motivation* — fournit une explication de la facilitation et de l'inhibition sociales qui va au-delà des phénomènes eux-mêmes en incluant des concepts tels que « l'excitation » et la « réponse dominante », ainsi que des processus tels que l'effet de l'excitation sur la réponse dominante.

En dehors de la science, qualifier une idée de théorie implique souvent qu'elle n'a pas été testée — qu'elle n'est peut-être rien de plus qu'une simple supposition. En science cependant, le terme théorie n'a pas cette implication. Une théorie est simplement une explication ou une interprétation d'un ensemble de phénomènes. Elle peut ne pas avoir été testée, mais elle peut aussi avoir fait l'objet de nombreux tests approfondis, être bien étayée et être acceptée par la communauté scientifique comme une description exacte du monde. La théorie de l'évolution par sélection naturelle, par exemple, est une théorie parce qu'elle explique la diversité de la vie sur terre — et non pas parce qu'elle n'a pas été testée ou soutenue par la recherche scientifique. Au contraire, les preuves de cette théorie sont extrêmement positives et la quasi-totalité des scientifiques acceptent ses postulats de base comme étant exacts. De même, la « théorie des germes » des maladies est une théorie parce qu'elle explique l'origine de diverses maladies, et non parce

qu'il existe un doute sur le fait que de nombreuses maladies sont causées par des micro-organismes qui infectent le corps.

Une **hypothèse** en revanche est une prédiction spécifique à propos d'un nouveau phénomène qui devrait être observé si une théorie particulière est exacte. Il s'agit d'une explication qui repose sur seulement quelques concepts clés. Les hypothèses sont souvent des prédictions spécifiques sur ce qui se passera dans une étude particulière. Elles sont élaborées en tenant compte des données existantes, et en utilisant le raisonnement pour déduire ce qui se passe dans le contexte spécifique qui nous intéresse. Les hypothèses sont souvent, mais pas toujours, dérivées de théories. Ainsi une hypothèse est souvent une prédiction faite à partir d'une théorie mais certaines hypothèses sont a-théoriques, et ce n'est qu'après un ensemble d'observations qu'une théorie est élaborée. Cela s'explique par le fait que les théories sont de nature générale et qu'elles expliquent des corpus de données plus importants. Ainsi, si notre question de recherche est vraiment originale, il se peut que nous devions recueillir des données et faire des observations avant de pouvoir élaborer une théorie plus générale.

Les théories et les hypothèses ont toujours cette relation « si-alors ». « *Si* la théorie de la motivation est correcte, *alors* les blattes devraient courir plus vite dans un couloir droit, et plus lentement dans un couloir à embranchements, en présence d'autres blattes ». Bien que les hypothèses soient généralement exprimées sous forme d'affirmations, elles peuvent toujours être reformulées sous forme de questions. « Les blattes courent-elles plus vite sur une piste droite en présence d'autres blattes? » Ainsi, dériver des hypothèses à partir de théories est un excellent moyen de générer des questions de recherche intéressantes.

Mais comment les personnes en recherche peuvent-elles dériver des hypothèses à partir de théories? L'une des méthodes consiste à formuler une question de recherche à l'aide des techniques présentées dans ce chapitre, puis à se demander si une théorie donne implicitement une réponse à cette question. Par exemple vous pourriez vous demander si l'écriture expressive sur des expériences positives améliore la santé autant que l'écriture expressive sur des expériences traumatiques. Bien que cette question soit intéressante en soi, vous pourriez alors vous demander si la théorie de l'habituation — l'idée selon laquelle l'écriture expressive amène les gens à s'habituer aux pensées et aux sentiments négatifs — implique une réponse. Dans ce cas, il semble évident que, si la théorie de l'accoutumance est correcte, l'écriture expressive sur des expériences positives ne devrait pas être efficace, car elle n'amènerait pas les personnes à s'habituer à des pensées et des sentiments négatifs. Une deuxième façon de dériver des hypothèses à partir de théories consiste à se concentrer sur un élément de la théorie qui n'a pas encore été directement observé. Par exemple une personne en recherche pourrait se concentrer sur le processus d'habituation — en émettant l'hypothèse que les personnes devraient montrer moins de signes de détresse émotionnelle à chaque nouvelle séance d'écriture.

Parmi les meilleures hypothèses figurent celles qui permettent de départager des théories concurrentes. Par exemple, Norbert Schwarz et ses collègues ont étudié deux théories sur la façon dont les gens portent des jugements sur eux-mêmes, comme la manière dont ils se perçoivent en matière d'assertivité (Schwarz, Bless, Strack, Klumpp, Rittenaurer-Schatka et Simons, 1991). Selon les deux théories, ces jugements se fondent sur des exemples pertinents que les gens évoquent. Cependant selon une théorie, les gens fondent leurs jugements sur le nombre d'exemples évoqués tandis que, selon l'autre, ils fondent leurs jugements sur la facilité avec laquelle ces exemples sont évoqués. Pour tester ces théories, les personnes menant la recherche ont demandé à des personnes de se rappeler six fois où elles avaient été assertives (ce qui est facile pour la plupart des gens) ou douze fois (ce qui est difficile pour la plupart des gens). Puis, on leur a demandé de juger leur propre assertivité. Remarquons que la théorie du nombre d'exemples implique que

les personnes qui se souviennent de douze exemples devraient se juger plus assertives parce qu'elles se souviennent de plus d'exemples, alors que la théorie de la facilité d'évocation implique que les personnes qui se souviennent de six exemples devraient se juger plus assertives parce que l'évocation des exemples était plus facile. Les deux théories ont donc fait des prédictions opposées, de sorte qu'une seule d'entre elles a pu être confirmée. Le résultat surprenant a été que les personnes qui se souvenaient de moins d'exemples se jugeaient plus sûres d'elles, plus assertives — ce qui constitue une preuve particulièrement convaincante en faveur de la théorie de la facilité de récupération par rapport à la théorie du nombre d'exemples.

Test de la théorie

La principale façon dont les personnes en recherche scientifique utilisent les théories est parfois appelée la **méthode hypothético-déductive** (bien que ce terme soit beaucoup plus souvent utilisé par les philosophes des sciences que par les scientifiques eux-mêmes ou elles-mêmes). Les scientifiques partent d'un ensemble de phénomènes et élaborent soit une théorie pour les expliquer ou les interpréter, soit choisissent une théorie existante sur laquelle travailler. Ces scientifiques font ensuite une prédiction sur un nouveau phénomène qui devrait être observé si la théorie est correcte. Là encore, cette prédiction est appelée hypothèse. Les personnes en recherche mènent ensuite une étude empirique pour tester l'hypothèse. Enfin, ces personnes réévaluent la théorie à la lumière des nouveaux résultats et la révisent si nécessaire. Ce processus est généralement conceptualisé comme un cycle, car les scientifiques peuvent alors dériver une nouvelle hypothèse à partir de la théorie révisée, mener une nouvelle étude empirique pour tester l'hypothèse, et ainsi de suite. Comme le montre la Figure 2.3, cette approche s'intègre parfaitement au modèle de la recherche scientifique en psychologie présenté plus haut dans le livre — créant ainsi un modèle plus détaillé de recherche « motivée par la théorie » ou « guidée par la théorie ».

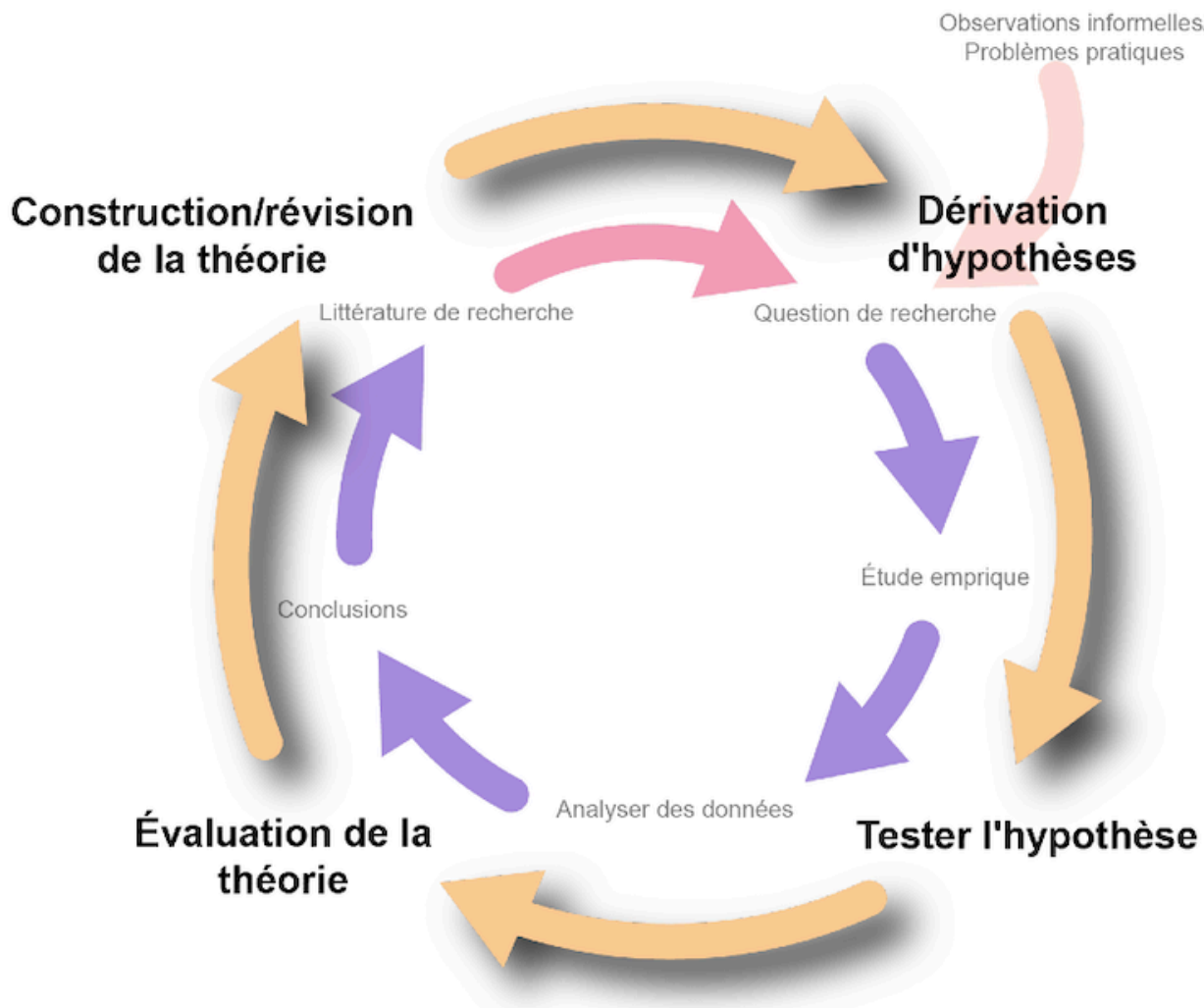


Figure 2.3. La méthode hypothético-déductive combinée au modèle général de la recherche scientifique en psychologie. Ensemble, ils forment un modèle de recherche motivée par la théorie.

Prenons l'exemple des recherches de Zajonc sur la facilitation et l'inhibition sociales. Il est parti d'un ensemble de résultats quelque peu contradictoires provenant de la littérature de recherche. Il a ensuite construit sa théorie de l'activation, selon laquelle le fait d'être observé par d'autres personnes pendant l'exécution d'une tâche provoque une excitation physiologique qui accroît la tendance d'un organisme à adopter la réponse dominante. Cette théorie prédit une facilitation sociale pour les tâches bien apprises, et une inhibition sociale pour les tâches mal apprises. Ce chercheur disposait désormais d'une théorie qui organisait les résultats antérieurs de manière significative — mais il devait encore la tester. Il a émis l'hypothèse que, si sa théorie était correcte, il devrait observer que la présence d'autres personnes améliore les performances dans une tâche de laboratoire simple, mais inhibe les performances dans une version difficile de cette même tâche de laboratoire. Pour tester cette hypothèse, l'une des études qu'il a menées a utilisé des blattes comme sujets (Zajonc, Heingartner et Herman, 1969). Les blattes devaient soit courir dans un couloir droit (une tâche facile pour une blatte), soit traverser un labyrinthe en forme de croix (une tâche difficile pour une blatte) pour atteindre une chambre sombre lorsqu'une lumière était projetée sur

elles. Elles réalisaient ces tâches soit seules, soit en présence d'autres blattes placées dans des « boîtes d'audience » en plastique transparent. Zajonc a constaté que les blattes dans le couloir droit atteignaient leur but plus rapidement en présence d'autres blattes, alors que les blattes dans le labyrinthe en forme de croix atteignaient leur but plus lentement en présence d'autres blattes. Il a ainsi confirmé son hypothèse et apporté un soutien à sa théorie de l'activation. (Zajonc a également montré que la théorie de l'activation existait chez les humains [Zajonc et Sales, 1966] dans de nombreuses autres études par la suite).

Intégrer la théorie dans votre recherche

Lorsque vous rédigez votre rapport de recherche ou préparez votre présentation, sachez qu'il existe deux façons principales pour les personnes en recherche d'inclure la théorie. La première consiste à poser une question de recherche, y répondre en menant une nouvelle étude, puis proposer une ou plusieurs théories (généralement plusieurs) pour expliquer ou interpréter les résultats. Ce format fonctionne bien pour des questions de recherche appliquée et pour des questions de recherche auxquelles les théories existantes n'y répondent pas. La deuxième méthode consiste à décrire une ou plusieurs théories existantes, à dériver une hypothèse de l'une de ces théories, à tester l'hypothèse dans le cadre d'une nouvelle étude, et enfin à réévaluer la théorie. Ce format fonctionne bien lorsqu'il existe une théorie répondant à la question de recherche — en particulier si l'hypothèse qui en découle est surprenante ou contredit une hypothèse dérivée d'une autre théorie.

L'utilisation de théories dans vos recherches vous aidera non seulement à trouver des idées d'expériences et des projets potentiels, mais cela donnera également une légitimité à votre travail. Les psychologues se sont intéressés à une grande variété de comportements humains et ont développé de nombreuses théories au fil du temps. Le recours à des théories établies vous aidera à innover en recherche, sans vous empêcher de développer vos propres idées.

Caractéristiques d'une bonne hypothèse

Une bonne hypothèse présente trois caractéristiques générales. Premièrement une bonne hypothèse doit être **testable et falsifiable**. Nous devons être en mesure de tester l'hypothèse en utilisant les méthodes en science et, pour reprendre le critère de falsifiabilité de Popper, il doit être possible de recueillir des données susceptibles d'infirmer l'hypothèse si elle est effectivement fausse. Deuxièmement une bonne hypothèse doit être **logique**. Comme nous l'avons vu plus haut, les hypothèses ne sont pas de simple intuition proposée au hasard. Les hypothèses doivent s'appuyer sur des théories ou des observations antérieures et sur un raisonnement logique. En règle générale, nous partons d'une théorie générale et vaste, et utilisons un raisonnement déductif pour générer une hypothèse plus spécifique à tester à partir de cette théorie. Toutefois lorsqu'il n'existe pas de théorie pour étayer notre hypothèse, nous recourons parfois au raisonnement inductif, qui consiste à partir d'observations ou de résultats de recherche spécifiques, pour formuler une hypothèse plus générale. Enfin l'hypothèse doit être formulée de manière **positive**. En d'autres termes, l'hypothèse doit énoncer l'existence d'une relation ou d'un effet, plutôt que d'indiquer qu'une relation

ou un effet n'existe pas. En tant que scientifiques, nous ne cherchons pas à démontrer que les relations n'existent pas ou que les effets ne se produisent pas, donc nos hypothèses ne doivent pas être formulées de manière à suggérer qu'une relation ou un effet est absent. La nature même de la science consiste à supposer qu'une chose n'existe pas, puis à chercher des preuves pour démontrer le contraire, afin de montrer qu'elle existe vraiment. Cette démarche peut vous sembler contre-intuitive, mais c'est la nature même de la méthode scientifique. La raison sous-jacente dépasse le cadre de ce chapitre, mais elle est liée à la théorie statistique.

14. Tirer des conclusions et présenter les résultats

Objectifs d'apprentissage

1. Identifier les conclusions que les personnes en recherche peuvent tirer des résultats de leurs études.
2. Décrire pourquoi les scientifiques évitent le terme « preuve scientifique ».
3. Expliquer les différentes façons dont les scientifiques partagent leurs découvertes.

Tirer des conclusions

Étant donné que les statistiques sont de nature probabiliste et que les résultats peuvent refléter des erreurs de type I ou de type II, nous ne pouvons pas utiliser les résultats d'une seule étude pour conclure avec certitude qu'une théorie est vraie. Les théories sont plutôt soutenues, réfutées ou modifiées en fonction des résultats de la recherche.

Si les résultats sont statistiquement significatifs et cohérents, avec l'hypothèse et la théorie à partir de laquelle l'hypothèse a été formulée, alors les personnes en recherche peuvent conclure que la théorie est soutenue. Non seulement la théorie a fait une prédiction exacte, mais il y a maintenant un nouveau phénomène que la théorie explique. Si une hypothèse est infirmée dans le cadre d'une étude empirique systématique, alors la théorie est affaiblie. Elle a fait une prédiction inexacte et il y a maintenant un nouveau phénomène qu'elle ne permet pas d'expliquer.

Bien que cela semble simple, il y a quelques complications. Tout d'abord, confirmer une hypothèse peut renforcer une théorie, mais ne peut jamais la prouver. En fait les scientifiques évitent d'utiliser le mot « prouver » quand vient le temps de parler et d'écrire à propos des théories. Une des raisons de cette précaution est que le résultat peut refléter une erreur de type I. Une autre raison est qu'il peut y avoir d'autres théories plausibles qui impliquent la même hypothèse, ce qui signifie que la confirmation de l'hypothèse renforce toutes ces théories de la même manière. Une troisième raison est qu'il est toujours possible qu'un autre test de l'hypothèse, ou un test d'une nouvelle hypothèse dérivée de la théorie soient infirmés. Cette difficulté est une version du célèbre « problème de l'induction » en philosophie. On ne peut pas prouver de manière définitive un principe général (par exemple, « Tous les cygnes sont blancs ») simplement en observant des cas qui le confirment (par exemple, des cygnes blancs) — peu importe leur nombre. Il est toujours possible qu'un cas non confirmé (par exemple, un cygne noir) apparaisse un jour. C'est pourquoi les scientifiques ont

tendance à considérer les théories — même celles qui ont fait leurs preuves — comme susceptibles d'être révisées à la lumière d'observations nouvelles et inattendues.

Une deuxième complication est liée à la signification de la non-confirmation d'une hypothèse. Selon la méthode hypothético-déductive dans sa version la plus stricte, l'infirmité d'une hypothèse réfute la théorie dont elle est issue. En effet, en logique formelle, les prémisses « si A alors B » et « pas B » conduisent nécessairement à la conclusion « pas A ». Si A est la théorie et B l'hypothèse (« si A alors B »), la non-confirmation de l'hypothèse (« pas B ») doit signifier que la théorie est incorrecte (« pas A »). Toutefois en pratique, les scientifiques n'abandonnent pas leurs théories aussi facilement. Une raison est qu'une hypothèse non confirmée peut être une occasion manquée (le résultat d'une erreur de type II) ou le résultat d'un mauvais devis de recherche. La personne en recherche n'a peut-être pas réussi à manipuler correctement la variable indépendante ou à mesurer adéquatement la variable dépendante.

Une hypothèse non confirmée peut également signifier qu'une hypothèse non formulée, mais relativement mineure, de la théorie n'a pas été respectée. Par exemple, si Zajonc n'avait pas trouvé de facilitation sociale chez les blattes, il aurait pu conclure que la théorie de l'activation est toujours correcte, mais qu'elle ne s'applique qu'aux animaux dotés d'un système nerveux suffisamment complexe. Autrement dit, les résultats d'une étude peuvent être utilisés pour modifier une théorie. Cela ne signifie pas que les personnes en recherche sont libres d'ignorer les infirmités de leurs théories. Si ces personnes ne peuvent pas améliorer leurs devis de recherche ou modifier leurs théories pour tenir compte des infirmités répétées, elles doivent finalement abandonner leurs théories et les remplacer par d'autres, plus performantes.

L'essentiel ici est que, puisque les statistiques sont de nature probabiliste et que toutes les études comportent des limites, il n'existe pas de preuve scientifique, mais seulement des indices scientifiques.

Communication des résultats

La dernière étape du processus de recherche consiste à communiquer les résultats. Comme décrit dans la section Analyse documentaire de la recherche de ce chapitre, les résultats sont généralement présentés dans des articles de revues évalués par les pairs, et lors de congrès.

La manière la plus prestigieuse de communiquer ses résultats est de rédiger un article et de le faire publier dans une revue scientifique évaluée par les pairs. Les articles publiés dans les revues américaines ou canadiennes en psychologie doivent généralement respecter le style de rédaction de l'American Psychological Association (style APA). Vous apprendrez probablement les principaux éléments de ce style de rédaction dans ce cours.

Une autre façon de communiquer ses résultats est de rédiger un chapitre de livre publié dans un livre édité. Idéalement, la personne responsable de l'édition du livre soumet le chapitre à un processus d'évaluation par les pairs, mais ce n'est pas toujours le cas, et parfois des scientifiques reçoivent des invitations directement par la personne responsable de l'édition à rédiger des chapitres de livre.

Une manière plus conviviale de diffuser ses résultats consiste à faire une présentation lors d'un congrès. Il peut s'agir d'une communication orale ou d'une communication par affiches. Les communications orales impliquent de prendre la parole devant un public de scientifiques et de faire un exposé qui peut durer de 10 minutes à 1 heure (selon le congrès), suivie d'une période de questions du public. Les communications par affiches consistent à résumer l'étude sur une grande affiche qui donne un aperçu de l'objectif, de la méthode,

des résultats et de la discussion. La personne responsable de la communication reste près de son affiche pendant une heure ou deux et en discute avec les gens qui passent. En fait présenter son travail lors d'un congrès est un excellent moyen d'obtenir une rétroaction de la part de ses pairs, avant de soumettre son travail au processus plus rigoureux d'évaluation par les pairs, nécessaire pour publier un article dans une revue scientifique.

16. Références

- Adair, J. G. et Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. (L'explosion des connaissances, des références et des citations : La réponse unique de la psychologie à une crise.). *American Psychologist*, 58, 15-23. <https://doi.org/10.1037/0003-066X.58.1.15>
- Collet, C., Guillot, A. et Petit, C. (2010). Phoning while driving I: A review of epidemiological, psychological, behavioral and physiological studies. (Téléphoner au volant I : examen des études épidémiologiques, psychologiques, comportementales et physiologiques). *Ergonomics*, 53, 589-601. <https://doi.org/10.1080/00140131003672023>
- Drews, F. A., Pasupathi, M. et Strayer, D. L. (2004). Passenger and cell-phone conversations in simulated driving. (Conversations entre passagers et téléphones portables lors d'une simulation de conduite). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 2210-2212. <https://doi.org/10.1037/a0013119>
- Milgram, S. (1963). Behavioral study of obedience. (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67, 371-378. <https://doi.org/10.1037/h0040525>
- Mueller, P. A., et Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. (La plume est plus puissante que le clavier : les avantages de la prise de notes manuscrite par rapport à la prise de notes sur ordinateur portable.) *Psychological Science*, 25(6), 1159-1168. <https://psycnet.apa.org/doi/10.1177/0956797614524581>
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H. et Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. (Facilité de récupération de l'information : Un autre regard sur l'heuristique de disponibilité.). *Journal of Personality and Social Psychology*, 61(2), 195-202. <https://doi.org/10.1037/0022-3514.61.2.195>
- Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius*. (Créativité : au-delà du mythe du génie). New York, NY : Freeman.
- Zajonc, R. B. (1965). Social facilitation: A solution is suggested for an old unresolved social psychological problem. (Facilitation sociale : Une solution est proposée pour un vieux problème socio-psychologique non résolu.). *Science*, 149, 269-274. <https://doi.org/10.1126/science.149.3681.269>
- Zajonc, R. B., Heingartner, A. et Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. (Renforcement social et altération des performances chez le cafard). *Journal of Personality and Social Psychology*, 13(2), 83-92. <https://doi.org/10.1037/h0028063>
- Zajonc, R. B. et Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. (Facilitation sociale des réponses dominantes et subordonnées). *Journal of Experimental Social Psychology*, 2(2), 160-168. [https://doi.org/10.1016/0022-1031\(66\)90077-1](https://doi.org/10.1016/0022-1031(66)90077-1)

13. Analyser les données

Objectifs d'apprentissage

1. Distinguer les statistiques descriptives des statistiques inférentielles.
2. Identifier les différents types de statistiques descriptives que les personnes en recherche utilisent pour faire un sommaire de leurs données.
3. Décrire l'objectif des statistiques inférentielles.
4. Distinguer les erreurs de type I et de type II.

Une fois que l'étude est terminée et que les observations ont été faites et enregistrées, les personnes en recherche doivent analyser les données et tirer leurs conclusions. En règle générale, les données sont analysées à l'aide de statistiques descriptives et inférentielles. Les statistiques descriptives sont utilisées pour résumer les données et les statistiques inférentielles sont utilisées pour généraliser les résultats de l'échantillon à la population. En d'autres termes, les statistiques inférentielles sont utilisées pour tirer des conclusions quant à savoir si une théorie a été soutenue ou réfutée, ou qu'elle nécessite une modification.

Statistiques descriptives

Les **statistiques descriptives** sont utilisées pour organiser ou résumer un ensemble de données. Par exemple il y a les pourcentages, les mesures de la tendance centrale (moyenne, médiane, mode), les mesures de la dispersion (étendue, écart-type, variance) et les coefficients de corrélation.

Les mesures de la tendance centrale sont utilisées pour décrire le mode, la moyenne et le centre d'une distribution de scores. Le **mode** est le score le plus fréquent dans une distribution. La **médiane** est le point central d'une distribution de scores. La **moyenne** est la moyenne d'une distribution de scores.

Les mesures de dispersion sont également considérées comme des statistiques descriptives. Elles sont utilisées pour décrire le degré de dispersion d'un ensemble de scores. En d'autres mots, les scores sont-ils tous similaires et regroupés autour de la moyenne ou bien existe-t-il une grande variabilité dans les scores? L'**étendue** est une mesure de dispersion qui mesure la distance entre le score le plus élevé et le score le plus bas d'une distribution. L'**écart-type** est une mesure de dispersion plus sophistiquée qui mesure la distance moyenne des scores par rapport à la moyenne. La **variance** est simplement l'écart-type au carré. Donc elle mesure également la distance entre les scores et la moyenne, mais dans une unité de mesure différente.

En règle générale, les moyennes et les écarts-types sont calculés pour les études en recherche

expérimentale dans lesquelles une variable indépendante a été manipulée pour former deux groupes ou plus, et une variable dépendante a été mesurée quantitativement. Les moyennes de chaque groupe expérimental ou condition sont calculées séparément, puis comparées afin de voir si elles diffèrent.

Pour la recherche non expérimentale, de simples pourcentages peuvent être calculés afin de décrire le pourcentage de personnes ayant un certain comportement ou une certaine croyance. Mais le plus souvent, la recherche non expérimentale implique le calcul de la corrélation entre deux variables. Un **coefficient de corrélation** décrit la force et la direction de la relation entre deux variables. Les valeurs d'un coefficient de corrélation peuvent varier de -1,00 (la plus forte relation négative possible) à +1,00 (la plus forte relation positive possible). Une valeur de 0 signifie qu'il n'existe aucune relation entre les deux variables. Les coefficients de corrélation positifs indiquent que plus les valeurs d'une variable augmentent, plus les valeurs de l'autre variable augmentent également. Un bon exemple de corrélation positive est la corrélation entre la taille et le poids, car lorsque la taille augmente, le poids tend également à augmenter. Les coefficients de corrélation négatifs indiquent que, lorsque la valeur d'une variable augmente, les valeurs de l'autre variable diminuent. Un exemple de corrélation négative est la corrélation entre les événements stressants de la vie et le bonheur, car lorsque le stress augmente, le bonheur est susceptible de diminuer.

Statistiques inférentielles

Comme vous l'avez appris dans la section de ce chapitre portant sur l'échantillonnage, les personnes en recherche utilisent généralement un échantillon d'une population, mais ces personnes souhaitent ultimement pouvoir généraliser leurs résultats de l'échantillon à une population plus large. Les personnes en recherche veulent généralement déduire ce qu'est la population à partir de l'échantillon étudié. Les **statistiques inférentielles** sont utilisées à cette fin. Les statistiques inférentielles permettent aux personnes en recherche de tirer des conclusions sur une population à partir des données d'un échantillon. Les statistiques inférentielles sont cruciales, car les effets (c'est-à-dire les différences entre les moyennes ou le coefficient de corrélation) que les personnes en recherche trouvent dans une étude peuvent être simplement dus à une variabilité aléatoire, ou ils peuvent être dus à un effet réel (c'est-à-dire qu'ils peuvent refléter une relation réelle entre les variables ou un effet réel d'une variable indépendante sur une variable dépendante).

Les personnes en recherche utilisent les statistiques inférentielles pour déterminer si leurs effets sont statistiquement significatifs. Un effet **statistiquement significatif** est un effet qui a peu de chances d'être dû au hasard et qui représente vraisemblablement un effet réel dans la population. Plus précisément, les résultats qui ont moins de 5% de chances d'être attribuables à une erreur aléatoire sont généralement considérés comme statistiquement significatifs. Lorsqu'un effet est statistiquement significatif, il est approprié de généraliser les résultats de l'échantillon à la population. En revanche, si les statistiques inférentielles révèlent qu'il y a plus de 5% de chances qu'un effet soit dû uniquement au hasard, alors il faut conclure que le résultat n'est pas statistiquement significatif.

Il est important de garder à l'esprit que les statistiques sont de nature probabiliste. Elles permettent aux personnes en recherche de déterminer si la probabilité que leurs résultats soient dus au hasard est faible, mais elles n'offrent aucune certitude absolue. Il faut espérer que, lorsque nous concluons qu'un effet est statistiquement significatif, il s'agit d'un effet réel que nous trouverions si nous testions l'ensemble de la population. Et inversement, nous espérons que, lorsque nous concluons qu'un effet n'est pas statistiquement

significatif, c'est qu'il n'y a pas d'effet. Si nous testions l'ensemble de la population, nous ne trouverions pas d'effet. Le seuil de 5 % est fixé à 5 % pour garantir une forte probabilité que nous prenions une décision correcte et que notre conclusion sur la signification statistique reflète fidèlement la réalité.

Cependant des erreurs restent toujours possibles. Plus précisément, deux types d'erreurs peuvent être commises. Premièrement, les personnes en recherche peuvent commettre une **erreur de type I**, qui correspond à un faux positif. Cela se produit lorsqu'une personne en recherche conclut que les résultats sont statistiquement significatifs (on affirme donc qu'il y a un effet dans la population) alors qu'en réalité, il n'y a pas d'effet réel dans la population et les résultats sont dus au hasard (il s'agit d'un coup de chance). Lorsque le seuil est fixé à 5%, ce qui est la convention, la personne en recherche a 5% de chances ou moins de commettre une erreur de type I. Vous vous demandez peut-être pourquoi les personnes en recherche ne fixent pas un seuil encore plus bas pour réduire les risques d'erreur de type I. La raison est que lorsque le risque d'erreur de type I est réduit, le risque d'**erreur de type II** est augmenté. Une erreur de type II correspond à un faux négatif, c'est-à-dire une occasion manquée. Cela se produit lorsqu'une personne en recherche conclut que les résultats ne sont pas statistiquement significatifs alors qu'en réalité, il existe un effet réel dans la population et qu'il n'a pas été détecté. Une fois encore, ces erreurs de type II sont plus susceptibles de se produire lorsque le seuil est trop bas (par exemple, fixé à 1 % au lieu de 5 %) et/ou lorsque l'échantillon est trop petit.

15. Points clés et exercices

Points clés

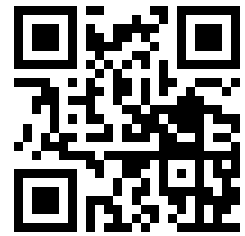
- La recherche en psychologie peut être décrite par un modèle cyclique simple. Une question de recherche fondée sur la littérature de recherche, conduit à une étude empirique, dont les résultats sont publiés et deviennent à leur tour partie intégrante de la littérature de recherche.
- La littérature de recherche en psychologie comprend l'ensemble des travaux de recherche publiés dans le domaine de la psychologie, principalement des articles parus dans des revues scientifiques et des livres savants.
- Au début du processus de recherche, il est important de procéder à une analyse de la littérature de recherche sur votre sujet afin d'affiner votre question de recherche, d'identifier les méthodes appropriées, de situer votre question dans le contexte des recherches déjà existantes et de vous préparer à rédiger un bon rapport de recherche.
- Il existe plusieurs stratégies pour trouver des recherches antérieures portant sur votre sujet. L'une des meilleures consiste à utiliser PsycINFO, une base de données qui répertorie des millions d'articles, de livres et de chapitres de livres en psychologie et dans des domaines connexes.
- Les questions de recherche, exprimées en termes de variables et de relations entre les variables, peuvent être suggérées par d'autres personnes en recherche ou bien générées en posant plusieurs questions générales à propos du comportement ou de la caractéristique psychologique d'intérêt.
- Avant de concevoir une étude et de collecter des données pour répondre à une question de recherche, il est important d'évaluer l'intérêt de cette question. Les facteurs qui influencent l'intérêt sont le degré d'incertitude de la réponse, la mesure avec laquelle elle comble une lacune dans la littérature de recherche et l'existence d'implications pratiques importantes.
- Il est également important d'évaluer la faisabilité de répondre à une question de recherche. Les facteurs qui influencent la faisabilité sont le temps, l'argent, les connaissances et les compétences techniques, ainsi que l'accès à des équipements spécialisés et à des personnes qui participent à la recherche.
- Une théorie est de nature générale et explique un ensemble plus vastes de données. Une hypothèse est plus spécifique et fait des prédictions quant aux résultats d'une étude particulière.
- Travailler avec des théories n'est pas une « cerise sur le gâteau ». C'est un ingrédient fondamental de la recherche en psychologie.
- Comme les autres scientifiques, les psychologues utilisent la méthode hypothético-déductive. Ces personnes élaborent des théories pour expliquer ou interpréter des phénomènes (ou travaillent avec des

théories existantes), en dérivent des hypothèses, testent ces hypothèses, puis réévaluent les théories à la lumière des nouveaux résultats.

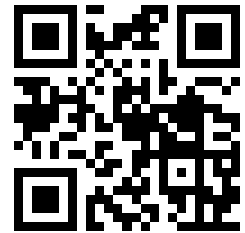
- Les variables varient d'une personne ou d'une situation à l'autre et peuvent être quantitatives (par exemple l'âge) ou catégorielles (par exemple la matière d'un cours).
- Un échantillon est un petit sous-ensemble d'une population plus large qui est sélectionné pour participer à l'étude. Il existe de nombreuses façons d'échantillonner les personnes participantes, notamment l'échantillonnage de commodité et l'échantillonnage aléatoire simple.
- La recherche expérimentale consiste à manipuler une variable indépendante pour en observer les effets sur une variable dépendante mesurée, tandis que la recherche non expérimentale consiste à mesurer les variables telles qu'elles se présentent naturellement (c'est-à-dire sans manipulation).
- La recherche peut être menée sur le terrain ou en laboratoire. Les expériences en laboratoire ont tendance à avoir une grande validité interne (ce qui nous permet de tirer des conclusions causales solides), tandis que les études sur le terrain ont souvent une plus grande validité externe (ce qui nous permet de généraliser aux situations du monde réel).
- La moyenne, la médiane et le mode sont des mesures de la tendance centrale utilisées pour décrire les scores typiques, moyens ou centraux dans une distribution. L'étendue, l'écart-type et la variance sont des mesures de la dispersion des scores. Ces mesures de tendance centrale et de dispersion sont des statistiques descriptives importantes.
- Les statistiques inférentielles permettent aux personnes en recherche de déterminer si leurs résultats sont statistiquement significatifs, c'est-à-dire s'il est peu probable qu'ils soient uniquement dus au hasard et s'ils sont donc susceptibles de représenter un effet réel dans la population.
- Puisque les statistiques sont de nature probabiliste, nous ne savons jamais avec certitude si nos conclusions sont correctes. Nous pouvons commettre des erreurs de type I (conclure qu'un effet est réel alors qu'il ne l'est pas) ou des erreurs de type II (conclure qu'il n'y a pas d'effet alors qu'il y a un effet réel dans la population).
- Les théories peuvent être soutenues sans pour autant être prouvées. De même, le fait qu'il y ait non-confirmation d'une hypothèse ne signifie pas nécessairement que la théorie ait été réfutée.
- La dernière étape du processus de recherche consiste à présenter les résultats lors de congrès, dans des articles de revues scientifiques et/ou dans des livres savants.
-

Exercices

- **Pratique** : Trouvez la description d'une étude empirique dans une revue scientifique ou dans l'un des blogues de psychologie scientifique. Rédigez ensuite une brève description de la recherche en fonction du modèle cyclique présenté ici. Une ou deux phrases pour chaque partie du cycle devraient suffire.
- Visionnez la vidéo [TED Ed](#), dans laquelle David H. Schwartz présente une introduction à deux types d'études empiriques ainsi que certaines méthodes utilisées par les scientifiques pour accroître la fidélité de leurs résultats.
- **Pratique** : Utilisez les techniques présentées dans cette section pour trouver dix (10) articles de revues scientifiques et chapitres de livres sur l'une des idées de recherche suivantes : la mémoire des odeurs, la conduite agressive, les causes du trouble de la personnalité narcissique, les fonctions du sillon intrapariétal ou les préjugés à l'égard des personnes ayant des handicaps physiques.
- Regardez le clip vidéo suivant produit par [l'UBCiSchool](#) sur la façon de lire un article académique (sans perdre la tête)
- **Pratique** : Générez trois idées de recherche basées sur chacun des éléments suivants: observations informelles, problèmes pratiques et sujets abordés dans des numéros récents de revues scientifiques.
- **Pratique** : Formulez une question de recherche empirique sur chacun des comportements ou caractéristiques psychologiques suivants : la course à pied de longue durée, se faire tatouer, l'anxiété sociale, l'intimidation et la mémoire des événements de la petite enfance.
- **Pratique** : Évaluez l'intérêt de chacune des questions de recherche que vous avez générées dans l'exercice 2 sur la base des critères discutés dans cette section.
- **Pratique** : Trouvez le numéro d'une revue qui publie de courts articles empiriques (par exemple, *Psychological Science*, *Psychonomic Bulletin and Review*, *Personality and Social Psychology Bulletin*). Choisissez trois études et évaluez chacune d'entre elles en fonction de la faisabilité pour vous de les reproduire avec les ressources dont vous disposez actuellement. Utilisez l'échelle d'évaluation suivante: (1) Vous pourriez la reproduire essentiellement telle qu'elle a été rapportée. (2) Vous pourriez la reproduire avec quelques simplifications. (3) Vous ne pourriez pas la reproduire. Justifiez votre évaluation.
- **Pratique** : Trouvez un article empirique récent dans une revue scientifique. Lisez son introduction et surlignez en différentes couleurs les descriptions des théories et des hypothèses.
- **Pratique** : En utilisant l'article de recherche que vous avez trouvé dans une revue scientifique, déterminez



Lecture sur version imprimée?
Allez sur <https://youtu.be/GUpdzHJHU8> ou scannez ce code QR avec votre cellulaire



Lecture sur version imprimée?
Allez sur https://youtu.be/SKxm2HF_-k0 ou scannez ce code QR avec votre cellulaire

s'il s'agit d'une étude expérimentale ou non expérimentale. Si c'est une étude expérimentale, identifiez les variables indépendantes et dépendantes.

- **Pratique** : En utilisant l'article de recherche que vous avez trouvé dans une revue scientifique, identifiez les statistiques descriptives rapportées.
- **Pratique** : Décrivez pourquoi les théories peuvent être soutenues, mais elles ne peuvent pas être prouvées.

Attributions des médias

- "[Not all scientific studies are created equal – David H. Schwartz](#)" (« Toutes les études scientifiques ne naissent pas égales ») par [TED-Ed](#). Standard YouTube Licence.
- "[How to Read an Academic Paper](#)" (« Comment lire un article académique (sans perdre la tête) ») par [UBCiSchool](#). CC BY (Attribution)

10. Générer de bonnes questions de recherche

Objectifs d'apprentissage

1. Décrire quelques techniques permettant de transformer des idées de recherche en questions de recherche empirique et utiliser ces techniques pour générer des questions.
2. Expliquer ce qui rend une question de recherche intéressante et évaluer les questions de recherche en fonction de leur intérêt.

Générer des questions de recherche testables empiriquement

Une fois que vous avez une idée de recherche, vous devez l'utiliser pour générer une ou plusieurs questions de recherche testables empiriquement, c'est-à-dire des questions exprimées en termes d'une seule variable ou d'une relation entre des variables. Une façon d'y parvenir est d'examiner attentivement la section de discussion d'un article de recherche récent sur le sujet. Il s'agit de la dernière grande section de l'article, dans laquelle les personnes en recherche résument leurs résultats, les interprètent dans le contexte des recherches antérieures et suggèrent des orientations pour les recherches futures. Ces suggestions prennent souvent la forme de questions de recherche spécifiques, auxquelles vous pouvez ensuite essayer de répondre par des recherches supplémentaires. Il s'agit là d'une bonne stratégie, car il est probable que les questions suggérées aient déjà été identifiées comme intéressantes et importantes par des personnes d'expérience en recherche.

Mais il se peut aussi que vous souhaitiez générer vos propres questions de recherche. Comment procéder? Tout d'abord, si vous avez en tête un comportement particulier ou une caractéristique psychologique, vous pouvez simplement le conceptualiser comme une variable, et demander quelle est sa fréquence ou son intensité. Combien de mots parle-t-on en moyenne par jour? Quelle est la précision de nos souvenirs d'événements traumatisants? Quel est le pourcentage de personnes qui ont cherché une aide professionnelle pour une dépression? Si la question n'a jamais fait l'objet d'une étude scientifique — ce que vous apprendrez lors de votre analyse documentaire — alors elle peut être intéressante et valoir la peine d'être étudiée.

Si la recherche scientifique a déjà répondu à la question de la fréquence ou de l'intensité d'un comportement ou d'une caractéristique, vous devriez alors envisager de la transformer en une question sur la relation entre ce comportement ou cette caractéristique et une autre variable. Pour ce faire, vous pouvez

vous poser la série suivante de questions plus générales et noter toutes les réponses qui vous viennent à l'esprit.

- Quelles sont les causes possibles de ce comportement ou de cette caractéristique?
- Quels sont les effets possibles de ce comportement ou de cette caractéristique?
- Quels sont les types de personnes susceptibles de présenter plus ou moins le comportement ou la caractéristique en question?
- Quels types de situations pourraient susciter plus ou moins le comportement ou la caractéristique en question?

En général, chaque réponse que vous notez peut être conceptualisée comme une seconde variable, suggérant une question sur une relation. Si vous vous intéressez à la loquacité, par exemple, vous pourriez vous demander si la taille de la famille est une cause possible de cette caractéristique psychologique. Existe-t-il une relation entre la taille de la famille et le caractère loquace? Vous pourriez aussi vous rendre compte que les gens semblent plus loquaces dans les groupes de même sexe que dans les groupes mixtes. Existe-t-il une différence entre le niveau moyen de loquacité des personnes appartenant à des groupes homosexuels et celui des personnes appartenant à des groupes mixtes? Cette approche devrait vous permettre de générer de nombreuses questions empiriques différentes sur presque n'importe quel comportement ou caractéristique psychologique.

Si au cours de ce processus, vous trouvez une question qui n'a jamais fait l'objet d'une étude scientifique — ce que vous apprendrez dans votre analyse documentaire — alors elle peut être intéressante et mérite d'être étudiée. Mais que se passe-t-il si vous découvrez qu'elle a été étudiée scientifiquement? Bien que les novices en recherche aient souvent envie d'abandonner et de passer à une nouvelle question à ce stade, ce n'est pas nécessairement une bonne stratégie. D'une part, le fait que la question ait été étudiée scientifiquement et que la recherche ait été publiée suggère qu'elle intéresse la communauté scientifique. D'autre part, la question peut très certainement être affinée de manière à ce que sa réponse apporte toujours quelque chose de nouveau à la littérature de recherche. Là encore, se poser une série de questions plus générales sur la relation est une bonne stratégie.

- Existe-t-il d'autres façons de définir et de mesurer les variables?
- Y a-t-il des types de personnes pour lesquelles la relation pourrait être plus forte ou plus faible?
- Existe-t-il des situations dans lesquelles la relation pourrait être plus forte ou plus faible, y compris des situations ayant une importance pratique?

Par exemple des recherches ont montré que les femmes et les hommes parlent à peu près le même nombre de mots par jour, mais c'était lorsque la loquacité était mesurée en nombre de mots prononcés par jour

parmi des personnes étudiant à l'université au Canada, aux États-Unis et au Mexique. On peut toujours se demander si d'autres façons de mesurer la loquacité — par exemple, le nombre de personnes différentes à qui l'on parle chaque jour — aboutissent au même résultat. Nous pouvons également nous demander si l'étude de personnes âgées ou de personnes issues d'autres cultures produit le même résultat. Une fois de plus, cette approche devrait vous aider à générer de nombreuses questions de recherche différentes sur presque n'importe quelle relation.

Évaluer les questions de recherche

Les personnes en recherche génèrent généralement beaucoup plus de questions de recherche qu'elles ne tentent d'y répondre. Cela signifie que ces personnes doivent disposer d'un moyen d'évaluer les questions de recherche générées afin de pouvoir choisir celles qui seront poursuivies. Dans cette section, nous examinons deux critères d'évaluation des questions de recherche : l'intérêt de la question et la faisabilité de la réponse.

Intérêt

À quelle fréquence les gens lacent-ils leurs chaussures? Les personnes ressentent-elles de la douleur lorsque vous leur donnez un coup de poing dans la mâchoire? Les femmes ont-elles plus tendance à se maquiller que les hommes? Les personnes préfèrent-elles les glaces à la vanille ou au chocolat? Bien qu'il soit assez simple de concevoir une étude et de collecter des données pour répondre à ces questions, vous ne le feriez probablement pas parce qu'elles ne sont pas intéressantes. Il ne s'agit pas ici de savoir si une question de recherche est intéressante pour nous personnellement, mais si elle est intéressante pour les gens en général et, en particulier, pour la communauté scientifique. Mais qu'est-ce qui rend une question de recherche intéressante dans ce sens? Nous examinons ici trois facteurs qui influencent l'**intérêt scientifique** (*interestingness* en anglais) d'une question de recherche : la réponse est incertaine, la réponse comble une lacune dans la littérature de recherche, et elle a des implications pratiques importantes.

Premièrement, une question de recherche est intéressante dans la mesure où sa réponse est incertaine. Il est évident que les questions auxquelles la recherche scientifique a répondu ne sont plus intéressantes pour faire l'objet de nouvelles recherches empiriques. Mais le fait qu'une question n'ait pas trouvé de réponse dans la recherche scientifique ne la rend pas nécessairement intéressante. Il doit y avoir une chance raisonnable que la réponse à la question soit quelque chose que nous ne connaissions pas déjà. Mais comment évaluer cette probabilité avant de collecter des données? L'une des approches consiste à essayer de trouver des raisons de s'attendre à des réponses différentes à la question, en particulier celles qui semblent aller à l'encontre du bon sens. Si vous pouvez trouver des raisons de vous attendre à au moins deux réponses différentes, alors la question peut être intéressante. Si vous avez des raisons de n'attendre qu'une seule réponse, c'est qu'elle ne l'est probablement pas. La question de savoir si les femmes sont plus loquaces que les hommes est intéressante parce qu'il y a des raisons de s'attendre aux deux réponses. L'existence même du stéréotype suggère que la réponse pourrait être oui, mais le fait que les capacités verbales des femmes et des hommes soient relativement similaires suggère que la réponse pourrait être non. La question

de savoir si les gens ressentent de la douleur lorsqu'on leur donne un coup de poing dans la mâchoire n'est pas intéressante parce qu'il n'y a absolument aucune raison de penser que la réponse pourrait être autre chose qu'un oui retentissant.

Un deuxième facteur important à considérer pour déterminer si une question de recherche est intéressante est de savoir si la réponse à cette question comblera une lacune dans la littérature de recherche. Là encore, cela signifie en partie que la question n'a pas encore trouvé de réponse dans la recherche scientifique. Mais cela signifie également que la question est, d'une certaine manière, naturelle pour les personnes qui connaissent bien la littérature de recherche. Par exemple la question de savoir, si le fait de prendre des notes de cours à la main peut contribuer à améliorer les résultats aux examens de personnes aux études, viendrait probablement à l'esprit de toute personne familière avec les recherches sur la prise de notes et l'inefficacité du traitement superficiel sur l'apprentissage.

Un dernier facteur à considérer pour déterminer si une question de recherche est intéressante est de savoir si sa réponse a des implications pratiques importantes. Là encore la question de savoir, si le fait de prendre des notes à la main améliore l'apprentissage, a des répercussions importantes pour l'éducation, notamment sur les politiques de classe concernant l'utilisation de la technologie. La question de savoir si l'utilisation du téléphone cellulaire nuit à la conduite est intéressante parce qu'elle concerne la sécurité personnelle de toute personne se déplaçant en voiture et le débat sur la question de savoir si l'utilisation du téléphone cellulaire devrait être restreinte par la loi.

Faisabilité

Un deuxième critère important pour l'évaluation des questions de recherche est la **faisabilité** d'y répondre avec succès. De nombreux facteurs influent sur la faisabilité, notamment le temps, l'argent, l'équipement et le matériel, les connaissances et les compétences techniques, et l'accès aux personnes participantes requises par la recherche. Il est évident que les personnes en recherche doivent tenir compte de ces facteurs afin de ne pas perdre de temps et d'efforts à entreprendre des recherches ne pouvant pas être menées à bien.

En parcourant un échantillon de revues scientifiques en psychologie, vous trouverez de nombreuses études compliquées et difficiles à réaliser. Il s'agit notamment d'études longitudinales dans lesquelles les personnes participantes sont suivies pendant de nombreuses années, d'études en neuro-imagerie dans lesquelles l'activité cérébrale des personnes participantes est mesurée pendant que ces personnes effectuent diverses tâches mentales, et d'études non expérimentales complexes impliquant plusieurs variables et des analyses statistiques compliquées. Il faut toutefois garder à l'esprit que ces recherches sont généralement menées par des équipes en recherche hautement qualifiées dont les travaux sont souvent financés en partie par des subventions publiques et privées. N'oubliez pas non plus que la recherche n'a pas besoin d'être compliquée ou difficile pour produire des résultats intéressants et importants. En parcourant un échantillon de revues scientifiques, vous découvrirez des études relativement simples et faciles à réaliser, impliquant par exemple un échantillon de commodité de personnes inscrites à l'université et une tâche papier-crayon.

Enfin il est généralement bon d'utiliser des méthodes qui ont déjà été utilisées avec succès par d'autres personnes en recherche. Par exemple, si vous souhaitez manipuler l'humeur des gens pour rendre certains d'entre eux heureux, il serait judicieux d'utiliser l'une des nombreuses approches qui ont été utilisées avec

succès par d'autres équipes de recherche (par exemple, leur faire un compliment). C'est une bonne chose non seulement pour des raisons de faisabilité — l'approche est « éprouvée » — mais aussi parce qu'elle assure une plus grande continuité avec les recherches antérieures. Il est ainsi plus facile de comparer vos résultats avec ceux d'autres personnes en recherche et de comprendre les implications de leurs recherches pour les vôtres, et vice versa.

12. Concevoir une étude de recherche

Objectifs d'apprentissage

1. Définir le concept de variable, distinguer les variables quantitatives des variables catégorielles et donner des exemples de variables susceptibles d'intéresser les psychologues.
2. Expliquer la différence entre une population et un échantillon.
3. Distinguer la recherche expérimentale de la recherche non expérimentale.
4. Distinguer les études en laboratoire, les études sur le terrain et les expériences sur le terrain.

Identification et définition des variables et de la population

Variables et définitions opérationnelles

Une partie du processus de formulation d'une hypothèse consiste notamment à identifier les variables que l'on souhaite étudier, et à les définir de manière opérationnelle afin de pouvoir les mesurer. Les questions de recherche en psychologie portent sur des variables. Une **variable** est une quantité ou une qualité qui varie selon les personnes ou les situations. Par exemple, la taille des personnes inscrites à un cours universitaire est variable, car elle varie d'une personne à l'autre. Le choix de programme des personnes aux études est également une variable, tant que tout le monde dans la classe n'est pas inscrit dans le même programme. Presque tout dans notre monde varie, si bien qu'il est beaucoup plus difficile de trouver des exemples de constantes (des choses qui ne varient pas). Un exemple rare de constante est la vitesse de la lumière. Les variables peuvent être quantitatives ou catégorielles. Une **variable quantitative** est une quantité, comme la taille, qui est généralement mesurée en attribuant un nombre à chaque personne. D'autres exemples de variables quantitatives incluent le niveau de loquacité d'une personne, son degré de dépression ou le nombre de frères et sœurs qu'elle a. Une **variable catégorielle** est une qualité, comme le choix de programme, et est généralement mesurée en attribuant une catégorie à chaque personne (par exemple psychologie, anglais, science infirmière, etc.). D'autres exemples incluent la nationalité des personnes, leur profession et le fait qu'elles suivent ou non une psychothérapie.

Après avoir formulé son hypothèse et sélectionné les variables à manipuler et mesurer, la personne en recherche doit trouver les moyens de mesurer réellement les variables qui l'intéressent. Cela nécessite une **définition opérationnelle** — c'est-à-dire une définition précisant la manière dont elle doit être mesurée.

La plupart des variables que les personnes en recherche souhaitent étudier ne peuvent être directement observées ou mesurées, ce qui pose un problème puisque l'empirisme (l'observation) est au cœur de la méthode scientifique. La définition opérationnelle d'une variable consiste à transformer un concept abstrait, comme la dépression, qui ne peut être observé directement, en quelque chose qui peut être observé et mesuré directement. La plupart des variables peuvent être définies opérationnellement de multiples façons. Par exemple, la dépression peut être définie opérationnellement comme les scores des personnes sur une échelle de dépression papier-crayon, comme l'inventaire de dépression de Beck, le nombre de symptômes dépressifs qu'elles éprouvent ou le fait qu'elles aient été diagnostiquées comme souffrant d'un trouble dépressif caractérisé. Il est judicieux pour les personnes en recherche de choisir une définition opérationnelle qui a été fréquemment utilisée dans la littérature de recherche.

Échantillon et mesure

En plus d'identifier les variables à manipuler et à mesurer, et d'en avoir une définition opérationnelle, les personnes en recherche doivent identifier la population d'intérêt. Les personnes menant des recherches en psychologie souhaitent généralement tirer des conclusions sur un très grand groupe de personnes. C'est ce qu'on appelle la **population**. Il peut s'agir de toutes les personnes adolescentes canadiennes et américaines, des enfants autistes, des athlètes professionnelles ou même simplement de l'ensemble des êtres humains — en fonction des intérêts et des objectifs des personnes en recherche. Mais ces personnes menant des recherches en psychologie n'étudient généralement qu'un petit sous-ensemble ou un **échantillon** de la population. Par exemple, une chercheuse peut mesurer le degré de loquacité de quelques centaines de personnes étudiant à l'université dans l'intention de tirer des conclusions sur le degré de loquacité des hommes et des femmes en général. Il est donc essentiel d'utiliser un échantillon représentatif — un échantillon similaire à la population sur les aspects importants.

Une méthode pour obtenir un échantillon est l'**échantillonnage aléatoire simple**, dans lequel chaque membre de la population a une chance égale d'être sélectionnée pour faire partie de l'échantillon. Par exemple, un enquêteur pourrait utiliser une liste de l'électorat inscrit dans une ville (la population), sélectionner au hasard 100 membres sur la liste (l'échantillon) et demander à ces 100 personnes pour quel candidat elles ont l'intention de voter. Malheureusement, l'échantillonnage aléatoire est difficile, voire impossible, dans la plupart des recherches en psychologie, car les populations y sont moins clairement définies que la liste de l'électorat d'une ville. Comment, en effet, donner à toutes les personnes adolescentes canadiennes et américaines ou à tous les enfants autistes une chance égale d'être sélectionnée pour un échantillon? L'alternative la plus courante est l'échantillonnage de commodité, où l'échantillon est constitué de personnes se trouvant à proximité et disposé à participer (comme les personnes inscrites au cours d'introduction à la psychologie). Le problème évident de l'**échantillonnage de commodité** est que l'échantillon peut ne pas être représentatif de la population et qu'il peut donc être moins approprié de généraliser les résultats de l'échantillon à la population.

Recherche non expérimentale et recherche expérimentale

L'étape suivante pour une personne en recherche est de décider du type d'approche à utiliser pour collecter les données. Comme vous l'apprendrez dans votre cours de méthodes de recherche, il existe de nombreuses approches différentes qui peuvent être divisées de plusieurs façons. L'une des distinctions fondamentales est celle qui existe entre la recherche expérimentale et la recherche non expérimentale.

Recherche expérimentale

Les personnes en recherche qui souhaitent tester des hypothèses sur des relations de cause à effet entre des variables (dont l'objectif est d'expliquer) doivent utiliser une méthode expérimentale. En effet, la méthode expérimentale est la seule permettant de déterminer des relations de cause à effet. Avec la **méthode expérimentale**, les personnes en recherche manipulent d'abord une ou plusieurs variables tout en tentant de contrôler les variables externes, puis mesurent comment les variables manipulées influencent les réponses des personnes participantes.

Les termes « **variable indépendante** » et « **variable dépendante** » sont utilisés dans le contexte de la recherche expérimentale. La variable indépendante est la variable manipulée par la personne responsable de l'expérimentation (c'est la cause présumée) et la variable dépendante est la variable mesurée par la personne responsable de l'expérimentation (c'est l'effet présumé). Les **variables externes** sont toutes les variables autres que la variable dépendante. Les **variables confondantes** sont un type spécifique de variables externes qui varient systématiquement en même temps que les variables étudiées, et qui apportent donc une explication alternative aux résultats. Lorsque les personnes en recherche conçoivent une expérience, elles doivent veiller à contrôler les variables confondantes; elles doivent s'assurer que les variables externes ne deviennent pas des variables confondantes, car pour tirer une conclusion causale, elles doivent s'assurer que les explications alternatives des résultats ont été écartées.

Par exemple si nous manipulons l'éclairage dans une pièce et examinons les effets de cette manipulation sur la productivité des personnes au travail, les conditions d'éclairage (lumières vives ou tamisées) seront considérées comme la variable indépendante et la productivité des personnes au travail, comme la variable dépendante. Si les lumières vives sont bruyantes, alors ce bruit sera une variable confondante, puisque le bruit serait présent lorsque les lumières sont vives et serait absent lorsque les lumières sont tamisées. Si le bruit varie systématiquement avec la lumière, nous ne saurons pas si la différence de productivité de la main-d'œuvre dans les deux conditions d'éclairage est due au bruit ou à la lumière. Les variables confondantes sont donc indésirables, elles nuisent à notre capacité de tirer des conclusions causales sur la nature de la relation entre les variables. Toutefois s'il y a du bruit dans la pièce à la fois lorsque les lumières sont allumées et lorsqu'elles sont éteintes, le bruit n'est qu'une variable externe (c'est tout simplement une variable autre que la variable indépendante ou dépendante) et nous ne nous préoccupons pas beaucoup des variables externes. En effet, si une variable ne varie pas systématiquement avec la variable indépendante manipulée, elle ne peut pas être une explication alternative des résultats.

Recherche non expérimentale

Les personnes en recherche qui s'intéressent uniquement à décrire les caractéristiques des personnes, décrire les relations entre les variables et utiliser ces relations pour faire des prédictions peuvent utiliser la **recherche non expérimentale**. Dans le cadre de la recherche non expérimentale, la personne en recherche se contente de mesurer les variables telles qu'elles se présentent naturellement, elles ne sont pas manipulées. Par exemple si je me contente de mesurer le nombre d'accidents mortels de la circulation survenus au Canada et aux États-Unis l'année dernière et impliquant l'utilisation d'un téléphone cellulaire, mais que je ne manipule pas l'utilisation du téléphone cellulaire, il s'agit alors d'une recherche non expérimentale. De plus, si je me place à un carrefour très fréquenté et que je prends note du sexe des automobilistes, et du fait que ces personnes utilisaient ou non un téléphone cellulaire au moment de traverser le carrefour afin de déterminer si les hommes ou les femmes sont plus susceptibles d'utiliser un téléphone cellulaire au volant, il s'agirait encore d'une recherche non expérimentale. Il est important de souligner que « non expérimental » ne signifie pas non scientifique. La recherche non expérimentale est de nature scientifique. Elle peut être utilisée pour remplir deux des trois objectifs de la science (décrire et prédire). Toutefois, contrairement à la recherche expérimentale, cette méthode ne permet pas de tirer des conclusions causales; elle ne permet pas d'affirmer qu'une variable cause des changements au niveau d'une autre variable.

Recherche en laboratoire ou sur le terrain

La prochaine grande distinction entre les méthodes de recherche concerne les études en laboratoire et les études sur le terrain. Une étude en laboratoire est une étude menée dans un environnement de **laboratoire**. En revanche, une étude sur **le terrain** est une étude menée dans le monde réel, dans un environnement naturel.

Les expériences en laboratoire ont généralement une grande **validité interne**. La validité interne fait référence au degré de confiance avec lequel nous pouvons déduire une relation de cause à effet entre des variables. Lorsque nous menons une étude expérimentale en laboratoire, la validité interne est très élevée, car nous manipulons une variable tout en contrôlant toutes les autres variables externes. Lorsque nous manipulons une variable indépendante et observons un effet sur une variable dépendante, et que nous contrôlons tout le reste de sorte que la seule différence entre nos groupes ou conditions expérimentales soit la variable manipulée, nous pouvons avoir confiance que c'est la variable indépendante qui cause le changement survenu dans la variable dépendante. En revanche comme les études sur le terrain sont menées dans le monde réel, la personne responsable de l'expérimentation a généralement moins de contrôle sur l'environnement et les variables externes potentielles, ce qui diminue la validité interne et permet moins de conclure à l'existence d'une relation de cause à effet.

Cependant il existe généralement un compromis entre la validité interne et la **validité externe**. La validité externe réfère simplement au degré de généralisation des résultats à d'autres circonstances ou contextes, comme le monde réel. Lorsque la validité interne est élevée, la validité externe tend à être faible; et lorsque la validité interne est faible, la validité externe tend à être élevée. Ainsi les études en laboratoire ont généralement une faible validité externe, tandis que les études sur le terrain ont généralement une validité

externe élevée. Étant donné que les études sur le terrain sont menées dans un environnement réel, il est beaucoup plus juste de généraliser les résultats à cet environnement réel que lorsque la recherche est menée dans le cadre plus stérile et artificiel du laboratoire.

Enfin il existe des études sur le terrain qui sont de nature non expérimentale parce que rien n'y est manipulé. Mais il existe aussi des expériences sur le terrain où une variable indépendante est manipulée dans un environnement naturel et où les variables externes sont contrôlées. En fonction de leur qualité globale et du niveau de contrôle des variables externes, ces **expériences sur le terrain** peuvent avoir à la fois une validité externe élevée et une validité interne élevée.

9. Trouver un sujet de recherche

Objectifs d'apprentissage

1. Apprendre quelques sources fréquentes d'idées de recherche.
2. Définir la littérature de recherche en psychologie et donner des exemples de sources qui font partie de la littérature, et de sources qui n'en font pas partie.
3. Décrire et utiliser plusieurs méthodes pour trouver des recherches antérieures sur une idée ou une question de recherche particulière.

Une bonne recherche doit commencer par une bonne question de recherche. Pourtant trouver de bonnes questions de recherche est une tâche que les personnes novices en recherche trouvent souvent difficile et stressante. L'une des raisons est qu'il s'agit d'un processus créatif qui peut sembler mystérieux — voire magique — où les personnes expérimentées en recherche semblent tirer de nulle part des questions de recherche intéressantes. Cependant la recherche en psychologie sur la créativité a montré qu'elle n'est ni aussi mystérieuse, ni aussi magique qu'elle le paraît. Elle est en grande partie le fruit de stratégies de réflexion ordinaires et de persévérance (Weisberg, 1993). Cette section présente quelques stratégies assez simples pour trouver des idées générales de recherche, transformer ces idées en questions empiriques testables, et enfin évaluer ces questions en matière d'intérêt et de faisabilité.

Trouver l'inspiration

Les questions de recherche commencent souvent par des idées de recherche plus générales — principalement axées sur un comportement ou une caractéristique psychologique : la loquacité, l'apprentissage, la dépression, le saut à l'élastique, etc. Avant d'examiner comment transformer ces idées en questions empiriques, il convient de s'interroger sur leur origine. Trois des sources d'inspiration les plus fréquentes sont les observations informelles, les problèmes pratiques et les recherches antérieures.

Les observations informelles comprennent les observations directes de notre propre comportement et de celui des autres, ainsi que les observations de seconde main provenant de sources non scientifiques, telles que les journaux, les livres, les blogues, etc. Par exemple, vous pouvez remarquer que vous semblez toujours être dans la file d'attente la plus lente à l'épicerie. Se pourrait-il que la plupart des gens pensent la même chose? Vous pouvez aussi lire dans un journal local qu'une famille locale reçoit de l'argent et de la nourriture d'autres personnes après que sa maison ait brûlé, et vous commencez alors à vous demander qui fait de tels dons et pourquoi. Certaines des recherches les plus célèbres en psychologie ont été inspirées par des observations informelles. Les célèbres recherches de Stanley Milgram sur l'obéissance à l'autorité, par

exemple, ont été inspirées en partie par des reportages journalistiques sur les procès de criminels de guerre nazis accusés, dont beaucoup affirmaient qu'ils ne faisaient qu'obéir à des ordres. Cela l'a amené à se demander dans quelle mesure les gens ordinaires commettent des actes immoraux simplement parce qu'ils en reçoivent l'ordre d'une figure d'autorité (Milgram, 1963).

Les problèmes pratiques peuvent également inspirer des idées de recherche, menant directement à la recherche appliquée dans des domaines tels que le droit, la santé, l'éducation et le sport. Le fait de prendre des notes de cours à la main améliore-t-il les résultats d'examens de personnes aux études? Quelle est l'efficacité de la psychothérapie contre la dépression par rapport à la pharmacothérapie? Dans quelle mesure les téléphones cellulaires nuisent-ils à la capacité de conduire? Comment apprendre aux enfants à lire plus efficacement? Quelle est la meilleure préparation mentale pour courir un marathon?



Lecture sur version imprimée? Allez sur <https://youtu.be/t1ZwgDeXzeQ> ou scannez ce code QR avec votre cellulaire

Cependant, la source d'inspiration la plus courante pour les nouvelles idées de recherche est probablement la recherche antérieure. Rappelons que la science est une sorte de collaboration à grande échelle dans laquelle de nombreux scientifiques différents lisent et évaluent les travaux des autres, et mènent de nouvelles études en s'appuyant sur ces travaux. Bien entendu, les personnes expérimentées en recherche connaissent les recherches antérieures dans leur domaine d'expertise et ont probablement une longue liste d'idées. Les personnes novices en recherche peuvent donc trouver l'inspiration en consultant une autre personne plus expérimentée en recherche (par exemple, les personnes aux études peuvent consulter un membre du corps professoral). Mais elles peuvent également trouver l'inspiration en se procurant un exemplaire de presque toutes les revues scientifiques et en lisant les titres et les résumés. Dans un numéro typique de *Psychological Science*, par exemple, vous trouverez des articles sur la

perception des formes, l'antisémitisme, les files d'attente dans les services de police, le sens de la mort, l'apprentissage d'une seconde langue, les personnes qui recherchent des expériences émotionnelles négatives, et bien d'autres sujets. Si vous pouvez restreindre vos intérêts à un sujet (par exemple, la mémoire) ou à un domaine particulier (par exemple, les soins de santé), vous pouvez également consulter des revues plus spécifiques, telles que *Memory and Cognition*, *Health Psychology* ou *Revue québécoise de psychologie*.



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=49#oembed-2>

Analyse documentaire de la recherche

Une fois encore, l'une des sources d'inspiration les plus fréquentes est la recherche antérieure. Il est donc important d'analyser la littérature dès le début du processus de recherche. La **littérature de recherche**, quel que soit le domaine, est l'ensemble des travaux de recherche publiés dans ce domaine. Analyser la littérature de recherche signifie trouver, lire et résumer les travaux de recherche publiés en rapport avec votre sujet

d'intérêt. En plus de vous aider à découvrir de nouvelles questions de recherche, l'analyse documentaire au début du processus de recherche peut vous être utile à plusieurs égards.

- Elle peut vous indiquer si une question de recherche a déjà reçu une réponse.
- Elle peut vous aider à évaluer l'intérêt d'une question de recherche.
- Elle peut vous donner des idées sur la manière de mener votre propre étude.
- Elle peut vous indiquer comment votre étude s'inscrit dans la littérature de recherche.

La littérature de recherche en psychologie est énorme — comprenant des millions d'articles et de livres savants datant des débuts de la discipline — et elle continue de croître. Bien que ses limites soient quelque peu floues, la littérature de recherche n'inclut certainement pas les livres de développement personnel et autres ouvrages de psychologie populaire, les entrées de dictionnaires et d'encyclopédies, les sites web et autres sources similaires destinées principalement au grand public. Ces sources sont considérées comme peu fiables parce qu'elles ne sont pas examinées par d'autres personnes en recherche, et qu'elles ne reposent souvent que sur le bon sens ou l'expérience personnelle. Wikipédia contient beaucoup d'informations précieuses mais, comme ses auteurs et autrices sont anonymes et n'ont pas forcément de formation ou d'expertise formelle dans le domaine concerné, et que son contenu change continuellement, elle ne peut pas servir de base à une recherche scientifique solide. Pour nos besoins, il est utile de définir la littérature de recherche comme étant presque entièrement constituée de deux types de sources : les articles de revues scientifiques et les livres savants en psychologie, et dans les domaines connexes.

Revues scientifiques

Les **revues scientifiques** sont des périodiques qui publient des articles de recherche originaux. Il existe des milliers de revues scientifiques qui publient des recherches en psychologie et dans des domaines connexes. Elles sont généralement publiées mensuellement ou trimestriellement sous forme de numéros individuels, chacun contenant plusieurs articles. Les numéros sont organisés en volumes, qui comprennent généralement tous les numéros d'une année civile. Certaines revues sont publiées uniquement en version imprimée, d'autres à la fois sur papier et sous forme électronique, et d'autres encore uniquement en version électronique.

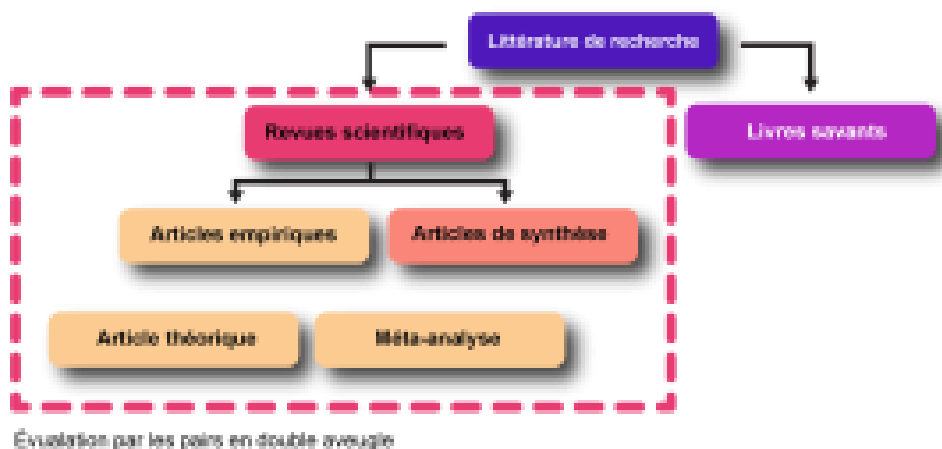
La plupart des articles publiés dans les revues scientifiques sont de deux types fondamentaux: les articles empiriques et les articles de synthèse. Les **articles empiriques** décrivent une ou plusieurs nouvelles études empiriques menées par les auteurs et autrices. Ils introduisent une question de recherche, expliquent pourquoi elle est intéressante, passent en revue les recherches antérieures, décrivent leur méthode et leurs résultats, et tirent leurs conclusions. Les **articles de synthèse** résument les recherches précédemment publiées sur un sujet et présentent généralement de nouvelles façons d'organiser ou d'expliquer les résultats. Lorsqu'un article de synthèse est consacré principalement à la présentation d'une nouvelle théorie, on parle souvent d'**article théorique**. Lorsqu'un article de synthèse fournit un résumé statistique de tous les résultats précédents, on parle de **méta-analyse**.

La plupart des revues scientifiques en psychologie font l'objet d'une **évaluation par les pairs en double aveugle**. Les personnes en recherche qui souhaitent publier leurs travaux dans la revue soumettent un document à la direction de la rédaction — qui est généralement aussi une personne établie en recherche — laquelle l'envoie à son tour à deux ou trois personnes expertes du sujet pour une évaluation. Chaque responsable de l'évaluation lit le document, rédige une évaluation critique mais constructive, et renvoie cette évaluation à la direction de la rédaction accompagnée de recommandations quant à la publication ou non du document. La direction de la rédaction décide alors d'accepter l'article pour publication, de demander aux auteurs ou autrices d'y apporter des modifications et de le soumettre à nouveau pour un examen plus approfondi, ou de le rejeter purement et simplement. Dans tous les cas, la direction de la rédaction transmet les commentaires écrits par les responsables de l'évaluation aux gens ayant soumis le manuscrit afin que ces dernières personnes puissent le réviser en conséquence. L'ensemble du processus se déroule en double aveugle, car les gens faisant l'évaluation ne connaissent pas l'identité de la/des personne-s ayant soumis le document et vice versa. L'examen par les pairs en double aveugle est utile, car il permet de s'assurer que les travaux répondent aux normes fondamentales du domaine avant qu'ils ne soient publiés dans la littérature de recherche. Toutefois, afin d'accroître la transparence et la responsabilité, certaines revues en libre accès plus récentes (par exemple, [Frontiers in Psychology](#)) utilisent un processus ouvert d'évaluation par les pairs dans lequel l'identité des personnes faisant l'évaluation (qui reste cachée pendant le processus d'évaluation par les pairs) est ensuite publiée avec l'article de la revue.

Livres savants

Les **livres savants** sont des ouvrages écrits par des personnes en recherche et des personnes praticiennes, principalement à l'intention d'autres personnes travaillant dans ces domaines. Une **monographie** est rédigée par une seule personne ou un petit groupe de personnes, et présente généralement un sujet de manière cohérente à l'instar d'un article de synthèse. Les **volumes édités** ont une ou quelques personne-s responsable-s de l'édition qui recrutent plusieurs autrices et auteurs pour rédiger des chapitres distincts sur différents aspects d'un même sujet. Bien que les livres édités puissent également présenter le sujet de manière cohérente, il n'est pas rare que chaque chapitre adopte un point de vue différent, voire que les auteurs et autrices de différents chapitres soient ouvertement en désaccord les un-e-s avec les autres. En général, les livres savants sont soumis à un processus d'évaluation par les pairs similaire à celui utilisé par les revues scientifiques.

Figure 2.2. Analyse documentaire de la recherche avec la littérature de recherche



Stratégies de recherche documentaire

Utilisation de PsycINFO et d'autres bases de données

La principale méthode utilisée pour rechercher la littérature de recherche consiste à utiliser une ou plusieurs bases de données (ou index électroniques). Celles-ci comprennent **Academic Search Premier**, **JSTOR** et **ProQuest** pour toutes les disciplines universitaires, **ERIC** pour l'éducation et **PubMed** pour la médecine et les domaines connexes. La plus importante pour nos besoins est cependant **PsycINFO**, qui est produite par l'American Psychological Association (APA). PsycINFO est si complet — il couvre des milliers de revues scientifiques et de livres savants remontant à plus de 100 ans — que pour la plupart des besoins, son contenu est synonyme de littérature de recherche en psychologie. Comme la plupart des bases de données de ce type, vous avez accès à PsycINFO grâce à un accès limité à la communauté de votre université, à partir de la page web de la bibliothèque universitaire. Notons que PsycInfo a pris récemment le nom de APA PsycINFO.

PsycINFO se compose d'enregistrements individuels (notices bibliographiques) pour chaque article, chapitre de livre ou livre dans la base de données. Chaque notice comprend des informations de base sur la publication, un résumé de l'ouvrage (comme celui présenté au début de ce chapitre) et une liste des autres articles et ouvrages cités par ce dernier (section Références). Une interface informatique permet de saisir un ou plusieurs termes de recherche et renvoie toutes les notices qui contiennent ces termes de recherche. (Ces interfaces sont fournies par différents fournisseurs et peuvent donc se présenter différemment selon la bibliothèque que vous utilisez.) Chaque enregistrement contient également des listes de mots-clés qui décrivent le contenu de l'ouvrage ainsi qu'une liste de termes d'indexation. Les termes d'indexation sont particulièrement utiles, car ils sont normalisés. Les recherches sur les différences entre les hommes et les femmes, par exemple, sont toujours indexées sous « différences entre les sexes » (en anglais *"Human Sex Differences"*). Les recherches sur la prise de notes sont toujours indexées sous le terme « stratégies

d'apprentissage » (en anglais *“Learning Strategies”*). Si vous ne connaissez pas les termes d'index appropriés, PsycINFO comprend un thésaurus qui peut vous aider à les trouver.

Étant donné que PsycINFO contient près de quatre millions d'enregistrements, il se peut que vous deviez essayer plusieurs termes de recherche dans différentes combinaisons et à différents niveaux de spécificité avant de trouver ce que vous cherchez. Imaginons, par exemple, que vous vous intéressez à la question de savoir si les hommes et les femmes diffèrent en matière de capacité à se souvenir d'expériences vécues lorsqu'ils/elles étaient très jeunes. Si vous connaissez le terme de recherche « mémoire », vous obtiendrez beaucoup trop d'informations pour pouvoir les consulter individuellement. C'est là qu'intervient le thésaurus. En saisissant « mémoire » (*“Memory”* en anglais) dans le thésaurus, on obtient plusieurs termes d'indexation plus spécifiques, dont « souvenirs précoces » (*“Early Memories”* en anglais). Bien que la recherche de « premiers souvenirs » (*“First Memories”* en anglais) comme termes de l'index renvoie encore un nombre trop important de documents pour être consultée individuellement, la combinaison de ce terme avec « différences entre les sexes » comme deuxième terme de recherche renvoie moins d'articles, dont beaucoup sont très pertinents pour le sujet. Veuillez noter que la recherche dans PsycINFO doit se faire avec des termes en anglais.



Lecture sur version imprimée? Allez sur <https://youtu.be/t1ZwgDeXzeQ> ou scannez ce code QR avec votre cellulaire

Selon le/la prestataire de service (généralement les prestataires américains ProQuest ou EBSCOhost) proposant l'interface avec PsycINFO, vous pourrez peut-être sauvegarder, imprimer ou envoyer par courrier électronique les enregistrements PsycINFO pertinents. Les enregistrements peuvent même contenir des liens vers la copie en texte intégral des œuvres elles-mêmes (en version html ou en pdf). (PsycARTICLES est une base de données qui donne accès au texte intégral des articles de toutes les revues publiées par l'APA.) Si ce n'est pas le cas et que vous souhaitez obtenir une copie de l'ouvrage, vous devrez vous renseigner pour savoir si votre bibliothèque possède la revue en version imprimée, si le livre et la copie papier se trouvent dans les rayonnages de la bibliothèque universitaire, ou si vous pouvez l'emprunter d'une autre bibliothèque grâce au service de prêt entre bibliothèques (PEB). N'hésitez pas à demander de l'aide à un-e bibliothécaire.



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=49#oembed-3>

Utiliser d'autres techniques de recherche

En plus de la saisie de termes de recherche dans PsycINFO et d'autres bases de données, il existe plusieurs autres techniques de recherche dans la littérature de recherche. Tout d'abord, si vous disposez d'un bon article ou d'un bon chapitre de livre sur votre sujet — de préférence un article de synthèse récent — vous pouvez consulter la liste des références de cet article pour trouver d'autres articles, livres et chapitres de livres pertinents. En fait, vous devriez faire de même avec tout article ou chapitre de livre pertinent que vous trouvez. Vous pouvez également partir d'un article ou d'un chapitre de livre classique sur votre sujet, trouver sa notice dans PsycINFO (en entrant le nom de l'auteur ou autrice, ou le titre de l'article comme terme de recherche) et, à partir de là, établir un lien avec une liste d'autres travaux dans PsycINFO qui citent cet article classique. Cette méthode



*Lecture sur version
imprimée? Scannez ce code
QR avec votre cellulaire*

fonctionne, car d'autres personnes en recherche travaillant sur votre sujet sont susceptibles de connaître l'article classique et de le citer dans leurs propres travaux. Vous pouvez également effectuer une recherche générale sur Internet en utilisant des termes de recherche liés à votre sujet ou le nom d'une personne qui effectue des recherches sur votre sujet. Cela peut vous conduire directement à des travaux qui font partie de la littérature de recherche (par exemple des articles dans des revues en libre accès ou publiés sur les sites web de personnes responsables de recherches). Le moteur de recherche **Google Scholar** est particulièrement utile à cette fin. Une recherche générale sur Internet peut également vous conduire à des sites web qui ne font pas partie de la littérature de recherche, mais qui peuvent fournir des références à des travaux qui en font partie. Enfin, vous pouvez vous adresser à des personnes (par exemple, la personne qui vous enseigne ou d'autres membres de la faculté de psychologie) qui connaissent votre sujet et peuvent vous suggérer des articles et des chapitres de livres pertinents.



One or more interactive elements has been excluded from this version of the text. You can view them online here:

<https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=49#oembed-1>

Ce qu'il faut rechercher

Lorsque vous effectuez une analyse documentaire, il faut être sélectif. Tous les articles, chapitres de livres et ouvrages, en rapport avec votre idée ou question de recherche, ne valent pas la peine d'être obtenus, lus et intégrés dans votre analyse.

Vous devez plutôt vous concentrer sur les sources qui vous aident à réaliser quatre choses fondamentales : (a) affiner votre question de recherche, (b) identifier les méthodes de recherche appropriées, (c) placer votre recherche dans le contexte de recherches antérieures, (d) rédiger un rapport de recherche efficace. Plusieurs principes de base peuvent vous aider à trouver les sources les plus utiles.

Tout d'abord, il est préférable de se concentrer sur les recherches récentes, tout en gardant à l'esprit que ce qui est considéré comme récent dépend du sujet. Pour les sujets plus récents qui font l'objet d'études actives, « récent » peut signifier publié au cours de l'année ou des deux dernières années. Pour les sujets plus anciens qui font l'objet de moins d'attention, le terme « récent » peut signifier qu'ils ont été publiés au cours des dix dernières années. Vous aurez une idée de ce qui est considéré comme récent pour votre sujet lorsque vous commencerez votre recherche documentaire. Une bonne règle générale consiste toutefois à commencer par les sources publiées au cours des cinq dernières années. La principale exception à cette règle concerne les articles classiques qui figurent dans la liste de référence de presque toutes les autres sources. Si d'autres personnes en recherche estiment que ce travail est important, même s'il est ancien, vous devez l'inclure dans votre analyse.

Deuxièmement, vous devez rechercher des articles de synthèse sur votre sujet, car ils vous donneront une vue d'ensemble utile, en abordant souvent des définitions, des théories, des tendances, des controverses et des résultats importants, ce qui vous donnera une bonne idée de la place de votre propre recherche dans la littérature documentaire. Vous devez également rechercher des articles empiriques traitant de votre question ou de questions similaires, qui peuvent vous donner des idées sur la manière de mesurer vos variables et de collecter vos données. En règle générale, il est bon d'utiliser des méthodes que d'autres ont déjà utilisées avec succès, sauf si vous avez de bonnes raisons de ne pas le faire. Enfin vous devez rechercher des sources qui fournissent des informations susceptibles de vous aider à démontrer l'intérêt de votre question de recherche. Pour une étude portant sur les effets de l'utilisation du téléphone cellulaire sur l'aptitude à la conduite par exemple, vous pourriez rechercher des informations sur la fréquence de l'utilisation du téléphone cellulaire, la fréquence et le coût des accidents de la route, etc.

Combien de sources sont suffisantes pour votre analyse documentaire? Il s'agit d'une question difficile, car elle dépend de l'étendue de l'étude de votre sujet et de vos propres objectifs. Une étude a montré que, dans diverses revues scientifiques de psychologie, le nombre moyen de sources citées par article était d'environ 50 (Adair et Vohra, 2003). Cela donne une idée approximative de ce que les personnes professionnelles en recherche considèrent comme adéquat. En tant que personne aux études, il se peut que le nombre minimum de références à inclure soit moindre, mais les principes de sélection des références les plus utiles restent les mêmes.

Attributions aux médias

"How to Develop a Good Research Topic" (Comment développer un bon sujet de recherche) par KStateLibraries. CC BY (Attribution)

["La recherche avancée dans Google Scholar"](#) par ULiege Library. Licence YouTube standard.

"Sample PsycINFO Search on EBSCOhost" (Échantillon de recherche PsycINFO sur EBSCOhost) par APA Publishing Training. Licence YouTube standard.

8. Un modèle de recherche scientifique en psychologie

Objectifs d'apprentissage

1. Examiner un modèle général de recherche scientifique en psychologie.

La Figure 2.1 présente un modèle simple de recherche scientifique en psychologie. Les personnes responsables formulent une question de recherche, mènent une étude empirique conçue pour répondre à la question, analysent les données obtenues, tirent des conclusions à propos de la réponse à la question et publient les résultats afin qu'ils fassent partie de la littérature de recherche (c'est-à-dire l'ensemble des recherches publiées dans ce domaine). La littérature de recherche étant l'une des principales sources de nouvelles questions de recherche, ce processus peut être considéré comme un cycle. De nouvelles recherches mènent à de nouvelles questions, qui mènent à de nouvelles recherches, et ainsi de suite. La Figure 2.1 montre également que les questions de recherche peuvent naître en dehors de ce cycle, soit par des observations informelles, soit par des problèmes pratiques à résoudre. Mais même dans ces cas, la personne responsable d'une recherche commencera par consulter la littérature de recherche pour voir si la question a déjà reçu une réponse, et pour l'affiner en fonction de ce que les recherches précédentes ont déjà trouvé.

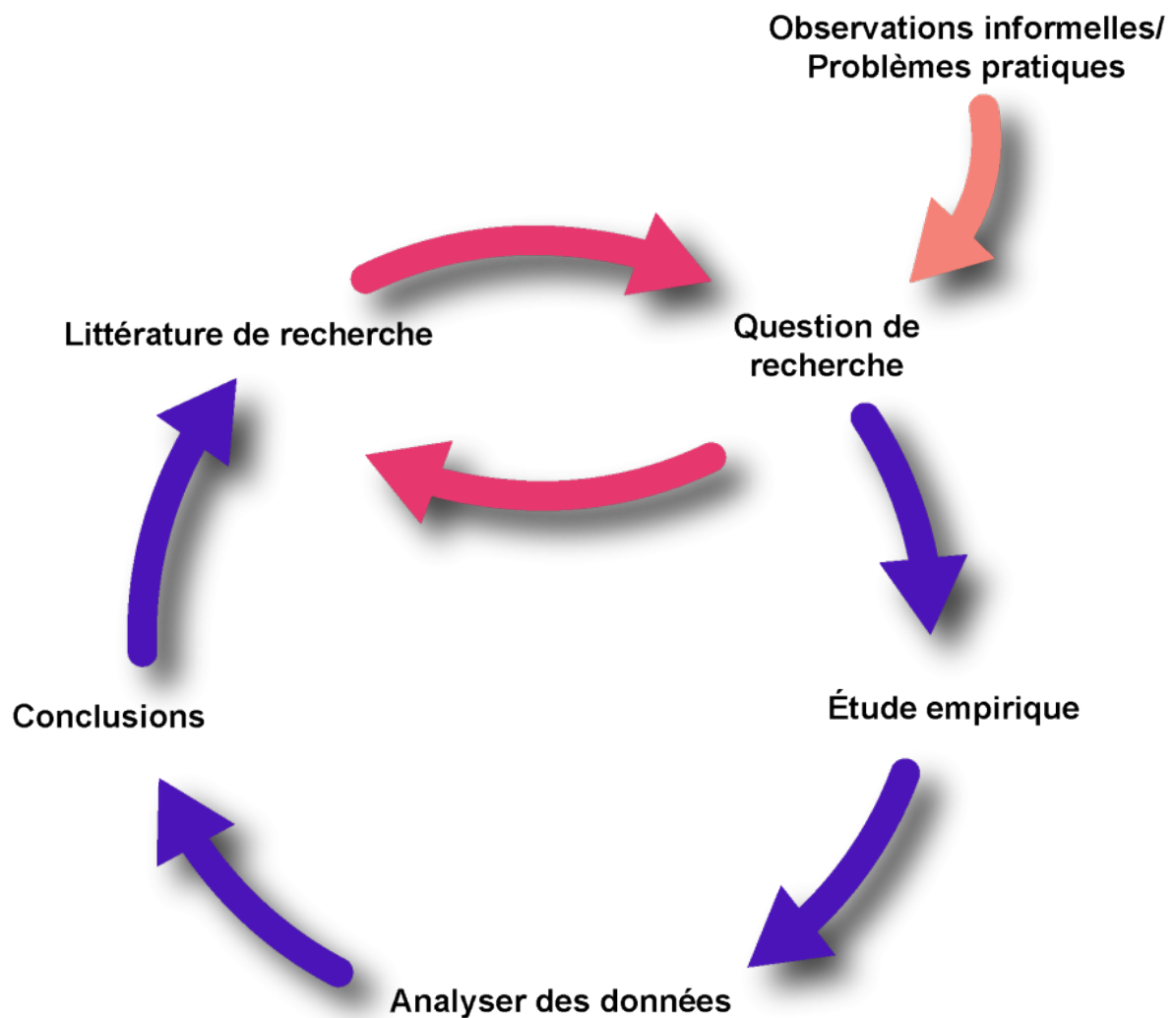


Figure 2.1. Un modèle simple de recherche scientifique en psychologie

La recherche de Mehl et de ses collègues est bien décrite par ce modèle. Leur question de recherche — à savoir si les femmes sont plus loquaces que les hommes — leur a été suggérée à la fois par les stéréotypes des gens et par les affirmations publiées dans la littérature de recherche au sujet de la loquacité relative des femmes et des hommes. En consultant la littérature de recherche, cette équipe de recherche a toutefois constaté que cette question n'avait pas fait l'objet d'études scientifiques adéquates. Cette équipe a alors mené une étude empirique minutieuse, analysé les résultats (il y a très peu de différences entre les femmes et les hommes), tiré leurs conclusions et publié leurs travaux afin que ces résultats fassent partie de la littérature de recherche. La publication de leur article n'est cependant pas la fin de l'histoire, car leur travail suggère plusieurs nouvelles questions (sur la fidélité du résultat, sur les différences culturelles potentielles, etc.) qui seront probablement reprises par cette équipe et par d'autres personnes inspirées par leur travail.



One or more interactive elements has been excluded from this version of the text. You can view them online here:
<https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=43#oembed-1>

Un autre exemple, lorsque les téléphones cellulaires se sont répandus dans les années 1990, les gens ont commencé à se demander si, et dans quelle mesure, l'utilisation du téléphone cellulaire avait un effet négatif sur la conduite. De nombreux psychologues ont décidé d'aborder cette question de manière scientifique (Collet, Guillot et Petit, 2010). Des recherches publiées précédemment, il ressortait clairement que le fait de s'engager dans une tâche verbale simple nuit aux performances d'une tâche perceptive ou motrice effectuée en même temps. Mais personne n'avait étudié l'effet spécifique de l'utilisation d'un téléphone cellulaire sur la conduite. Dans des conditions soigneusement contrôlées, ces personnes en recherche ont comparé les performances de conduite de personnes utilisant un téléphone cellulaire à celles de personnes n'utilisant pas de téléphone cellulaire, à la fois en laboratoire et sur la route. Cette équipe a constaté que la capacité à détecter les dangers de la route, le temps de réaction et le contrôle du véhicule étaient tous altérés par l'utilisation du téléphone cellulaire. Chaque nouvelle étude a été publiée et s'est inscrite dans la littérature de recherche croissante sur ce sujet. Par exemple, d'autres équipes de recherche ont par la suite démontré que les conversations par téléphone cellulaire comportent un plus grand risque que les conversations avec un passager conscient des conditions de conduite, qui deviennent souvent un sujet de conversation (Drews, Pasupathi et Strayer, 2004).



Lecture sur version imprimée? Scannez ce code QR avec votre cellulaire.

Attributions aux médias

- [«Understanding driver distraction»](#) (Comprendre la distraction au volant) par [American Psychological Association](#). Licence standard de YouTube.

CHAPITRE 3-L'ÉTHIQUE DE LA RECHERCHE



Image 3. Vaccination

En 1998, la revue médicale *The Lancet* a publié un article qui a intéressé de nombreux psychologues. Des personnes en recherche affirment avoir mis en évidence une relation statistique entre l'administration du vaccin combiné contre la rougeole, les oreillons et la rubéole (ROR) et le développement de l'autisme, suggérant en outre que le vaccin pouvait être à l'origine de l'autisme. En conséquence, de nombreux parents ont décidé de ne pas faire vacciner leurs enfants, ce qui les a évidemment exposés à une augmentation du risque de rougeole, d'oreillons et de rubéole. Toutefois, les études suivantes faites par d'autres personnes en recherche n'ont pas réussi à établir un lien statistique entre le vaccin ROR et l'autisme — et il est aujourd'hui largement admis par la communauté scientifique qu'il n'y a pas de lien. En outre, plusieurs problèmes graves concernant la recherche initiale ont été découverts. Parmi eux, le chercheur principal avait tout intérêt à tirer profit de ses conclusions, car il avait breveté un vaccin concurrent contre la rougeole. Il avait également utilisé des méthodes biaisées pour sélectionner et tester ses personnes participantes à la recherche, et leur avait

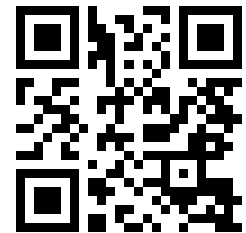
fait subir des procédures non approuvées et inutiles d'un point de vue médical. En 2010, *The Lancet* a rétracté l'article et le chercheur principal s'est vu retirer le droit d'exercer la médecine (Burns, 2010).

En 2011, Diederik Stapel, un psychologue social éminent et respecté de l'Université de Tilburg, aux Pays-Bas, a été reconnu coupable d'un crime académique audacieux : la fabrication de données (Jump, 2011). À la suite d'une enquête menée par plusieurs universités, Stapel a avoué avoir inventé les données d'au moins 55 études qu'il avait publiées dans des revues scientifiques depuis 2004. Cette révélation a choqué la communauté scientifique, y compris des collègues qui avaient consacré du temps et des ressources précieuses à concevoir et à mener des recherches reposant sur des résultats frauduleusement publiés par Stapel.

Encore plus tragique, Stapel a révélé avoir commis la même fraude dans 10 thèses de doctorat qu'il avait dirigées, ce qui a nui à la carrière universitaire de ses personnes ayant été sous sa supervision. À un niveau plus général toutefois, ses actions ont porté un coup sérieux au code d'honneur auquel les scientifiques adhèrent. La science est après tout un processus collectif de découverte qui exige que les personnes en recherche soient honnêtes au sujet de leurs travaux et de leurs résultats — que leurs hypothèses soient ou non confirmées par les données recueillies. Rompre cette confiance aussi gravement que Stapel l'a fait, mine les fondements mêmes de ce processus. Il va sans dire que Stapel a été suspendu de son poste à l'Université de Tilburg. De plus l'American Psychological Association lui a retiré le prix *Career Trajectory Award* qui lui avait été décerné en 2009, et le gouvernement néerlandais a lancé une enquête sur son mauvais usage des fonds de recherche. Stapel a depuis rendu le doctorat qu'il avait reçu de l'Université d'Amsterdam, notant que son « comportement des dernières années n'est pas compatible avec les devoirs associés au doctorat ». Il a également présenté ses excuses à ses collègues, déclarant : « J'ai échoué en tant que scientifique et chercheur. J'en éprouve de la honte et un profond regret. » (Carey, 2011)

En psychologie politique, une affaire litigieuse de données frauduleuses a entraîné la rétractation d'un article dans la prestigieuse revue *Science*, ainsi que le retrait d'une offre d'emploi de l'Université de Princeton. Michael LaCour, étudiant diplômé en science politique, avait publié avec Donald Green, professeur reconnu à l'Université Columbia, un résultat surprenant : interagir avec une personne homosexuelle faisant du porte-à-porte peut changer l'opinion de l'électorat sur l'égalité des droits pour les personnes homosexuelles. Mais il y avait une multitude d'éléments fabriqués par LaCour à propos de subventions, de prix et de l'approbation éthique. Bien que Green ait demandé la rétractation de l'article de *Science* sans consulter son co-auteur, LaCour maintient la validité de ses données (Singal, 2015).

Dans ce chapitre, nous allons explorer l'éthique de la recherche scientifique en psychologie. Nous commencerons par un cadre général de réflexion sur l'éthique de la recherche scientifique en psychologie. Nous examinerons ensuite certains codes éthiques spécifiques pour les scientifiques en biomédecine et en sciences du comportement — en nous concentrant sur le code d'éthique de l'American Psychological Association et l'Énoncé de politiques des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2 (2022). Enfin, nous examinerons quelques conseils pratiques pour mener une recherche éthique en psychologie.



Reading in print? Scan this QR code to view the video on your mobile device. Or go to youtu.be/o651YAVaYc



One or more interactive elements has been excluded from this version of the text. You can

view them online here:

<https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=67#oembed-1>

Attributions aux médias

[Non, l'autisme n'est pas causé par les vaccins](#) : par [France Culture](#) License Standard YouTube.

17. Fondements moraux de la recherche éthique

Objectifs d'apprentissage

1. Décrire un cadre simple de réflexion à propos des questions éthiques dans la recherche en psychologie.
2. Donner des exemples de plusieurs questions éthiques qui se posent dans la recherche en psychologie, y compris celles qui affectent les participantes et participants à la recherche, la communauté scientifique et la société en général.

L'**éthique** est la branche de la philosophie qui s'intéresse à la moralité — c'est-à-dire à ce que signifie un comportement moral et à la manière dont les gens peuvent atteindre cet objectif. Elle peut également se référer à un ensemble de principes et de pratiques qui fournissent une orientation morale dans un domaine particulier. Il existe une éthique des affaires, de la médecine, de l'enseignement et, bien sûr, de la recherche scientifique. Comme l'illustre le premier exemple, plusieurs types de problèmes éthiques peuvent se poser dans le cadre de la recherche scientifique, en particulier lorsqu'elle implique des êtres humains. C'est pourquoi il est utile de commencer par un cadre général pour réfléchir à ces questions.

Un cadre de réflexion sur l'éthique de la recherche

Le Tableau 3.1 présente un cadre de réflexion sur les questions éthiques de la recherche en psychologie. Les lignes du Tableau 3.1 représentent trois principes directeurs qui s'appliquent à la recherche scientifique : respect des personnes, préoccupation pour le bien-être et justice. (Ces principes sont adaptés de ceux du code éthique de l'Énoncé de politiques des trois conseils [EPTC 2]). Les colonnes du Tableau 3.1 représentent trois groupes de personnes concernées par la recherche scientifique : la communauté participante, la communauté à la recherche, la communauté scientifique et la société en général. L'idée est qu'un examen approfondi de l'éthique de tout projet de recherche doit tenir compte de la manière dont chacun des trois principes directeurs s'applique à chacun des trois groupes de personnes. Les lignes directrices de l'EPTC 2 suivent les trois principes directeurs suivants :

1. **Respect des personnes:** Il s'agit notamment de respecter l'autonomie des personnes participantes à la recherche en assurant un consentement libre, éclairé et continu, ainsi que de protéger les personnes «

incapables d'exercer leur autonomie en raison de leur jeune âge, d'une maladie, de troubles cognitifs ou d'autres problèmes de santé mentale ».

2. **Préoccupation pour le bien-être:** Il s'agit notamment de veiller à ce que les personnes participantes ne soient pas exposées à des risques inutiles, de prendre en compte leur vie privée et de préserver leur confidentialité, ainsi que de fournir aux personnes participantes « suffisamment de renseignements pour que ces dernières puissent évaluer convenablement les avantages potentiels et les risques de leur participation à la recherche ».
3. **Justice:** Il s'agit de l'obligation de traiter les personnes de manière juste et équitable, notamment en tenant compte de la vulnérabilité des personnes participantes et en veillant à ce que les personnes historiquement marginalisées (y compris les minorités ethnoculturelles) ne soient pas injustement exclues des possibilités de recherche.

Tableau 3.1 Cadre de réflexion sur les questions éthiques dans la recherche scientifique

Principes directeurs	Qui est concerné?		
	La communauté participante	La communauté scientifique	La société
Peser les risques et les avantages			
Agir de façon responsable et intègre			
En quête de justice			
Respecter les droits et la dignité			

Principes directeurs

Il peut sembler évident de distinguer ce qui est éthique de ce qui ne l'est pas, mais deux études récentes nous montrent que les scientifiques brouillent souvent les pistes dans l'intérêt de la science. Prenons l'exemple de l'étude originale de Stanley Milgram sur l'obéissance à l'autorité (Milgram, 1963). Les gens sont informés de leur participation à une étude des effets de la punition sur l'apprentissage. Ces gens reçoivent pour instruction de donner des décharges électriques à une autre personne participante chaque fois que cette personne répond de manière incorrecte à une tâche d'apprentissage. À chaque réponse incorrecte, la décharge devient plus forte, ce qui finit par amener l'autre personne participante (qui se trouve dans la pièce voisine) à protester, à se plaindre de son cœur, à crier de douleur, puis à se taire et à cesser de répondre. Si la première personne hésite ou exprime une inquiétude, le chercheur lui répond de continuer. En réalité, l'autre personne participante est **complice** de la personne responsable de l'étude — une aide qui prétend être un·e vrai·e participant·e — et les protestations, plaintes et cris, entendus par la vraie personne participante, s'avèrent être un enregistrement audio activé lorsque cette dernière appuie sur l'interrupteur pour administrer les « décharges ». Le résultat surprenant de cette étude est que la plupart des personnes participantes véridiques ont continué à administrer les décharges malgré les protestations, les plaintes et les cris de l'aide complice.

Un autre exemple tragique est l'étude sur la syphilis de Tuskegee, menée par le service de santé publique des États-Unis de 1932 à 1972 (Reverby, 2009). Les participants à cette étude étaient des hommes afro-américains pauvres des

environs de Tuskegee, en Alabama, à qui l'on a dit qu'ils étaient traités pour du « mauvais sang ». Même s'ils recevaient des soins médicaux gratuits, ils n'étaient pas traités pour leur syphilis. Au lieu de cela, on les observait pour voir comment la maladie allait se développer chez les patients non traités. Même si la pénicilline était le traitement standard de la syphilis dans les années 1940, ces participants continuaient à se voir refuser un traitement sans avoir la possibilité de quitter l'étude. Il a fallu l'intervention de journalistes et de militants et la révélation de ces détails au grand public pour que l'étude soit finalement interrompue.

Bien qu'avantageux pour la communauté scientifique, ces deux exemples d'expériences non éthiques ont amené les personnes responsables de recherches à développer des cadres éthiques régissant leurs activités. Examinons de plus près chacun des principes directeurs de l'EPTC 2 (2022) et la manière dont ils peuvent être appliqués à chacun des trois groupes.

Respect des personnes

Les personnes responsables de recherches doivent respecter les droits et la dignité des personnes en tant qu'êtres humains. L'un de ces éléments est de respecter leur **autonomie** — le droit de faire ses propres choix et d'entreprendre ses propres actions sans contrainte. Le concept de **consentement éclairé** revêt ici une importance fondamentale. Cela signifie que les personnes menant une recherche doivent obtenir et documenter l'accord des personnes pour participer à une étude après les avoir informées de tout ce qui peut raisonnablement influencer sur leur décision. Prenons l'exemple des participants à l'étude Tuskegee. Bien qu'ils acceptent de participer à l'étude, ils ne sont pas informés qu'ils sont atteints de la syphilis et que le traitement leur sera refusé. Si on les avait informés de ce fait fondamental, ils n'auraient probablement pas accepté de participer à l'étude. De même si l'on avait informé les personnes participantes dans l'étude de Milgram qu'elles risquaient d'être « réduites à l'état d'épave bégayante et vacillante », il est probable que nombre d'entre elles n'auraient pas accepté d'y participer. Les personnes participantes n'ont pas donné un véritable consentement éclairé dans aucune de ces deux études.

Il peut y avoir des populations vulnérables dont il faut tenir compte et se demander dans quelle mesure elles sont capables d'accorder leur consentement. Les enfants sont un exemple où, légalement, ils ne sont pas en mesure d'accorder leur consentement en raison de leur possible manque de compréhension. Souvent si les enfants peuvent comprendre le langage, les personnes responsables de la recherche leur demanderont leur accord après avoir reçu le consentement de leurs parents ou de leurs tuteurs ou tutrices. Les adultes dont les capacités sont réduites constituent également un autre groupe qui peut ne pas être en mesure de donner son consentement en raison de limitations mentales ou physiques (par exemple les patient·e·s dans le coma). La recherche avec ces populations peut être extrêmement précieuse, et des mesures supplémentaires doivent être mises en place pour assurer le respect de leurs intérêts et de leur bien-être. Le consentement doit être donné volontairement et de manière continue. Les personnes menant des recherches ont le devoir d'informer les personnes participantes de toute modification du projet susceptible d'affecter leur volonté de consentir à leur participation.

Préoccupation pour le bien-être

La recherche scientifique en psychologie est éthique seulement si ses risques sont surpassés par ses avantages. Parmi les risques encourus par les personnes participantes à la recherche, il est possible qu'un traitement ne soit pas utile, voire nocif; qu'une procédure entraîne des dommages physiques ou psychologiques, et une violation de leur droit à la protection privée. Parmi les avantages potentiels, il y a le fait de recevoir un traitement utile, d'apprendre des choses sur la psychologie, d'éprouver la satisfaction de contribuer aux connaissances scientifiques et de recevoir de l'argent ou un crédit de cours en échange de leur participation. La recherche scientifique peut présenter des risques et des avantages pour la communauté scientifique et la société aussi (Rosenthal, 1994). Un risque pour la science arrive quand une question de recherche est inintéressante ou une étude est mal conçue; le temps, l'argent et les efforts consacrés à cette recherche auraient pu être consacrés à des recherches plus productives. Le risque pour la société est que les résultats de la recherche soient mal compris ou mal appliqués, avec des conséquences néfastes. La recherche faisant l'association erronée entre le vaccin contre la rougeole, la rubéole et les oreillons (RRO) et l'autisme a engendré ces deux types de conséquences négatives. Bien entendu, les avantages de la recherche scientifique pour la science et la société sont qu'elle fait progresser les connaissances scientifiques et peut contribuer au bien-être de la société.

La comparaison entre les risques et les avantages d'une recherche n'est pas toujours aisée, car les risques et les avantages ne sont pas toujours directement comparables. Par exemple, il est fréquent que les risques d'une étude concernent principalement les personnes participantes à la recherche, alors que les avantages concernent principalement la science ou la société. Bien que les conclusions de l'étude de Milgram soient considérées comme l'un des résultats les plus importants de la psychologie — avec des implications pour comprendre des événements, tels que l'Holocauste ou les mauvais traitements infligés aux prisonniers par les soldats américains à Abu Ghraib — cette connaissance a été obtenue au prix d'un stress psychologique important chez les personnes participantes à la recherche qui croyaient administrer de vraies décharges électriques.

Un autre élément de préoccupation pour le bien-être des personnes participantes est le **respect de leur vie privée** — leur droit de décider quelles informations les concernant sont partagées avec d'autres. Les personnes menant des recherches doivent respecter la **confidentialité**, c'est-à-dire s'engager à ne pas divulguer les informations personnelles des personnes participantes sans leur consentement ou une autorisation légale appropriée. Dans l'idéal, les personnes participantes peuvent conserver l'anonymat, c'est-à-dire que leur nom et d'autres informations permettant de les identifier ne sont pas du tout collectés.

Cela en valait-il la peine?

Une grande partie du débat sur l'éthique de l'étude de Milgram sur l'obéissance concerne la question du bien-fondé des connaissances scientifiques acquises par rapport aux dommages causés aux personnes participantes à l'étude. Pour mieux comprendre ce préjudice, il suffit de se reporter à la description qu'en fait Milgram lui-même (1963).

Dans un grand nombre de cas, le degré de tension a atteint des extrêmes rarement observés dans les études

sociopsychologiques en laboratoire. Les sujets ont été observés en train de transpirer, de trembler, de bégayer, de se mordre les lèvres, de gémir et de s'enfoncer les ongles dans la chair... Quatorze des 40 sujets ont montré des signes évidents de rires nerveux et de sourires. Ces rires semblaient tout à fait déplacés, voire bizarres. Des crises incontrôlables [de rire] ont été observées chez trois sujets. À une occasion, nous avons observé une crise si violemment convulsive qu'il a été nécessaire d'interrompre l'expérience (p. 375).

Milgram a également noté qu'une autre personne responsable de l'observation avait rapporté qu'en l'espace de 20 minutes, une personne participante « en était réduite à un état de ruine tremblotante, bégayante, s'approchant rapidement du point d'effondrement nerveux » (p. 377).

Il faut reconnaître que Milgram s'est donné beaucoup de mal pour débriefer ses personnes participantes — notamment en ramenant leur état mental à la normale — et pour montrer que la plupart d'entre elles pensaient que la recherche était utile et que ces personnes étaient heureuses d'y avoir participé. Pourtant, cette recherche serait considérée comme contraire à l'éthique selon les normes d'aujourd'hui.

En quête de justice

Les personnes responsables de recherches doivent les mener de manière juste. Elles doivent traiter leurs personnes participantes correctement, par exemple en leur accordant une compensation adéquate pour leur participation, et en veillant à ce que les avantages et les risques soient distribués entre toutes les personnes participantes. Par exemple dans le cadre d'une étude portant sur une nouvelle psychothérapie potentiellement bénéfique, des personnes participantes peuvent recevoir la psychothérapie tandis que d'autres serviront de groupe contrôle ne recevant aucun traitement. Si la psychothérapie s'avère efficace, il serait juste de la proposer aux personnes participantes du groupe contrôle à la fin de l'étude.

À un niveau sociétal plus large, les membres de certains groupes ont historiquement fait face à plus que leur juste part des risques de la recherche scientifique, notamment les personnes institutionnalisées, handicapées ou appartenant à des minorités raciales ou ethniques. Dans l'étude de Tuskegee, les hommes afro-américains pauvres étaient ciblés pour l'étude et étaient particulièrement vulnérables en raison de leur statut dans la société. Il est aujourd'hui largement reconnu que les personnes menant des recherches doivent prendre en compte les questions de justice et d'équité au niveau de la société.

Pour être justes, les personnes menant des recherches doivent agir de manière responsable et intègre. Ces gens doivent mener leurs recherches de manière approfondie et compétente, respecter leurs obligations professionnelles et faire preuve de sincérité. L'intégrité est importante, car elle favorise la confiance, qui est un élément essentiel de toute relation humaine efficace. Les personnes participantes doivent pouvoir croire que les personnes menant des recherches sont honnêtes avec elles (par exemple sur ce que l'étude implique), tiendront leurs promesses (par exemple de maintenir la confidentialité) et mèneront leur recherche de manière à en maximiser les avantages et à en minimiser les risques. Une question importante à cet égard est le recours à la duperie. Il est difficile, voire impossible, de répondre à certaines questions de recherche (comme celle de Milgram) sans tromper les personnes participantes. Agir avec intégrité peut donc entrer en conflit avec le fait de mener des recherches qui font progresser les connaissances scientifiques et profitent à la société. Nous verrons bientôt comment les psychologues gèrent généralement ce conflit.

La communauté scientifique et la société doivent également être en mesure de croire que les personnes mènent leurs recherches avec rigueur et compétence et les rapportent de manière honnête. Une fois encore,

l'exemple du début du chapitre illustre ce qui peut se produire lorsque cette confiance n'est pas respectée. Dans ce cas, d'autres scientifiques ont gaspillé des ressources sur des recherches de suivi inutiles, et des personnes ont évité le vaccin ROR, exposant ainsi leurs enfants à un risque accru de rougeole, d'oreillons et de rubéole.

« Ils ont été trahis »

En 1997 — 65 ans après le début de l'étude Tuskegee sur la syphilis et 25 ans après sa fin — le président Bill Clinton a présenté des excuses officielles, au nom du gouvernement américain, aux personnes concernées. Voici un extrait traduit de ces excuses :

Aujourd'hui, l'Amérique se souvient des centaines d'hommes utilisés pour la recherche à leur insu et sans leur consentement. Nous nous souvenons d'eux et des membres de leur famille. Des hommes qui étaient pauvres et afro-américains, sans ressource et avec peu d'alternatives, pensaient avoir trouvé un espoir lorsqu'on leur a offert des soins médicaux gratuits par le Service de santé publique des États-Unis. Ils ont été trahis.

Le texte intégral des excuses est disponible à l'adresse suivante [\[en anglais\]](#)

Conflit éthique inévitable

Il est peut-être déjà clair que les conflits éthiques dans la recherche en psychologie sont inévitables. Étant donné que peu de recherches en psychologie, voire aucune, ne sont totalement exemptes de risques, il y aura presque toujours un conflit entre les risques et les avantages. Une recherche avantageuse pour un groupe (par exemple, la communauté scientifique) peut être néfaste pour un autre (par exemple, les personnes participantes à la recherche), ce qui rend les compromis particulièrement difficiles. Nous avons également vu que le fait de dire toute la vérité aux personnes participantes à la recherche peut rendre difficile voire impossible, la réalisation d'études scientifiquement valables sur des questions importantes.

Bien entendu, de nombreux conflits éthiques sont assez faciles à résoudre. Presque tout le monde est d'accord pour dire qu'il n'est pas justifié de tromper les personnes participantes à une recherche et de leur faire subir un préjudice physique pour combler une petite lacune dans la littérature de recherche. Mais de nombreux conflits éthiques ne sont pas faciles à résoudre, et des personnes menant des recherches avec de bonnes intentions et de bonnes compétences peuvent être en désaccord sur la manière de les résoudre. Prenons l'exemple d'une étude sur « l'espace personnel » dans des toilettes publiques pour hommes (Middlemist, Knowles et Matter, 1976). On observe secrètement les hommes pour voir s'il leur faut plus de temps pour commencer à uriner lorsqu'un autre homme (un complice des personnes responsables de la recherche) se trouve à l'urinoir voisin. Si certains critiques estiment qu'il s'agit d'une atteinte injustifiée à la dignité humaine (Koocher, 1977), les personnes responsables de la recherche ont soigneusement examiné les conflits éthiques, les ont résolus de leur mieux et ont conclu que les avantages de la recherche l'emportaient sur les risques (Middlemist, Knowles et Matter, 1977). Par exemple, les personnes responsables de la recherche ont interviewé quelques individus participants préliminaires et ont constaté qu'aucun d'entre eux n'était gêné par le fait d'avoir été observé.

Ce qu'il faut retenir, c'est que même s'il n'est pas possible d'éliminer complètement les conflits éthiques, il est possible de les gérer de manière responsable et constructive. En général, cela signifie qu'il faut réfléchir soigneusement et en profondeur aux questions éthiques soulevées, minimiser les risques et peser les risques par rapport aux avantages. Cela signifie également être capable d'expliquer ses décisions éthiques aux autres, de demander une rétroaction à leur sujet et, en fin de compte, d'en assumer la responsabilité.

18. Des principes moraux aux codes éthiques

Objectifs d'apprentissage

1. Décrire l'histoire des codes d'éthique pour la recherche scientifique avec des êtres humains.
2. Résumer l'énoncé de politique des trois conseils sur l'éthique de la recherche avec des êtres humains — en particulier en ce qui concerne le consentement éclairé, la duperie, le débrefage, la recherche avec des animaux non humains et l'intégrité scientifique.

Les principes directeurs du respect des personnes, de préoccupation pour le bien-être et de justice constituent un point de départ utile pour réfléchir à l'éthique de la recherche en psychologie, car tout le monde est d'accord sur ces principes. Toutefois comme nous l'avons vu, même les personnes qui s'accordent sur ces principes directeurs peuvent être en désaccord sur des questions éthiques spécifiques qui se posent au cours de la recherche. C'est pourquoi il existe également des codes éthiques plus détaillés et applicables fournissant des conseils sur des questions importantes qui se posent fréquemment. Dans cette section, nous commençons par un aperçu historique de ces codes éthiques, puis nous examinons de près celui qui est le plus pertinent en recherche en psychologie au Canada — l'énoncé de politique des trois conseils (EPTC 2 (2022)).

Aperçu historique

Un des premiers codes éthiques est le **code de Nuremberg**, un ensemble de dix principes rédigés en 1947 à l'occasion des procès de médecins nazis accusés d'avoir mené des recherches d'une cruauté choquante sur des prisonniers de camps de concentration pendant la Seconde Guerre mondiale. Le code de Nuremberg a fourni une norme permettant de comparer le comportement des hommes jugés, dont beaucoup ont finalement été reconnus coupables et emprisonnés ou condamnés à mort. Ce code était particulièrement clair sur l'importance de peser soigneusement les risques par rapport aux avantages et sur la nécessité d'un consentement éclairé. La **déclaration d'Helsinki** est un code d'éthique similaire créé par l'Organisation mondiale de la santé en 1964. Parmi les normes qu'elle a ajoutées au code de Nuremberg figure celle selon laquelle la recherche avec des êtres humains doit être fondée sur un **protocole** écrit — une description détaillée de la recherche — qui est examiné par un comité indépendant. La déclaration d'Helsinki a été révisée à plusieurs reprises, la dernière fois en 2013.

Aux États-Unis, les inquiétudes suscitées par l'étude Tuskegee et d'autres études ont conduit à la

publication, en 1978, d'un ensemble de lignes directrices fédérales appelées **rapport Belmont** (*Belmont Report* en anglais). Celui-ci reconnaît explicitement le principe de la recherche sur la **justice**, y compris l'importance de mener la recherche de manière à répartir équitablement les risques et les avantages entre les différents groupes au niveau de la société. Il reconnaît également l'importance du **respect des personnes**, c'est-à-dire l'autonomie des gens et la protection des personnes dont l'autonomie est réduite (par exemple, les prisonniers, les enfants), et se traduit par la nécessité d'un consentement éclairé. Enfin, il reconnaît le principe de bienfaits, qui souligne l'importance de maximiser les avantages de la recherche tout en minimisant les inconvénients pour les participantes et participants ainsi que pour la société. Finalement, il reconnaît le principe des **bienfaits**, lequel souligne l'importance de maximiser les avantages de la recherche tout en minimisant les inconvénients pour les personnes participantes et la société. Le rapport Belmont a joué un rôle important dans l'élaboration de lignes directrices nationales en matière d'éthique de la recherche aux États-Unis et au Canada.

Les liens qui suivent vous permettent de lire dans leur intégralité les codes d'éthique dont il est question dans cette section. Ils sont tous fortement recommandés et sont courts et faciles à lire en français.

- Le Code de Nuremberg : frq.gouv.qc.ca/app/uploads/2021/03/code_nuremberg_1947.pdf
- La Déclaration d'Helsinki de L'Association Médicale Mondiale (AMM) : www.frq.gouv.qc.ca/app/uploads/2021/03/declaration_helsinki_2013.pdf
- Le Rapport Belmont : frq.gouv.qc.ca/app/uploads/2021/04/rapport_belmont_1974.pdf

Au Canada, les scientifiques et les institutions de recherche doivent respecter le code d'éthique formellement établi dans l'**énoncé de politique des trois conseils : éthique de la recherche avec des êtres humains** — EPTC 2 (2022). Le terme « Tri-Council » [en anglais] et le terme « les trois conseils » en français sont couramment utilisés pour décrire collectivement les trois organismes subventionnaires de la recherche qui sont financés par le gouvernement fédéral canadien : le Conseil de recherches en sciences humaines du Canada (CRSH), les Instituts de recherche en santé du Canada (IRSC) et le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG). La première édition de l'Énoncé de politique des trois Conseils (EPTC) a été publiée en 1998. Elle remplaçait alors toutes les lignes directrices antérieures qui avaient été élaborées par des agences et des institutions individuelles. En 2010, une deuxième édition (EPTC 2) a été publiée pour consolider les principes directeurs, clarifier et mettre à jour les lignes directrices, actualiser la terminologie et clarifier la portée et les responsabilités des comités institutionnels d'éthique de la recherche (CÉR). La dernière version de la politique a été publiée en 2022. Les lignes directrices de l'EPTC 2 (2022) reposent sur les trois principes fondamentaux déjà évoqués : le respect des personnes, la préoccupation pour le bien-être et la justice.

Un tutoriel en ligne détaillé des lignes directrices spécifiques contenues dans l'EPTC 2 (2022) est disponible sur le site web suivant : tcps2core.ca/welcome?lang=fr

La réussite du tutoriel prend jusqu'à 4 heures et donne à la personne utilisatrice un certificat qui est maintenant exigé par de nombreuses universités et autres institutions de recherche avant qu'une proposition de recherche ne soit évaluée par le **comité d'éthique de la recherche (CÉR)** de l'institution — un comité

responsable de l'examen des protocoles de recherche pour les problèmes éthiques potentiels. Un CÉR doit être composé d'au moins cinq personnes aux profils variés, dont au moins deux membres spécialisés dans les disciplines de recherche concernées, au moins un membre compétent en matière d'éthique et au moins un membre de la communauté qui n'a pas d'affiliation avec l'établissement. Le CÉR s'assure que les risques de la recherche proposée sont réduits au minimum, que les avantages l'emportent sur les risques, que la recherche est menée de manière équitable et que la procédure de consentement éclairé est adéquate.

L'EPTC 2 (2022) comporte deux niveaux d'évaluation de l'éthique de la recherche. L'**évaluation de l'éthique de la recherche par le CÉR en comité plénier** (*Full REB review* en anglais) est l'évaluation par défaut pour les recherches avec des êtres humains. Cependant, pour la recherche dans laquelle «la probabilité et l'ampleur des préjudices éventuels, découlant de la participation à la recherche, ne sont pas plus grandes que celles des préjudices inhérents aux aspects de la vie quotidienne de la personne participante qui sont associés à la recherche», le CÉR pourrait la considérer comme une **recherche à risque minimal**. Dans ce cas (comme c'est parfois le cas pour les enquêtes, les études basées sur des questionnaires et les études d'observation naturaliste), le CÉR peut déléguer l'évaluation de l'éthique de la recherche à un ou plusieurs de ses membres. Une autre exception est faite dans le cas de recherches menées par des personnes étudiantes effectuant une recherche dans le cadre d'un cours, auquel cas le CÉR peut déléguer l'évaluation à la faculté ou au département concerné. Notons par ailleurs, que certaines recherches satisfaisant des critères bien établis sont exemptées de l'évaluation par un CÉR (articles 2.2 à 2.4 dans l'EPTC 2 2022).



Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou allez sur <https://ccac.ca/fr/lignes-directrices-et-politiques/principes-fondamentaux.html>

Au Canada, bien que les scientifiques et les établissements de recherche suivent officiellement les lignes directrices détaillées dans l'Énoncé de politique des trois Conseils 2 (EPTC 2 2022), les personnes étudiant en psychologie devraient également connaître le Code d'éthique de l'American Psychological Association (APA), qui partage de nombreux principes et procédures avec l'EPTC 2 2022.

Les principes éthiques de l'APA pour les psychologues et leur code de conduite (également connus sous le nom de code de déontologie de l'APA, en anglais *APA Ethics Code*) ont été publiés pour la première fois en 1953, ils ont été révisés à plusieurs reprises. La dernière révision a été faite en 2002. Il comprend environ 150 normes éthiques spécifiques que les psychologues et leurs étudiant.e.s sont censés respecter. Une grande partie du code de déontologie de l'APA concerne la pratique clinique de la psychologie — la publicité des services, la fixation et la perception des honoraires, les relations personnelles avec les clients, etc. En ce qui concerne la recherche, la partie la plus pertinente pour notre objectif c'est la norme 8 : recherche et publication (en anglais, *Standard 8: Research and Publication*) présentée dans l'encadré un peu plus loin.

Une différence globale entre l'éthique canadienne et l'éthique américaine est d'ordre organisationnel. Aux États-Unis, l'APA supervise la pratique clinique en psychologie, la recherche sur des êtres humains ainsi que la recherche sur des participants animaux non humains. Au Canada, l'évaluation éthique de la recherche sur les animaux est supervisée par le Conseil canadien de protection des animaux. La pratique clinique des psychologues doit suivre le Code canadien d'éthique pour les psychologues, lequel est géré par la Société canadienne de psychologie, et la conduite éthique de la recherche avec des êtres humains est sous la responsabilité des trois organismes de recherche fédéraux.



Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou allez sur <https://cpa.ca/fr/aproposdelascp/comites/ethics/codeofethics/>

- Lignes directrices et politiques du [Conseil canadien de protection des animaux](#)
- Code canadien d'éthique pour les psychologues de la [Société canadienne de psychologie](#)

Dans les pages qui suivent, nous allons examiner plus en détail certains des aspects les plus importants qu'ont en commun l'EPTC 2 (2022) et l'APA – le consentement éclairé, la duperie, le débriefage et l'intégrité scientifique.

Code éthique de l'APA

Norme 8 : Recherche et publication

8.01 Approbation institutionnelle

Lorsque l'approbation de l'institution est requise, les chercheuses et chercheurs en psychologie fournissent des informations précises sur leurs propositions de recherche et obtiennent l'approbation avant de mener la recherche. Ils mènent la recherche conformément au protocole de recherche approuvé.

8.02 Consentement éclairé à la recherche

- Lors de l'obtention du consentement éclairé, conformément à la norme 3.10, Consentement éclairé, les psychologues informent la personne participante **(1)** du but de la recherche, de sa durée prévue et des procédures; **(2)** de leur droit de refuser de participer et de se retirer de la recherche une fois que la participation a commencé; **(3)** des conséquences prévisibles d'un refus ou d'un retrait; **(4)** les facteurs raisonnablement prévisibles susceptibles d'influencer leur volonté de participer, tels que les risques potentiels, l'inconfort ou les effets indésirables; **(5)** les éventuels bénéfices de la recherche; **(6)** les limites de la confidentialité; **(7)** les incitations à la participation; et **(8)** les personnes à contacter pour toute question concernant la recherche et les droits de la personne participante à la recherche. Les psychologues donnent

aux personnes participantes potentielles la possibilité de poser des questions et d'obtenir des réponses. (Voir aussi les normes 8.03, Consentement éclairé à l'enregistrement de voix et d'images dans la recherche; 8.05, Se dispenser du consentement éclairé à la recherche; et 8.07, Duperie dans la recherche).

- b. Les psychologues qui mènent des recherches interventionnelles impliquant l'utilisation de traitements expérimentaux précisent à la personne participante, dès le début de la recherche, **(1)** la nature expérimentale du traitement, **(2)** les services qui seront ou non offerts au-x groupe-s contrôle-s le cas échéant, **(3)** les moyens par lesquels l'attribution des conditions traitement et contrôle aux personnes sera faite, **(4)** les options de traitements alternatifs disponibles si une personne ne souhaite pas participer à la recherche ou souhaite se retirer une fois que l'étude a commencé, et **(5)** la compensation ou les coûts monétaires de la participation, y compris, le cas échéant, si un remboursement aura été sollicité par la personne participante ou par une autre personne payeuse. (Voir également la norme 8.02a, Consentement éclairé à la recherche).

8.03 Consentement éclairé pour l'enregistrement de voix et d'images dans la recherche

Les psychologues obtiennent le consentement éclairé des personnes participantes à la recherche avant d'enregistrer leur voix ou leur image pour la collecte de données, sauf si **(1)** la recherche consiste uniquement en des observations naturalistes dans des lieux publics, et qu'il n'est pas prévu que l'enregistrement soit utilisé d'une manière susceptible d'entraîner une identification personnelle ou un préjudice, ou **(2)** la conception de la recherche inclut la duperie, et que le consentement à l'utilisation de l'enregistrement est obtenu au cours du débriefage. (Voir aussi la norme 8.07, duperie dans la recherche).

8.04 Personnes clientes/patientes, étudiantes et participantes subordonnées à la recherche

- a. Lorsque les psychologues mènent des recherches auxquelles participent des personnes clientes/patientes, étudiantes ou subordonnées, ces psychologues prennent des mesures pour protéger les personnes participantes potentielles contre les conséquences négatives d'un refus ou d'un retrait de participation.
- b. Lorsque la participation à la recherche est une exigence du cours ou une possibilité d'obtenir des crédits supplémentaires, la personne participante potentielle se voit offrir le choix d'activités alternatives équitables.

8.05 Exemptions de consentement éclairé pour la recherche

Les psychologues peuvent ne pas obtenir le consentement éclairé uniquement **(1)** lorsque la recherche ne risque pas de créer une détresse ou un préjudice, et qu'elle concerne **(a)** l'étude de pratiques éducatives normales, de programmes d'études ou de méthodes de gestion de classe dans un contexte éducatif; **(b)** uniquement des questionnaires anonymes, des observations naturalistes ou des recherches dans les archives pour lesquelles la divulgation des réponses n'exposerait pas les personnes participantes à un risque de responsabilité pénale ou civile ou ne porterait pas atteinte à leur situation financière, à leur employabilité ou à leur réputation, et pour lesquelles la confidentialité est protégée; ou **(c)** l'étude de facteurs liés à l'efficacité professionnelle ou organisationnelle menée dans des environnements organisationnels pour lesquels il n'y a pas

de risque pour l'employabilité des personnes participantes, et dont la confidentialité est protégée ou (2) lorsque la loi ou les réglementations fédérales ou institutionnelles l'autorisent.

8.06 Proposer des incitations à la participation à la recherche

- a. Les psychologues font des efforts raisonnables pour éviter d'offrir des incitations financières ou autres incitations excessives ou inappropriées, pour la participation à la recherche, lorsque ces incitations sont susceptibles de contraindre à la participation.
- b. Lorsque les psychologues offrent des services professionnels en guise d'incitation à la participation à la recherche, les psychologues clarifient la nature des services, ainsi que les risques, les obligations et les limites. (Voir également la norme 6.05, Troc avec les personnes clientes/patientes).
- c. Les psychologues ne mènent pas d'étude impliquant la duperie à moins d'avoir déterminé que l'utilisation de techniques de duperie est justifiée par la valeur prospective significative de l'étude au plan scientifique, éducatif ou appliqué, et qu'il n'est pas possible d'utiliser d'autres procédures efficaces sans duperie.
- d. Les psychologues ne dupent pas les personnes participantes potentielles au sujet d'une recherche dont on peut raisonnablement s'attendre à ce qu'elle provoque une douleur physique ou une détresse émotionnelle grave.
- e. Les psychologues expliquent toute duperie faisant partie intégrante du devis et de la conduite d'une expérience aux personnes participantes dès que possible, de préférence à la fin de leur participation, mais au plus tard à la fin de la collecte des données, et permettent aux personnes participantes de retirer leurs données. (Voir également la norme 8.08, Débriefage).

8.08 Débriefage

- a. Les psychologues offrent aux personnes participantes la possibilité d'obtenir rapidement des informations appropriées sur les sujets suivants: la nature, les résultats et les conclusions de la recherche, et ces psychologues prennent des mesures raisonnables pour corriger toute idée fausse que les personnes participantes pourraient avoir et dont les psychologues ont connaissance.
- b. Si des valeurs scientifiques ou humaines justifient de retarder ou de ne pas divulguer ces informations, les psychologues prennent des mesures raisonnables pour réduire le risque de préjudice.
- c. Lorsque les psychologues se rendent compte que les procédures de recherche ont porté préjudice à une personne participante, ces psychologues prennent des mesures raisonnables pour minimiser ce préjudice.

8.09 Soins et utilisation sans cruauté des animaux dans la recherche

- a. Les psychologues acquièrent, soignent, utilisent et éliminent les animaux conformément aux lois et réglementations fédérales, nationales et locales en vigueur, ainsi qu'aux normes professionnelles.
- b. Des psychologues formés aux méthodes de recherche et expérimentés dans le soin des animaux de laboratoire supervisent toutes les procédures impliquant des animaux et sont responsables de veiller à ce que leur confort, leur santé et leur traitement sans cruauté soient dûment pris en compte.
- c. Les psychologues s'assurent que toutes les personnes sous leur supervision utilisant des animaux ont reçu une formation sur les méthodes de recherche et sur les soins, l'entretien et la manipulation des espèces utilisées, dans la mesure appropriée de leur rôle. (Voir également la norme 2.05, Délégation du travail à d'autres personnes).

- d. Les psychologues font des efforts raisonnables pour minimiser l'inconfort, l'infection, la maladie et la douleur des sujets animaux.
- e. Les psychologues n'utilisent une procédure soumettant les animaux à la douleur, au stress ou à la privation que lorsqu'aucune procédure alternative n'est disponible et que l'objectif est justifié par sa valeur scientifique, éducative ou appliquée potentielle.
- f. Les psychologues pratiquent des interventions chirurgicales sous anesthésie appropriée et suivent des techniques visant à éviter les infections, et à minimiser la douleur pendant et après l'opération.
- g. Lorsqu'il convient de mettre fin à la vie d'un animal, les psychologues procèdent rapidement, en s'efforçant de minimiser la douleur et en respectant les procédures en vigueur.

8.10 Communication des résultats de la recherche

- a. Les psychologues ne fabriquent pas de données. (Voir également la norme 5.01a, Éviter les déclarations fausses ou déceptives).
- b. Si les psychologues découvrent des erreurs importantes dans leurs données publiées, ces psychologues prennent des mesures raisonnables pour corriger ces erreurs dans un rectificatif, une rétractation, un erratum ou tout autre moyen de publication approprié.

8.11 Plagiat

Les psychologues ne présentent pas des parties du travail ou des données d'une autre personne comme étant les leurs, même si l'autre travail ou la source de données est cité occasionnellement.

8.12 Crédit de publication

- a. Les psychologues assument la responsabilité et le crédit, y compris le crédit d'auteur ou d'autrice, seulement pour leur travail réellement effectué ou auquel ces psychologues ont contribué de manière substantielle. (Voir également la norme 8.12b, Crédit de publication).
- b. La qualité d'auteur ou autrice principale et les autres crédits de publication reflètent fidèlement les contributions scientifiques ou professionnelles relatives des personnes impliquées, quel que soit leur statut relatif. Le simple fait d'occuper un poste institutionnel, comme être à la direction d'un département, ne justifie pas l'attribution de la qualité d'auteur ou d'autrice. Les contributions mineures à la recherche ou à la rédaction de publications sont reconnues de manière appropriée, par exemple dans des notes de bas de page ou dans une déclaration d'introduction.
- c. Sauf circonstances exceptionnelles, une personne étudiante est auteur ou autrice principale de tout article à plusieurs auteurs qui repose en grande partie sur la thèse de doctorat de cette personne étudiante. Les guides pédagogiques discutent des crédits de publication avec les personnes étudiantes dès que possible, et tout au long du processus de recherche et de publication, le cas échéant. (Voir également la norme 8.12b, Crédit de publication).

8.13 Double publication de données

Les psychologues ne publient pas, en tant que données originales, des données qui ont déjà été publiées. Cela n'empêche pas de republier des données lorsqu'elles sont accompagnées d'une mention appropriée.

8.14 Partage des données de recherche à des fins de vérification

- a. Après la publication des résultats de la recherche, les psychologues ne cachent pas les données sur lesquelles se fondent leurs conclusions à d'autres personnes professionnelles compétentes qui cherchent à vérifier les affirmations substantielles par une nouvelle analyse, et qui ont l'intention d'utiliser ces données uniquement à cette fin, à condition que la confidentialité des personnes participantes puisse être protégée, et à moins que des droits légaux concernant des données exclusives ne s'opposent à leur divulgation. Cela n'empêche pas les psychologues d'exiger que ces personnes ou groupes prennent en charge les coûts liés à la transmission de ces informations.
- b. Les psychologues qui demandent des données à d'autres psychologues pour vérifier les affirmations substantielles par le biais d'une nouvelle analyse ne peuvent utiliser les données partagées que dans le but déclaré. Les psychologues qui en font la demande doivent obtenir un accord écrit préalable pour toute autre utilisation des données.

8.15 Examen

Les psychologues qui examinent les documents soumis pour présentation, publication, subvention ou examen de propositions de recherche respectent la confidentialité et les droits de propriété des personnes qui ont soumis ces informations.

Source: Vous pouvez lire tout le [code éthique de l'APA](#)

Consentement éclairé

Le consentement éclairé consiste à obtenir et à documenter l'accord des personnes pour participer à une étude, après les avoir informées de tout ce qui est raisonnablement susceptible d'influer sur leur décision. Bien informer les personnes participantes inclut les détails de la procédure, les risques et les avantages de la recherche, le fait qu'elles ont le droit de refuser de participer ou de se retirer de l'étude, les conséquences d'une telle décision et toute limite légale à la confidentialité. Par exemple certains endroits exigent des personnes menant des recherches, qui apprennent qu'un enfant a été maltraité ou qu'un autre crime a été commis, de signaler cette information aux autorités.

Bien que le processus d'obtention du consentement éclairé consiste souvent à faire lire et signer un **formulaire de consentement** aux personnes participantes, il est important de comprendre qu'il ne s'agit pas uniquement de cela. Bien que le fait de faire lire et signer un formulaire de consentement aux personnes participantes puisse être suffisant, lorsqu'il s'agit d'adultes compétents ayant les capacités et la motivation nécessaires, de nombreuses personnes participantes ne lisent pas réellement les formulaires de

consentement ou les lisent, mais ne les comprennent pas. Par exemple les personnes participantes confondent souvent les formulaires de consentement avec des documents juridiques, et croient à tort qu'en les signant, elles renoncent à leur droit de poursuivre le chercheur en justice (Mann, 1994). Par conséquent même avec des adultes compétents, il est bon d'informer les personnes participantes des risques et des avantages, de leur montrer la procédure, de leur demander s'il y a des questions, et de leur rappeler qu'il est possible de se retirer à tout moment, en plus de leur faire lire et signer un formulaire de consentement.

Notons également qu'il existe des situations dans lesquelles le consentement éclairé n'est pas nécessaire. Il s'agit notamment des situations où la recherche ne devrait pas causer de préjudice et où la procédure est simple, ou encore lorsque l'étude est menée dans le cadre des activités ordinaires des personnes. Par exemple si vous voulez vous asseoir à l'extérieur d'un bâtiment public et observer si les gens tiennent la porte ouverte aux personnes derrière eux, vous n'aurez pas besoin d'obtenir leur consentement éclairé. De même si un membre de corps professoral souhaite comparer deux méthodes d'enseignement légitimes dans deux sections de son cours de méthodologie de recherche, il n'est pas nécessaire d'obtenir le consentement éclairé des personnes inscrites à ses cours, à moins de planifier publier leurs résultats sur l'apprentissage dans une revue scientifique.

Duperie

La **duperie** des personnes participantes à une recherche en psychologie peut prendre diverses formes : désinformation des personnes participantes sur l'objectif d'une étude, recours à des aides complices, utilisation d'un équipement factice, comme le générateur de décharges de Milgram, et présentation aux personnes participantes de fausses informations sur leurs performances (par exemple, leur dire qu'elles ont obtenu un piètre résultats à un test alors qu'elles ont bien réussi). La duperie consiste également à ne pas informer les personnes participantes du devis complet ou de l'objectif réel de la recherche, même s'il n'y a pas désinformation active (Sieber, Iannuzzo et Rodriguez, 1995). Par exemple une étude sur l'apprentissage incident — l'apprentissage sans effort conscient — demande aux personnes participantes de lire une liste de mots en vue d'un « test de mémoire » ultérieur. Bien que les personnes participantes soient susceptibles de penser que le test de mémoire leur demandera de se souvenir des mots, il pourrait plutôt leur demander de se souvenir du contenu de la pièce ou de l'apparence des auxiliaires de recherche.

Certaines personnes responsables de recherches affirment que la duperie de personnes participantes à la recherche est rarement, voire jamais, justifiée d'un point de vue éthique. Elles font valoir, entre autres, qu'elle empêche les personnes participantes de donner un consentement véritablement éclairé, qu'elle ne respecte pas leur dignité en tant qu'êtres humains, qu'elle risque de les perturber, de les rendre méfiantes et donc moins honnêtes dans leurs réponses, et qu'elle nuit à la réputation des personnes menant des recherches dans ce domaine (Baumrind, 1985).

Il convient toutefois de noter que l'EPTC 2 (2022) et le code éthique de l'APA adoptent une approche plus modérée — autorisant la duperie lorsque les avantages de l'étude l'emportent sur les risques, que l'on ne peut raisonnablement s'attendre à ce que les personnes participantes subissent un préjudice, qu'il est impossible de répondre à la question de recherche sans recourir à la duperie, et que l'information concernant la duperie est donnée, dès que possible, aux personnes participantes. Cette approche reconnaît que toutes les formes de duperie ne sont pas aussi mauvaises les unes que les autres. Comparez par exemple l'étude de

Milgram, dans laquelle il trompe ses personnes participantes de plusieurs manières importantes entraînant un stress psychologique grave, avec une étude d'apprentissage incident dans laquelle un « test de mémoire » s'avère légèrement différent de ce à quoi les personnes participantes s'attendaient. Elle reconnaît également qu'il peut être difficile, voire impossible, de répondre à certaines questions de recherche importantes sur le plan scientifique et social sans tromper les personnes participantes. Savoir que le sujet de l'étude est jusqu'à quel point on obéit à l'autorité, on agit de manière agressive envers un pair ou on aide un étranger, est susceptible de modifier la manière dont les gens se comportent, de sorte que les résultats ne peuvent plus être généralisés dans le monde réel.

Débriefage

Le **débriefage** consiste à informer, dès que possible, les personnes participantes à la recherche de l'objectif de l'étude, de révéler toute duperie et de corriger toute autre idée fausse qu'il est possible d'avoir à la suite de leur participation. Le débriefage consiste également à minimiser les dommages pouvant survenir. Par exemple une expérience portant sur les effets d'une humeur triste sur la mémoire pourrait consister à induire une humeur triste chez les personnes participantes en leur faisant avoir des pensées tristes, en leur faisant regarder une vidéo triste ou écouter de la musique triste. Le débriefage est le moment de ramener l'humeur des personnes participantes à la normale en leur faisant penser à des choses joyeuses, en leur faisant regarder une vidéo joyeuse ou écouter de la musique joyeuse.

Intégrité scientifique

L'intégrité scientifique comprend les points évidents selon lesquels les personnes menant des recherches ne doivent pas fabriquer des données ou plagier. Le plagiat consiste à utiliser les mots ou les idées d'autrui sans le consentement de l'auteur ou de l'autrice. Une mention correcte signifie généralement que les citations directes sont mises entre guillemets et que la source de la citation ou de l'idée utilisée est citée. L'autoplagiat est également considéré comme contraire à l'éthique et consiste à publier le même matériel plus d'une fois. En d'autres termes, les personnes en recherche ne doivent pas reprendre des passages de leurs travaux déjà publiés, tout comme les personnes aux études ne doivent pas rendre le même travail dans plusieurs cours.

Les autres normes contiennent des points moins évidents, mais tout aussi importants. Les personnes menant des recherches ne doivent pas publier les mêmes données une deuxième fois, comme si elles étaient nouvelles. Elles doivent partager leurs données avec d'autres scientifiques et, en tant que personnes responsables de l'évaluation, les scientifiques doivent préserver la confidentialité des recherches examinées, mais non publiées. Il convient de noter que les noms des auteurs ou des autrices d'une recherche publiée — et l'ordre dans lequel ces noms apparaissent — doivent refléter l'importance de la contribution de chacun à la recherche. Il est contraire à l'éthique par exemple d'inclure comme auteur ou autrice une personne qui n'a apporté qu'une contribution mineure à la recherche (par exemple l'analyse de certaines données) ou qu'un membre du corps enseignant se désigne comme premier auteur d'une recherche qui a été menée en grande partie par une personne étudiant sous sa supervision.

19. Mettre l'éthique en pratique

Objectifs d'apprentissage

1. Décrire plusieurs stratégies permettant d'identifier et de minimiser les risques et la duperie dans la recherche en psychologie.
2. Élaborer des procédures complètes de consentement éclairé et de débriefage, y compris un formulaire de consentement.

Dans cette section, nous examinons quelques conseils pratiques pour mener une recherche éthique en psychologie. Une fois de plus, il est important de se rappeler que les questions éthiques se posent bien avant que vous ne commenciez à collecter des données, et qu'elles continuent à se poser jusqu'à la publication et au-delà.

Connaître et accepter ses responsabilités éthiques

Comme l'énoncé de politique des trois conseils sur l'éthique de la recherche avec des êtres humains — EPTC 2 (2022) le mentionne au début de son chapitre 1, « Étant donné l'importance primordiale de la recherche et de la participation d'êtres humains à la recherche, nous devons, en tant que société, prendre tous les moyens possibles pour que la recherche soit menée de façon éthique afin de préserver la confiance du public. » C'est pourquoi la toute première chose à faire en tant que personne nouvellement responsable de mener une recherche est de connaître et d'accepter ses responsabilités éthiques. Cela implique au minimum de lire et de comprendre l'EPTC 2 (2022), de faire la distinction entre une recherche à risque minimal et une recherche requérant l'évaluation en comité plénier, et de connaître les politiques et procédures spécifiques de votre établissement, y compris la manière de préparer et de soumettre un protocole de recherche pour évaluation par le comité d'éthique de la recherche (CÉR). Si vous effectuez une recherche dans le cadre d'un cours obligatoire, il peut exister des normes, des politiques et des procédures spécifiques à ce cours. Si une norme, une politique ou une procédure n'est pas claire — ou si vous n'êtes pas sûr de la marche à suivre en cas de problème éthique — vous devez demander des éclaircissements. Vous pouvez le faire en consultant les codes d'éthique pertinents, en lisant comment des problèmes similaires ont été résolus par d'autres, ou en consultant des personnes ayant plus d'expérience en recherche, votre CÉR ou la personne responsable du cours. Ultiment c'est vous, en tant que personne menant une recherche, qui devez en assumer la responsabilité éthique.

Identifier et minimiser les risques

Lorsque vous concevez votre étude, vous devez identifier et minimiser les risques pour les personnes participantes. Commencez par dresser la liste de tous les risques, y compris les risques de dommages physiques et psychologiques, et de violation de la confidentialité. Rappelez-vous qu'il est facile pour les personnes menant des recherches de considérer les risques comme moins graves que les personnes participantes, voire de les ignorer complètement. Par exemple une personne en recherche veut tester la sensibilité des gens aux images violentes en leur montrant des photographies horribles de scènes de crimes et d'accidents. Cependant parce que cette personne était membre de l'équipe technique en médecine d'urgence, elle a sous-estimé le caractère dérangeant de ces images pour la plupart des gens. N'oubliez pas non plus que certains risques ne s'appliquent qu'à certaines personnes participantes. Par exemple alors que la plupart des gens n'ont aucun problème à répondre à une enquête sur leur peur de différents crimes, les gens qui ont été victimes de l'un de ces crimes pourraient être perturbés. C'est pourquoi vous devez demander l'avis de diverses personnes, y compris les membres de votre équipe de recherche, des personnes ayant plus d'expérience en recherche, et même des non-scientifiques qui peuvent être plus en mesure d'adopter le point de vue des personnes participantes.

Une fois les risques identifiés, il est souvent possible d'en réduire ou d'en éliminer un grand nombre. L'une des solutions consiste à modifier le devis de recherche. Par exemple vous pouvez raccourcir ou simplifier la procédure pour éviter l'ennui et la frustration. Vous pouvez peut-être remplacer des stimuli perturbants ou choquants (par exemple, des photos de scènes d'accident très explicites) par des stimuli moins perturbants ou choquants (par exemple, des photos plus douces du type de celles que les gens sont susceptibles de voir dans les journaux). Un bon exemple de modification d'un devis de recherche est la reproduction en 2009 de l'étude de Milgram menée par Jerry Burger. Au lieu de permettre à ses personnes participantes de continuer à administrer des décharges jusqu'au maximum de 450 V, ce chercheur interrompt toujours la procédure juste avant le point où la décharge administrée serait de 150 V (Burger, 2009). Cela se justifie par le fait que, dans l'étude de Milgram, **(a)** les réactions négatives graves des personnes participantes se produisaient après ce point et **(b)** la plupart des personnes participantes qui administraient la décharge de 150 V continuent jusqu'au maximum de 450 V. Le chercheur peut donc comparer directement ses résultats avec ceux de Milgram pour chaque point jusqu'à la décharge de 150 V et peut également estimer combien de ses personnes participantes auraient continué jusqu'au maximum — mais sans les soumettre au stress sévère de Milgram. (Les résultats, soit dit en passant, ont montré que ces personnes participantes contemporaines étaient tout aussi obéissantes que celles de Milgram).

Une deuxième façon de minimiser les risques est d'utiliser une procédure de **présélection** afin d'identifier et d'éliminer les personnes participantes à haut risque. Vous pouvez le faire en partie dans le cadre du processus de consentement éclairé. Par exemple vous pouvez avertir les personnes participantes qu'une enquête comprend des questions sur leur peur de la criminalité. Vous les informez que, dès l'arrivée de pensées qui pourraient potentiellement les perturber, ces personnes sont libres de se retirer de l'étude. La présélection peut également consister à collecter des données afin d'identifier et d'éliminer des personnes participantes. Par exemple Burger utilise une procédure approfondie de présélection comprenant plusieurs questionnaires, et un entretien avec un psychologue clinicien afin d'identifier et d'éliminer les gens présentant des problèmes physiques ou psychologiques les exposant à un risque élevé.

Une troisième façon de minimiser les risques est de prendre des mesures actives pour maintenir la

confidentialité. Quand cela est possible, vous devez conserver les formulaires de consentement signés séparément de toutes les données que vous collectez, et de manière à ce qu'aucun nom d'individus ne soit associé à ses données. En outre au-delà du sexe et de l'âge des personnes, vous ne devez collecter que les informations personnelles dont vous avez réellement besoin pour répondre à votre question de recherche. Si l'orientation sexuelle ou l'appartenance ethnique d'une personne n'est pas clairement pertinente pour votre question de recherche par exemple, ne lui posez pas de questions à ce sujet. Sachez également que certaines procédures de collecte de données peuvent entraîner des violations involontaires de la confidentialité. Lorsque les personnes participantes répondent à une enquête orale dans un centre commercial ou remplissent un questionnaire dans une salle de classe, il est possible que leurs réponses soient entendues ou vues par d'autres personnes. Si les réponses sont personnelles, il est préférable d'administrer l'enquête ou le questionnaire individuellement en privé ou d'utiliser d'autres techniques pour empêcher le partage involontaire d'informations personnelles.

Identifier et minimiser la duperie

N'oubliez pas que la duperie peut prendre diverses formes, qui n'impliquent pas toutes d'induire activement les personnes participantes en erreur. Il est également trompeur de leur permettre de faire des suppositions erronées (par exemple sur le contenu d'un « test de mémoire ») ou simplement de ne pas divulguer des informations sur la conception ou l'objectif de l'étude. Il est préférable d'identifier et de minimiser toutes les formes de duperie.

N'oubliez pas que, selon les codes éthiques de l'EPTC 2 (2022) et de l'APA, la duperie n'est acceptable d'un point de vue éthique que s'il n'y a aucun moyen de répondre à votre question de recherche sans elle. Par conséquent si votre devis de recherche inclut une forme quelconque de duperie active, vous devez vous demander si elle est vraiment nécessaire. Imaginons par exemple que vous souhaitez savoir si l'âge des membres du corps professoral universitaire influence les attentes des personnes aux études quant à leur capacité à enseigner. Pour ce faire, vous pouvez dire aux personnes participantes que vous allez leur montrer des photos de membres du corps professoral et que vous allez leur demander d'évaluer, pour chaque membre, leur capacité à enseigner. Mais si les photos ne représentent pas vraiment des membres du corps professoral, mais plutôt des membres de votre propre famille et des camarades, il s'agit alors d'une duperie. Cette duperie peut être facilement éliminée en demandant aux personnes participantes d'imaginer que les photos représentent des membres du corps professoral universitaire, et de les évaluer comme s'il en était ainsi.

En général on considère qu'il est acceptable d'attendre le débriefage avant de révéler votre question de recherche, à condition que vous décriviez la procédure, les risques et les avantages au cours du processus de consentement éclairé. Par exemple vous n'avez pas à dire aux personnes participantes que vous voulez savoir si l'âge des membres du corps professoral affecte les attentes des gens à leur égard avant la fin de l'étude. Non seulement cette information n'est pas susceptible d'influencer la décision des personnes de participer ou non à l'étude, mais elle peut également invalider les résultats. Les personnes participantes qui savent que l'âge est la variable indépendante peuvent évaluer différemment les « membres du corps professoral » plus âgés et plus jeunes parce qu'elles pensent que c'est ce que vous attendez d'elles. À l'inverse, elles pourraient veiller à leur attribuer la même note afin de ne pas donner l'impression d'avoir

des préjugés. Mais même cette forme extrêmement légère de duperie peut être minimisée en informant les personnes participantes — oralement, par écrit ou les deux — que, bien que vous ayez décrit avec précision la procédure, les risques et les avantages, vous attendrez jusqu'à la fin pour révéler la question de recherche. En substance, les personnes participantes consentent à la duperie ou à ce que des informations leur soient cachées jusqu'à plus tard.

Peser les risques et les avantages

Une fois que les risques de la recherche sont identifiés et minimisés, vous devez peser les risques avec les avantages. Pour ce faire, il est nécessaire d'identifier tous les avantages. N'oubliez pas de prendre en compte les avantages pour les personnes participantes à la recherche, pour la science et pour la société. Si vous êtes une personne étudiante menant une recherche, n'oubliez pas que l'un des avantages est la connaissance que vous allez acquérir sur la manière de mener une recherche scientifique en psychologie — connaissance que vous pourrez ensuite utiliser pour terminer vos études, et réussir dans vos études supérieures ou dans votre carrière.

Si la recherche ne présente qu'un risque minimal — pas plus élevé que dans la vie quotidienne des gens ou lors d'examens physiques ou psychologiques de routine — un avantage, même minime, pour les personnes participantes, la science ou la société est généralement considéré comme suffisant pour la justifier. Si elle comporte plus qu'un risque minimal, alors il doit y avoir plus d'avantages. Si la recherche risque de perturber des personnes participantes, par exemple, il devient d'autant plus important que l'étude soit bien conçue, et qu'elle puisse répondre à une question de recherche scientifiquement intéressante ou avoir des implications pratiques claires. Il est contraire à l'éthique de soumettre des personnes à la douleur, à la peur ou à l'embarras sans autre raison que de satisfaire une curiosité personnelle. En général les recherches en psychologie susceptibles de causer des dommages plus que mineurs ou d'une durée plus que brève sont rarement considérées comme justifiées par leurs avantages. Considérons par exemple que l'étude de Milgram — aussi intéressants et importants que furent ses résultats — serait jugée comme contraire à l'éthique selon les normes d'aujourd'hui.

Créer des procédures de consentement éclairé et de débriefage

Une fois que vous avez choisi un devis de recherche, vous devez élaborer vos procédures de consentement éclairé et de débriefage. Commencez par décider si le consentement éclairé est nécessaire (p. ex. vous allez effectuer des observations dans un lieu public). Si le consentement éclairé est nécessaire, il y a plusieurs choses à faire. Tout d'abord, lorsque vous recrutez des personnes participantes — que ce soit à l'aide du bouche-à-oreille, d'annonces ou d'un groupe de sujets — fournissez-leur autant d'informations que possible sur l'étude. Cela leur permettra d'éviter l'étude qui risquerait de leur être désagréable. Deuxièmement, préparez un script ou un ensemble de «points de discussion» pour vous aider à expliquer l'étude aux personnes dans un langage simple de tous les jours. Ces points devraient comprendre une description de la procédure, des risques et des avantages, et de leur droit de se retirer à tout moment. Troisièmement, créez

un formulaire de consentement éclairé, qui couvre tous les droits des personnes participantes ainsi que ce à quoi ces personnes peuvent s'attendre dans l'étude, et qui peut être lu et signé après que l'étude leur ait été décrite. Votre université, votre département ou votre professeur-e peut avoir un modèle de formulaire de consentement type que vous pouvez adapter à votre propre étude. Si ce n'est pas le cas, une recherche sur Internet vous permettra de trouver plusieurs modèles. N'oubliez pas que, le cas échéant, les parties orale et écrite du processus de consentement éclairé doivent inclure le fait que vous retenez une partie de l'information concernant le devis ou l'objectif de l'étude, mais que vous les révélez au cours du débriefage.

Le débriefage est similaire au consentement éclairé en ce sens que vous ne pouvez pas forcément vous attendre à ce que les personnes participantes lisent et comprennent les formulaires de débriefage écrits. Il est donc préférable de rédiger un script ou une série de points de discussion afin de pouvoir expliquer l'étude dans un langage simple et courant. Au cours du débriefage, vous devez révéler la question de recherche et le devis complet de l'étude. Par exemple si une seule condition est complétée par la personne participante, il faut lui expliquer ce qui s'est passé dans les autres conditions. S'il y a eu duperie des personnes participantes, il faut le révéler dès que possible, vous excuser pour la duperie, expliquer pourquoi elle était nécessaire et corriger toute idée fausse que ces personnes pourraient avoir à la suite de cette duperie. Le débriefage est également un moment propice pour offrir des avantages supplémentaires aux personnes participantes en leur donnant des informations pratiques pertinentes ou en les orientant vers d'autres sources d'aide. Par exemple dans le cadre d'une étude sur les attitudes à l'égard de la violence domestique, vous pouvez fournir des brochures sur la violence domestique et des informations sur l'orientation vers le centre de conseil de l'université pour ceux et celles qui le souhaitent.

N'oubliez pas de prévoir suffisamment de temps pour les processus de consentement éclairé et de débriefage. Ils ne peuvent être efficaces si vous devez les mener à la hâte.

Obtenir l'approbation

L'étape suivante consiste à obtenir l'approbation de l'établissement pour votre recherche en fonction des politiques et procédures spécifiques de votre établissement ou de votre cours, par l'intermédiaire du CÉR (comité d'éthique de la recherche) de votre institution. Pour ce faire, il faut généralement rédiger un protocole décrivant l'objectif de l'étude, le devis de recherche et la procédure, les risques et les avantages, les mesures prises pour minimiser les risques, ainsi que les procédures de consentement éclairé et de débriefage. Ne considérez pas la procédure d'approbation institutionnelle comme un simple obstacle à surmonter, mais comme une occasion de réfléchir à l'éthique de votre recherche et de consulter d'autres personnes susceptibles d'avoir plus d'expérience ou des points de vue différents des vôtres. Si le CÉR a des questions ou des préoccupations concernant votre recherche, répondez-y rapidement et en toute bonne foi. Cela peut même signifier qu'il faut apporter de nouvelles modifications au devis et à la procédure de votre recherche avant de soumettre à nouveau votre protocole.

Assurer le suivi

Votre souci de l'éthique ne doit pas s'arrêter au moment où votre étude reçoit l'approbation de l'institution. Il devient alors important de s'en tenir au protocole que vous avez soumis ou de demander une approbation supplémentaire pour tout ce qui n'est pas une modification mineure. Au cours de la recherche, vous devez surveiller les réactions imprévues de vos personnes participantes et demander leur avis lors du débriefage. L'une des critiques formulées à l'encontre de l'étude de Milgram est que, même s'il ne savait pas à l'avance que les personnes participantes auraient des réactions négatives aussi sévères, il l'a certainement su après avoir testé quelques personnes et aurait dû procéder à des ajustements à ce moment-là (Baumrind, 1985). Soyez également attentif aux violations potentielles de la confidentialité. Conservez les formulaires de consentement et les données en lieu sûr et séparément, et assurez-vous que personne, intentionnellement ou non, n'ait accès aux informations personnelles des personnes participantes.

Enfin vous devez maintenir votre intégrité tout au long du processus de publication et au-delà. Abordez rapidement avec les personnes collaborant à la recherche, la question des crédits de publication — qui seront auteurs ou autrices de la recherche et l'ordre des auteurs et autrices) et évitez le plagiat dans vos écrits. N'oubliez pas que votre objectif scientifique est d'apprendre comment fonctionne réellement le monde, et que votre devoir scientifique est de présenter vos résultats de manière honnête et exacte. Ne soyez donc pas tenté de fabriquer des données ou de modifier vos résultats de quelque manière que ce soit. En outre les résultats inattendus sont souvent aussi intéressants, voire plus, que les résultats attendus.

20. Points clés et exercices

Points clés

- La recherche en psychologie soulève une grande variété de questions éthiques. Pour les résoudre, il faut examiner comment chacun des trois principes directeurs (le respect des personnes, la préoccupation pour le bien-être et la justice) s'applique à chacun des trois groupes de personnes (les personnes participantes à la recherche, la science et la société).
- Les conflits éthiques dans la recherche en psychologie sont inévitables. Les responsables de recherches doivent réfléchir aux questions éthiques soulevées par leur recherche, minimiser les risques, peser les risques versus les avantages, être en mesure d'expliquer leurs décisions éthiques, demander à d'autres personnes une rétroaction sur ces décisions et, en fin de compte, en assumer la responsabilité.
- Il existe plusieurs codes d'éthique écrits pour la recherche avec des êtres humains participants qui fournissent des conseils spécifiques concernant les enjeux éthiques fréquemment rencontrés. Il s'agit notamment du code de Nuremberg, de la déclaration d'Helsinki et du rapport Belmont.
- L'énoncé de politique des Trois conseils est le code d'éthique suivi par les personnes menant des recherches et les institutions de recherche au Canada.
- Le code éthique de l'APA est aussi un code éthique important pour les personnes menant des recherches en psychologie. Il comprend de nombreuses normes qui s'appliquent principalement à la pratique clinique, mais la norme 8 concerne le consentement éclairé, la duperie, le débriefage, l'utilisation de sujets animaux non humains et l'intégrité scientifique en recherche.
- Les recherches menées dans les universités, les hôpitaux et les autres institutions, qui reçoivent un soutien du gouvernement fédéral, doivent être examinées par un comité d'éthique de recherche (CÉR) — un comité de l'institution qui examine les protocoles de recherche pour s'assurer qu'ils sont conformes aux normes éthiques.
- Le consentement éclairé est le processus qui consiste à obtenir et à documenter l'accord des personnes pour participer à une étude, après les avoir informées de tout ce qui pourrait raisonnablement influencer sur leur décision. Bien que cela implique souvent de leur faire lire et signer un formulaire de consentement, ce n'est pas équivalent à la lecture et à la signature d'un formulaire de consentement.
- Bien que des responsables de recherches affirment que la duperie des personnes participantes à la recherche n'est jamais justifiée d'un point de vue éthique, le EPTC et le code éthique de l'APA autorisent son utilisation lorsque les avantages l'emportent sur les risques, que l'on ne peut raisonnablement s'attendre à ce que les personnes participantes soient lésées, qu'il n'y a aucun moyen de mener l'étude sans duperie et que les personnes participantes sont informées de la duperie le plus tôt possible.

- En tant que personne qui mène une recherche, il est de votre responsabilité de connaître et d'accepter les responsabilités éthiques qui vous incombent.
- Vous pouvez prendre plusieurs mesures concrètes pour minimiser les risques et la duperie dans votre recherche. Il s'agit notamment de modifier le devis de votre recherche, de procéder à une présélection afin d'identifier et d'éliminer les gens à haut risque, et de fournir aux personnes participantes autant d'informations que possible lors du consentement éclairé et du débriefage.
- Vos responsabilités éthiques se poursuivent au-delà de l'approbation du CÉR. Vous devez surveiller les réactions des personnes participantes, être attentif aux violations potentielles de la confidentialité et maintenir l'intégrité scientifique tout au long du processus de publication.

Exercices

- **Pratique** : Imaginez une étude testant l'efficacité d'un nouveau médicament pour traiter les troubles obsessionnels compulsifs. Donnez un exemple hypothétique d'une question éthique de chaque cellule du Tableau 3.1 « Cadre de réflexion sur les questions éthiques dans la recherche scientifique » qui pourraient se poser dans le cadre de cette recherche.
- **Discussion** : Il a été avancé que les personnes menant des recherches ne sont pas éthiquement responsables de la mauvaise interprétation ou de la mauvaise utilisation de leurs recherches par d'autres personnes. Êtes-vous d'accord? Pourquoi ou pourquoi pas?
- **Pratique** : Lisez le code de Nuremberg, le rapport Belmont et la norme 8 du code éthique de l'APA. Citez cinq similitudes et cinq différences spécifiques entre eux.
- **Pratique** : Complétez le tutoriel en ligne de l'EPTC 2 (2022).
- **Discussion** : Dans une étude sur les effets du dégoût sur le jugement moral, les personnes participantes doivent juger de la moralité d'actes dégoûtants, y compris des personnes mangeant un animal mort et des baisers passionnés entre un frère et une sœur (Haidt, Koller et Dias, 1993). Si vous faisiez partie du CÉR chargé d'évaluer ce protocole, quelles seraient vos préoccupations à son égard ? Faites référence aux principes directeurs appropriés pour étayer votre argumentation.
- **Discussion** : Comment pourriez-vous mener une étude pour connaître à quel point les gens obéissent à l'autorité d'une manière qui minimise autant que possible les risques et la duperie? (Remarque : une telle étude ne devrait pas ressembler du tout à celle de Milgram).
- **Pratique** : Trouvez une étude dans une revue scientifique et créez un formulaire de consentement pour cette étude. Veuillez inclure tous les droits des personnes participantes et les détails de la procédure afin qu'elles puissent prendre une décision éclairée.

21. Références

- Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited (Recherche utilisant la duperie intentionnelle : Questions éthiques revisitées). *American Psychologist*, 40(2), 165-174. <https://doi.org/10.1037/0003-066X.40.2.165>
- Bowd, A. D. et Shapiro, K. J. (1993). The case against animal laboratory research in psychology (Les arguments contre la recherche animale en laboratoire dans le domaine de la psychologie). *Journal of Social Issues*, 49(1), 133-142. <https://doi.org/10.1111/j.1540-4560.1993.tb00913.x>
- Burger, J. M. (2009). Replicating Milgram: Would people still obey today? (Reproduction de Milgram : Les gens obéiraient-ils encore aujourd'hui ?). *American Psychologist*, 64(1), 1-11. <https://doi.org/10.1037/a0010932>
- Burns, J. F. (2010, 24 mai). British medical council bars doctor who linked vaccine to autism (Le conseil médical britannique exclut le médecin qui a établi un lien entre les vaccins et l'autisme). *The New York Times*. <http://www.nytimes.com/2010/05/25/health/policy/25autism.html>
- Carey, B. (2011, November 2). Fraud case seen as a red flag for psychology research. (Une affaire de fraude considérée comme un signal d'alarme pour la recherche en psychologie.) *The New York Times*. Repéré au <http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html>
- Conseil de recherches en sciences humaines du Canada, Conseil de recherches en sciences naturelles et en génie du Canada, Instituts de recherche en santé du Canada. (2022). *Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2 (2022)*. Gouvernement du Canada.
- Haidt, J., Koller, S. et Dias, M. (1993) Affect, culture, and morality, or is it wrong to eat your dog? (Affect, culture et moralité, ou est-ce mal de manger son chien ?) *Journal of Personality and Social Psychology*, 65(4), 613-628. <https://doi.org/10.1037/0022-3514.65.4.613>
- Jump, P. (2011, November 28). A star's collapse. (Une étoile s'écroule). *Inside Higher Ed*. Repéré au <https://www.insidehighered.com/news/2011/11/28/scholars-analyze-case-massive-research-fraud>
- Koocher, G. P. (1977). Bathroom behavior and human dignity (Comportement dans les toilettes et dignité humaine). *Journal of Personality and Social Psychology*, 35(2), 120-121. <https://doi.org/10.1037/0022-3514.35.2.120>
- Mann, T. (1994). Informed consent for psychological research: Do subjects comprehend consent forms and understand their legal rights? (Consentement éclairé pour la recherche en psychologie : Les sujets comprennent-ils les formulaires de consentement et leurs droits légaux ?) *Psychological Science*, 5(3), 140-143. <https://doi.org/10.1111/j.1467-9280.1994.tb00650.x>
- Milgram, S. (1963). Behavioral study of obedience (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67(4), 371-378. <https://doi.org/10.1037/h0040525>

- Middlemist, R. D., Knowles, E. S. et Matter, C. F. (1976). Personal space invasions in the lavatory : Suggestive evidence for arousal (Envahissement de l'espace personnel dans les toilettes : preuves suggestives de pulsion). *Journal of Personality and Social Psychology*, 33(5), 541-546. <https://doi.org/10.1037/0022-3514.33.5.541>
- Miller, N. E. (1985). The value of behavioral research on animals (La valeur de la recherche comportementale sur les animaux). *American Psychologist*, 40(4), 423-440. <https://doi.org/10.1037/0003-066X.40.4.423>
- Reverby, S. M. (2009). *Examining Tuskegee : The infamous syphilis study and its legacy* (Examiner Tuskegee : la tristement célèbre étude sur la syphilis et son héritage). University of North Carolina Press.
- Rosenthal, R. M. (1994). Science and ethics in conducting, analyzing, and reporting psychological research (Science et éthique dans la conduite, l'analyse et le rapport de la recherche en psychologie). *Psychological Science*, 5(3), 127-133. <https://doi.org/10.1111/j.1467-9280.1994.tb00646.x>
- Sieber, J. E., Iannuzzo, R. et Rodriguez, B. (1995). Deception Methods in Psychology: Have They Changed in 23 Years? (Les méthodes de duperie en psychologie : ont-elles changé en 23 ans ?) *Ethics Behavior*, 5(1), 67-85. https://doi.org/10.1207/s15327019eb0501_5
- Singal, J. (2015, May 29). The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud. *New York Magazine*. Répéré au <http://nymag.com/scienceofus/2015/05/how-a-grad-student-uncovered-a-huge-fraud.html>

CHAPITRE 4-LA MESURE PSYCHOLOGIQUE



Les chercheuses Tara MacDonald et Alanna Martineau se sont intéressées à l'effet de l'humeur des étudiantes universitaires sur leur intention d'entretenir des relations sexuelles non protégées (MacDonald & Martineau, 2002). Dans une étude empirique soigneusement conçue, elles ont constaté que le fait d'être d'humeur négative augmentait l'intention d'entretenir des rapports sexuels non protégés — mais seulement pour les étudiantes qui avaient une faible estime de soi. Bien que la réalisation d'une étude de ce type comporte de nombreux défis, l'un des principaux est la mesure des variables pertinentes. Dans cette étude, les chercheuses devaient savoir si chacune de leurs participantes avait une bonne ou une mauvaise estime de soi, ce qui nécessitait bien sûr de mesurer leur estime de soi. Elles devaient également s'assurer que leur tentative de mettre les participantes dans un état d'esprit négatif (en leur faisant penser à des choses négatives) était couronnée de succès, ce qui nécessitait de mesurer leur humeur. Enfin elles devaient vérifier si l'estime de soi et l'humeur étaient liées aux intentions des participantes d'entretenir des rapports sexuels non protégés, ce qui nécessitait de mesurer ces intentions.

Pour les personnes étudiantes s'initiant à la recherche en psychologie, le défi que représente la mesure de telles variables peut sembler insurmontable. Est-il vraiment possible de mesurer des choses aussi intangibles que l'estime de soi, l'humeur ou l'intention de faire quelque chose? La réponse est un oui retentissant et, dans ce chapitre, nous examinerons de près la nature des variables étudiées par les psychologues et la manière dont elles peuvent être mesurées. Nous examinerons également quelques questions pratiques de la mesure psychologique.

Avez-vous le sentiment d'être une personne de valeur?

L'échelle de l'estime de soi de Rosenberg (Rosenberg, 1989, pour la traduction canadienne-française Vallières et Vallerand, 1990) est l'une des mesures les plus courantes de l'estime de soi et celle que MacDonald et Martineau ont utilisée dans leur étude. Les personnes participantes répondent à chacun des 10 éléments suivants par une évaluation sur une échelle de 4 points : tout à fait en accord, plutôt en accord, plutôt en désaccord, tout à fait en désaccord. Coter les éléments 1, 2, 4, 6 et 7 en attribuant 3 points à chaque réponse tout à fait en accord, 2 à chaque réponse plutôt en accord, 1 à chaque réponse plutôt en désaccord et 0 à chaque réponse tout à fait en désaccord. Inverser la cotation des éléments 3, 5, 8, 9 et 10 en attribuant 0 point à chaque réponse tout à fait en accord, 1 point à chaque réponse plutôt en accord et ainsi de suite. Le score global est le nombre total de points.

1. Je pense que je suis une personne de valeur, au moins égale à n'importe qui d'autre.
2. Je pense que je possède un certain nombre de belles qualités.
3. Tout bien considéré, je suis porté.e à me considérer comme un.e raté.e.
4. Je suis capable de faire les choses aussi bien que la majorité des gens.
5. Je sens peu de raisons d'être fier.e de moi.
6. J'ai une attitude positive vis-à-vis de moi-même.

7. Dans l'ensemble, je suis satisfait.e de moi.
8. J'aimerais avoir plus de respect pour moi-même.
9. Parfois, je me sens vraiment inutile.
10. Il m'arrive de penser que je suis un.e bon.ne à rien.



Pour voir les étapes de
[traduction et validation de l'échelle de l'estime de soi Rosenberg](https://onlinelibrary.wiley.com/doi/epdf/10.1080/00207599008247865)

*Lecture sur version
imprimée? Scannez ce code
QR avec votre téléphone
cellulaire ou allez sur
[https://onlinelibrary.wiley.co
m/doi/epdf/10.1080/
00207599008247865](https://onlinelibrary.wiley.com/doi/epdf/10.1080/00207599008247865)*

22. Comprendre les mesures psychologiques

Objectifs d'apprentissage

1. Définir la notion de mesure et donner plusieurs exemples de mesure en psychologie.
2. Expliquer ce qu'est un construit psychologique et donner plusieurs exemples.
3. Distinguer les définitions conceptuelles des définitions opérationnelles, donner des exemples de chacune et créer des définitions opérationnelles simples.
4. Distinguer les quatre niveaux de mesure, donner des exemples pour chacun et expliquer pourquoi cette distinction est importante.

Qu'est-ce que la mesure?

La **mesure** est l'attribution de scores à des personnes de manière que ces scores représentent une certaine caractéristique de ces personnes. Cette définition très générale est conforme aux types de mesures que tout le monde connaît — par exemple, se peser en montant sur un pèse-personne ou vérifier la température interne d'une dinde en train de rôtir à l'aide d'un thermomètre à viande. Elle est également conforme aux mesures effectuées dans les autres sciences. En physique, par exemple, on peut mesurer l'énergie potentielle d'un objet dans le champ gravitationnel de la Terre en déterminant sa masse et sa hauteur (ce qui nécessite bien sûr de mesurer ces variables), puis en les multipliant par l'accélération gravitationnelle de la Terre ($9,8 \text{ m/s}^2$). Le résultat de cette procédure est un score qui représente l'énergie potentielle de l'objet.

Cette définition générale de la mesure est conforme également à la mesure en psychologie (la mesure en psychologie est souvent appelée la **psychométrie**.) Imaginons, par exemple, que des psychologues en cognition veuillent mesurer la capacité de la mémoire de travail d'une personne — c'est-à-dire sa capacité à garder à l'esprit plusieurs éléments d'information en même temps. Pour ce faire, ces gens pourraient utiliser une tâche d'empan de chiffres inversés, dans laquelle il y a lecture de deux chiffres à la personne et on lui demande de les répéter dans l'ordre inverse. On répète ensuite cette opération plusieurs fois, en augmentant la longueur de la liste d'un chiffre chaque fois, jusqu'à ce que la personne commette une erreur. La liste la plus longue pour laquelle la personne a répondu correctement est le score et représente sa capacité de mémoire de travail. Ou encore, imaginons des psychologues en clinique qui s'intéressent au degré de dépression d'une personne. On lui fait passer l'inventaire de dépression de Beck, un questionnaire d'autoévaluation de 21 items dans lequel la personne évalue dans quelle mesure elle s'est sentie triste, moins

énergique et a ressenti d'autres symptômes de dépression au cours des deux dernières semaines. La somme de ces 21 items constitue le score et représente le niveau actuel de dépression de la personne.

Le point important est que cette mesure ne nécessite pas d'instruments ou de procédures particulières. Ce qu'elle exige, c'est une procédure systématique pour attribuer des scores aux personnes ou objets, de manière à ce que ces scores représentent la caractéristique en question.

Construits psychologiques

De nombreuses variables étudiées par les personnes menant des recherches en psychologie sont simples et faciles à mesurer. Il s'agit notamment de l'âge, de la taille, du poids et de l'ordre de naissance. Vous pouvez demander aux gens quel âge ils ont et être raisonnablement sûrs qu'ils le savent et qu'ils vous le diront. Même si les gens ne savent pas ou ne veulent pas vous dire combien ils pèsent, vous pouvez leur demander de monter sur un pèse-personne. D'autres variables étudiées par les personnes menant des recherches en psychologie — peut-être la majorité — ne sont pas aussi simples à mesurer. Nous ne pouvons pas évaluer avec précision le niveau d'intelligence d'une personne en la regardant, et nous ne pouvons certainement pas mesurer son estime de soi sur un pèse-personne. Ces variables s'appellent des **construits** et comprennent les traits de personnalité (par exemple l'extraversion), les états émotionnels (par exemple la peur), les attitudes (par exemple envers les impôts) et les capacités (par exemple les qualités athlétiques).

Les construits psychologiques ne peuvent pas être observés directement. L'une des raisons en est qu'ils représentent souvent des tendances à penser, à ressentir ou à agir d'une certaine manière. Par exemple, dire qu'une étudiante universitaire est très extravertie ne signifie pas nécessairement qu'elle se comporte de manière extravertie en ce moment même. En fait elle pourrait être assise tranquillement, seule, en train de lire un livre. Cela signifie plutôt qu'elle a une tendance générale à se comporter de manière extravertie (par exemple en sortant, en appréciant les interactions sociales) dans diverses situations. Une autre raison pour laquelle les construits psychologiques ne peuvent être observés directement, c'est qu'ils impliquent souvent des processus internes. La peur, par exemple, implique l'activation de certaines structures du système nerveux central et périphérique, ainsi que certains types de pensées, de sentiments et de comportements — dont aucun n'est nécessairement évident pour une personne externe qui observe. Notons également que ni l'extraversion, ni la peur ne se « réduisent à » une pensée, un sentiment, un acte, une structure ou un processus physiologique particulier. Au contraire, chacun est une sorte de synthèse d'un ensemble complexe de comportements et de processus internes.

The Big Five

Le modèle des Big Five est un ensemble de cinq grandes dimensions qui rendent compte d'une grande partie des variations de la personnalité humaine. Chacun des Big Five peut même être défini en termes de six

construits plus spécifiques appelés « facettes » (Costa et McCrae, 1992; Traduction et adaptation canadienne-française par Le Corff, Y., s.d.; Le Corff, Y. et Busque-Carrier, M., 2016)

Ouverture	Aux Idées	Envers les rêveries	Envers l'esthétique	Aux actions	Aux sentiments	Aux valeurs
Conscience						
Extraversion						
Agréabilité						
Névrose						

Tableau 4.1 Facettes du modèle de personnalité Big Five

Ouverture	Aux Idées	Envers les rêveries	Envers l'esthétique	Aux actions	Aux sentiments	Aux valeurs
Conscience	Compétence	Ordre	Sens du devoir	Recherche de réussite	Autodiscipline	Réflexion
Extraversion	Grégarisme	Autorité	Activité	Recherche de sensations	Émotions positives	Chaleur
Agréabilité	Confiance	Droiture	Altruisme	Soumission	Modestie	Sensibilité
Névrose	Anxiété	Colère/Hostilité	Dépression	Conscience de soi	Impulsivité	Vulnérabilité

La **définition conceptuelle** d'un construit psychologique décrit les comportements et les processus internes qui composent ce construit, ainsi que la manière dont il est lié à d'autres variables. Par exemple une définition conceptuelle de névrose (une autre facette des Big Five) serait qu'il s'agit de la tendance à éprouver des émotions négatives, telles que l'anxiété, la colère et la tristesse dans diverses situations. Cette définition pourrait également inclure le fait que la névrose a une forte composante génétique, qu'elle reste relativement stable dans le temps et qu'elle est positivement corrélée à la tendance à ressentir de la douleur et d'autres symptômes physiques.

Les personnes aux études se demandent parfois pourquoi, lorsque les personnes en recherche veulent comprendre un construit tel que l'estime de soi ou la névrose, elles ne regardent pas simplement dans le dictionnaire. L'une des raisons est que de nombreux construits scientifiques n'ont pas d'équivalent dans le langage courant (par exemple la capacité en mémoire de travail). Plus important encore, les personnes menant des recherches ont pour mission d'élaborer des définitions plus détaillées et plus précises — et qui décrivent avec plus d'exactitude la façon dont le monde est constitué — que les définitions informelles du dictionnaire. Comme nous le verrons, ces personnes le font en proposant des définitions conceptuelles, en les testant empiriquement et en les révisant si nécessaire. Parfois, elles les rejettent complètement. C'est pourquoi la littérature de recherche comprend souvent différentes définitions conceptuelles d'un même construit. Dans certains cas, une ancienne définition conceptuelle a été remplacée par une nouvelle qui convient et fonctionne mieux. Dans d'autres cas, les personnes menant des recherches sont encore en train de décider laquelle des différentes définitions conceptuelles est la meilleure.

Définitions opérationnelles

Une **définition opérationnelle** décrit précisément la manière dont une variable doit être mesurée. Ces mesures se classent généralement en trois grandes catégories. Les **mesures d'autoévaluation** sont celles où les personnes participantes font état de leurs propres pensées, sentiments et actions, comme dans le cas de l'échelle de l'estime de soi de Rosenberg (Rosenberg, 1965). Les **mesures comportementales** correspondent à l'observation et à l'enregistrement d'un autre aspect du comportement des personnes participantes. Il s'agit d'une catégorie extrêmement étendue qui comprend l'observation du comportement des personnes à la fois dans des tâches de laboratoire très structurées et dans des contextes plus naturels. Un bon exemple de la première, mesurant la capacité en mémoire de travail, est la tâche de mémoire de chiffres inversés. Un bon exemple pour les tâches dans un contexte plus naturel est associé à une célèbre définition opérationnelle de l'agression physique du chercheur Albert Bandura et de ses collègues (Bandura, Ross et Ross, 1961). L'équipe de recherche a laissé chaque enfant jouer pendant 20 minutes dans une pièce où se trouvait un sac de frappe en forme de clown appelé poupée Bobo. Chacun des enfants est filmé et leurs nombres d'actes d'agression physique commis sont comptés. Il s'agissait notamment de frapper la poupée avec un maillet, de lui donner des coups de poing et de lui donner des coups de pied. Leur définition opérationnelle est donc le nombre d'actes spécifiquement définis que l'enfant a commis pendant la période de 20 minutes. Enfin, les **mesures physiologiques** sont celles qui impliquent l'enregistrement d'une grande variété de processus physiologiques, y compris la fréquence cardiaque et la pression sanguine, la réponse galvanique de la peau, les niveaux d'hormones, l'activité électrique et le flux sanguin dans le cerveau.

Pour une variable ou un construit donné, il existe plusieurs définitions opérationnelles. Le stress en est un bon exemple. Une définition conceptuelle approximative décrit le stress comme étant une réponse adaptative à un danger ou à une menace perçue, qui implique des composantes physiologiques, cognitives, affectives et comportementales. Mais les personnes menant des recherches l'ont défini de manière opérationnelle de plusieurs façons. L'échelle d'ajustement social (Holmes et Rahe, 1967, pour voir l'échelle en français, Paquette Houde, Abdel Baki, Lecompte, Lussier-Valade et Ngô, 2018, p. 103) est un questionnaire d'autoévaluation dans lequel les personnes identifient les événements stressants qu'elles ont vécus au cours de l'année précédente et, pour chacun d'entre eux, des points y sont associés en fonction de sa gravité. Par exemple, un homme qui a divorcé (73 points), changé d'emploi (36 points) et modifié ses habitudes de sommeil (16 points) au cours de l'année précédente obtiendrait un score total de 125 points. L'échelle de tracasseries quotidiennes et d'encouragements (et d'expériences positives) (DeLongis, Coyne, Dakof, Folkman et Lazarus, 1982) est similaire, mais se concentre sur les facteurs de stress quotidiens, tels que le fait d'égarer des objets et de s'inquiéter de son poids. L'échelle du stress perçu (Cohen, Kamarck et Mermelstein, 1983) est une autre mesure d'auto-évaluation qui se concentre sur les sentiments de stress des personnes (par exemple, « À quelle fréquence vous sentez-vous nerveux·se et stressé·e? ») Les personnes menant des recherches ont également défini le stress de manière opérationnelle en fonction de plusieurs variables physiologiques, notamment la tension artérielle et les niveaux de cortisol, l'hormone du stress (pour plus de détails sur la critique de la mesure du stress, [Guillet et Hermand, 2006](#)).



Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou allez sur <https://numerique.banq.qc.ca/patrimoine/details/52327/3533838>

Lorsque les psychologues utilisent plusieurs définitions opérationnelles d'un même construit, que ce soit au sein d'une étude ou d'une étude à l'autre, il y a utilisation d' **opérations convergentes**. L'idée est que les différentes définitions opérationnelles « convergent » ou se rejoignent sur le même construit. Lorsque les scores basés sur plusieurs définitions opérationnelles différentes sont étroitement liés les uns aux autres et produisent des modèles de résultats similaires, cela constitue une bonne preuve que le construit est mesuré de manière efficace et qu'il est utile. Les différentes mesures du stress, par exemple, sont toutes corrélées entre elles et il a été démontré qu'elles étaient toutes corrélées avec d'autres variables, telles que le fonctionnement du système immunitaire (également mesuré de diverses manières) (Segerstrom et Miller, 2004). C'est ce qui permet aux personnes menant des recherches de tirer des conclusions générales utiles, telles que « le stress est négativement corrélé avec le fonctionnement du système immunitaire », par opposition à des conclusions plus spécifiques et moins utiles, telles que « les scores des personnes sur l'échelle du stress perçu sont négativement corrélés avec leur nombre de globules blancs ».



Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou allez sur https://hal.science/hal-00755690/file/Article_Critique_de_L_A_valuation_du_stress.pdf

Types d'échelles

Le chercheur en psychologie Stanley Smith Stevens dans l'article intitulé "On the Theory of Scales of Measurement" a suggéré que des scores peuvent être attribués aux personnes d'une manière qui communique plus ou moins d'informations quantitatives sur la variable d'intérêt (Stevens, 1946). Par exemple, les officiels d'une course de 100 mètres pourraient simplement classer les coureurs au moment où ils franchissent la ligne d'arrivée (premier, deuxième, etc.), ou ils pourraient chronométrer chaque coureur au dixième de seconde près à l'aide d'un chronomètre (11,5 s., 12,1 s., etc.). Dans les deux cas, ils mesurent le temps des coureurs en attribuant systématiquement des scores pour représenter ces temps. Mais alors que la procédure de classement communique le fait que le coureur arrivé en deuxième position a mis plus de temps que le coureur arrivé en première position, la procédure du chronomètre communique en plus combien de temps supplémentaire le deuxième coureur a pris. Stevens a en fait proposé quatre **types d'échelles** correspondant à quatre types d'informations pouvant être communiquées par un ensemble de scores. Ces types d'échelles diffèrent également suivant les procédures statistiques utilisées.

L'**échelle nominale** est utilisée pour les variables catégorielles. On attribue un chiffre à chaque personne qui permet de déterminer si deux personnes sont identiques ou différentes au niveau de la variable mesurée. Par exemple si vous demandez à vos personnes participantes leur état matrimonial, vous utilisez l'échelle nominale. De même si vous demandez à vos personnes participantes d'indiquer à quelle ethnie elles s'identifient parmi plusieurs, vous utilisez à nouveau l'échelle nominale. Le point essentiel des échelles nominales est qu'elles n'impliquent aucun ordre parmi les réponses, aucune distinction quantitative. Par exemple lorsque l'on classe les personnes en fonction de leur couleur préférée, le vert n'est pas placé « devant » le bleu. Les réponses sont simplement catégorisées. Les échelles nominales représentent donc le niveau de mesure le plus bas.

Les trois autres types d'échelles sont utilisés pour les variables quantitatives. L'**échelle ordinale** consiste à assigner des scores qui représentent l'ordre des personnes. Les rangs indiquent non seulement si deux personnes sont identiques ou différentes en matière de la variable mesurée, mais aussi si une personne se trouve plus haut ou plus bas sur cette variable. Par exemple un chercheur qui souhaite mesurer la satisfaction des consommateurs et consommatrices à l'égard de leur four à micro-ondes peut leur demander de préciser leur sentiment : *très insatisfait*, *plutôt insatisfait*, *plutôt satisfait* ou *très satisfait*. Les choix de réponse de cette échelle sont classés par ordre de grandeur, allant du moins satisfait au plus satisfait. Contrairement aux échelles nominales, les échelles ordinales permettent de comparer le degré avec lequel deux personnes évaluent la variable d'intérêt. Par exemple le classement du niveau de satisfaction permet d'affirmer qu'une personne est plus satisfaite qu'une autre de son four à micro-ondes. Ceci vient du fait que la première personne a utilisé un choix de réponses figurant plus loin dans la liste que celui choisi par la seconde personne.

Par ailleurs les échelles ordinales ne permettent pas de saisir d'importantes informations qui seront présentes dans les autres types d'échelles que nous examinerons. En particulier, on ne peut pas supposer que la différence entre deux choix de réponses disponibles dans une échelle ordinale est la même que la différence entre deux autres choix de réponses. En revenant à l'exemple des coureurs, on ne peut pas supposer que l'écart entre les coureurs de la première et de la deuxième place est égal à l'écart entre les coureurs de la deuxième et de la troisième place. Ainsi dans l'exemple de l'échelle de satisfaction pour leur four à micro-ondes, la différence entre les réponses *très insatisfait* et *plutôt insatisfait* n'est probablement

pas équivalente à la différence entre *plutôt insatisfait* et *plutôt satisfait*. Rien dans la mesure utilisée ne nous permet de déterminer si les deux différences reflètent la même différence de satisfaction psychologique. Les personnes spécialisées en statistique expriment ce point en disant que les différences entre des valeurs d'échelle adjacentes ne représentent pas nécessairement des intervalles égaux sur l'échelle sous-jacente qui donne lieu aux mesures. (Dans notre cas, l'échelle sous-jacente est le véritable sentiment de satisfaction, que nous essayons de mesurer.)

L'**échelle d'intervalle** implique l'attribution de scores à l'aide d'échelles numériques où les intervalles ont toujours la même interprétation. Prenons l'exemple des échelles de température, Fahrenheit ou Celsius. La différence entre 30 et 40 degrés représente la même différence de température que la différence entre 80 et 90 degrés. Cela s'explique par le fait que chaque intervalle de 10 degrés a la même signification physique (en matière d'énergie cinétique des molécules).

Les échelles d'intervalles ne sont cependant pas parfaites. Elles n'ont pas de véritable point zéro, de zéro absolu et ce même si l'une des valeurs de l'échelle porte le nom « zéro ». L'échelle Fahrenheit illustre ce problème. Le zéro degré Fahrenheit (0 °F) ne représente pas l'absence totale de température (en d'autres mots, l'absence d'énergie cinétique moléculaire). En réalité l'étiquette « zéro » est appliquée à une température donnée pour des raisons tout à fait accidentelles liées à l'histoire de la mesure de la température. Étant donné qu'une échelle d'intervalle n'a pas de véritable zéro absolu, il n'est pas logique de calculer des rapports de températures. Par exemple le ratio 40 sur 20 degrés Fahrenheit n'est pas identique au ratio 100 sur 50 degrés Fahrenheit; aucune propriété physique intéressante n'est conservée entre ces deux rapports. Après tout si l'étiquette « zéro » était appliquée à la température nommée 10 degrés par Daniel Gabriel Fahrenheit (l'inventeur de l'échelle de température du même nom), les deux ratios seraient maintenant 30 sur 10 et 90 sur 40, ce qui n'est plus la même chose! De ce fait il n'est pas logique de dire que 80 degrés sont « deux fois plus chauds » que 40 degrés. Une telle affirmation dépendrait d'une décision arbitraire concernant le point de départ de l'échelle de température, à savoir la température qu'il convient d'appeler « zéro » (alors que ladite affirmation a pour but de faire une déclaration plus fondamentale sur la réalité physique sous-jacente).

En psychologie, on considère souvent que le quotient intellectuel (QI) est mesuré avec l'échelle d'intervalle. Bien qu'il soit techniquement possible d'obtenir un score de 0 à un test de QI, un tel score n'indique pas l'absence totale de QI. De plus une personne ayant un QI de 140 n'a pas deux fois le QI d'une personne ayant un QI de 70. En revanche la différence entre deux QIs, 80 et 100, est la même que la différence entre les deux QIs suivants, 120 et 140.

Finalement, l'**échelle de rapport** implique l'attribution de scores de telle sorte qu'il existe un zéro absolu représentant l'absence totale de quantité. La taille mesurée en mètres et le poids mesuré en kilogrammes en sont de bons exemples. Il en va de même pour le dénombrement d'objets ou d'événements discrets, comme le nombre de frères et sœurs d'une personne ou le nombre de questions auxquelles une personne aux études répond correctement lors d'un examen. On peut considérer qu'une échelle de rapport est la synthèse des trois échelles précédentes. Comme une échelle nominale, elle fournit un nom ou une catégorie pour chaque objet (les nombres servent d'étiquettes). Comme une échelle ordinale, les objets sont ordonnés (en fonction de l'ordre des nombres). Comme une échelle d'intervalle, la même différence à deux endroits sur l'échelle a la même signification. En revanche, l'échelle de rapport permet en plus de donner la même signification à deux rapports identiques, même s'ils se situent à deux endroits sur l'échelle (voir Tableau 4.1).

L'échelle de température de Fahrenheit comporte un point zéro arbitraire et n'est donc pas une échelle de rapport. En revanche le zéro sur l'échelle de Kelvin est un zéro absolu. L'échelle de Kelvin est une échelle de

rapport. Par exemple si une température est deux fois plus élevée qu'une autre sur l'échelle de Kelvin, alors elle a deux fois plus d'énergie cinétique que l'autre température.

Un autre exemple d'échelle de rapport est la somme d'argent que vous avez dans votre poche en ce moment (25 cents, 50 cents, etc.) L'argent est mesuré sur une échelle de rapport, car en plus de posséder les propriétés d'une échelle d'intervalle, il possède un zéro absolu : si vous avez zéro argent, c'est que vous n'avez pas d'argent. Avec cette échelle, il est aussi logique de dire qu'une personne ayant 50 cents possède deux fois plus d'argent qu'une personne ayant 25 cents.

Les types d'échelles de mesure de Stevens sont importants pour au moins deux raisons. Premièrement, ils soulignent la généralité du concept de mesure. Bien que l'on ne considère généralement pas la catégorisation ou le classement des personnes comme des mesures, elles le sont en fait, tant qu'ils sont effectués de manière à représenter une certaine caractéristique des gens. Deuxièmement, les types d'échelles peuvent servir de guide approximatif pour les procédures statistiques qui peuvent être utilisées avec les données et les conclusions qui peuvent en être tirées. Dans le cas d'une échelle nominale par exemple, la seule mesure de tendance centrale disponible est le mode. Dans le cas des échelles ordinales, la médiane ou le mode peuvent être utilisés comme indicateurs de la tendance centrale. Les types d'échelles d'intervalle et de rapport sont généralement considérés comme les plus souhaitables, car ils permettent de calculer n'importe quel indicateur de tendance centrale (c'est-à-dire la moyenne, la médiane ou le mode). En outre l'échelle des rapports est le seul niveau permettant de se prononcer concernant les ratios de scores. Encore une fois on ne peut pas dire qu'une personne ayant un QI de 140 est deux fois plus intelligente qu'une personne ayant un QI de 70 parce que le QI est mesuré avec une échelle d'intervalle. En revanche on peut dire qu'une personne ayant six membres dans sa fratrie en a deux fois plus qu'une personne ayant trois membres dans sa fratrie, car le nombre de membres dans une fratrie est mesuré avec l'échelle de rapport. Pour une autre description de ces échelles en français, consultez Hogan, Stephenson et Parent (2007)

Tableau 4.1 Résumé des types d'échelles et de leurs propriétés

Échelle	Classe et attribut des numéros	Place en ordre de grandeur	Utilise des intervalles égaux	Comprends un zéro absolu
NOMINALE	X			
ORDINALE	X	X		
INTERVALLE	X	X	X	
DE RAPPORT	X	X	X	X

23. Fidélité et validité des mesures

Objectifs d'apprentissage

1. Définir la fidélité, y compris les différents types et la manière dont ils sont évalués.
2. Définir la validité, y compris les différents types et la manière dont ils sont évalués.
3. Décrire les types de preuves qui seraient pertinentes pour évaluer la fidélité et la validité d'une mesure particulière.

Là encore la mesure consiste à attribuer des scores à des gens de manière à ce qu'ils représentent une certaine caractéristique de ces personnes. Mais comment les personnes menant des recherches savent-elles que les scores représentent réellement la caractéristique en particulier lorsqu'il s'agit d'un construit comme l'intelligence, l'estime de soi, la dépression ou la capacité de mémoire de travail? La réponse est qu'elles mènent des recherches en utilisant la mesure pour confirmer que les scores ont un sens d'après leur compréhension du construit mesuré. Il s'agit là d'un point extrêmement important. Les psychologues ne supposent pas simplement que leurs mesures fonctionnent. Au lieu de cela, ces personnes collectent des données pour démontrer qu'ils fonctionnent. Si leurs recherches ne démontrent pas que la mesure fonctionne, elles cessent de l'utiliser.

À titre d'exemple informel, imaginons que vous suivez un régime depuis un mois. Vos vêtements semblent plus amples et plusieurs amis vous demandent si vous avez perdu du poids. Si à ce stade, votre pèse-personne indique que vous avez perdu 10 livres, cela serait logique de continuer à utiliser ce pèse-personne. Mais s'il indique que vous avez pris 10 livres, vous allez conclure, avec raison, qu'il ne fonctionne pas bien, vous tenterez alors de le réparer ou de vous en débarrasser. Lorsque les psychologues évaluent une méthode de mesure, deux dimensions générales sont considérées: la fidélité et la validité.

Fidélité

La **fidélité** fait référence à la cohérence d'une mesure. Les chercheurs en psychologie considèrent trois types de cohérence : au fil du temps (fidélité test-retest), entre les éléments (cohérence interne) et entre les différents utilisateurs (fidélité interjuges).

Fidélité test-retest

Lorsque les personnes en recherche mesurent un concept prétendument stable dans le temps, les scores qu'elles obtiennent devraient également être stables dans le temps. La **fidélité test-retest** vérifie dans quelle mesure cela est effectivement le cas. Par exemple on considère généralement que l'intelligence est constante dans le temps. Une personne très intelligente aujourd'hui le sera encore la semaine prochaine. Cela signifie que toute bonne mesure de l'intelligence devrait produire à peu près les mêmes résultats pour cette personne la semaine prochaine qu'aujourd'hui. Il est clair qu'une mesure qui produit des résultats très inconsistants dans le temps ne peut pas être une très bonne mesure d'un construit qui est censé être stable.

Pour évaluer la fidélité test-retest, il faut utiliser la mesure sur un groupe de personnes à un moment donné, l'utiliser à nouveau sur le même groupe de personnes à un moment plus tard, puis examiner la corrélation test-retest entre les deux séries de résultats. Pour ce faire, on représente généralement les données sous forme de diagramme de dispersion, puis on calcule le coefficient de corrélation. La Figure 4.2 montre la corrélation entre deux séries de scores de plusieurs personnes étudiant à l'université sur l'échelle d'estime de soi de Rosenberg, administrée deux fois à une semaine d'intervalle. Le coefficient de corrélation pour ces données est de + 0,95. En général, une corrélation test-retest de + 0,80 ou plus est considérée comme indiquant une bonne fidélité.

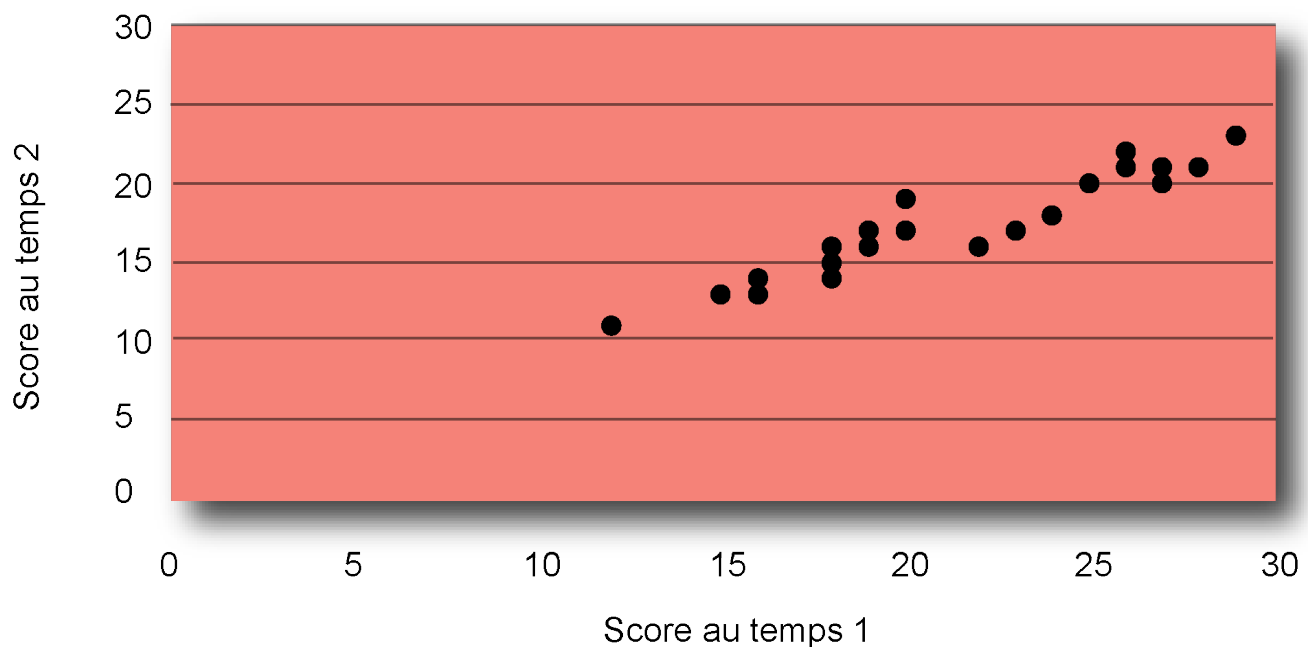


Figure 4.2. Corrélation test-retest entre deux séries de scores de personnes étudiant à l'université sur l'échelle d'estime de soi de Rosenberg, administrés à deux reprises à une semaine d'intervalle

Là encore obtenir des corrélations test-retest élevées est logique lorsque le construit mesuré est supposé être stable dans le temps. Ceci est attendu pour les construits suivants: l'intelligence, l'estime de soi et les dimensions de la personnalité du Big Five. Mais d'autres construits ne sont pas censés être stables au fil du

temps. La nature même de l'humeur, par exemple, est qu'elle change. Ainsi une mesure de l'humeur qui produirait une faible corrélation test-retest sur une période d'un mois ne serait pas une source d'inquiétude.

Cohérence interne

Un autre type de fidélité est la **cohérence interne**, qui est la cohérence entre les réponses des personnes pour l'ensemble des items d'une mesure comportant plusieurs items. En général, tous les items de ces mesures sont censés refléter le même construit sous-jacent, de sorte que les scores des personnes sur ces éléments devraient être corrélés entre eux. Sur l'échelle d'estime de soi de Rosenberg, les personnes qui reconnaissent être des personnes de valeur devraient avoir tendance à reconnaître qu'elles sont dotées de plusieurs bonnes qualités. Si les réponses des personnes aux différents items ne sont pas corrélées entre elles, il n'y a plus de sens à affirmer qu'ils mesurent tous le même construit sous-jacent. Cela vaut aussi bien pour les mesures comportementales et physiologiques que pour les mesures d'auto-évaluation. Par exemple les gens peuvent faire une série de paris dans un jeu de roulette simulé pour mesurer leur niveau de recherche du risque. Cette mesure serait cohérente sur le plan interne dans la mesure où les mises des participants seraient systématiquement élevées ou faibles d'un essai à l'autre.

Tout comme la fidélité test-retest, la cohérence interne ne peut être évaluée qu'en collectant et en analysant des données. L'une des approches consiste à examiner une **corrélation moitié-moitié**. Cela implique de diviser les items d'un test en deux ensembles, comme a) la première et la seconde moitié des items ou b) les items pairs et impairs. Un score est ensuite calculé pour chaque ensemble d'items et la relation entre les deux ensembles de scores est examinée. Par exemple la Figure 4.3 montre la corrélation moitié-moitié entre les scores de plusieurs personnes étudiant à l'université aux items pairs, et leurs scores aux items impairs de l'échelle d'estime de soi de Rosenberg. Le coefficient de corrélation pour ces données est de + 0,88. Une corrélation moitié-moitié de + 0,80 ou plus est généralement considérée comme une bonne cohérence interne.

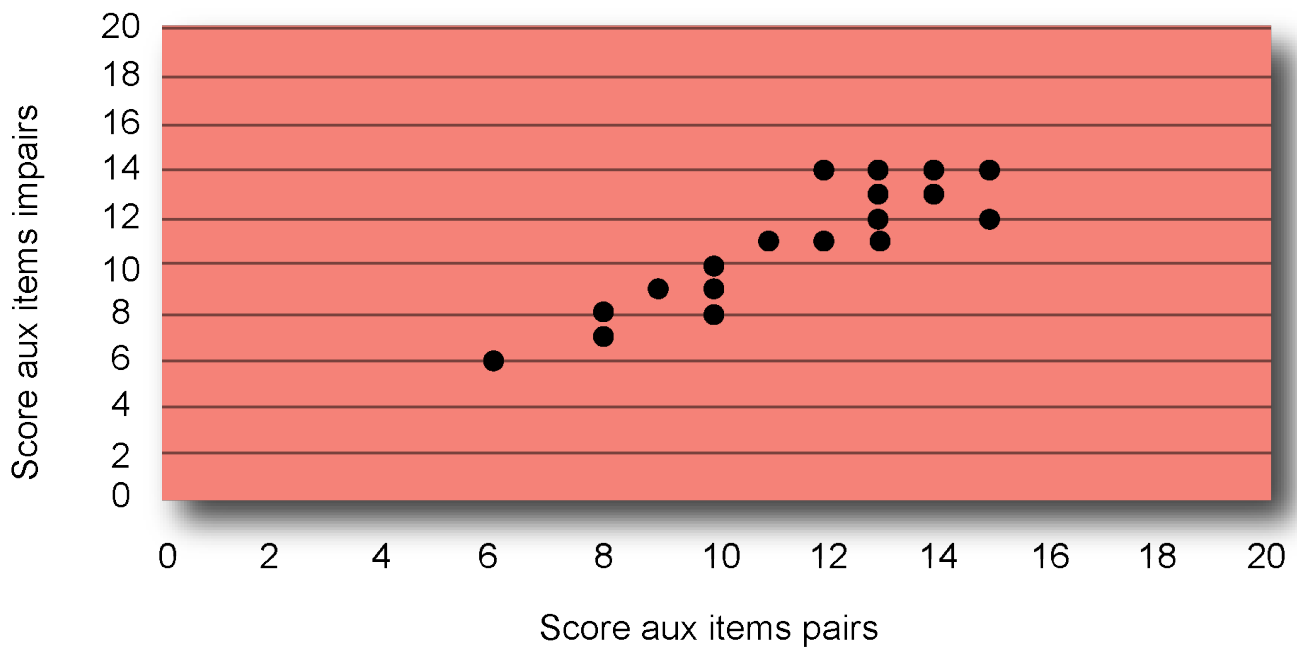


Figure 4.3. Corrélation moitié-moitié entre les scores de plusieurs personnes étudiant à l'université aux items pairs et leurs scores aux items impairs de l'échelle d'estime de soi de Rosenberg

La mesure la plus courante de la cohérence interne utilisée par les personnes menant des recherches en psychologie est une statistique appelée l' **α de Cronbach** (α est la lettre grecque alpha). Conceptuellement, α est la moyenne de toutes les corrélations moitié-moitié possibles pour un ensemble d'items. Par exemple il y a 252 façons de diviser un ensemble de dix items en deux ensembles de cinq. Le α de Cronbach serait la moyenne des 252 corrélations moitié-moitié. Notons que ce n'est pas la façon dont α est réellement calculé, mais c'est une façon correcte d'interpréter la signification de cette statistique. Une fois encore, une valeur de + 0,80 ou plus est généralement considérée comme indiquant une bonne cohérence interne.

Fidélité interjuges

De nombreuses mesures comportementales impliquent un jugement substantiel de la part de la personne responsable de l'observation ou de l'évaluation. La **fidélité interjuges** est le degré de concordance entre les jugements des différentes personnes responsables de l'observation. Par exemple si vous souhaitez mesurer les compétences sociales de la population estudiantine universitaire, vous pouvez enregistrer ces gens sur vidéo lors de leur interaction avec une autre personne aux études, au moment de leur première rencontre. Vous pouvez ensuite demander à deux personnes responsables de l'observation ou plus de regarder ces vidéos afin d'évaluer le niveau de compétences sociales de chaque personne aux études. Dans la mesure où chaque personne participante possède effectivement un certain niveau de compétences sociales pouvant être détecté par une personne responsable de l'observation attentive, les évaluations des différentes personnes responsables de l'observation devraient être fortement corrélées les unes aux autres. La fidélité interjuges a également été mesurée dans l'étude de Bandura sur la poupée Bobo. Dans ce cas, les

évaluations des personnes responsables de l'observation, concernant le nombre d'actes d'agression commis par un enfant donné lorsqu'il jouait avec la poupée Bobo, auraient dû être fortement corrélées positivement. La fidélité interjuges est souvent évaluée à l'aide de l' α de Cronbach lorsque les jugements sont quantitatifs ou d'une statistique analogue appelée κ de Cohen (lettre grecque kappa) lorsqu'ils sont catégoriels.

Validité

La **validité** est le degré auquel les scores issus d'une méthode de mesure représentent la variable qu'ils sont censés représenter. Mais comment les personnes en recherche font-elles ce jugement ? Nous avons déjà examiné un facteur — la fidélité. Lorsqu'une mesure présente une bonne fidélité test-retest et une bonne cohérence interne, les personnes en recherche devraient être plus confiantes dans le fait que les scores représentent ce qu'ils sont censés représenter. Il faut toutefois aller plus loin, car une mesure peut être extrêmement fidèle, mais n'avoir aucune validité. Pour prendre un exemple absurde, imaginons une personne qui pense que la longueur de l'index des gens reflète leur estime de soi et qui, par conséquent, essaie de mesurer l'estime de soi en tenant une règle sur l'index des personnes. Bien que cette mesure ait une très bonne fidélité test-retest, elle n'a absolument aucune validité. Le fait que l'index d'une personne soit plus long d'un centimètre que celui d'une autre n'indique en rien laquelle des deux a une meilleure estime de soi.

Les discussions sur la validité la divisent généralement en plusieurs « types » distincts. Mais une bonne façon d'interpréter ces types est de dire qu'il s'agit d'autres types de preuves — en plus de la fidélité — qui doivent être prises en compte pour juger de la validité d'une mesure. Nous examinerons ici trois types fondamentaux : la validité apparente, la validité de contenu et la validité de critère.

Validité apparente

La **validité apparente** cherche à vérifier si la méthode de mesure semble à première vue mesurer le construit étudié. La plupart des gens s'attendent à ce qu'un questionnaire sur l'estime de soi comprenne des questions sur le fait de se considérer comme une personne de valeur, et sur le fait de penser qu'ils ont de bonnes qualités. Un questionnaire comprenant ce type de questions aurait donc une bonne validité apparente. La méthode de mesure de l'estime de soi par la longueur de l'index, en revanche, semble n'avoir aucun rapport avec l'estime de soi et a donc une faible validité apparente. Bien que la validité apparente puisse être évaluée quantitativement — par exemple en demandant à un large échantillon de personnes d'évaluer une mesure pour savoir si elle semble mesurer ce qu'elle est censée mesurer — elle est généralement évaluée de manière informelle.

La validité apparente est au mieux une preuve très faible qu'une méthode de mesure évalue ce qu'elle est censée mesurer. L'une des raisons est qu'elle repose sur les intuitions des gens concernant le comportement humain, qui sont souvent incorrectes. Il est également vrai que de nombreuses mesures établies en psychologie fonctionnent très bien malgré leur manque de validité apparente. Le Minnesota Multiphasic Personality Inventory-2 (MMPI-2) mesure de nombreuses caractéristiques de la personnalité et de nombreux troubles en demandant aux personnes de décider si chacune des 567 affirmations s'applique à elles —

alors que de nombreuses affirmations n'ont aucune relation évidente avec le construit qu'elles mesurent. Par exemple, les énoncés « J'aime les histoires policières ou mystérieuses » et « La vue du sang ne m'effraie pas et ne me rend pas malade » mesurent tous deux la suppression de l'agressivité. Dans ce cas, ce ne sont pas les réponses littérales des personnes participantes à ces questions qui sont intéressantes, mais plutôt le fait de savoir si le schéma des réponses des personnes participantes à une série de questions correspond à celui des gens qui ont tendance à réprimer leur agressivité.

Validité de contenu

La **validité de contenu** correspond à la mesure dans laquelle un instrument de mesure « couvre » le construit d'intérêt. Par exemple si une personne menant une recherche définit conceptuellement l'anxiété liée à un examen, comme impliquant à la fois l'activation du système nerveux sympathique (conduisant à de la nervosité) et des pensées négatives, alors sa mesure de l'anxiété liée à un examen devrait inclure des éléments concernant à la fois les sentiments nerveux et les pensées négatives. Par ailleurs, les attitudes sont généralement définies comme impliquant des pensées, des sentiments et des actions à l'égard de quelque chose. Selon cette définition conceptuelle, une personne a une attitude positive à l'égard de l'exercice physique dans la mesure où elle a des pensées positives à ce sujet, se sent bien à l'idée de faire de l'exercice et fait effectivement de l'exercice. Ainsi pour avoir une bonne validité de contenu, une mesure de l'attitude des gens à l'égard de l'exercice devrait refléter ces trois aspects. Comme la validité apparente, la validité de contenu n'est généralement pas évaluée quantitativement. Elle est plutôt évaluée en vérifiant soigneusement la méthode de mesure par rapport à la définition conceptuelle du construit psychologique.

Validité de critère

La **validité de critère** correspond à la mesure dans laquelle les résultats d'une mesure sont corrélés avec d'autres variables (appelés critères) avec lesquelles on s'attendrait à ce qu'ils soient corrélés. Par exemple les résultats d'une nouvelle mesure de l'anxiété liée à un examen devraient être négativement corrélés avec les résultats d'un examen scolaire important. S'il était constaté que les scores des personnes étaient en fait négativement corrélés avec leurs résultats à l'examen, cela constituerait un élément de preuve que ces scores représentent réellement l'anxiété liée à l'examen. Mais si l'on constatait que les personnes obtenaient les mêmes résultats à l'examen, quel que soit leur niveau d'anxiété, cela remettrait en doute la validité de la mesure.

Un **critère** peut être n'importe quelle variable dont on a des raisons de penser qu'elle devrait être corrélée avec le construit mesuré, et il y en aura généralement plusieurs. Par exemple on peut s'attendre à ce que les scores d'anxiété aux examens soient négativement corrélés avec les résultats aux examens et les notes de cours, et positivement corrélés avec l'anxiété générale et la tension artérielle pendant un examen. Ou encore, imaginons qu'une personne menant une recherche développe une nouvelle mesure de la prise de risque physique. Les scores obtenus par les gens sur cette mesure devraient être corrélés avec leur participation à des activités « extrêmes », telles que le snowboard et l'escalade, le nombre de contraventions pour excès

de vitesse reçues et même le nombre d'os cassés qu'ils ont eus au fil des ans. Lorsque le critère est mesuré en même temps que le construit, la validité du critère est appelée **validité concomitante**; tandis que lorsque le critère est mesuré à un moment donné dans le futur (après que le construit a été mesuré), on parle de **validité prédictive** (parce que les résultats de la mesure ont « prédit » un résultat futur).

Les critères peuvent également inclure d'autres mesures du même construit. Par exemple on pourrait s'attendre à ce que de nouvelles mesures de l'anxiété liée à l'examen ou de la prise de risque physique soient positivement corrélées avec des mesures existantes et établies des mêmes construits. Il s'agit de la **validité convergente**.

L'évaluation de la validité convergente nécessite la collecte de données à l'aide de la mesure. C'est ce qu'ont fait les chercheurs John Cacioppo et Richard Petty en créant leur échelle d'auto-évaluation du besoin de cognition (*Need for Cognition Scale* en anglais), afin de mesurer l'importance que les gens valorisent la réflexion et s'y impliquent (Cacioppo et Petty, 1982). Dans une série d'études, ils ont montré que les scores des personnes étaient positivement corrélés avec leurs résultats à un test standardisé de réussite scolaire, et que leurs scores étaient négativement corrélés avec leurs résultats à une mesure du dogmatisme (qui représente une tendance à l'obéissance). Depuis sa création, l'échelle du besoin de cognition a été utilisée dans des centaines d'études et a montré qu'elle était en corrélation avec une grande variété d'autres variables, y compris l'efficacité d'une publicité, l'intérêt envers la politique et les décisions des jurés (Petty, Briñol, Loersch et McCaslin, 2009). Elle a été traduite en français auprès de personnes adolescentes de 10 à 14 ans (Ginet et Py, 2000). Cet article illustre bien les procédures utilisées pour démontrer les qualités psychométriques d'un outil en matière de fidélité et de validité.

Validité discriminante

Concernant la **validité discriminante**, les résultats d'une mesure ne sont pas corrélés avec des mesures de variables conceptuellement distinctes. Par exemple l'estime de soi est une attitude générale à l'égard de soi-même qui est relativement stable dans le temps. Ce n'est pas la même chose que l'humeur, qui est la façon dont on se sent bien ou mal en ce moment. Par conséquent, les scores obtenus par les personnes sur une nouvelle mesure de l'estime de soi ne devraient pas être très fortement corrélés avec leur humeur. Si la nouvelle mesure de l'estime de soi était fortement corrélée avec une mesure de l'humeur, il est possible que la nouvelle mesure ne mesure pas vraiment l'estime de soi, mais plutôt l'humeur.

Lorsqu'ils ont créé l'échelle du besoin de cognition, Cacioppo et Petty ont également apporté la preuve de la validité discriminante en montrant que les scores des personnes n'étaient pas corrélés avec certaines autres variables. Par exemple ils n'ont trouvé qu'une faible corrélation entre le besoin de cognition des gens et une mesure de leur style cognitif, c'est-à-dire leur tendance à penser de manière analytique en divisant les idées en petites parties ou de manière holistique en ayant une « vue d'ensemble ». Il n'y avait pas non plus de corrélation entre le besoin de cognition des gens et les mesures de leur anxiété liée aux examens, et de leur tendance à répondre de manière socialement souhaitable. Toutes ces faibles corrélations montrent que la mesure reflète un construit distinct.

24. Stratégies pratiques pour la mesure psychologique

Objectifs d'apprentissage

1. Préciser les quatre grandes étapes du processus de mesure.
2. Expliquer comment se décide l'utilisation d'une mesure existante ou de créer sa propre mesure.
3. Décrire les multiples stratégies permettant d'identifier et de localiser les mesures existantes des construits psychologiques.
4. Décrire plusieurs principes généraux pour la création de nouvelles mesures et pour la mise en œuvre des mesures existantes et nouvelles.
5. Créer un plan simple pour évaluer la fidélité et la validité d'une mesure existante ou nouvelle.

Jusqu'à présent dans ce chapitre, nous avons examiné plusieurs idées fondamentales concernant la nature des construits psychologiques et leur mesure. Imaginons maintenant que vous soyez dans la position de devoir mesurer un construit psychologique pour un projet de recherche. Comment devez-vous procéder? D'une manière générale, le processus de mesure comporte quatre étapes : **(a)** la définition conceptuelle du construit, **(b)** la définition opérationnelle du construit, **(c)** la mise en œuvre de la mesure et **(d)** l'évaluation de la mesure. Dans cette section, nous examinerons successivement chacune de ces étapes.

Définition conceptuelle du construit

Disposer d'une définition conceptuelle claire et complète d'un construit est une condition préalable à une bonne mesure. Cela vous permet de prendre des décisions judicieuses sur la manière exacte de mesurer le construit. Si vous n'aviez qu'une vague idée de vouloir mesurer la « mémoire » des gens, par exemple, vous n'auriez aucun moyen de choisir entre leur demander de se souvenir d'une liste de mots de vocabulaire, d'une série de photos, d'une compétence nouvellement acquise, d'une expérience vécue il y a longtemps, ou de leur demander de se souvenir d'effectuer une tâche à un moment ultérieur. Comme les psychologues conçoivent aujourd'hui la mémoire comme un ensemble de systèmes semi-indépendants, vous allez devoir préciser ce que vous entendez par « mémoire ». Si vous vous intéressez à la mémoire épisodique à long terme (la mémoire des expériences passées), il est logique de demander aux personnes participantes de se souvenir d'une liste de mots apprise la semaine dernière, mais non d'essayer de se souvenir d'exécuter une

tâche dans le futur. En général, rien ne remplace la lecture de la littérature de recherche sur un construit, et l'attention portée à la manière dont les autres ont défini le construit.

Définition opérationnelle du construit

Une fois que vous disposez d'une définition conceptuelle du construit que vous souhaitez étudier, il est temps de le définir de manière opérationnelle. Rappelons qu'une définition opérationnelle est une définition de la variable en des termes qui précisent comment elle sera mesurée. Étant donné que la plupart des variables sont des concepts relativement abstraits qui ne peuvent être observés directement (par exemple, le stress) et que l'observation est au cœur de la méthode scientifique, les définitions conceptuelles doivent être transformées en quelque chose qui peut être directement observé et mesuré. La plupart des variables peuvent être définies opérationnellement de différentes manières. Par exemple, le stress peut être défini de manière opérationnelle comme les scores des personnes sur une échelle de stress, tel que l'échelle du stress perçu (Cohen, Kamarck et Mermelstein, 1983), les concentrations de cortisol dans leur salive, ou le nombre d'événements stressants vécus récemment. Comme décrit ci-dessous, la définition opérationnelle de votre/vos variables d'intérêt peut impliquer l'utilisation d'une mesure existante ou la création de votre propre mesure.

Utilisation d'une mesure existante

C'est généralement une bonne idée d'utiliser une mesure existante déjà utilisée avec succès dans des recherches antérieures. Parmi les avantages, on peut citer : **(a)** vous économisez du temps et des efforts nécessaires pour créer votre propre mesure, **(b)** il existe déjà des preuves de la validité de la mesure (si elle a été utilisée avec succès) et **(c)** vos résultats peuvent être plus facilement comparés et combinés avec les résultats antérieurs. En fait, s'il existe déjà une mesure fidèle et valide d'un construit, les autres personnes menant des recherches peuvent s'attendre à ce qu'on l'utilise, à moins d'avoir une bonne raison clairement énoncée de ne pas le faire.

Si vous choisissez d'utiliser une mesure existante, il se peut que vous deviez encore choisir entre plusieurs possibilités. Vous pourriez choisir la plus courante, celle dont la fidélité et la validité sont les mieux démontrées, celle qui mesure le mieux un aspect particulier d'un construit qui vous intéresse (par exemple une mesure physiologique du stress si vous vous intéressez plus particulièrement à sa physiologie sous-jacente), ou même celle qui serait la plus facile à utiliser. Par exemple le Ten-Item Personality Inventory (TIPI) est un questionnaire d'auto-évaluation qui mesure les cinq grandes (Big Five) dimensions de la personnalité avec seulement 10 items (Gosling, Rentfrow et Swann, 2003). Il n'est pas aussi fidèle ou valide que des mesures plus longues et plus complètes, mais une personne menant une recherche peut choisir de l'utiliser lorsque le temps de passation est très limité.

Lorsqu'une mesure existante est créée principalement pour être utilisée dans le cadre d'une recherche scientifique, elle est généralement décrite en détail dans un article de recherche publié, et peut être utilisée librement dans le cadre de votre propre recherche — avec une citation appropriée. Il se peut que des gens

menant des recherches ultérieurement avec cette même mesure ne la décrivent que brièvement, mais qu'ils fournissent une référence à l'article original, auquel cas vous devrez obtenir les détails dans l'article original. L'American Psychological Association publie également le Directory of Unpublished Experimental Measures (Dossier de mesures expérimentales non publiées) et [PsycTESTS](#), qui sont de vastes catalogues/collections de mesures utilisées dans des recherches précédentes. De nombreuses mesures existantes — en particulier celles qui ont des applications en psychologie clinique — sont protégées par des droits de propriété. Cela signifie que la maison d'édition détient les droits et que vous devez les acheter. Il s'agit notamment de nombreux tests d'intelligence, de l'inventaire de dépression de Beck et de l'inventaire multiphasique de personnalité du Minnesota (MMPI). Vous trouverez des détails sur beaucoup de ces mesures et sur la manière de les obtenir dans d'autres livres de référence, notamment Tests in Print et Mental Measurements Yearbook. Il y a de fortes chances que vous puissiez trouver ces ouvrages de référence dans les collections de la bibliothèque de votre université ou dans la testothèque du département de psychologie. Pour les tests en français, il y a notamment l'[Institut de recherches psychologiques](#), [Pearson Canada](#) et [Psychometrics](#) qui décrivent ou même donnent accès à des tests.

Créer sa propre mesure

Au lieu d'utiliser une mesure existante, vous pourriez vouloir créer votre propre mesure. Il se peut qu'il n'existe aucune mesure du concept qui vous intéresse ou que les mesures existantes soient trop difficiles ou trop longues à utiliser. Il se peut aussi que vous souhaitiez utiliser une nouvelle mesure spécifiquement pour voir si elle fonctionne de la même manière que les mesures existantes, c'est-à-dire pour évaluer la validité convergente. Dans cette section, nous examinons certaines questions générales relatives à la création de nouvelles mesures qui s'appliquent également aux mesures d'auto-évaluation, comportementales et physiologiques. Des lignes directrices plus détaillées pour la création de mesures d'auto-évaluation sont présentées au chapitre 7.

Tout d'abord, sachez que la plupart des nouvelles mesures en psychologie sont en fait des variations de mesures existantes, de sorte que vous devez toujours chercher des idées dans la littérature de recherche. Vous pouvez peut-être modifier un questionnaire existant, créer une version papier et crayon d'une mesure qui est normalement informatisée (ou vice versa), ou adapter une mesure qui a été traditionnellement utilisée à d'autres fins. Par exemple la célèbre tâche de Stroop (Stroop, 1935) — qui consiste à nommer rapidement les couleurs dans lesquelles différents mots de couleur sont imprimés — a été adaptée à l'étude de l'anxiété sociale. Les personnes souffrant d'anxiété sociale sont moins efficaces à nommer les couleurs lorsque les mots ont des connotations sociales négatives comme « stupide » (Amir, Freshman, et Foa, 2002).

Lorsque vous créez une nouvelle mesure, vous devriez viser la simplicité. N'oubliez pas que l'intérêt élevé que vous avez envers votre projet de recherche n'est peut-être pas partagé par vos personnes participantes. Il est possible que leur capacité à comprendre et à exécuter la tâche que vous leur confiez soit très variable. Vous devez créer un ensemble d'instructions claires utilisant un langage simple que vous pouvez présenter par écrit ou lire à haute voix (ou les deux). C'est aussi une bonne idée d'inclure un ou plusieurs exercices pratiques afin que les personnes participantes puissent se familiariser avec la tâche et avoir l'occasion de poser des questions avant de poursuivre. Il est également préférable que la mesure soit brève afin d'éviter

d'ennuyer ou de frustrer vos personnes participantes au point que leurs réponses deviennent moins fidèles et valides.

Cependant le besoin de brièveté doit être pondéré par le fait qu'il est presque toujours préférable pour une mesure d'inclure plusieurs items plutôt qu'un seul. Il y a deux raisons à cela. La première est une question de validité de contenu. Plusieurs items sont souvent nécessaires pour couvrir un construit de manière appropriée. La deuxième est une question de fidélité. Les réponses à un seul item peuvent être influencées par toutes sortes de facteurs non pertinents — une mauvaise compréhension de l'item en question, une distraction momentanée ou une simple erreur telle que cocher la mauvaise réponse. Mais lorsque plusieurs réponses sont additionnées ou moyennées, les effets de ces facteurs non pertinents tendent à s'annuler pour produire des scores plus fidèles. Cependant n'oublions pas que les questions multiples doivent être structurées de manière à pouvoir être combinées en une seule note globale, par leur addition ou le calcul de leur moyenne. Pour mesurer la « responsabilité financière », une étudiante pourrait vouloir connaître le revenu annuel des gens, obtenir leur cote de crédit et leur demander d'évaluer à quel point ils se considèrent « économes », mais il n'y a pas de moyen évident pour combiner ces réponses et obtenir un score global. Pour créer une véritable mesure à items multiples, l'étudiante pourrait plutôt demander aux gens d'évaluer à quel point dix affirmations sur la responsabilité financière les décrivent sur la même échelle de cinq points.

Finalement la meilleure façon de s'assurer que votre mesure comporte des instructions claires, contient suffisamment de pratique et est d'une longueur appropriée, est de tester plusieurs personnes. Observez-les pendant qu'elles accomplissent la tâche, chronométrez-les et demandez-leur ensuite de commenter la facilité ou la difficulté de la tâche, la clarté des instructions et toute autre question que vous vous posez. Il est évident qu'il est préférable de découvrir les problèmes liés à une mesure avant d'entreprendre une collecte de données à grande échelle.

Mise en œuvre de la mesure

Vous allez vouloir mettre en œuvre toute mesure de manière à en maximiser sa fidélité et sa validité. Dans la plupart des cas, il est préférable de tester tout le monde dans des conditions similaires, qui sont idéalement calmes et exemptes de distractions. Les personnes participantes sont souvent évaluées en groupe pour des raisons d'efficacité, mais il faut savoir que cela peut créer des distractions qui réduisent la fidélité et la validité de la mesure. Comme toujours, il est bon de s'inspirer des recherches antérieures. Si d'autres ont testé avec succès des personnes en groupe à l'aide d'une mesure particulière, vous devriez envisager de le faire également.

Il faut aussi être conscient que les personnes participantes peuvent réagir de diverses manières aux mesures, ce qui réduit la fidélité et la validité des résultats. Bien que des personnes désagréables puissent intentionnellement réagir de manière à perturber l'étude, la réactivité des gens est plus susceptible de prendre le sens inverse. Les personnes accommodantes peuvent répondre suivant la manière dont elles pensent qu'on s'attend qu'elles le fassent. Des gens pourraient s'engager dans des **réponses socialement souhaitables**, en faisant ou en disant des choses parce qu'ils pensent que c'est la chose socialement appropriée. Par exemple les personnes ayant une faible estime de soi indiquent être une personne de valeur, non pas par conviction de leur valeur, mais par impression que c'est la réponse socialement appropriée. Ces personnes ne veulent pas mal paraître aux yeux de l'équipe de recherche. En outre les études de recherche

peuvent comporter des **caractéristiques intégrées de la demande**: des indices subtils qui révèlent la manière dont la personne en recherche s'attend à ce que les personnes participantes se comportent. Par exemple une participante dont l'attitude à l'égard de l'exercice physique est mesurée immédiatement après qu'on lui ait demandé de lire un passage sur les dangers de la maladie cardiaque pourrait raisonnablement conclure que ce passage est destiné à améliorer son attitude. En conséquence elle pourrait répondre plus favorablement, car elle pense que la personne menant la recherche s'attend à ce qu'elle le fasse. Enfin vos propres attentes peuvent fausser les comportements des personnes participantes de manière involontaire.

Plusieurs précautions peuvent être prises pour minimiser ce type de réactivité. L'une d'entre elles consiste à rendre la procédure aussi claire et brève que possible afin que les personnes participantes ne soient pas tentées d'exprimer leurs frustrations sur vos résultats. Une autre consiste à garantir l'anonymat des personnes participantes et à leur expliquer clairement que vous le faites. Lorsque vous faites passer ces tests en groupe, veillez à ce que les gens soient suffisamment éloignés les uns des autres pour les empêcher de voir les réponses des autres. Donnez-leur tous le même type de matériel d'écriture afin de prévenir leur identification, par exemple par l'utilisation du stylo rose à paillettes. Vous pouvez même leur permettre de sceller les questionnaires remplis dans des enveloppes individuelles ou de les déposer dans une boîte de dépôt où ils seront immédiatement mélangés aux autres questionnaires. Bien que le consentement éclairé exige d'informer les personnes participantes de ce qui sera fait pendant l'étude, il n'exige pas de révéler votre hypothèse ni d'autres informations qui pourraient leur suggérer la manière dont vous vous attendez à ce qu'elles répondent. Un questionnaire conçu pour mesurer la responsabilité financière ne doit pas nécessairement s'intituler « Êtes-vous financièrement responsable? ». Il peut s'intituler « Questionnaire sur l'argent » ou ne pas avoir de titre du tout. Enfin les effets de vos attentes peuvent être minimisés en organisant la passation de la mesure par une personne assistante qui est « à l'aveugle », c'est-à-dire qui n'est pas au courant de l'objectif ni de l'hypothèse testée. Que cela soit possible ou non, il est important de standardiser toutes les interactions entre les personnes en recherche et les personnes participantes — par exemple en lisant toujours la même série d'instructions, mot pour mot.

Évaluation de la mesure

Une fois que vous avez utilisé votre mesure auprès d'un échantillon de personnes et que vous disposez d'un ensemble de scores, il est possible de l'évaluer de manière plus approfondie en matière de fidélité et de validité. Même si la mesure a été largement utilisée par d'autres personnes en recherche et a déjà montré des preuves de fidélité et de validité, vous ne devez pas présumer qu'elle a fonctionné comme prévu pour votre échantillon particulier et dans vos conditions particulières de passation. Quoi qu'il en soit, vous disposez désormais de données supplémentaires concernant la fidélité et la validité de la mesure, et il serait logique d'ajouter ces preuves à la littérature de recherche.

Dans la plupart des devis de recherche, il n'est pas possible d'évaluer la fidélité test-retest, car les personnes participantes ne sont testées qu'une seule fois. Pour une nouvelle mesure, vous pourriez concevoir une étude spécifiquement destinée à évaluer sa fidélité test-retest en testant le même groupe de gens à deux moments distincts. Dans d'autres cas, une étude conçue pour répondre à une question différente permet tout de même d'évaluer la fidélité test-retest. Par exemple un membre du corps professoral en psychologie peut mesurer l'attitude de sa cohorte étudiante à l'égard de la pensée critique en utilisant

la même mesure au début et à la fin du semestre pour voir s'il y a un changement sur le plan de la pensée critique. Même s'il n'y a pas de changement à ce niveau, il peut toujours examiner la corrélation entre les résultats de son échantillon aux deux moments pour évaluer la fidélité test-retest de la mesure. Il est également d'usage d'évaluer la cohérence interne de toute mesure à items multiples — généralement en examinant la corrélation moitié-moitié (*split-half* en anglais) ou l' α de Cronbach.

La validité de critère peut être évaluée de différentes manières. Par exemple si votre étude comprend plusieurs mesures du même construit ou des mesures de construits conceptuellement distincts, vous devez examiner les corrélations entre ces mesures pour vous assurer qu'elles correspondent à vos attentes. Notez également qu'une manipulation expérimentale réussie fournit également une preuve de la validité du critère. Rappelons que MacDonald et Martineau ont manipulé l'humeur des personnes participantes en leur faisant penser soit à des choses positives, soit à des choses négatives, et qu'après la manipulation, leur mesure de l'humeur a montré une différence distincte entre les deux groupes. Cela prouve à la fois que leur manipulation de l'humeur a fonctionné et que leur mesure de l'humeur est valide.

Mais que se passe-t-il si les nouvelles données recueillies remettent en question la fidélité ou la validité de votre mesure? La réponse courte est qu'il faut se demander pourquoi. Il se peut qu'il y ait un problème avec votre mesure ou avec la manière dont vous l'avez administrée. Il se peut que votre définition conceptuelle soit erronée. Il se peut que votre manipulation expérimentale ait échoué. Par exemple, si une mesure de l'humeur ne montre aucune différence entre les personnes à qui vous avez demandé de penser à des choses positives et celles à qui vous avez demandé de penser à des choses négatives, c'est peut-être parce que les gens n'ont pas réellement pensé aux choses auxquelles ils étaient censés penser ou que les choses n'ont pas réellement affecté leur humeur. En bref, il faut « retourner à la planche à dessin » pour réviser la mesure, revoir la définition conceptuelle ou essayer une nouvelle manipulation.

25. Points clés et exercices

Points clés

- La mesure est l'attribution de scores à des personnes de manière à ce que ces scores représentent une certaine caractéristique de ces personnes. La mesure psychologique peut être réalisée de différentes manières, notamment par des mesures d'auto-évaluation comportementales et physiologiques.
- Les construits psychologiques tels que l'intelligence, l'estime de soi et la dépression ne sont pas directement observables parce qu'ils représentent des tendances comportementales ou des modèles complexes de comportements et de processus internes. Un objectif important de la recherche scientifique est de définir conceptuellement les construits psychologiques de manière à les décrire avec précision.
- Pour toute définition conceptuelle d'un construit, il y aura plusieurs définitions opérationnelles possibles ou de manières de le mesurer. L'utilisation de plusieurs définitions opérationnelles ou opérations convergentes est une stratégie fréquente dans la recherche en psychologie.
- Les variables peuvent être mesurées avec 4 types d'échelles différentes — nominale, ordinale, intervalle et rapport — qui communiquent des quantités croissantes d'informations quantitatives. Le type d'échelle affecte le type de statistiques pouvant être utilisé et les conclusions pouvant être tirées.
- Les personnes menant des recherches en psychologie ne présument pas simplement que leurs mesures fonctionnent. Les scientifiques mènent plutôt des recherches pour démontrer leur efficacité et si leur efficacité n'est pas démontrée, leur utilisation est arrêtée.
- Les personnes en recherche évaluent leurs mesures en fonction de deux critères distincts : la fidélité et la validité. La fidélité est la cohérence dans le temps (fidélité test-retest), entre les items (cohérence interne) et entre les scientifiques (fidélité interjuges). La validité est le degré auquel les scores représentent réellement la variable qu'ils sont censés représenter.
- La validité est un jugement basé sur différents types de preuves. Les éléments de preuve pertinents comprennent la fidélité de la mesure, le fait qu'elle couvre le construit d'intérêt et le fait que les scores qu'elle produit sont corrélés avec d'autres variables avec lesquelles ils sont censés être corrélés et le fait que ces scores ne sont pas corrélés avec des variables conceptuellement distinctes.
- Une bonne mesure commence par une définition conceptuelle claire du construit à mesurer. Cela se fait à la fois par une réflexion claire et détaillée et par une analyse de la littérature de recherche.
- Vous avez souvent la possibilité d'utiliser une mesure existante ou d'en créer une nouvelle. Vous devriez prendre cette décision en fonction de la disponibilité des mesures existantes et de leur adéquation à vos objectifs.

- Plusieurs étapes simples peuvent être suivies lors de la création de nouvelles mesures et de la mise en œuvre de mesures existantes et nouvelles, afin de maximiser la fidélité et la validité de ces mesures.
- Une fois que vous avez utilisé une mesure, vous devriez réévaluer sa fidélité et sa validité sur la base de vos nouvelles données. N'oubliez pas que l'évaluation de la fidélité et de la validité est un processus continu.

Exercices

- **Pratique** : Complétez [l'échelle d'estime de soi de Rosenberg](#) et calculez votre score global.
- **Pratique** : Réfléchissez à trois définitions opérationnelles de la jalousie sexuelle, de la prise de décision et de l'anxiété sociale. Considérez la possibilité d'utiliser des mesures d'auto-évaluation, comportementales et physiologiques. Soyez aussi précis que possible.
- **Pratique** : Pour chacune des variables suivantes, déterminez le type d'échelle utilisé.
 - Un professeur d'université mesure le temps pris pour terminer l'examen pour sa cohorte étudiante en regardant la pile d'examens à la fin de la période. En commençant par le bas de la pile, il attribue la note 1 au premier examen remis, la note 2 au deuxième, et ainsi de suite.
 - Une chercheuse consulte les dossiers médicaux de ses personnes participantes et compte le nombre de fois où ils ont vu un médecin au cours de l'année écoulée.
 - Dans une étude, on demande aux personnes participantes si elles sont droitières ou gauchères.
- **Pratique** : Demandez à plusieurs collègues de classe de remplir l'échelle d'estime de soi de Rosenberg. Évaluez ensuite sa cohérence interne en établissant un diagramme de dispersion pour montrer la corrélation moitié-moitié (éléments pairs et éléments impairs). Calculez également le coefficient de corrélation si vous savez comment le faire.
- **Discussion** : Repensez au dernier examen universitaire que vous avez passé et considérez l'examen comme une mesure psychologique. Quel construit pensez-vous qu'il était censé mesurer? Commentez sa validité apparente et sa validité de contenu. Quelles données pourriez-vous recueillir pour évaluer sa fidélité et sa validité de critère?
- **Pratique** : Rédigez votre propre définition conceptuelle de la confiance en soi, de l'irritabilité et de l'athlétisme.
- **Pratique** : Choisissez un construit (jalousie sexuelle, confiance en soi, etc.) et trouvez deux mesures de ce construit dans la littérature de recherche. Si vous meniez votre propre étude, laquelle de ces mesures utiliseriez-vous (ou aucune) et pourquoi?

26. Bibliographie

- Amir, N., Freshman, M. et Foa, E. (2002). Enhanced Stroop interference for threat in social phobia (Interférence Stroop accrue pour la menace dans la phobie sociale). *Journal of Anxiety Disorders*, 16(1), 1–9. [https://doi.org/10.1016/S08876185\(01\)00084-6](https://doi.org/10.1016/S08876185(01)00084-6)
- Bandura, A., Ross, D. et Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models (Transmission de l'agressivité par l'imitation de modèles agressifs). *Journal of Abnormal and Social Psychology*, 63(3), 575–582. <https://doi.org/10.1037/h0045925>
- Cacioppo, J. T. et Petty, R. E. (1982). The need for cognition (Le besoin de cognition). *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cohen, S., Kamarck, T. et Mermelstein, R. (1983). A global measure of perceived stress (Une mesure globale de la perception du stress). *Journal of Health and Social Behavior*, 24(4), 386–396. <https://doi.org/10.2307/2136404>
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. (Une mesure globale de la perception du stress). *Journal of Health and Social Behavior*, 24(4), 386–396. <https://doi.org/10.2307/2136404>
- Costa, P. T., Jr., et McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory (L'évaluation de la personnalité normale dans la pratique clinique : L'inventaire de personnalité NEO). *Psychological Assessment*, 4(1), 5–13. <https://doi.org/10.1037/1040-3590.4.1.5>
- Delongis, A., Coyne, J. C., Dakof, G., Folkman, S. et Lazarus, R. S. (1982). Relationships of daily hassles, uplifts, and major life events to health status (Relation entre les tracas quotidiens, les satisfactions et les événements majeurs de la vie et l'état de santé). *Health Psychology*, 1(2), 119–136. <https://doi.org/10.1037/0278-6133.1.2.119>
- Ginet, A. et Py, J. (2000). Le besoin de cognition : une échelle française pour enfants et ses conséquences au plan sociocognitif. *L'Année psychologique*, 100(4), 585–628. <https://doi.org/10.3406/psy.2000.28665>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains (Une mesure très succincte des cinq grands domaines de la personnalité). *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Guillet, L. et Hermand, D. (2006). Critique de la mesure du stress. *L'Année psychologique*, 106, 129–164. <https://hal.science/hal-00755690>
- Hogan, T. P., Stephenson, R. et Parent, N. (2007). *Introduction à la psychométrie* (2e éd.). Chenelière éducation.
- Holmes, T. H. et Rahe, R. H. (1967). The Social Readjustment Rating Scale (L'échelle d'évaluation de la réadaptation sociale). *Journal of Psychosomatic Research*, 11(2), 213–218. [https://doi.org/10.1016/0022-3999\(67\)90010-4](https://doi.org/10.1016/0022-3999(67)90010-4)

- MacDonald, T. K., et Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? (L'estime de soi, l'humeur et l'intention d'utiliser des préservatifs : Quand une faible estime de soi conduit-elle à des comportements de santé à risque ?) *Journal of Experimental Social Psychology*, 38(3), 299–306. <https://doi.org/10.1006/jesp.2001.1505>
- Le Corff, Y. et Busque-Carrier, M. (2016). Structural validity of the NEO personality inventory 3 (NEO-PI-3) in a French-Canadian sample (Validité structurelle de l'inventaire de personnalité NEO 3 (NEO-PI-3) auprès d'un échantillon canadien-français). *International Journal of Arts & Sciences*, 9(3), 461-472.
- Le Corff, Y. (s.d.). NEO-PI3™ : *Inventaire de personnalité : Traduction et adaptation canadienne-française*. <https://www.irpcanada.com/neo-pi3>
- Paquette Houde, C., Abdel Baki, A., Lecomte, T., Lussier, Valade, M. et Ngô, T. L. (2018). *Guide de pratique pour le traitement cognitif-comportemental des troubles psychotiques* (1ère éd.). Jean Goulet – tccmontreal. <https://collections.banq.qc.ca/ark:/52327/3533838>
- Petty, R. E., Briñol, P., Loersch, C. et McCaslin, M. J. (2009). The need for cognition. Dans M. R. Leary et R. H. Hoyle (dir.) *Handbook of Individual Differences in Social Behavior* (p.318-329). Guilford Press
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press
- Rosenberg, M. (1989). *Society and the adolescent self-image* (éd. rév.). Wesleyan University Press.
- Segerstrom, S. E. et Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry (Le stress psychologique et le système immunitaire humain : Une étude méta-analytique de 30 ans de recherche). *Psychological Bulletin*, 130(4), 601–630. <https://doi.org/10.1037/0033-2909.130.4.601>
- Stevens, S. S. (1946). On the theory of scales of measurement (Sur la théorie des échelles de mesure). *Science*, 103(2684), 677–680.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions (Études de l'interférence dans les réactions verbales en série). *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Vallière, E. F., et Vallerand, R. J. (1990). Traduction et validation canadienne-française de l'échelle de l'estime de soi de Rosenberg. *International Journal of Psychology*, 25, 305-316. <https://doi.org/10.1080/00207599008247865>

CHAPITRE 5-LA RECHERCHE EXPÉRIMENTALE



À la fin des années 1960, les psychologues sociaux John Darley et Bibb Latané ont proposé une hypothèse contre-intuitive. Plus il y a de témoins d'un accident ou d'un crime, moins il y a de chances que l'un d'entre eux vienne en aide à la victime (Darley et Latané, 1968).

Ils ont aussi suggéré la théorie selon laquelle ce phénomène se produit parce que chaque témoin se sent moins responsable d'aider — un processus appelé « dilution de la responsabilité ». Darley et Latané ont noté que leurs idées étaient cohérentes avec de nombreux cas réels. Par exemple, une femme de New York nommée Catherine « Kitty » Genovese a été agressée et assassinée alors que plusieurs témoins n'ont manifestement rien fait pour l'aider. Mais Darley et Latané ont également compris que de tels cas isolés ne fournissaient pas de preuves convaincantes de leur hypothèse de « l'effet du témoin ». Il n'y avait aucun moyen de savoir par exemple si l'une des personnes témoins du meurtre de Kitty Genovese aurait aidé, si ces témoins avaient été moins nombreux.

Pour tester leur hypothèse, Darley et Latané ont donc créé une situation d'urgence simulée dans un laboratoire auprès de personnes étudiant à l'université. Chaque personne participante à l'étude a été isolée dans une petite pièce et on lui a dit qu'il y aurait une discussion sur la vie universitaire avec d'autres personnes étudiantes par un système d'interphone. Cependant au début de la discussion, l'une des personnes a commencé à avoir ce qui semblait être une crise d'épilepsie. Sur l'interphone, on entend ce qui suit : « J'aurais vraiment-euh-besoin d'aide, alors si quelqu'un pouvait-euh-me donner un peu d'aide, quelqu'un pourrait-il m-m'aider-euh (bruits d'étouffement)... Je vais mourir, je vais... m-mourir, a-aidez-moi... euh-euh-crise d'épilepsie-euh [étouffement, puis silence] » (Darley et Latané, 1968, p. 379, traduction).

En réalité il n'y avait pas d'autres personnes étudiantes. Ces commentaires avaient été préenregistrés et diffusés afin de donner l'impression d'une véritable situation d'urgence. La clé de l'étude réside dans le fait que certaines personnes participantes ont été informées que la discussion ne concernait qu'une seule autre personne étudiante (la victime), d'autres ont été informées qu'elle concernait deux autres personnes étudiantes, et d'autres encore ont été informés qu'elle concernait cinq autres personnes étudiantes. Comme il s'agissait de la seule différence entre ces trois groupes de personnes participantes, toute différence dans leur tendance à aider la victime devait en être la cause. Et bien sûr, la probabilité que la personne participante quitte la pièce pour aller chercher de l'aide pour la « victime » a diminué de 85 % à 62 % et à 31 % à mesure que le nombre de « témoins » augmentait.

La parabole des 38 témoins

L'histoire de Kitty Genovese a été racontée et répétée dans de nombreux manuels de psychologie. La version standard est qu'il y avait 38 témoins du crime, que tous ont regardé (ou écouté) pendant une longue période, et qu'aucun d'entre eux n'a fait quoi que ce soit pour aider la victime. Cependant de récentes recherches suggèrent que l'histoire standard est inexacte à de nombreux égards (Manning, Levine et Collins, 2007). Par exemple seuls six témoins oculaires ont témoigné au procès, aucune de ces personnes ne savait qu'elle assistait à une agression mortelle, et plusieurs rapports font état de témoins qui ont appelé la police ou qui sont même venus en aide à Kitty Genovese. Bien que l'histoire standard ait inspiré une longue série de recherches sur « l'effet du témoin » et la dilution de la responsabilité, elle a peut-être aussi détourné l'attention des personnes en recherche et des personnes aux études de d'autres questions tout aussi intéressantes et importantes de la psychologie de l'aide — notamment les conditions dans lesquelles les gens réagissent en fait collectivement à des situations d'urgence. Pour en savoir plus sur les processus conduisant à l'inaction, les situations pouvant réduire l'effet du témoin, ou même l'empêcher, et la description plus exacte de ce qu'ont vu et entendu les 38 témoins, lisez l'article de Peggy Chekroun (2008).



*Lecture sur version imprimée?
Scannez ce code QR avec votre
téléphone cellulaire ou allez
sur
<https://psychologiescientifique.org/wp-content/uploads/2018/02/Chekroun-2008-Pourquoi-les-individus-aident-ils-moins-autrui-lor.pdf>*

La recherche menée par Darley et Latané correspond à un type particulier d'étude appelée « expérience ». Les expériences sont utilisées pour déterminer non seulement s'il existe une relation significative entre deux variables, mais aussi si la relation est causale et si elle est soutenue par une analyse statistique. C'est pourquoi les expériences sont l'un des outils les plus courants et les plus utiles dans la boîte à outils de la personne menant des recherches en psychologie. Dans ce chapitre, nous examinons les expériences en détail. Nous examinerons tout d'abord ce qui distingue les expériences des autres types d'études et pourquoi elles soutiennent des conclusions causales alors que d'autres types d'études ne le font pas. Nous examinerons ensuite deux devis de base pour concevoir une expérience : les devis intersujets et les devis intrasujets, et nous discuterons de leurs avantages et de leurs inconvénients. Enfin nous examinerons plusieurs questions pratiques importantes qui se posent lors de la réalisation d'expériences.

27. Expérience : notions de base

Objectifs d'apprentissage

1. Expliquer ce qu'est une expérience et reconnaître des exemples d'études qui sont des expériences, et d'autres qui n'en sont pas.
2. Faire la distinction entre la manipulation de la variable indépendante et le contrôle des variables externes et expliquer l'importance de chacune d'entre elles.
3. Reconnaître des exemples de variables confondantes et expliquer comment elles affectent la validité interne d'une étude.
4. Définir ce qu'est une condition contrôle, expliquer son rôle dans la recherche sur l'efficacité des traitements et décrire d'autres types de conditions contrôles.

Qu'est-ce qu'une expérience?

Comme nous l'avons vu plus tôt dans ce livre, une **expérience** est un type d'étude conçue spécifiquement pour répondre à la question de savoir s'il existe un lien de causalité entre deux variables. En d'autres termes, il s'agit de savoir si la modification d'une variable (appelée **variable indépendante**) entraîne une modification d'une autre variable (appelée **variable dépendante**). Les expériences ont deux caractéristiques fondamentales. La première est que les personnes en recherche manipulent ou font varier systématiquement le niveau de la variable indépendante. Les différents niveaux de la variable indépendante sont appelés **conditions**. Par exemple dans l'expérience de Darley et Latané, la variable indépendante était le nombre de témoins que les personnes participantes croyaient présents. L'équipe de recherche a manipulé cette variable indépendante en disant aux personnes participantes qu'il y avait soit une, soit deux, soit cinq autres personnes aux études impliquées dans la discussion, créant ainsi trois conditions. Pour une nouvelle personne en recherche, il est facile de confondre ces termes en croyant qu'il y a trois variables indépendantes dans cette situation : une, deux ou cinq personnes étudiantes impliquées dans la discussion, alors qu'il n'y a en fait qu'une seule variable indépendante (le nombre de témoins) avec trois différents niveaux ou conditions (une, deux ou cinq personnes). La deuxième caractéristique fondamentale d'une expérience est que les scientifiques exercent un **contrôle** sur des variables autres que la variable indépendante et la variable dépendante, ou que ces gens minimisent la variabilité de ces variables. Ces autres variables sont appelées **variables externes**. Darley et Latané ont testé toutes leurs personnes participantes dans la même pièce, les ont exposés à la même situation d'urgence, etc. Ils ont également assigné aléatoirement leurs personnes participantes aux conditions afin que les trois groupes soient similaires les uns aux autres au départ. Notons que, même si les mots manipulation et contrôle ont des

significations similaires dans le langage courant, les scientifiques font une distinction claire entre eux. On manipule la variable indépendante en modifiant systématiquement ses niveaux et on contrôle les autres variables en les maintenant constantes.

Manipulation de la variable indépendante

Là encore **manipuler** une variable indépendante signifie en modifier systématiquement le niveau de manière à ce que différents groupes de personnes participantes soient exposés à différents niveaux de cette variable, ou exposer le même groupe de personnes à différents niveaux à différents moments. Par exemple pour déterminer si l'écriture expressive affecte la santé des gens, une personne menant une recherche peut demander à des personnes participantes d'écrire sur des expériences traumatisantes et à d'autres, d'écrire sur des expériences neutres. Les différents niveaux de la variable indépendante sont appelés conditions, et les scientifiques donnent souvent aux conditions des noms courts et descriptifs pour qu'il soit facile d'en parler et d'écrire à leur sujet. Dans ce cas, les conditions peuvent être appelées condition traumatique et condition neutre.

Notons que la manipulation d'une variable indépendante doit impliquer l'intervention active de l'équipe de recherche. Comparer des groupes de personnes, qui diffèrent sur la variable indépendante avant le début de l'étude, n'est pas la même chose que de manipuler cette variable. Par exemple l'équipe de recherche qui compare la santé des personnes, qui tiennent déjà un journal à celle des personnes qui n'en tiennent pas, n'a pas manipulé cette variable et n'a donc pas mené d'expérience. Cette distinction est importante, car les groupes qui diffèrent déjà d'une manière au début d'une étude sont susceptibles de différer également d'autres manières. Par exemple les personnes qui choisissent de tenir un journal peuvent également être plus consciencieuses, plus introverties ou moins stressées que les autres. Par conséquent toute différence observée entre les deux groupes, en termes de santé, peut avoir été causée par le fait qu'elles tiennent ou non un journal, ou causée par toute autre différence entre les personnes qui tiennent un journal et celles qui n'en tiennent pas. Ainsi la manipulation active de la variable indépendante est donc cruciale pour éliminer les autres explications possibles des résultats.

Bien entendu il existe de nombreuses situations dans lesquelles la variable indépendante ne peut pas être manipulée pour des raisons pratiques ou éthiques et où une expérience n'est donc pas possible. Par exemple il n'est pas possible de manipuler le fait qu'une personne ait ou non une expérience significative de maladie précoce, ce qui rend impossible la réalisation d'une expérience sur l'effet des expériences de maladie précoce sur le développement de l'hypocondrie. Cette mise en garde ne signifie pas qu'il est impossible d'étudier la relation entre les expériences de maladie précoce et l'hypocondrie, mais seulement qu'il faut le faire en utilisant des approches non expérimentales. Nous discuterons de ce type de méthodologie en détail plus loin dans le livre.

Les variables indépendantes peuvent être manipulées pour créer deux conditions; les expériences qui impliquent une variable indépendante unique avec deux conditions sont souvent désignées comme un **devis à facteur unique à deux niveaux**. Cependant il est parfois possible d'obtenir de meilleures informations en ajoutant davantage de conditions à une expérience. Lorsqu'une expérience comporte une variable indépendante qui est manipulée pour produire plus de deux conditions, on parle de **devis à facteur unique à plusieurs niveaux**. Ainsi plutôt que de comparer une condition dans laquelle il y a un témoin à une

condition dans laquelle il y a cinq témoins (ce qui représenterait un plan à facteur unique de deux niveaux), l'expérience de Darley et Latané a utilisé un plan à un facteur et à plusieurs niveaux, en manipulant la variable indépendante pour produire trois conditions (une condition à un témoin, une condition à deux témoins et une condition à cinq témoins).

Contrôle des variables externes

Comme nous l'avons vu précédemment dans ce chapitre, une variable externe est tout ce qui varie dans le contexte d'une étude autre que les variables indépendantes et dépendantes. Dans une expérience sur l'effet de l'écriture expressive sur la santé par exemple, les variables externes incluent les variables des personnes participantes (différences individuelles) telles que leur capacité d'écriture, leur régime alimentaire et leur sexe. Seraient également incluses des variables situationnelles ou de tâches, telles que le moment de la journée où les personnes participantes écrivent, si elles écrivent à la main ou sur un ordinateur, et le type de temps qu'il fait. Les variables externes posent un problème, car beaucoup d'entre elles sont susceptibles d'avoir un effet sur la variable dépendante. Par exemple la santé des personnes participantes sera affectée par de nombreux facteurs autres que le fait de s'adonner ou non à l'écriture expressive. Ce facteur d'influence peut rendre difficile la séparation de l'effet de la variable indépendante des effets des variables externes, c'est pourquoi il est important de contrôler les variables externes en les maintenant constantes.

Les variables externes en tant que « bruit »

Les variables externes rendent difficile la détection de l'effet de la variable indépendante de deux manières. La première consiste à ajouter de la variabilité ou du « bruit » aux données. Imaginons une expérience simple sur l'effet de l'humeur (heureuse ou triste) sur le nombre d'événements heureux de l'enfance dont les gens sont capables de se souvenir. Les personnes participantes sont mises dans un état d'esprit négatif ou positif (en leur montrant un clip vidéo joyeux ou triste), puis on leur demande de se rappeler autant d'événements heureux de leur enfance que possible. Les deux colonnes les plus à gauche du Tableau 5.1 montrent à quoi les données pourraient ressembler s'il n'y avait pas de variables externes et si le nombre d'événements heureux de l'enfance dont se souviennent les personnes participantes n'était affecté que par leur humeur. Chaque personne participante dans la condition d'humeur joyeuse se souvient d'exactly quatre événements heureux de son enfance, et chaque personne participante dans la condition d'humeur triste s'en souvient d'exactly trois. L'effet de l'humeur est ici tout à fait évident. En réalité cependant, les données ressembleraient probablement davantage à celles des deux colonnes les plus à droite du Tableau 5.1. Même dans la condition d'humeur joyeuse, des personnes participantes se rappelleraient moins de souvenirs heureux parce qu'elles en ont moins à leur disposition, qu'elles utilisent des stratégies de rappel moins efficaces ou qu'elles sont moins motivées. Et même dans la condition d'humeur triste, des personnes participantes se rappelleraient plus de souvenirs d'enfance heureux parce qu'elles ont davantage de souvenirs heureux sur lesquels s'appuyer, qu'elles utilisent des stratégies de rappel plus efficaces ou qu'elles sont plus motivées. Bien que la différence moyenne entre les deux groupes soit la même que dans

les données idéalisées, cette différence est beaucoup moins évidente dans le contexte de la plus grande variabilité des données. L'une des raisons pour lesquelles les scientifiques essaient de contrôler les variables externes est de faire en sorte que leurs données ressemblent davantage aux données idéalisées du Tableau 5.1, ce qui permet de détecter plus facilement l'effet de la variable indépendante (bien que les données réelles ne ressemblent jamais tout à fait à cela).

Tableau 5.1 Données hypothétiques sans bruit et données réalistes avec bruit

Données "sans bruit" idéalisées		Données "bruyantes" réalistes	
Condition d'humeur positive	Condition d'humeur négative	Condition d'humeur positive	Condition d'humeur négative
4	3	3	1
4	3	6	3
4	3	2	4
4	3	4	0
4	3	5	5
4	3	2	7
4	3	3	2
4	3	1	5
4	3	6	1
4	3	8	2
$M = 4$	$M = 3$	$M = 4$	$M = 3$

Une façon de contrôler les variables externes est de les maintenir constantes. Cette technique peut consister à maintenir constantes les variables de la situation ou de la tâche en testant toutes les personnes participantes au même endroit, en leur donnant des instructions identiques, en les traitant de la même manière, etc. Il peut également s'agir de maintenir les variables des personnes participantes constantes. Par exemple, de nombreuses études sur le langage se limitent aux personnes participantes droitières, dont les zones langagières sont généralement isolées dans l'hémisphère cérébral gauche (Knecht, Dräger, Deppe, Bobe, Lohmann, Flöel, Ringelstein et Henningsen, 2000). Les individus gauchers sont plus susceptibles d'avoir leurs zones langagières isolées dans leur hémisphère cérébral droit ou réparties dans les deux hémisphères, ce qui peut modifier la façon dont ils traitent le langage et donc ajouter du bruit aux données.

En principe, les personnes menant des recherches peuvent contrôler les variables externes en limitant les personnes participantes à une catégorie très spécifique de personnes, comme âgé de 20 ans, hétérosexuel, de sexe féminin, programme en psychologie et droitier. L'inconvénient évident de cette approche est qu'elle réduirait la validité externe de l'étude — en particulier, la mesure dans laquelle les résultats peuvent être généralisés au-delà des personnes effectivement étudiées. Par exemple, il n'est pas certain que les résultats obtenus avec un échantillon de jeunes femmes lesbiennes s'appliquent à des hommes homosexuels plus âgés. Dans de nombreuses situations, les avantages d'un échantillon diversifié (augmentation de la validité externe) l'emportent sur la réduction du bruit obtenue par un échantillon homogène.

Variables externes en tant que variables confondantes

La deuxième façon dont les variables externes peuvent rendre difficile la détection de l'effet de la variable indépendante est en devenant des variables confondantes. Une **variable confondante** est une variable externe qui diffère en moyenne entre les niveaux de la variable indépendante (c'est-à-dire une variable externe qui varie systématiquement avec la variable indépendante). Par exemple, dans presque toutes les expériences, les quotients intellectuels (QI) des personnes participantes seront une variable externe. Mais, tant qu'il y a des personnes participantes avec des QI inférieurs et supérieurs dans chaque condition, de sorte que le QI moyen est à peu près égal entre les conditions, cette variation est probablement acceptable (et peut même être souhaitable). Ce qui serait problématique, en revanche, ce serait que les personnes participantes d'une condition aient en moyenne un QI nettement inférieur et que les personnes participantes d'une autre condition aient en moyenne un QI nettement supérieur. Dans ce cas, le QI serait une variable confondante.

Les variables confondantes ne sont pas souhaitables puisque, comme le terme le dit, elles apportent de la confusion. Parce qu'elles diffèrent systématiquement d'une condition à l'autre — tout comme la variable indépendante — elles fournissent une explication alternative à toute différence observée dans la variable dépendante. La Figure 5.1 présente les résultats d'une étude hypothétique, dans laquelle les personnes participantes en condition d'humeur positive ont obtenu de meilleurs résultats à une tâche de mémoire que les personnes en condition d'humeur négative. Mais si le QI est une variable confondante — les personnes participantes à la condition d'humeur positive ayant en moyenne un QI plus élevé que les personnes à la condition d'humeur négative — il n'est pas clair si c'est l'humeur positive ou le QI plus élevé qui est à l'origine du score plus élevé des gens dans la première condition. L'une des façons d'éviter les variables confondantes consiste à maintenir constantes les variables externes. Par exemple, on pourrait éviter que le QI ne devienne une variable confondante en limitant les personnes participantes à celles dont le QI est exactement égal à 100. Mais cette approche n'est pas toujours souhaitable pour les raisons que nous avons déjà évoquées. Une deuxième approche beaucoup plus générale, l'assignation aléatoire aux conditions, sera examinée en détail sous peu.

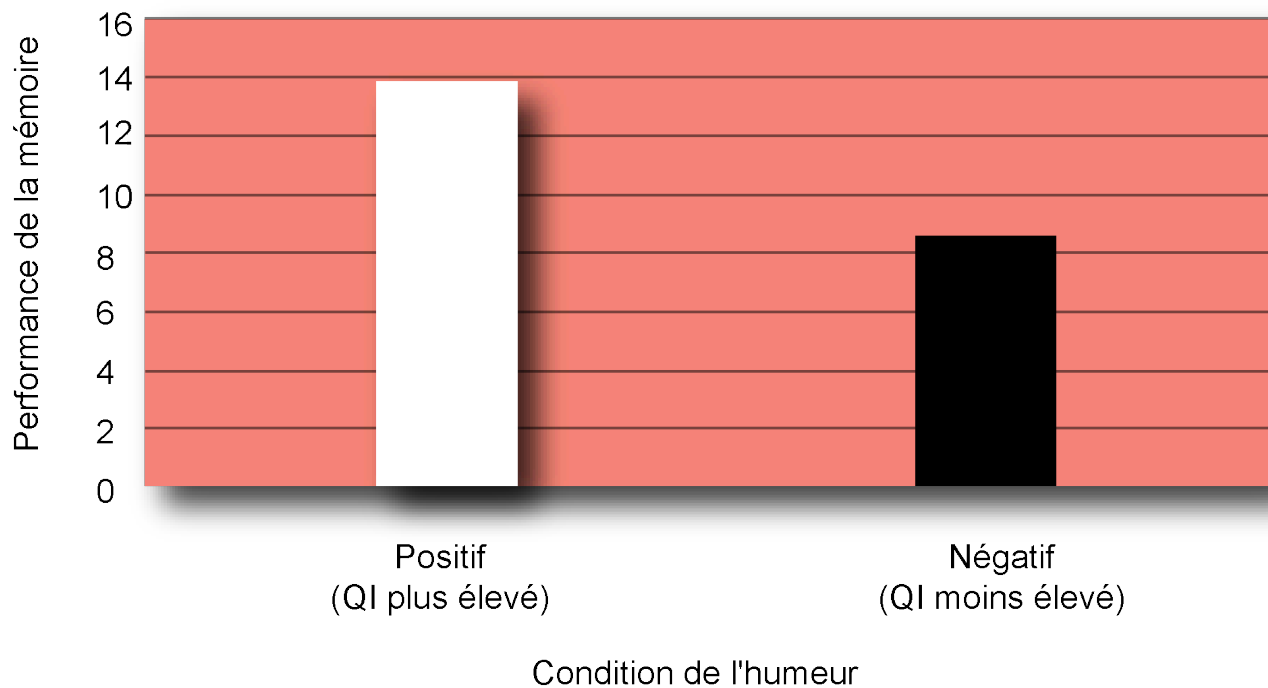


Figure 5.1. Résultats hypothétiques d'une étude sur l'effet de l'humeur sur la mémoire. Le QI étant également différent d'une condition à l'autre, il s'agit d'une variable confondante.

Conditions traitement et contrôle

Dans la recherche en psychologie, un **traitement** est une intervention destinée à modifier le comportement des personnes pour le rendre meilleur. Cette intervention comprend les psychothérapies et les traitements médicaux des troubles psychologiques, mais aussi les interventions destinées à améliorer l'apprentissage, à promouvoir la conservation, à réduire les préjugés, etc. Pour déterminer si un traitement est efficace, chaque personne participante est assignée aléatoirement soit à une **condition traitement**, dans laquelle on reçoit le traitement, soit à une **condition contrôle**, dans laquelle on ne reçoit pas le traitement. Si les membres avec traitement se retrouvent dans une meilleure situation que les membres du groupe contrôle, par exemple, ces gens sont plus dynamiques, plus rapides, plus altruistes, expriment moins de préjugés, alors les scientifiques peuvent conclure que le traitement est efficace. Dans la recherche sur l'efficacité des psychothérapies et des traitements médicaux, ce type d'expérience est souvent appelé **essai clinique randomisé**.

Il existe différents types de conditions contrôles. Dans une **condition contrôle sans traitement**, les personnes participantes ne reçoivent aucun traitement. Cette approche pose toutefois un problème en raison de l'existence d'effets placebo. Un **placébo** est un traitement simulé dépourvu de tout ingrédient ou élément actif qui devrait le rendre efficace, et un **effet placebo** est un effet positif d'un tel traitement. De nombreux remèdes populaires qui semblent fonctionner — comme manger de la soupe au poulet pour soigner un rhume ou placer du savon sous les draps pour faire cesser les crampes nocturnes dans les jambes — ne sont probablement rien d'autre que des placebo. Bien que les effets placebo ne soient pas bien compris, ils sont probablement liés principalement aux attentes des personnes quant à l'amélioration

de leur état. Le fait de s'attendre à une amélioration peut entraîner une réduction du stress, de l'anxiété et de la dépression, ce qui peut modifier les perceptions et même améliorer le fonctionnement du système immunitaire (Price, Finniss, & Benedetti, 2008).

Les effets placebo sont intéressants en eux-mêmes (voir la note « Le puissant placebo »), mais ils posent également un sérieux problème aux scientifiques qui veulent déterminer si un traitement est efficace. La Figure 5.2 présente quelques résultats hypothétiques dans lesquels les personnes participantes à une condition traitement se sont améliorées davantage en moyenne que les personnes participantes à une condition contrôle sans traitement. Si ces conditions (les deux barres les plus à gauche de la Figure 5.2) étaient les seules conditions de cette expérience, on ne pourrait pas conclure que le traitement a fonctionné. Il se pourrait plutôt que les personnes avec traitement montrent plus d'amélioration en raison de leurs attentes, alors que les personnes de la condition contrôle ne s'attendaient pas à une amélioration.

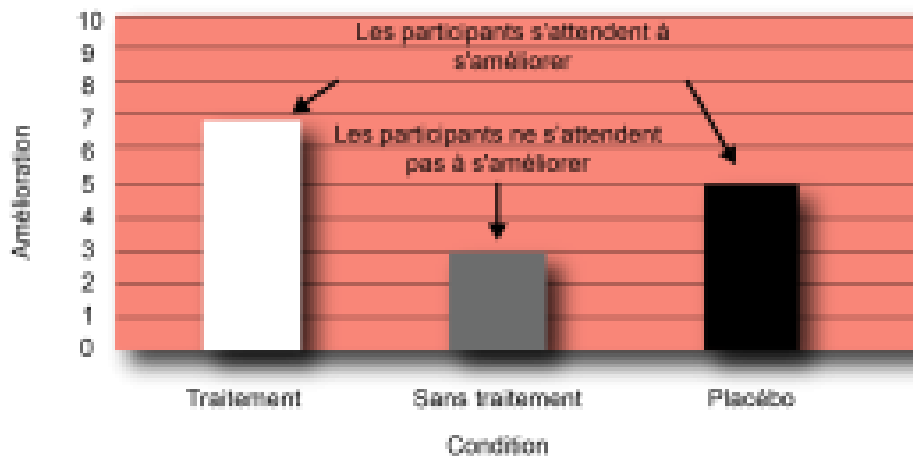


Figure 5.2. Résultats hypothétiques d'une étude comprenant des conditions traitement, sans traitement et placebo

Heureusement, il existe plusieurs solutions à ce problème. L'une d'entre elles consiste à inclure une **condition contrôle du placebo**, dans laquelle les gens reçoivent un placebo qui ressemble beaucoup au traitement, mais qui ne contient pas l'ingrédient actif ou l'élément considéré comme responsable de l'efficacité du traitement. Lorsque les personnes dans la condition traitement prennent une pilule par exemple, les autres dans une condition contrôle placebo prennent une pilule d'apparence identique, mais dépourvue de l'ingrédient actif du traitement (une « pilule de sucre »). Dans les recherches sur l'efficacité de la psychothérapie, le placebo peut consister à aller voir un psychothérapeute et à parler de ses problèmes de manière non structurée. L'idée est que si les membres du groupe traitement et du groupe contrôle du placebo s'attendent à une amélioration, alors toute amélioration dans le groupe traitement supérieure à celle du groupe contrôle du placebo résulte du traitement et non des attentes des membres participants. C'est cette différence qui ressort de la comparaison des deux barres extérieures de la Figure 5.4.

Bien entendu le principe du consentement éclairé exige que les personnes participantes soient informées de leur assignation soit à un traitement, soit à une condition contrôle du placebo. Même ces personnes ne peuvent connaître la condition assignée avant la fin de l'expérience. Dans de nombreux cas, les personnes

participantes de la condition contrôle sans traitement, se voient ensuite offrir l'opportunité de recevoir le vrai traitement. Une autre approche consiste à utiliser une **condition contrôle sur liste d'attente**, où les personnes participantes sont informées du traitement, mais doivent d'abord attendre que les gens de la condition traitement l'aient reçu avant d'y avoir accès. Cette divulgation permet aux scientifiques de comparer les personnes participantes qui reçoivent le traitement avec des personnes qui ne le reçoivent pas actuellement, mais qui s'attendent à une amélioration (éventuellement). Une dernière solution au problème des effets placebo consiste à exclure complètement la condition contrôle et à comparer tout nouveau traitement avec le meilleur traitement alternatif disponible. Par exemple un nouveau traitement pour la phobie simple pourrait être comparé à la thérapie d'exposition standard. Étant donné que les personnes participantes aux deux conditions reçoivent un traitement, leurs attentes en matière d'amélioration devraient être similaires. Cette approche est également judicieuse car, une fois qu'il existe un traitement efficace, la question intéressante concernant un nouveau traitement n'est pas simplement de savoir s'il fonctionne, mais s'il est plus efficace que ce qui existe déjà.

Le puissant placebo

Beaucoup de personnes ne sont pas surprises que les placebo puissent avoir un effet positif sur des troubles qui semblent fondamentalement psychologiques, comme la dépression, l'anxiété et l'insomnie. Cependant les placebo peuvent également avoir un effet positif sur des troubles que la plupart des gens considèrent comme fondamentalement physiologiques. Il s'agit notamment de l'asthme, des ulcères et des verrues (Shapiro et Shapiro, 1999). Il existe même des preuves que la chirurgie placebo — également appelée « chirurgie fictive » — peut être aussi efficace que la chirurgie réelle.

Le chercheur médical J. Bruce Moseley et ses collègues ont mené une étude sur l'efficacité de deux chirurgies arthroscopiques pour l'arthrose du genou (Moseley, O'Malley, Petersen, Menke, Brody, Kuykendall, Hollingsworth et Wray, 2002). Les personnes participantes au contrôle ont été préparées à l'intervention chirurgicale, ont reçu un tranquillisant et ont même subi trois petites incisions au genou. Mais elles n'ont pas reçu la chirurgie arthroscopique réelle. Notons que le CÉR aurait soigneusement examiné le recours à la duperie dans ce cas et jugé que les avantages de son utilisation l'emportaient sur les risques et qu'il n'y avait pas d'autre moyen de répondre à la question de recherche (sur l'efficacité d'une procédure placebo) sans cela. Le résultat surprenant a été que toutes les personnes participantes se sont améliorées en termes de douleur et de fonction du genou, et que le groupe ayant subi l'intervention chirurgicale fictive s'est amélioré tout autant que les groupes traités. Selon l'équipe de recherche, « Cette étude fournit des preuves solides que le lavage arthroscopique avec ou sans débridement [les procédures chirurgicales utilisées] n'est pas meilleur et semble équivalent à une procédure placebo dans l'amélioration de la douleur du genou et son fonctionnement d'après les déclarations [traduction] » (p. 85).

28. Plan expérimental

Objectifs d'apprentissage

1. Expliquer la différence entre les expériences intersujets et intrasujets, énumérer quelques-uns des avantages et des inconvénients de chaque approche et décider de l'approche à utiliser pour répondre à une question de recherche particulière.
2. Définir l'assignation aléatoire, la distinguer de l'échantillonnage aléatoire, expliquer son rôle dans la recherche expérimentale et utiliser quelques stratégies simples pour la mettre en œuvre.
3. Définir plusieurs types d'effet de l'ordre, donner des exemples de chacun d'entre eux et expliquer comment le contrebalancement permet de les contrôler.

Dans cette section, nous examinerons différentes façons de concevoir une expérience. Nous distinguerons principalement les approches dans lesquelles chaque personne participante fait l'expérience d'un niveau de la variable indépendante et les approches dans lesquelles chaque personne participante fait l'expérience de tous les niveaux de la variable indépendante. Les premières sont appelées expériences intersujets et les deuxièmes sont appelées expériences intrasujets.

Expériences intersujets

Dans une **expérience intersujets**, chaque personne participante est testée dans une seule condition. Par exemple une personne en recherche qui dispose d'un échantillon de 100 personnes qui étudient à l'université peut demander à la moitié d'entre elles d'écrire sur un événement traumatisant et à l'autre moitié d'écrire sur un événement neutre. Ou encore une personne en recherche qui dispose d'un échantillon de 60 personnes souffrant d'agoraphobie sévère (peur des espaces ouverts) peut assigner 20 d'entre elles à recevoir chacun des trois traitements différents pour ce problème. Dans une expérience inter-sujets, il est essentiel que les personnes menant des recherches répartissent les personnes participantes aux conditions de manière à ce que les différents groupes soient, en moyenne, très similaires les uns aux autres. Par exemple les personnes d'une condition traumatique et d'une condition neutre doivent comprendre une proportion similaire d'hommes et de femmes, avoir des QI moyens similaires, des niveaux de motivation moyens similaires, des nombres moyens similaires de problèmes de santé, etc. Cet appariement consiste à contrôler ces variables externes des personnes participantes entre les conditions afin qu'elles ne deviennent pas des variables confondantes.

Assignation aléatoire

La méthode principale utilisée par les scientifiques pour contrôler les variables externes entre les conditions est l'**assignation aléatoire**, c'est-à-dire l'utilisation d'un processus aléatoire pour décider quelles personnes participantes sont testées dans quelles conditions. Ne confondez pas l'assignation aléatoire avec l'échantillonnage aléatoire. L'échantillonnage aléatoire est une méthode d'échantillonnage selon laquelle chacun des éléments qui fait partie d'une population est choisi au hasard. Il est rarement utilisé dans la recherche en psychologie. L'assignation aléatoire est une méthode permettant de répartir les personnes d'un échantillon aux différentes conditions, et c'est un élément important de toute recherche expérimentale en psychologie et également dans d'autres domaines.

Dans son sens le plus strict, l'assignation aléatoire doit répondre à deux critères. Le premier est que chaque personne participante a une chance égale d'être assignée à chaque condition (par exemple 50 % de chances d'être assigné à chacune des deux conditions). Le second est que chaque personne participante est assignée à une condition indépendamment des autres personnes participantes. Ainsi une façon de répartir les personnes participantes à deux conditions serait de tirer à pile ou face pour chacune d'entre elles. Si la pièce tombe sur pile, l'individu est assigné à la condition A, et si elle tombe sur face, l'individu est assigné à la condition B. Pour trois conditions, on pourrait utiliser un ordinateur pour générer un nombre entier aléatoire de 1 à 3 pour chaque personne participante. Si le nombre entier est 1, l'individu est assigné à la condition A; s'il est 2, l'individu est assigné à la condition B; et s'il est 3, l'individu est assigné à la condition C. En pratique, une séquence complète de conditions — une pour chaque personne participante prévue pour l'expérience — est généralement créée à l'avance, et chaque nouvel individu est assigné à la condition suivante dans la séquence au fur et à mesure qu'il est testé. Lorsque la procédure est informatisée, le programme informatique se charge souvent de l'assignation aléatoire.

Un des problèmes qui se pose par le tirage à pile ou face et d'autres procédures strictes d'affectation aléatoire c'est qu'ils sont susceptibles d'entraîner des tailles d'échantillon inégales dans les différentes conditions. L'inégalité des tailles d'échantillon n'est généralement pas un problème grave, et vous ne devez jamais jeter les données que vous avez déjà collectées pour obtenir des tailles d'échantillon égales. Toutefois pour un nombre fixe de personnes participantes, il est statistiquement plus efficace de les diviser en groupes de tailles égales. La pratique courante consiste donc à utiliser une sorte d'assignation aléatoire modifiée qui maintient le nombre de personnes participantes dans chaque groupe aussi similaire que possible. L'une de ces approches est la **randomisation des blocs** où toutes les conditions se produisent une fois dans la séquence avant que l'une d'entre elles ne soit répétée. Ensuite, elles se reproduisent toutes avant que l'une d'entre elles ne soit répétée à nouveau. Dans chacun de ces « blocs », les conditions se présentent dans un ordre aléatoire. Là encore, la séquence de conditions est habituellement générée avant que les personnes participantes ne soient testées, et chaque nouvelle personne participante est assignée à la condition suivante dans la séquence. Le Tableau 5.2 présente une telle séquence pour assigner neuf participants à trois conditions. Le site Internet Research Randomizer (www.randomizer.org) permet de générer des séquences de randomisation des blocs pour n'importe quel nombre de personnes participantes et de conditions. Là encore, lorsque la procédure est informatisée, le programme informatique se charge souvent de la randomisation par blocs.

Tableau 5.2 Séquence de randomisation des blocs pour assigner neuf personnes participantes à trois conditions

Personne participante	Condition
1	A
2	C
3	B
4	B
5	C
6	A
7	C
8	B
9	A

L'assignation aléatoire ne garantit pas le contrôle de toutes les variables externes d'une condition à l'autre. Le processus étant aléatoire, il est toujours possible que, par hasard, les individus d'une condition se révèlent en moyenne beaucoup plus âgés, moins fatigués, plus motivés ou moins déprimés que les individus d'une autre condition. Cependant il y a quelques raisons pour lesquelles cette possibilité n'est pas une préoccupation majeure. L'une d'entre elles est que l'assignation aléatoire fonctionne mieux qu'on pourrait le penser, en particulier pour les grands échantillons. Ensuite, les statistiques inférentielles que les scientifiques utilisent pour décider si une différence entre les groupes reflète une différence dans la population tiennent compte de la « faillibilité » de l'assignation aléatoire. Une autre raison est que, même si l'assignation aléatoire entraîne une variable confondante et produit donc des résultats trompeurs, il est probable que cette confusion soit détectée lorsque l'expérience est répétée. En résumé, l'assignation aléatoire des conditions — bien qu'elle ne soit pas infaillible en matière de contrôle des variables externes — est toujours considérée comme un point fort de la conception d'une recherche.

Groupes appariés

Une alternative à l'assignation aléatoire simple des personnes participantes aux conditions est l'utilisation d'une **conception de groupes appariés**. Dans ce cas, les personnes participantes aux différentes conditions sont appariées sur la variable dépendante ou sur une (des) variable(s) externe(s) avant la manipulation de la variable indépendante. Cela garantit que ces variables ne seront pas confondues dans les conditions expérimentales. Par exemple si nous voulons déterminer si l'écriture expressive affecte la santé des gens, nous pourrions commencer par mesurer diverses variables liées à la santé chez nos futures personnes participantes à la recherche. Nous pourrions ensuite utiliser ces informations pour classer les individus en fonction de leur état de santé. Ensuite les deux individus les plus sains seraient assignés au hasard à des conditions différentes (l'un serait assigné au hasard à la condition d'écriture d'expériences traumatiques et l'autre à la condition d'écriture neutre). Les deux individus suivants les plus sains seraient alors assignés aléatoirement à des conditions différentes, et ainsi de suite jusqu'aux deux personnes participantes les

moins saines. Cette méthode garantit que les personnes participantes, à la condition d'écriture sur les expériences traumatisantes, soient appariés aux personnes participantes à la condition d'écriture neutre, en ce qui concerne la santé au début de l'étude. Si à la fin de l'expérience, une différence de santé est détectée entre les deux conditions, nous saurons alors qu'elle est due à la manipulation de l'écriture et non à des différences de santé préexistantes.

Expériences intrasujets

Dans une **expérience intrasujet**, chaque personne participante est testée dans toutes les conditions. Prenons l'exemple d'une expérience sur l'effet de l'attrait physique de la personne accusée sur le jugement de sa culpabilité. Là encore dans une expérience intersujets, un groupe de personnes participantes se verrait présenter une personne accusée séduisante et serait invité à juger de sa culpabilité, tandis qu'un autre groupe de personnes participantes se verrait présenter une personne accusée peu séduisante et serait invité à juger de sa culpabilité. Dans une expérience intrasujet, en revanche, le même groupe de personnes participantes jugerait de la culpabilité d'un individu accusé séduisant et d'un individu accusé peu séduisant.

Le principal avantage de cette approche est qu'elle permet un contrôle maximal des variables externes des personnes participantes. Les individus de toutes les conditions ont le même QI moyen, le même statut socioéconomique, le même nombre de frères et sœurs et ainsi de suite, parce qu'il s'agit des mêmes personnes. Les expériences intrasujets permettent également d'utiliser des procédures statistiques qui éliminent l'effet de ces variables externes de l'individu sur la variable dépendante, ce qui rend les données moins « bruyantes » et l'effet de la variable indépendante plus facile à détecter. Nous examinerons cette idée plus en détail plus tard dans le livre. Cependant, toutes les expériences ne peuvent pas utiliser un plan d'expérience intrasujet et il ne serait pas souhaitable de le faire.

Effets de transfert et contrebalancement

Le principal inconvénient des modèles intrasujets est qu'ils peuvent entraîner des effets de l'ordre. Un **effet de l'ordre** se produit lorsque les réponses des personnes participantes dans les différentes conditions sont affectées par l'ordre des conditions auxquelles elles ont été exposées. Un type d'effet de l'ordre est l'effet de transfert. L'**effet de transfert** (*carryover effect*, en anglais) est un effet où le fait d'être testé dans une condition affecte le comportement de la personne participante dans les conditions ultérieures. Un type d'effet de transfert est l'**effet de pratique**, où les personnes participantes exécutent mieux une tâche dans des conditions ultérieures en raison de l'occasion de la pratiquer. Un autre type d'effet de transfert est l'**effet de fatigue**, où les personnes participantes exécutent moins bien une tâche dans les conditions ultérieures en raison de la fatigue ou de l'ennui. Le fait d'être testé dans une condition peut également modifier la façon dont les personnes participantes perçoivent les stimulations ou interprètent leur tâche dans les conditions ultérieures. Ce type d'effet est appelé l'**effet de contexte (ou effet de contraste)**. Par exemple des accusés d'apparence moyenne peuvent être jugés plus sévèrement lorsque les personnes participantes viennent de juger des accusés séduisants, que lorsque ces personnes viennent de juger des accusés peu séduisants.

Les expériences intrasujets permettent également aux personnes participantes de deviner plus facilement l'hypothèse. Par exemple un individu à qui l'on demande de juger la culpabilité des accusés séduisants, puis de juger la culpabilité des accusés peu séduisants, est susceptible de deviner que l'hypothèse est que l'attrait physique des accusés affecte les jugements de culpabilité. Cette connaissance pourrait conduire l'individu à juger plus sévèrement les accusés peu séduisants parce qu'il pense que c'est ce que l'on attend de lui. Ou bien ceci peut amener les individus à juger tous les accusés de la même manière dans un effort pour être « juste ».

Les effets de transfert peuvent être intéressants en eux-mêmes. (L'attrait d'une personne dépend-il de l'attrait d'autres personnes que nous avons vues récemment?) Mais lorsqu'ils ne sont pas au centre de la recherche, les effets de transfert peuvent être problématiques. Imaginons par exemple que les individus jugent la culpabilité d'un individu accusé séduisant, et ensuite la culpabilité d'un individu accusé peu séduisant. S'ils jugent plus sévèrement l'individu accusé peu séduisant, cela pourrait être dû à son manque d'attrait. Mais il se peut aussi qu'ils jugent cette personne plus sévèrement parce qu'ils s'ennuient ou sont fatigués. En d'autres termes, l'ordre des conditions est une variable confondante. La condition séduisante est toujours la première condition et la condition non séduisante, la seconde. Ainsi toute différence entre les conditions, en matière de variable dépendante, pourrait être causée par l'ordre des conditions et non, par la variable indépendante elle-même.

Il existe cependant une solution, au problème des effets de transfert qui peut être utilisée dans de nombreuses situations. Il s'agit du **contrebalancement**, qui consiste à tester différentes personnes participantes dans des ordres différents. La meilleure méthode de contrebalancement est le **contrebalancement complet** qui consiste à tester un nombre égal de personnes participantes dans chaque ordre possible des conditions. Par exemple la moitié des individus serait testée dans la condition individu accusé séduisant suivie de la condition individu accusé non séduisant, et l'autre moitié serait testée dans la condition individu accusé non séduisant suivie de la condition individu accusé séduisant. Avec trois conditions, il y aurait six ordres différents (ABC, ACB, BAC, BCA, CAB et CBA), de sorte que certaines personnes participantes seraient testées dans chacun des six ordres. Avec quatre conditions, il y aurait 24 ordres différents; avec cinq conditions, il y aurait 120 ordres possibles. Avec le contrebalancement, les personnes participantes sont affectées aux ordres de manière aléatoire, en utilisant les techniques que nous avons déjà évoquées. L'assignation aléatoire joue donc un rôle important dans les devis intrasujets comme dans les devis intersujets. Ici au lieu d'être réparties aléatoirement à des conditions, les personnes participantes sont affectées aléatoirement à différents ordres de conditions. En fait on peut affirmer avec certitude que, si une étude ne comporte pas d'assignation aléatoire sous une forme ou une autre, il ne s'agit pas d'une expérience.

Un moyen plus efficace de contrebalancer est d'utiliser un carré latin, qui randomise en ayant des lignes et des colonnes égales. Par exemple, si vous avez quatre traitements, vous devez avoir quatre versions. Comme un sudoku, aucun traitement ne peut se répéter en ligne ou en colonne. Pour quatre versions de quatre traitements, le carré latin se présenterait comme suit :

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

Figure 5.3. Carré latin

Le diagramme ci-dessus montre que le carré a été construit de manière à ce que chaque condition apparaisse à chaque position ordinale (A apparaît une fois en premier, une fois en deuxième, une fois en troisième et une fois en quatrième) et que chaque condition précède et suive chaque autre condition une fois. Un carré latin pour une expérience avec 6 conditions sera de dimension 6 x 6, un carré latin pour une expérience avec 8 conditions sera de dimension 8 x 8, et ainsi de suite. Ainsi, alors que le contrebalancement complet de 6 conditions nécessiterait 720 ordres, un carré latin ne nécessiterait que 6 ordres.

Enfin lorsque le nombre de conditions est élevé, les expériences peuvent utiliser le **contrebalancement**

aléatoire dans lequel l'ordre des conditions est déterminé de manière randomisée pour chaque participant. Cette technique permet de déterminer tous les ordres possibles des conditions, puis de sélectionner au hasard l'un de ces ordres pour chaque participant. Cette technique n'est pas aussi puissante que le contrebalancement complet ou le contrebalancement partiel comme le carré latin. L'utilisation d'un contrebalancement aléatoire entraînera davantage d'erreurs aléatoires, mais si les effets de l'ordre sont susceptibles d'être faibles et que le nombre de conditions est élevé, il s'agit d'une option à la disposition des scientifiques.

Il y a deux façons de comprendre ce qu'accomplit le contrebalancement. La première consiste à contrôler l'ordre des conditions de manière à ce qu'il ne soit plus une variable confondante. Au lieu que la condition séduisante soit toujours la première et que la condition non séduisante soit toujours la deuxième, la condition séduisante vient en premier pour certaines personnes participantes et, en deuxième pour d'autres. De même la condition non séduisante vient en premier pour certaines personnes participantes et, en deuxième pour d'autres. Par conséquent, toute différence globale de la variable dépendante entre les deux conditions ne peut pas être due à l'ordre des conditions. Une deuxième façon de comprendre ce qu'accomplit le contrebalancement est que s'il y a des effets de l'ordre, il permet de les détecter. On peut analyser les données séparément pour chaque ordre afin de voir s'il a un effet.

Lorsque 9 est « plus grand » que 221

Le chercheur Michael Birnbaum argumente que le manque de contexte fourni par les devis intersujets est souvent un problème plus important que les effets de contexte créés par les devis intrasujets. Pour illustrer ce problème, il a demandé à des personnes participantes d'évaluer la taille de deux nombres sur une échelle de 1 à 10 où 1 correspondait à « très très petit » et 10 à « très très grand ». Un groupe de personnes participantes a été invité à évaluer le chiffre 9 et un autre groupe a été invité à évaluer le chiffre 221 (Birnbaum, 1999). Les personnes participantes à ce modèle intersujets ont attribué au chiffre 9 une note moyenne de 5,13 et au chiffre 221 une note moyenne de 3,10. En d'autres termes, elles ont évalué 9 comme plus grand que 221! Selon Birnbaum, cette différence s'explique par le fait que les personnes participantes ont spontanément comparé le 9 à d'autres nombres à un chiffre (auquel cas il est relativement grand) et ont comparé le 221 à d'autres nombres à trois chiffres (auquel cas il est relativement petit).

Devis intrasujets simultanés

Jusqu'à présent, nous avons discuté de devis intrasujets où les personnes participantes sont testées une condition à la fois. Il existe cependant une autre approche, souvent utilisée lorsque les personnes participantes donnent plusieurs réponses dans chaque condition. Imaginons par exemple que les individus jugent la culpabilité de 10 accusés séduisants et de 10 accusés peu séduisants. Au lieu de demander aux individus de porter un jugement sur les 10 accusés d'un type, puis sur les 10 accusés de l'autre type, les personnes menant la recherche peuvent présenter les 20 accusés dans un ordre qui mélange les deux

types. Elles pourraient alors calculer la moyenne de chaque personne participante pour chaque type de personne accusée. Imaginons une expérience visant à déterminer si les hommes souffrant de troubles de l'anxiété sociale se souviennent mieux des adjectifs négatifs (par exemple, « stupide », « incompetent ») que des adjectifs positifs (par exemple, « heureux », « productif »). Les scientifiques pourraient demander aux personnes participantes d'étudier une liste unique comprenant les deux types de mots et leur demander ensuite d'essayer de se souvenir d'autant de mots que possible. Les scientifiques pourraient ensuite compter le nombre de mots de chaque type qui ont été rappelés.

Intersujets ou intrasujet?

Presque toutes les expériences peuvent être menées en utilisant soit un devis intersujets, soit un devis intrasujet. Cette possibilité signifie que les scientifiques doivent choisir entre les deux approches en fonction de leurs mérites relatifs dans une situation donnée.

Les expériences intersujets ont l'avantage d'être conceptuellement plus simples et de nécessiter moins de temps de test par personne participante. Elles permettent également d'éviter les effets de l'ordre sans nécessiter de contrebalancement. Les expériences intrasujets présentent l'avantage de contrôler les variables externes des personnes participantes, ce qui réduit généralement le bruit dans les données et facilite la détection de tout effet de la variable indépendante sur la variable dépendante. Les expériences intrasujets nécessitent également moins de personnes participantes que les expériences inter-sujets pour détecter un effet de la même taille.

En règle générale, s'il est possible de mener une expérience intrasujet (avec un contrebalancement approprié) dans le temps disponible pour chaque personne participante — et si vous n'avez pas d'inquiétudes sérieuses concernant les effets de l'ordre — ce devis est probablement la meilleure option. Si une expérience intrasujet est difficile ou impossible à réaliser, vous devriez plutôt envisager une expérience intersujets. Par exemple si vous testez des personnes dans la salle d'attente d'une clinique médicale ou dans la file d'attente d'une épicerie, vous n'aurez peut-être pas assez de temps pour tester chaque personne participante dans toutes les conditions et vous opterez donc pour un devis intersujets. Ou imaginons que vous essayez de réduire le niveau de préjugés des hommes en les faisant interagir avec une personne d'une autre race. Un devis intrasujet avec contrebalancement nécessiterait de tester certaines personnes participantes d'abord dans la condition traitement, et ensuite dans une condition contrôle. Mais si le traitement fonctionne et réduit le niveau de préjugés des personnes participantes, ces dernières ne pourraient plus être testées dans la condition contrôle. Cette difficulté s'applique à de nombreux devis de recherche impliquant un traitement destiné à produire un changement à long terme dans le comportement des personnes participantes (par exemple les études testant l'efficacité de la psychothérapie). Il est clair qu'un devis intersujets serait nécessaire dans ce cas.

N'oubliez pas non plus que l'utilisation d'un type de devis de recherche n'exclut pas l'utilisation de l'autre type dans une autre étude. Il n'y a aucune raison pour que la personne en recherche ne puisse pas utiliser à la fois un devis intersujets et un devis intrasujet pour répondre à la même question de recherche. En fait, les personnes menant des recherches professionnellement adoptent souvent ce type d'approche mixte.

29. Expérimentation et validité

Objectifs d'apprentissage

1. Expliquer ce qu'est la validité interne et pourquoi les expériences sont considérées comme ayant une validité interne élevée.
2. Expliquer ce qu'est la validité externe et évaluer des études en matière de leur validité externe.
3. Expliquer les concepts de validité de construit et de validité statistique.

Quatre grandes validités

Lorsque nous lisons des expériences de psychologie avec un regard critique, l'une des questions à se poser est la suivante : « Cette étude est-elle valide (exacte)? » Cependant cette question n'est pas aussi simple qu'il y paraît car en psychologie, il existe de nombreux types de validité. Les personnes menant des recherches sont concentrées sur quatre validités pour aider à évaluer si une expérience est valable (Judd et Kenny, 1981; Morling, 2014): validité interne, validité externe, validité de construction et validité statistique. Nous examinerons chaque validité en profondeur.

Validité interne

Le fait que deux variables soient statistiquement liées ne signifie pas nécessairement que l'une cause l'autre. Dans le cadre de votre formation en psychologie, vous avez probablement entendu l'expression « la corrélation n'implique pas la causalité ». Par exemple s'il est vrai que les personnes qui font régulièrement de l'exercice sont plus heureuses que celles qui n'en font pas régulièrement, cette implication ne signifierait pas nécessairement que l'exercice augmente le bonheur des gens. Cela pourrait plutôt signifier qu'un plus grand bonheur incite les gens à faire de l'exercice ou que quelque chose comme une meilleure santé physique incite les gens à faire de l'exercice et à être plus heureux.

L'objectif d'une expérience est cependant de montrer que deux variables sont statistiquement liées et de le faire d'une manière qui permet de conclure que la variable indépendante est à l'origine des différences observées dans la variable dépendante. La logique repose sur cette hypothèse : si les scientifiques créent deux ou plusieurs conditions très similaires et manipulent ensuite la variable indépendante pour produire une seule différence entre elles, alors toute différence ultérieure entre les conditions doit avoir été causée par

la variable indépendante. Par exemple comme la seule différence entre les conditions de Darley et Latané était le nombre de témoins que les personnes participantes croyaient impliqués dans la discussion, cette différence de croyance doit avoir été responsable des différences d'aide entre les conditions.

On dit d'une étude empirique qu'elle présente une grande **validité interne** si la manière dont elle a été menée permet de conclure que la variable indépendante est à l'origine des différences observées dans la variable dépendante. La validité interne des expériences est donc élevée parce que la manière dont elles sont menées — avec la manipulation de la variable indépendante et le contrôle des variables externes (par exemple par l'utilisation de l'assignation aléatoire pour minimiser les confusions) — fournit un soutien solide aux conclusions causales. En revanche, les devis de recherche non expérimentale (par exemple les modèles corrélacionnels), dans lesquels les variables sont mesurées mais ne sont pas manipulées par la personne responsable de l'expérimentation, ont une faible validité interne.

Validité externe

Par ailleurs la manière dont les expériences sont menées conduit parfois à un autre type de critique. Plus précisément la nécessité de manipuler la variable indépendante et de contrôler les variables extérieures signifie que les expériences sont souvent menées dans des conditions semblant artificielles (Bauman, McGraw, Bartels et Warren, 2014). Dans de nombreuses expériences en psychologie, les personnes participantes sont toutes des personnes inscrites au premier cycle à l'université, qui se rendent dans une salle de classe ou un laboratoire pour remplir une série de questionnaires papier-crayon ou pour effectuer une tâche informatisée soigneusement conçue. Prenons l'exemple d'une expérience dans laquelle la chercheuse Barbara Fredrickson et ses collègues ont demandé à des personnes inscrites au premier cycle de se rendre dans un laboratoire sur le campus et de faire un test de mathématiques, alors qu'elles portaient un maillot de bain (Fredrickson, Roberts, Noll, Quinn et Twenge, 1998). À première vue cette manipulation peut sembler ridicule. Quand les personnes inscrites au premier cycle auront-elles un jour à effectuer des tests de mathématiques en maillot de bain en dehors de cette expérience?

La question à laquelle nous sommes confrontés est celle de la **validité externe**. Une étude empirique a une validité externe élevée si la manière dont elle a été menée permet de généraliser les résultats à des personnes et des situations autres que celles qui ont été étudiées. En règle générale, la validité externe des études est plus élevée lorsque les personnes participantes et la situation étudiée sont similaires à celles que les scientifiques souhaitent généraliser, et que les personnes participantes rencontrent tous les jours, ce qui est souvent décrit comme le **réalisme banal**. Imaginons par exemple qu'une équipe de recherche s'intéresse à la manière dont les personnes consommatrices dans les grandes épiceries sont influencées par le fait que les céréales pour petit-déjeuner sont emballées dans des boîtes jaunes ou violettes. La validité externe et le réalisme banal de leur étude seraient élevés si l'équipe étudiait les décisions de personnes ordinaires faisant leurs courses hebdomadaires dans une vraie épicerie. Si la clientèle a acheté beaucoup plus de céréales dans des boîtes violettes, l'équipe aurait confiance que cette augmentation serait vraie pour d'autres clientèles dans d'autres magasins. Cependant la validité externe de leur étude serait relativement faible, si cette équipe étudiait un échantillon de personnes inscrites au premier cycle dans un laboratoire d'une université sélective, qui ne faisaient que juger l'attrait de diverses couleurs présentées sur un écran d'ordinateur; en revanche, cette étude présenterait un **réalisme psychologique** élevé lorsque le même

processus mental est utilisé à la fois dans le laboratoire et dans le monde réel. Si les personnes aux études jugent le violet plus attrayant que le jaune, l'équipe de recherche n'aurait pas une grande confiance que cette préférence soit pertinente pour les décisions d'achat de céréales de la clientèle de l'épicerie en raison de la faible validité externe, mais elle pourrait avoir confiance que le traitement visuel des couleurs est d'un grand réalisme psychologique.

Nous devons toutefois veiller à ne pas tirer la conclusion générale que les expériences ont une faible validité externe. Une des raisons est que les expériences ne doivent pas nécessairement sembler artificielles. Considérons que l'expérience de Darley et Latané a fourni une simulation raisonnablement bonne d'une situation d'urgence réelle. Ou encore, considérons les expériences sur le terrain qui sont menées entièrement en dehors du laboratoire. Dans l'une de ces expériences, Robert Cialdini et ses collègues ont étudié si la clientèle d'un hôtel choisissait de réutiliser leurs serviettes pour un deuxième jour plutôt que de les faire laver, afin d'économiser de l'eau et de l'énergie (Cialdini, 2005). Cette équipe de recherche a manipulé le message d'une carte laissée dans un large échantillon de chambres d'hôtel. Une version du message mettait l'accent sur le respect de l'environnement, une autre sur le fait que l'hôtel ferait don d'une partie des économies réalisées à une cause environnementale, et une troisième sur le fait que la majorité de la clientèle de l'hôtel choisit de réutiliser leur serviette. Le résultat a été que la clientèle ayant reçu le message selon lequel la majorité de la clientèle de l'hôtel choisit de réutiliser leurs serviettes a réutilisé leur propre serviette beaucoup plus souvent que la clientèle ayant reçu l'un ou l'autre des deux autres messages. Compte tenu de la manière dont cette équipe a mené leur étude, il semble très probable que leurs résultats seraient vrais pour d'autres clientèles dans d'autres hôtels.

Une deuxième raison pour ne pas tirer la conclusion générale que les expériences ont une faible validité externe est qu'elles sont souvent menées pour en savoir plus sur les processus psychologiques susceptibles d'opérer auprès d'une grande variété de personnes et, dans de nombreuses situations. Revenons à l'expérience de Fredrickson et ses collègues. Cette équipe de recherche a constaté que les femmes de leur étude, mais pas les hommes, obtenaient de moins bons résultats au test de mathématiques lorsqu'elles portaient des maillots de bain. Selon l'équipe, cette différence entre les sexes est due au fait que les femmes ont davantage tendance à s'objectiver — à se considérer du point de vue d'une personne extérieure qui observe — ce qui détourne leur attention d'autres tâches. Cette équipe affirme également que ce processus d'auto-objectivation et son effet sur l'attention sont susceptibles de se produire chez des femmes et dans diverses situations, même si aucune d'entre elles ne se retrouve un jour à passer un test de mathématiques en maillot de bain.

Validité de construit

Outre la généralisation des résultats d'une expérience, un autre élément à examiner de près dans une étude est la qualité des manipulations de l'expérience ou la **validité de construit**. La question de recherche de départ que Darley et Latané posaient est la suivante : « Le comportement d'aide se divise-t-il ? » Ces scientifiques ont émis l'hypothèse que les personnes participantes en laboratoire seraient moins susceptibles d'aider lorsqu'elles croient qu'il y avait davantage d'autres personnes susceptibles d'aider en plus d'elles-mêmes. Cette conversion de la question de recherche à la conception de l'expérience s'appelle l'**opérationnalisation** (voir chapitre 4 pour plus d'informations sur la définition opérationnelle). Darley et

Latané ont opérationnalisé la variable indépendante de la dilution de la responsabilité en augmentant le nombre d'aides potentiels. En évaluant ce modèle, nous dirions que la validité de construit est très élevée parce que les manipulations de l'expérience répondent très clairement à la question de recherche; il y avait une crise, un moyen pour la personne participante d'aider, et en augmentant le nombre d'autres témoins impliqués dans la discussion, ces scientifiques ont fourni un moyen de tester la dilution.

Que se passerait-il si le nombre de conditions dans l'étude de Darley et Latané changeait? Imaginons qu'il n'y ait que deux conditions : une personne aux études ou deux impliquées dans la discussion. Même si nous constatons une diminution de l'aide en ajoutant une autre personne, il ne s'agit peut-être pas d'une démonstration claire de la dilution de la responsabilité, mais simplement de la présence d'autres personnes. On pourrait penser qu'il s'agit d'une forme d'inhibition sociale, un concept de Bandura. La validité de construit serait plus faible. Cependant s'il y avait eu cinq conditions, il est possible que la diminution se poursuive avec l'augmentation du nombre de personnes participantes à la discussion ou peut-être qu'elle se stabilise après un certain nombre de personnes. Dans ce cas, nous pourrions développer une compréhension plus nuancée du phénomène. Mais en ajoutant encore d'autres conditions, la validité de construit ne pourrait peut-être pas augmenter. Lorsque vous concevez votre propre expérience, réfléchissez à la manière dont la question de recherche est opérationnalisée dans votre étude.

Validité statistique

La **validité statistique** concerne le traitement statistique approprié des données et la justesse des conclusions statistiques des scientifiques. Il existe de nombreux types de tests statistiques inférentiels (par exemple les tests t, l'ANOVA, la régression, la corrélation) et la validité statistique concerne l'utilisation du type de test approprié pour analyser les données. Lorsque les personnes menant la recherche envisagent le type de test approprié, elles doivent tenir compte du type d'échelle utilisée pour leur variable dépendante et de la conception de leur étude. En outre de nombreux tests de statistiques inférentielles reposent sur des postulats (par exemple les données sont normalement distribuées) et la validité statistique est menacée lorsque ces postulats ne sont pas respectés, mais que les statistiques sont néanmoins utilisées.

Une critique courante des expériences est qu'il n'y a pas suffisamment de personnes participantes dans l'étude. La raison principale de cette critique est qu'il est difficile de faire des généralisations sur une population à partir d'un petit échantillon. Au départ, il semble que cette critique concerne la validité externe, mais il existe des études pour lesquelles la petite taille de l'échantillon ne pose pas de problème (les chapitres suivants expliqueront comment les petits échantillons, même d'une seule personne, sont toujours très informatifs pour la recherche en psychologie). Par conséquent les échantillons de petite taille sont en fait une critique de la validité statistique. La validité statistique indique si les statistiques utilisées dans l'étude sont correctes et si elles soutiennent les conclusions tirées.

L'analyse statistique appropriée doit être effectuée sur les données afin de déterminer si la différence ou la relation prédite a bien été trouvée. Il est intéressant de noter que la probabilité de détecter un effet de la variable indépendante sur la variable dépendante dépend non seulement de l'existence réelle d'une relation entre ces variables, mais aussi du nombre de conditions et de la taille de l'échantillon. C'est pourquoi il est important de réaliser une analyse de puissance au moment de déterminer le devis d'une étude. L'analyse de

puissance est un calcul qui vous informe sur le nombre de personnes participantes à recruter pour détecter un effet d'une taille spécifique.

Hierarchisation des validités

Ces quatre grandes validités — interne, externe, de construit et statistique — sont utiles à garder à l'esprit lors de la lecture d'autres expériences et de la conception de ses propres recherches. Toutefois, les personnes menant des recherches doivent établir des priorités et il n'est souvent pas possible d'obtenir une validité élevée dans les quatre domaines. Dans l'étude de Cialdini sur l'utilisation des serviettes dans les hôtels, la validité externe était élevée, mais la validité statistique était plus modeste. Cette divergence n'invalide pas l'étude, mais elle montre les points à améliorer pour les futures études de suivi (Goldstein, Cialdini et Griskevicius, 2008). Morling (2014) souligne que de nombreuses études en psychologie ont une validité interne et de construit élevée, mais sacrifient parfois la validité externe.

30. Considérations pratiques

Objectifs d'apprentissage

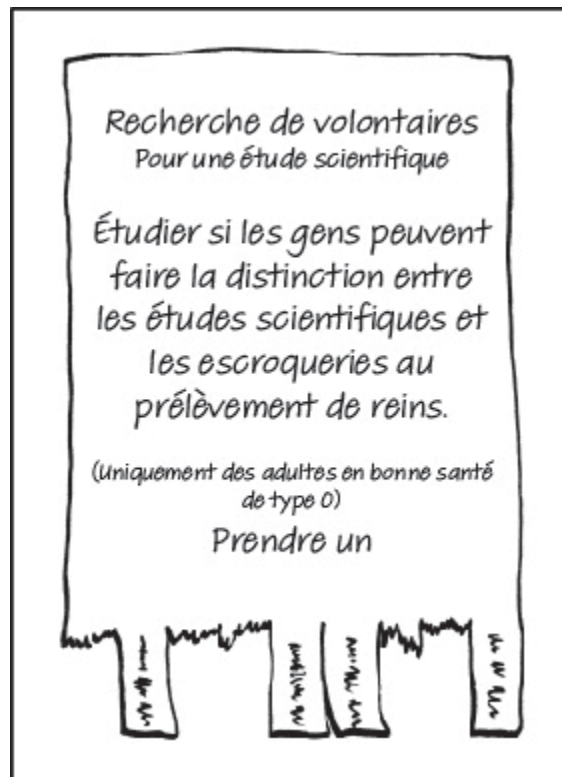
1. Décrire plusieurs stratégies de recrutement de personnes participantes à une expérience.
2. Expliquer pourquoi il est important de standardiser la procédure d'une expérience et indiquer plusieurs façons de le faire.
3. Expliquer ce qu'est une étude pilote et pourquoi elle est importante.

Les informations présentées jusqu'à présent dans ce chapitre sont suffisantes pour concevoir une expérience de base. Cependant, lorsque vient le moment de mener cette expérience, plusieurs questions pratiques supplémentaires se posent. Dans cette section, nous examinerons certaines de ces questions et comment les résoudre. La plupart de ces informations s'appliquent aussi bien aux études non expérimentales qu'aux études expérimentales.

Recrutement des personnes participantes

Bien entendu au début de tout projet de recherche, vous devez réfléchir à la manière dont vous obtiendrez vos personnes participantes. À moins d'avoir accès à des personnes qui souffrent de schizophrénie ou à des jeunes en délinquance incarcérées par exemple, il est inutile de concevoir une étude centrée sur ces populations. Mais même si vous prévoyez utiliser un échantillon de commodité, vous devrez recruter des personnes participantes pour votre étude.

Il existe plusieurs approches pour recruter des personnes participantes. L'une d'entre elles consiste à utiliser des personnes participantes d'une **réserve de sujets** (*subject pool*, en anglais) — officielle, c'est-à-dire un groupe établi de personnes qui ont accepté d'être contactées pour participer à des études de recherche. Par exemple dans de nombreux collèges et universités, il existe une réserve de sujets, composée de la communauté estudiantine inscrite à des cours d'introduction à la psychologie, qui doivent participer à un certain nombre d'études pour satisfaire aux exigences du cours. Les scientifiques publient des descriptions de leurs études et les personnes s'inscrivent pour y participer, généralement par un système en ligne. Les personnes participantes qui ne font pas partie des réserves de sujets peuvent également être recrutées par l'affichage ou la publication d'annonces ou par des appels personnels à des groupes représentatifs de la population étudiée. Par exemple une personne menant une recherche qui souhaite étudier les personnes âgées pourrait prendre la parole lors d'une réunion des personnes vivant dans une maison de retraite pour expliquer son étude et demander des volontaires.



"Étude scientifique" (traduit) récupéré à partir de <http://imgs.xkcd.com/comics/study.png> (CC-BY-NC 2.5)

Les personnes participantes volontaires

Même si les personnes participantes à une étude reçoivent une compensation sous la forme de crédits de cours, d'une petite somme d'argent ou d'une chance d'être traitées pour un problème psychologique, ces personnes restent essentiellement des volontaires. Cela vaut la peine d'être pris en compte, car des études montrent que les personnes se portant volontaires pour participer à une recherche en psychologie diffèrent de manière prévisible, de celles qui ne se portent pas volontaires. Plus précisément il existe de bonnes preuves qu'en moyenne, les volontaires présentent les caractéristiques suivantes par rapport aux non-volontaires (Rosenthal et Rosnow, 1976):

- Elles sont plus intéressées par le sujet de la recherche.
- Elles sont plus éduquées.
- Elles ont un plus grand besoin d'approbation.
- Elles ont un QI plus élevé.
- Elles sont plus sociables.
- Elles sont d'une classe sociale plus élevée.

Cette différence peut constituer un problème de validité externe s'il y a une raison de croire que les personnes participantes, qui présentent ces caractéristiques, sont susceptibles de se comporter différemment de la population générale. Par exemple en testant différentes méthodes de persuasion, un argument rationnel pourrait mieux fonctionner sur des volontaires que sur la population générale en raison de leur niveau d'éducation et de leur QI généralement plus élevés.

Dans de nombreuses expériences sur le terrain, il ne s'agit pas de recruter des personnes participantes, mais de les sélectionner. Par exemple l'équipe de recherche de Nicolas Guéguen et Marie-Agnès de Gail a mené une expérience sur le terrain sur l'effet d'un sourire sur l'aide, dans laquelle les personnes participantes faisaient partie de la clientèle d'un supermarché. Une personne complice descendant un escalier regardait directement une personne cliente montant l'escalier, et souriait ou ne souriait pas à cette personne. Peu de temps après, la personne cliente a croisé une autre complice, qui a laissé tomber des disquettes d'ordinateur sur le sol. La variable dépendante était l'arrêt ou non de la personne cliente pour aider à ramasser les disquettes (Guéguen et de Gail, 2003). Il y a deux aspects de cette étude qui méritent d'être abordés ici. Premièrement notons que cette clientèle n'a pas été « recrutée », ce qui signifie que le CÉR aurait pris soin de s'assurer que la dispense du consentement éclairé, dans ce cas, était acceptable (par exemple la situation ne devrait pas causer de préjudice, et l'étude aurait été menée dans le contexte d'activités ordinaires pour les gens). Deuxièmement même si le consentement éclairé n'était pas nécessaire, l'équipe de recherche a dû sélectionner les personnes parmi toute la clientèle qui prenait les escaliers ce jour-là. Il est extrêmement important que ce type de sélection soit effectué selon un ensemble de règles bien définies, établies avant le début de la collecte des données, et pouvant être expliquées clairement par la suite. Dans ce cas, à chaque descente d'escalier, la personne complice avait pour consigne de fixer du regard la première personne rencontrée semblant avoir entre 20 et 50 ans. Ce n'est que, si la personne soutenait son regard, qu'elle devient une personne participante à l'étude. Le but, d'une règle de sélection, bien définie est d'éviter les biais dans la sélection des personnes participantes. Par exemple si la personne complice était libre de choisir les gens de la clientèle regardés, cette personne pourrait choisir des personnes à l'air sympathique lorsqu'elle devait sourire et choisir des personnes à l'air antipathique lorsqu'elle devait éviter de sourire. Comme nous le verrons bientôt, de tels biais peuvent être totalement involontaires.

Normalisation de la procédure

Il est étonnamment facile d'introduire des variables externes au cours de la procédure. Par exemple une personne chargée de l'expérimentation pourrait donner des instructions claires à une personne participante, et des instructions vagues à une autre. Ou encore une chargée de l'expérimentation peut saluer chaleureusement les personnes participantes tandis qu'une autre les regarde à peine dans les yeux. Dans la mesure où ces variables affectent le comportement des personnes participantes, elles ajoutent du bruit aux données et rendent l'effet de la variable indépendante plus difficile à détecter. Si elles varient systématiquement d'une condition à l'autre, elles deviennent des variables confondantes et fournissent d'autres explications aux résultats. Par exemple si les personnes participantes d'un groupe traitement sont testées par une chargée de l'expérimentation chaleureuse et amicale et que les personnes participantes d'un groupe contrôle sont testées par une chargée de l'expérimentation froide et inamicale, ce qui semble être un

effet du traitement peut en fait être un effet du comportement de la personne chargée de l'expérimentation. Lorsqu'il y a plusieurs chargées de l'expérimentation, la possibilité d'introduire des variables externes est encore plus grande, mais elle est souvent nécessaire pour des raisons pratiques.

Le sexe de la personne chargée de l'expérimentation en tant que variable externe

Il est bien connu que le fait que les personnes participantes à une recherche soient des hommes ou des femmes peut avoir une influence sur les résultats d'une étude. Mais que dire du fait que la personne chargée d'une expérimentation soit un homme ou une femme? Il existe de nombreuses preuves que cela a également une importance. Les personnes chargées de l'expérimentation homme et femme ont des manières légèrement différentes d'interagir avec leurs personnes participantes, et bien sûr, les personnes participantes réagissent également différemment aux personnes chargées de l'expérimentation homme et femme (Rosenthal, 1976).

Par exemple dans une étude récente sur la perception de la douleur, les personnes participantes ont plongé leurs mains dans de l'eau glacée aussi longtemps qu'elles le pouvaient (Ibolya, Brake et Voss, 2004). Les participants masculins ont toléré la douleur plus longtemps lorsque la personne chargée de l'expérimentation était une femme, et les participants féminins l'ont tolérée plus longtemps lorsque la personne chargée de l'expérimentation était un homme.

Le chercheur Robert Rosenthal a consacré une grande partie de sa carrière à montrer que ce type de variation involontaire dans la procédure affecte effectivement le comportement des participants. En outre l'une des sources importantes de ces variations réside dans les attentes de la personne chargée de l'expérimentation quant à la manière dont les personnes participantes « devraient » se comporter au cours de l'expérience. Ce résultat est appelé **effet d'attente de la personne chargée de l'expérimentation** (Rosenthal, 1976). Par exemple si une personne chargée de l'expérimentation s'attend à ce que les personnes participantes d'un groupe traitement réussissent mieux une tâche que les participantes d'un groupe contrôle, elle peut sans le vouloir donner aux personnes du groupe traitement des instructions plus claires ou plus d'encouragements ou leur accorder plus de temps pour accomplir la tâche. Dans un exemple frappant, Robert Rosenthal et Kermit Fode ont demandé à plusieurs personnes étudiantes dans un cours de laboratoire en psychologie, d'entraîner des rats à parcourir un labyrinthe. Bien que les rats soient génétiquement similaires, certaines personnes aux études ont été informées qu'elles travaillaient avec des rats « intelligents dans les labyrinthes » qui avaient été élevés pour être de bons apprenants, et d'autres ont été informées qu'elles travaillaient avec des rats « imbéciles dans les labyrinthes » qui avaient été élevés pour être de mauvais apprenants. En effet en cinq jours d'entraînement, les rats « intelligents dans les labyrinthes » ont donné plus de réponses correctes, ont donné la bonne réponse plus rapidement et ont progressé plus régulièrement que les rats « imbéciles dans le labyrinthe » (Rosenthal & Fode, 1963). De toute évidence ce sont les attentes des personnes aux études concernant les performances des rats qui ont fait la différence. Mais comment? Certains indices proviennent des données recueillies à la fin de l'étude, montrant que les personnes étudiantes qui s'attendaient à ce que leurs rats apprennent rapidement avaient

une attitude plus positive à l'égard de leurs animaux et déclaraient se comporter avec eux de manière plus amicale (par exemple en les prenant plus souvent dans leurs mains).

La façon de minimiser les variations involontaires dans la procédure est de la standardiser autant que possible afin qu'elle soit exécutée de la même façon pour toutes les personnes participantes, quelle que soit la condition dans laquelle ils se trouvent. Voici plusieurs façons d'y parvenir :

- Créer un protocole écrit qui spécifie tout ce que les personnes chargées de l'expérimentation doivent faire et dire depuis le moment où elles accueillent les personnes participantes jusqu'à ce qu'elles les remercient à la fin.
- Créer des instructions standards que les personnes participantes lisent elles-mêmes ou qui leur sont lues mot pour mot par la personne chargée de l'expérimentation.
- Automatiser autant que possible le reste de la procédure en utilisant des logiciels prévus à cette fin ou même de simples diaporamas sur ordinateur.
- Anticiper les questions des personnes participantes et soit les soulever et y répondre dans les instructions, soit développer des réponses standards pour ces questions.
- Former ensemble plusieurs personnes chargées de l'expérimentation au protocole et leur demander de s'exercer les uns sur les autres.
- S'assurer que chaque personne chargée de l'expérimentation teste les personnes participantes dans toutes les conditions.

Une autre bonne pratique consiste à faire en sorte que les personnes chargées de l'expérimentation soient « aveugles » à la question de recherche ou à la condition dans laquelle chaque personne participante est testée. L'idée est de minimiser les effets d'attente de la personne chargée de l'expérimentation en minimisant leurs attentes. Par exemple dans une étude sur les médicaments où chaque personne participante reçoit le médicament ou un placebo, il arrive souvent que ni les personnes participantes ni les chargées de l'expérimentation qui interagissent avec elles, ne sachent dans quelle condition elles ont été placées. Étant donné que les personnes participantes et les chargées de l'expérimentation ne connaissent pas les conditions de l'expérience, cette technique est appelée **étude en double aveugle**. (Une étude en simple aveugle est une étude dans laquelle seules les personnes participantes ne connaissent pas la condition.) Bien entendu il arrive souvent qu'il ne soit pas possible d'être aveugle à la question de recherche. Par exemple si vous êtes à la fois la personne responsable de la recherche et la chargée de l'expérimentation, c'est impossible de rester aveugle à la question de recherche. De plus dans de nombreuses études, les personnes chargées de l'expérimentation doivent connaître la condition, comme elles sont responsables d'exécuter la procédure de manière différente dans des conditions différentes.



"Blokueur de placebo" (traduit en français) récupéré à partir de http://imgs.xkcd.com/comics/placebo_blocker.png (CC-BY-NC 2.5). [Description de l'image](#)

Enregistrement des données

Il est essentiel de bien enregistrer les données lorsque vous menez une expérience. Comme nous l'avons vu précédemment, il est courant que les personnes chargées de l'expérimentation établissent une séquence écrite de conditions avant le début de l'étude, puis testent chaque nouvelle personne participante dans la condition suivante de la séquence. Au fur et à mesure que vous les testez, il est bon d'ajouter à cette liste des informations démographiques de base, la date, l'heure et le lieu du test, ainsi que le nom de la personne chargée de cette session d'expérimentation. Il est également utile d'avoir un endroit où les personnes chargées de l'expérimentation peuvent noter des commentaires sur des événements inhabituels (par exemple une personne présente de la confusion ou un manque de motivation) ou des questions soulevées. Ce type d'information peut s'avérer utile ultérieurement si vous décidez d'analyser les différences entre les sexes ou les effets des différentes personnes chargées de l'expérimentation, ou si une question se pose à propos d'une personne participante ou d'une session d'expérimentation en particulier.

Comme l'identité des personnes participantes doit rester aussi confidentielle (ou anonyme) que possible, leurs noms et autres informations d'identification ne doivent donc pas être inclus dans leurs données. Afin d'identifier les personnes participantes de manière individuelle, il peut être utile d'attribuer un numéro d'identification à chacun au fur et à mesure que vous les testez. Il suffit généralement de les numéroter consécutivement en commençant par 1. Ce numéro peut également être inscrit sur toutes les feuilles de réponses ou questionnaires générés par les personnes participantes, ce qui permet de les garder plus facilement ensemble.

Vérification de la manipulation

Dans de nombreuses expériences, la variable indépendante est un construit qui peut seulement être manipulé indirectement. Par exemple des scientifiques peuvent essayer de manipuler indirectement le niveau de stress où des personnes participantes se font dire qu'elles ont cinq minutes pour préparer un bref discours, qu'elles auront à présenter devant un auditoire composé d'autres personnes participantes. Dans de telles situations, les personnes en recherche incluent souvent une **vérification de la manipulation** dans leur procédure. Une vérification de la manipulation est une mesure distincte du construit que les personnes en recherche tentent de manipuler. L'objectif d'une vérification de la manipulation est de confirmer que la variable indépendante a été manipulée avec succès. Par exemple les scientifiques qui tentent de manipuler les niveaux de stress des personnes participantes peuvent leur faire remplir un questionnaire papier-crayon sur le stress ou prendre leur tension artérielle — peut-être juste après la manipulation ou à la fin de la procédure — afin de vérifier si la variable a été manipulée.

Les vérifications de la manipulation sont particulièrement importantes lorsque les résultats d'une expérience s'avèrent nuls. Dans les cas où les résultats ne montrent aucun effet significatif de la manipulation de la variable indépendante sur la variable dépendante, une vérification de la manipulation peut aider les scientifiques à déterminer si le résultat nul est dû à une absence réelle d'effet de la variable indépendante sur la variable dépendante ou s'il est dû à un problème de manipulation de la variable indépendante. Imaginons par exemple que vous exposiez les personnes participantes à une musique de film joyeuse ou triste — dans l'intention de les mettre d'humeur joyeuse ou triste — mais que vous constatiez que cela n'avait aucun effet sur le nombre d'événements joyeux ou tristes de leur enfance dont ces personnes se souvenaient. Cela pourrait s'expliquer par le fait que l'humeur joyeuse ou triste n'a aucun effet sur les souvenirs des événements de l'enfance. Mais il se peut aussi que la musique n'ait pas réussi à mettre les personnes participantes dans des états joyeux ou tristes. Un contrôle de la manipulation — dans ce cas, une mesure de l'humeur des individus — permettrait de lever cette incertitude. S'il montrait que vous aviez réussi à manipuler l'humeur des personnes participantes, il semblerait alors qu'il n'y ait pas d'effet de l'humeur sur la mémoire des événements de l'enfance. Mais s'il montre que vous n'aviez pas réussi à manipuler leur humeur, il semblerait que vous ayez besoin d'une manipulation plus efficace pour répondre à votre question de recherche.

Les vérifications de la manipulation sont généralement effectuées à la fin de la procédure pour s'assurer que l'effet de la manipulation a duré pendant toute la procédure et pour éviter d'attirer inutilement l'attention sur la manipulation (pour éviter une caractéristique de la demande). Cependant les scientifiques font bien d'inclure une vérification de la manipulation dans un essai pilote de leur expérience afin d'éviter de consacrer beaucoup de temps et de ressources à une expérience vouée à l'échec, et de consacrer ce temps et cette énergie à la recherche d'une meilleure manipulation de la variable indépendante.

Étude pilote

Il est toujours judicieux de réaliser une **étude pilote** de votre expérience. Une étude pilote est une étude à petite échelle menée pour s'assurer qu'une nouvelle procédure fonctionne comme prévue. Dans le cadre d'une étude pilote, vous pouvez recruter des personnes participantes de manière formelle (par exemple à partir d'une réserve formelle de sujets) ou de manière informelle parmi les membres de votre famille, vos

amis, vos camarades de classe, etc. Le nombre de personnes participantes peut être modeste, mais il doit être suffisant pour vous donner la certitude que votre procédure fonctionne comme prévue. La réalisation d'une étude pilote permet de répondre à plusieurs questions importantes :

- Est-ce que les instructions sont comprises par les personnes participantes ?
- Quels types de malentendus trouve-t-on chez les personnes participantes? Quels types d'erreurs sont commises et quels types de questions sont posées?
- Les personnes participantes s'ennuient-elles ou sont-elles frustrées?
- Une manipulation indirecte est-elle efficace? (Vous devrez inclure une vérification de la manipulation.)
- Les personnes participantes peuvent-elles deviner la question ou l'hypothèse de recherche (y a-t-il des caractéristiques de la demande)?
- Combien de temps dure la procédure?
- Les programmes informatiques ou autres procédures automatisées fonctionnent-ils correctement?
- Les données sont-elles correctement enregistrées?

Bien entendu, pour répondre à certaines de ces questions, vous devrez observer attentivement les personnes participantes pendant la procédure et en parler avec elles par la suite. Ces personnes hésitent souvent à critiquer une étude devant les personnes menant une recherche. Faites-leur comprendre que leur participation sert à l'étude pilote et que vous voulez sincèrement connaître leurs commentaires, car ces derniers aideront à améliorer la procédure. Si la procédure fonctionne comme prévue, alors vous pouvez procéder à l'étude réelle. S'il y a des problèmes à résoudre, vous pouvez les résoudre, faire une étude pilote de la nouvelle procédure et poursuivre ce processus jusqu'à ce que vous soyez prêt à procéder.

Descriptions de l'image

Description de l'image du bloqueur de placebo : Bande dessinée qui représente deux bonhommes allumettes qui discutent.

Personne 1 : Certains scientifiques commencent à comprendre le mécanisme de l'effet placebo. Nous avons utilisé leurs travaux pour créer un nouveau médicament : un bloqueur de l'effet placebo. Il ne nous reste plus qu'à mener un essai. Nous allons faire deux groupes, leur donner chacun des placebos, puis donner à l'un le VRAI bloqueur de placebo, et à l'autre un... attend.

[Les deux personnes se grattent la tête]

Personne 2 : J'ai mal à la tête.

Personne 1 : Moi aussi. Vous voulez une pilule de sucre?

[\[Retour à l'image du bloqueur de placebo\]](#)

Attributions aux médias

- [Study](#) by XKCD [CC BY-NC \(Attribution NonCommercial\)](#)
- [Placebo blocker](#) by XKCD [CC BY-NC \(Attribution NonCommercial\)](#)

31. Points clés et exercices

Points clés

- Une expérience est un type d'étude empirique qui comprend la manipulation d'une variable indépendante, la mesure d'une variable dépendante et le contrôle de variables externes.
- Une variable externe est toute variable autre que les variables indépendantes et dépendantes. Une variable confondante est une variable externe qui varie systématiquement avec la variable indépendante.
- La recherche expérimentale sur l'efficacité d'un traitement nécessite une condition traitement et une condition contrôle, qui peut être une condition contrôle sans traitement, une condition contrôle placebo ou une condition contrôle avec liste d'attente. Les traitements expérimentaux peuvent également être comparés à la meilleure alternative disponible.
- Les expériences peuvent être menées en utilisant des devis d'expérience intersujets ou intrasujets. Pour décider lequel utiliser dans une situation donnée, il faut examiner attentivement les avantages et les inconvénients de chaque approche.
- L'assignation aléatoire des conditions dans les expériences intersujets ou le contrebalancement de l'ordre des conditions dans les expériences intrasujets sont des éléments fondamentaux de la recherche expérimentale. L'objectif de ces techniques est de contrôler les variables externes afin qu'elles ne deviennent pas des variables confondantes.
- Les études ont une validité interne élevée dans la mesure où la manière dont elles sont menées permet de conclure que la variable indépendante est à l'origine des différences observées dans la variable dépendante. La validité interne des expériences est généralement élevée en raison de la manipulation de la variable indépendante et du contrôle des variables externes.
- Les études ont une validité externe élevée dans la mesure où les résultats peuvent être généralisés à des personnes et des situations autres que celles étudiées. Bien que les expériences puissent sembler « artificielles » — et présenter une faible validité externe — il est important de se demander si les processus psychologiques étudiés sont susceptibles de fonctionner chez d'autres personnes et dans d'autres situations.
- Il existe plusieurs méthodes efficaces pour recruter des personnes participantes à votre expérience, notamment à l'aide de réserves formelles de sujets, d'annonces et d'appels personnels. Les expériences sur le terrain nécessitent des procédures de sélection des personnes participantes bien définies.
- Il est important de standardiser les procédures expérimentales afin de minimiser les variables externes, y compris les effets de l'attente de la personne chargée de l'expérimentation.

- Il est important de réaliser une ou plusieurs études pilotes à petite échelle pour s'assurer que la procédure fonctionne comme prévu.

Exercices

- **Pratique:** Citez cinq variables qui peuvent être manipulées par les scientifiques dans une expérience. Citez cinq variables qui ne peuvent pas être manipulées par les scientifiques dans une expérience.
- **Pratique:** Pour chacun des sujets suivants, décidez si ce sujet peut être étudié à l'aide d'un plan de recherche expérimentale et expliquez pourquoi.
 - Effet d'une lésion du lobe pariétal sur la capacité des personnes à effectuer des opérations arithmétiques de base.
 - Effet de la dépression clinique sur le nombre d'amitiés proches qu'ont les gens.
 - Effet d'une formation en groupe sur les compétences sociales de personnes adolescentes ayant un syndrome d'Asperger.
 - Effet de payer des personnes pour passer un test de QI sur leur performance à ce test.
- **Discussion:** Imaginez qu'une expérience montre que les personnes participantes, qui reçoivent une thérapie psychodynamique pour une phobie des chiens, s'améliorent davantage que les personnes participantes d'un groupe contrôle sans traitement. Expliquez le problème fondamental de ce devis de recherche et au moins deux façons de le corriger.
- **Discussion:** Pour chacun des sujets suivants, énumérez les avantages et les inconvénients du devis intersujets et du devis intrasujet et décidez lequel serait le préférable.
 - Vous souhaitez tester l'efficacité relative de deux programmes d'entraînement pour courir un marathon.
 - En utilisant des photographies de personnes comme stimuli, vous voulez voir si les personnes souriantes sont perçues comme plus intelligentes que les personnes qui ne sourient pas.
 - Dans le cadre d'une expérience sur le terrain, vous souhaitez vérifier si la tenue vestimentaire d'une personne mendiant (soigné ou négligé) influence la possibilité que les passants lui donnent ou non de l'argent.
 - Vous voulez voir si les noms concrets (par exemple chien) sont mieux mémorisés que les noms abstraits (par exemple vérité).
- **Pratique:** Citez deux façons de recruter des personnes participantes auprès de chacune des populations suivantes :

- les personnes âgées
 - les personnes au chômage
 - les personnes faisant régulièrement de l'exercice
 - la communauté estudiantine faisant une majeure en mathématiques
- **Discussion:** Imaginez une étude dans laquelle vous présentez visuellement aux personnes participantes une liste de 20 mots, un à la fois, attendez un court délai, puis vous leur demandez de rappeler le plus de mots possible. Dans la condition stressée, on leur dit qu'elles pourraient également être choisies pour faire un bref discours devant un petit public. Dans l'état non stressé, on ne leur dit pas qu'elles pourraient avoir à faire un discours. Quelles sont les différentes mesures spécifiques que vous pourriez prendre pour standardiser votre procédure?

32. Références

- Bauman, C.W., McGraw, A.P., Bartels, D.M. et Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology (Réexamen de la validité externe : Préoccupations concernant les problèmes de chariots et autres dilemmes sacrificiels en psychologie morale). *Social and Personality Psychology Compass*, 8(9), 536-554. <https://doi.org/10.1111/spc3.12131>
- Birnbaum, M.H. (1999). How to show that $g > 221$: Collect judgments in a between-subjects design (Comment montrer que $g > 221$: Recueillir des jugements dans un plan entre sujets). *Psychological Methods*, 4(3), 243-249. <https://doi.org/10.1037/1082-989X.4.3.243>
- Cialdini, R. (2005, April). Don't throw in the towel: Use social influence research (Ne jetez pas l'éponge : Utiliser la recherche sur l'influence sociale). *APS Observer*. <https://www.psychologicalscience.org/index.php/publications/observer/2005/april-05/dont-throw-in-the-towel-use-social-influence-research.html>
- Chekroun P. (2008). Pourquoi les individus aident-ils moins autrui lorsqu'ils sont nombreux ?. *Revue électronique de Psychologie Sociale*, 2, 9-16. <https://psychologiescientifique.org/wp-content/uploads/2018/02/Chekroun-2008-Pourquoi-les-individus-aident-ils-moins-autrui-lor.pdf>
- Darley, J. M. et Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility (L'intervention des passants dans les situations d'urgence : Diffusion de la responsabilité). *Journal of Personality and Social Psychology*, 8(4, Pt.1), 377-383. <https://doi.org/10.1037/h0025589>
- Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M. et Twenge, J. M. (1998). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance (Ce maillot de bain devient vous : Différences entre les sexes dans l'auto-objectivation, la restriction alimentaire et les performances en mathématiques). *Journal of Personality and Social Psychology*, 75(1), 269-284. <https://doi.org/10.1037/0022-3514.75.1.269>
- Goldstein, N. J., Cialdini, R. B. et Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels (Une chambre avec un point de vue : L'utilisation des normes sociales pour motiver la conservation de l'environnement dans les hôtels). *Journal of Consumer Research*, 35(3), 472-482. <https://doi.org/10.1086/586910>
- Guéguen, N. et de Gail, Marie-Agnès. (2003). The effect of smiling on helping behavior: Smiling and good Samaritan behavior (L'effet du sourire sur le comportement d'aide : Le sourire et le comportement du bon samaritain). *Communication Reports*, 16(2), 133-140. <https://doi.org/10.1080/08934210309384496>
- Ibolya, K., Brake, A., et Voss, U. (2004). The effect of experimenter characteristics on pain reports in women and men (Effet des caractéristiques de l'expérimentateur sur les rapports de douleur chez les femmes et les hommes). *Pain*, 112(1-2), 142-147. <https://doi.org/10.1016/j.pain.2004.08.008>

- Judd, C.M. et Kenny, D.A. (1981). *Estimating the effects of social interventions (Estimation des effets des interventions sociales)*. Cambridge University Press.
- Knecht, S., Draeger, B., Deppe, M., Bobe, L., Lohmann, H., Floel, A., Ringelstein, E.-B. et Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans (La main et la dominance hémisphérique du langage chez les humains en bonne santé.). *Brain: A Journal of Neurology*, 123(12), 2512-2518. <https://doi.org/10.1093/brain/123.12.2512>
- Manning, R., Levine, M. et Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses (le meurtre de Kitty Genovese et la psychologie sociale de l'aide : La parabole des 38 témoins). *American Psychologist*, 62(6), 555-562. <https://doi.org/10.1037/0003-066X.62.6.555>
- Morling, B. (2014, Avril). Teach your students to be better consumers (Apprenez à vos étudiants à devenir de meilleurs consommateurs). *APS Observer*. <http://www.psychologicalscience.org/index.php/publications/observer/2014/april-14/teach-your-students-to-be-better-consumers.html>
- Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Hollingsworth, J. C., Carol M. Ashton, C. M. et Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee (Essai contrôlé de la chirurgie arthroscopique pour l'arthrose du genou). *The New England Journal of Medicine*, 347(2), 81-88. <https://doi.org/10.1056/NEJMoa013259>
- Price, D. D., Finniss, D. G. et Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought (Une revue complète de l'effet placebo : Avancées récentes et réflexions actuelles). *Annual Review of Psychology*, 59, 565-590. <https://doi.org/10.1146/annurev.psych.59.113006.095941>
- Rosenthal, R. (1976). *Experimenter effects in behavioral research (Effets de l'expérimentateur dans la recherche comportementale)* (éd. augmentée). Wiley
- Rosenthal, R., et Fode, K. (1963). The effect of experimenter bias on performance of the albino rat (L'effet du biais de l'expérimentateur sur la performance du rat albinos). *Behavioral Science*, 8(3), 183-189. <https://doi.org/10.1002/bs.3830080302>
- Rosenthal, R. et Rosnow, R. L. (1976). *The volunteer subject (Le sujet volontaire)*. Wiley.
- Shapiro, A. K. et Shapiro, E. (1999). *The powerful placebo: From ancient priest to modern physician (Le puissant placebo : Du prêtre antique au médecin moderne)*. Johns Hopkins University Press.

CHAPITRE 6-LA RECHERCHE NON EXPÉRIMENTALE



Image 6. Observation

Quel est le point commun entre les études classiques suivantes?

- Stanley Milgram a constaté qu'environ deux tiers des personnes participantes à ses recherches étaient prêtes à administrer des chocs dangereux à une autre personne simplement parce qu'une figure d'autorité leur avait dit de le faire (Milgram, 1963).
- Elizabeth Loftus et Jacqueline Pickrell ont montré qu'il est relativement facile « d'implanter » de faux souvenirs chez des personnes en les interrogeant de manière répétée sur des événements de leur enfance qui ne leur sont pas réellement arrivés (Loftus et Pickrell, 1995).
- John Cacioppo et Richard Petty ont évalué la validité de leur Need for Cognition Scale (échelle du besoin

de cognition) — une mesure du degré d'appréciation et de valorisation de la pensée — en comparant les scores de membres du corps professoral d'université à ceux de la main-d'œuvre d'usine (Cacioppo et Petty, 1982).

- David Rosenhan a découvert que les personnes complices qui se rendaient dans des hôpitaux psychiatriques en affirmant avoir entendu des voix disant des choses telles que « vide » et « bruit sourd » étaient étiquetées comme schizophrènes par le personnel de l'hôpital et y étaient maintenues même si elles se comportaient normalement dans tous les autres domaines (Rosenhan, 1973).

Pour les besoins de ce chapitre, la réponse est qu'il ne s'agit pas d'expériences. Dans ce chapitre, nous examinons de plus près la recherche non expérimentale. Nous commençons par une définition générale de celle-ci, accompagnée d'une discussion sur le moment et les raisons pour lesquels la recherche non expérimentale est plus appropriée que la recherche expérimentale. Nous examinerons ensuite séparément deux types importants de recherche non expérimentale : la recherche corrélacionnelle et la recherche par observation.

33. Aperçu de la recherche non expérimentale

Objectifs d'apprentissage

1. Définir la recherche non expérimentale, la distinguer clairement de la recherche expérimentale et donner plusieurs exemples.
2. Expliquer dans quelles circonstances une personne pourrait choisir de mener une recherche non expérimentale plutôt qu'une recherche expérimentale.

Qu'est-ce que la recherche non expérimentale?

La **recherche non expérimentale** est une recherche qui ne comporte pas de manipulation d'une variable indépendante. Plutôt que de manipuler une variable indépendante, les personnes qui mènent des recherches non expérimentales se contentent de mesurer les variables telles qu'elles se présentent naturellement (en laboratoire ou dans le monde réel).

La plupart des personnes menant des recherches en psychologie considèrent que la distinction entre la recherche expérimentale et la recherche non expérimentale est extrêmement importante. C'est que, même si la recherche expérimentale peut fournir des preuves solides que des changements dans une variable indépendante causent des différences dans une variable dépendante, la recherche non expérimentale ne peut généralement pas le faire. Toutefois comme nous le verrons, cette incapacité à tirer des conclusions causales ne signifie pas que la recherche non expérimentale est moins importante que la recherche expérimentale. Elle est simplement utilisée dans les cas où la recherche expérimentale n'est pas possible.

Quand utiliser la recherche non expérimentale?

Comme nous l'avons vu dans le chapitre précédent, la recherche expérimentale est appropriée lorsque les scientifiques ont une question de recherche spécifique ou une hypothèse sur une relation causale entre deux variables et qu'il est possible, faisable et éthique de manipuler la variable indépendante. Il va donc de soi que la recherche non expérimentale est appropriée — voire nécessaire — lorsque ces conditions ne sont pas

remplies. Il existe de nombreuses situations dans lesquelles la recherche non expérimentale est préférable, notamment dans les cas suivants :

- La question ou l'hypothèse de recherche porte sur une seule variable plutôt que sur une relation statistique entre deux variables (par exemple dans quelle mesure les premières impressions des gens sont-elles exactes?).
- La question de recherche porte sur une relation statistique non causale entre des variables (par exemple existe-t-il une corrélation entre l'intelligence verbale et l'intelligence mathématique?).
- La question de recherche porte sur une relation causale, mais la variable indépendante ne peut pas être manipulée ou les personnes participantes ne peuvent pas être affectées de manière aléatoire à des conditions ou à des ordres de conditions pour des raisons pratiques ou éthiques (par exemple les dommages causés à l'hippocampe d'une personne nuisent-ils à la formation de traces en mémoire à long terme?).
- La question de recherche est large et exploratoire, ou porte sur à quoi ressemble le fait de vivre une expérience particulière (par exemple à quoi ressemble le fait d'être une mère qui travaille et qui a été diagnostiquée comme souffrant de dépression?).

Là encore le choix entre l'approche expérimentale ou l'approche non expérimentale est généralement dicté par la nature de la question de recherche. Rappelons que les trois objectifs de la science sont de décrire, de prédire et d'expliquer. Si l'objectif est d'expliquer et la question de recherche porte sur les relations causales, l'approche expérimentale est généralement préférable. Si l'objectif est de décrire ou de prédire, une approche non expérimentale est appropriée. Mais les deux approches peuvent également être utilisées de manière complémentaire pour répondre à la même question de recherche. Par exemple dans l'étude originale (non expérimentale) de Milgram sur l'obéissance, il s'intéressait principalement à une variable — la mesure dans laquelle les personnes participantes obéissaient au scientifique lorsque cette personne leur demandait de donner une décharge à la personne complice — et il a observé toutes les personnes participantes accomplissant la même tâche dans les mêmes conditions. Cependant Milgram a ensuite mené des expériences pour explorer les facteurs qui influencent l'obéissance. Il a manipulé plusieurs variables indépendantes, telles que la distance entre la personne responsable de l'étude (scientifique) et la personne participante, la personne participante et la personne complice, et le lieu de l'étude (Milgram, 1974; aussi disponible en français, 2017).

Types de recherche non expérimentale

La recherche non expérimentale se divise en deux grandes catégories : la recherche corrélacionnelle et la recherche par observation.

Le type de recherche non expérimentale le plus courant en psychologie est la **recherche corrélacionnelle**. La recherche corrélacionnelle est considérée comme non expérimentale parce qu'elle se concentre sur la

relation statistique entre deux variables, mais n'inclut pas la manipulation d'une variable indépendante. Plus précisément dans la recherche corrélacionnelle, les scientifiques mesurent deux variables avec peu ou pas de tentative de contrôle des variables externes et évaluent ensuite la relation entre elles. Par exemple des personnes intéressées par la relation entre l'estime de soi et les résultats scolaires pourraient collecter des données sur l'estime de soi des personnes étudiant à l'université et leur moyenne cumulative générale pour voir si les deux variables sont statistiquement liées.

La **recherche par observation** est non expérimentale parce qu'elle se concentre sur l'observation du comportement dans un environnement naturel ou en laboratoire sans manipuler quoi que ce soit. L'étude originale de Milgram sur l'obéissance était en ce sens non expérimentale. Il était principalement intéressé par la mesure dans laquelle les personnes participantes obéissaient au scientifique lorsqu'on leur demandait de donner une décharge à la personne complice et il a observé toutes les personnes participantes accomplir la même tâche dans les mêmes conditions. L'étude de Loftus et Pickrell décrite au début de ce chapitre est également un bon exemple de recherche par observation. La variable était de savoir si les personnes participantes « se souvenaient » d'avoir vécu des événements légèrement traumatisants dans leur enfance (par exemple se perdre dans un centre commercial) qu'elles n'avaient pas réellement vécus, mais sur lesquels les scientifiques les avaient interrogés à plusieurs reprises. Dans cette étude particulière, près d'un tiers des personnes participantes se sont « souvenues » d'au moins un événement. (Comme l'étude originale de Milgram, cette étude a inspiré plusieurs expériences ultérieures sur les facteurs qui influencent les faux souvenirs).

Études transversales, longitudinales et transversales séquentielles

Lorsque les psychologues souhaitent étudier un changement dans le temps (par exemple lorsque les psychologues du développement souhaitent étudier le vieillissement), ils adoptent généralement l'une des trois approches non expérimentales suivantes : transversales, longitudinales, transversales séquentielles. Les **études transversales** consistent à comparer deux ou plusieurs groupes préexistants de personnes (par exemple des enfants à différents stades de développement). Ce qui rend cette approche non expérimentale, c'est qu'il n'y a pas de manipulation d'une variable indépendante et pas de répartition aléatoire des personnes participantes à des groupes. En utilisant ce modèle, les psychologues du développement comparent des groupes de personnes d'âges différents (par exemple de jeunes adultes âgés de 18 à 25 ans et des adultes plus âgés de 60 à 75 ans) sur diverses variables dépendantes (par exemple la mémoire, la dépression, la satisfaction à l'égard de la vie). Bien entendu la principale limitation de l'utilisation de ce modèle pour étudier les effets du vieillissement, c'est que des différences entre les groupes autres que l'âge, peuvent expliquer les différences dans la variable dépendante. Par exemple les différences entre les groupes peuvent refléter la génération dont les personnes proviennent (**effet de cohorte**) plutôt qu'un effet direct de l'âge. C'est pourquoi les **études longitudinales**, dans lesquelles un groupe de personnes est suivi au fil du temps à mesure qu'elles vieillissent, offrent un moyen supérieur d'étudier les effets du vieillissement. Cependant les études longitudinales prennent par définition plus de temps et nécessitent donc un investissement beaucoup plus grand de la part des scientifiques et des personnes participantes. Une troisième approche, connue sous le nom d'**études transversales séquentielles**, combine des éléments des études transversales et longitudinales. Plutôt que de mesurer les différences entre

des personnes appartenant à des groupes d'âge différents ou de suivre les mêmes personnes sur une longue période, les scientifiques qui adoptent cette approche choisissent une période plus courte au cours de laquelle on suit des personnes appartenant à des groupes d'âge différents. Par exemple il est possible de mesurer les changements sur une période de dix ans parmi les personnes participantes qui, au début de l'étude, appartiennent aux groupes d'âge suivants : 20 ans, 30 ans, 40 ans, 50 ans et 60 ans. Ce modèle est avantageux car les scientifiques bénéficient immédiatement de la possibilité de comparer les groupes d'âge après la première évaluation. De plus en suivant les différents groupes d'âge dans le temps, il est possible ensuite de déterminer si les différences initiales trouvées entre les groupes d'âge sont dues à de véritables effets d'âge ou à des effets de cohorte.

Les types de recherche examinés jusqu'à présent sont tous quantitatifs, ce qui signifie que les données sont des nombres analysés à l'aide de techniques statistiques. Mais comme vous l'apprendrez dans ce chapitre, de nombreuses recherches par observation sont de nature plus qualitative. Dans la recherche qualitative, les données sont généralement non numériques et ne peuvent donc pas être analysées à l'aide de techniques statistiques. L'étude d'observation de Rosenhan sur l'expérience des personnes placées dans des services psychiatriques était principalement qualitative. Les données étaient les notes prises par les « pseudo-patientes et pseudo-patients » — les personnes prétendant avoir entendu des voix — ainsi que leurs dossiers hospitaliers. L'analyse de Rosenhan consiste principalement en une description écrite des expériences des personnes pseudo-patientes, appuyée par plusieurs exemples concrets. Pour illustrer la tendance des membres du personnel de l'hôpital à « dépersonnaliser » ses personnes patientes, il a noté : « Lors de mon admission, j'ai passé, avec d'autres pseudo-patientes et pseudo-patients, les premiers examens physiques dans une salle semi-publique, où les membres du personnel vaquaient à leurs occupations comme si nous n'étions pas là » (Rosenhan, 1973, p. 256, traduction). Les données qualitatives nécessitent un ensemble distinct d'outils d'analyse en fonction de la question de recherche. Par exemple l'analyse thématique se concentre sur les thèmes qui émergent des données ou l'analyse des conversations se concentre sur la façon dont les mots ont été prononcés lors d'un entretien ou d'un **groupe de discussion** (*focus group*, en anglais).

La validité interne revisitée

Rappelons que la validité interne correspond au degré auquel le plan de l'étude appuie la conclusion que les changements de la variable indépendante sont à l'origine des différences observées dans la variable dépendante. La Figure 6.1 montre comment les recherches expérimentales, quasi-expérimentales et non expérimentales (corrélationnelles) varient sur le plan de la validité interne. La recherche expérimentale a tendance à avoir la plus grande validité interne parce que le recours à la manipulation (de la variable indépendante) et au contrôle (de variables externes) permet d'exclure d'autres explications pour les relations observées. Si le score moyen de la variable dépendante dans une expérience diffère d'une condition à l'autre, il est fort probable que la variable indépendante soit responsable de cette différence. La recherche non expérimentale (corrélationnelle) est la moins performante en termes de validité interne, car ses devis ne font pas appel à la manipulation ou au contrôle. La recherche quasi-expérimentale (qui sera décrite plus en détail dans un chapitre ultérieur) se situe au milieu, car elle contient certaines des caractéristiques

d'une véritable expérience, mais pas toutes. Par exemple elle pourrait omettre d'utiliser l'assignation aléatoire pour répartir les personnes participantes dans des groupes ou omettre d'utiliser le contrebalancement pour contrôler les effets potentiels de l'ordre. Imaginons par exemple que des scientifiques trouvent deux écoles similaires, mettent en place un programme de lutte contre l'intimidation dans l'une d'elles et constatent moins d'incidents liés à l'intimidation dans cette « école traitement » que dans « l'école contrôle ». Bien qu'une comparaison soit effectuée avec une condition contrôle, l'impossibilité d'assigner aléatoirement les enfants dans les écoles pourrait toujours signifier que les élèves de l'école traitement diffèrent des élèves de l'école contrôle, d'une autre manière, qui pourrait expliquer la différence dans les intimidations (par exemple il pourrait y avoir un effet de sélection).



Figure 6.1. Validité interne des études corrélationnelles, quasi-expérimentales et expérimentales. La validité interne des expériences est généralement élevée, celle des quasi-expériences plus faible et celle des études de corrélation (non expérimentales) encore plus faible.

Remarquons également dans la Figure 6.1 qu'il y a un certain chevauchement dans la validité interne des expériences, des quasi-expériences et des études corrélationnelles (non expérimentales). Par exemple une expérience mal conçue qui comprend de nombreuses variables confondantes peut avoir une validité interne plus faible qu'une quasi-expérience bien conçue sans variable confondante évidente. La validité interne n'est qu'un des nombreux types de validité que l'on peut considérer, comme indiqué au chapitre 5.

34. Recherche corrélacionnelle

Objectifs d'apprentissage

1. Définir la recherche corrélacionnelle et donner plusieurs exemples.
2. Expliquer pourquoi une personne pourrait choisir de mener une recherche corrélacionnelle plutôt qu'une recherche expérimentale ou un autre type de recherche non expérimentale.
3. Interpréter la force et la direction des différents coefficients de corrélation.
4. Expliquer pourquoi la corrélation ne signifie pas la causalité.

Qu'est-ce que la recherche corrélacionnelle?

La recherche corrélacionnelle est un type de recherche non expérimentale dans lequel les scientifiques mesurent deux variables (binaires ou continues) et évaluent la relation statistique (c'est-à-dire la corrélation) entre elles avec peu ou pas d'effort pour contrôler les variables externes. Il existe de nombreuses raisons pour lesquelles les scientifiques s'intéressent aux relations statistiques entre les variables choisissent de mener une étude corrélacionnelle plutôt qu'une expérience. La première est que ces personnes ne croient pas que la relation statistique soit une relation de causalité ou qu'elles ne s'intéressent pas aux relations de causalité. Rappelons que deux objectifs de la science sont de décrire et de prédire, et que la stratégie de recherche corrélacionnelle permet aux scientifiques d'atteindre ces deux objectifs. Plus précisément, cette stratégie peut être utilisée pour décrire la force et la direction de la relation entre deux variables et, s'il existe une relation entre les variables, les scientifiques peuvent utiliser les scores d'une variable pour prédire les scores de l'autre variable (en utilisant une technique statistique appelée régression, qui est examinée plus en détail dans la section sur les corrélations complexes de ce chapitre).

Une autre raison pour laquelle les scientifiques choisissent d'utiliser une étude corrélacionnelle plutôt qu'une expérience est que la relation statistique d'intérêt est censée être causale, mais que les scientifiques *ne peuvent pas* manipuler la variable indépendante parce que c'est impossible, peu pratique ou non éthique. Par exemple alors que des scientifiques peuvent s'intéresser à la relation entre la fréquence de consommation de cannabis et les capacités de mémorisation, ces personnes ne peuvent pas, d'un point de vue éthique, manipuler la fréquence de consommation de cannabis. Elles doivent donc s'appuyer sur la stratégie de la recherche corrélacionnelle; elles doivent simplement mesurer la fréquence à laquelle les personnes consomment du cannabis et mesurer leurs capacités de mémorisation à l'aide d'un test de

mémoire standardisé, puis déterminer si la fréquence à laquelle ces personnes consomment du cannabis est statistiquement liée à la performance au test de mémoire.

La corrélation est également utilisée pour établir la fidélité et la validité des mesures. Par exemple des personnes menant une recherche peuvent évaluer la validité d'un bref test d'extraversion en l'administrant à un grand groupe de personnes participantes en même temps qu'un test d'extraversion plus long dont la validité a déjà été démontrée. Ces scientifiques pourraient alors vérifier si les résultats des personnes participantes au test bref sont fortement corrélés à leurs résultats au test plus long. Aucun des résultats des tests n'est considéré comme la cause de l'autre, de sorte qu'il n'y a pas de variable indépendante à manipuler. En fait les termes de *variable indépendante* et *variable dépendante* ne s'appliquent pas à ce type de recherche.

Un autre point fort de la recherche corrélationnelle est que sa validité externe est souvent supérieure à celle de la recherche expérimentale. Rappelons qu'il existe généralement un compromis entre la validité interne et la validité externe. Au fur et à mesure que l'on ajoute des contrôles aux expériences, la validité interne augmente, mais souvent au détriment de la validité externe, car on introduit des conditions artificielles qui n'existent pas dans la réalité. En revanche les études corrélationnelles ont généralement une faible validité interne parce que rien n'est manipulé ou contrôlé, mais elles ont souvent une validité externe élevée. Comme rien n'est manipulé ou contrôlé par la personne chargée de l'expérimentation, les résultats sont plus susceptibles de refléter les relations qui existent dans le monde réel.

Enfin dans le prolongement de ce compromis entre validité interne et externe, la recherche corrélationnelle peut contribuer à fournir des preuves convergentes en faveur d'une théorie. Si une théorie est soutenue par une véritable expérience dont la validité interne est élevée ainsi que par une recherche corrélationnelle dont la validité externe est élevée, les scientifiques peuvent avoir davantage confiance dans la validité de leur théorie. À titre d'exemple concret, les études corrélationnelles établissant une relation entre le fait de regarder des émissions de télévision violentes et le comportement agressif ont été suivies d'études expérimentales confirmant l'existence d'une relation de causalité (Bushman et Huesmann, 2001).

La recherche corrélationnelle implique-t-elle toujours des variables quantitatives?

Une fausse idée très répandue chez les personnes débutant en recherche est que la recherche corrélationnelle doit porter sur deux variables quantitatives, telles que les scores obtenus à deux tests d'extraversion ou le nombre de tracas quotidiens et le nombre de symptômes ressentis par les personnes. Cependant la caractéristique principale de la recherche corrélationnelle est que les deux variables sont mesurées — aucune n'est manipulée — et cela est vrai peu importe que les variables soient quantitatives ou catégorielles. Imaginons par exemple que des personnes menant une recherche administrent l'échelle d'estime de soi de Rosenberg à 50 personnes étudiant dans des universités au Canada et 50 personnes étudiant dans des universités au Japon. Bien que cela ressemble à une expérience intersujets, il s'agit d'une recherche corrélationnelle, car les scientifiques n'ont pas manipulé les nationalités des personnes aux études. Il en va de même pour l'étude de Cacioppo et Petty, qui compare le besoin de cognition des membres du corps professoral universitaire et celui de la main-d'œuvre en usine. Il s'agit d'une étude corrélationnelle, car les personnes menant la recherche n'ont pas manipulé l'occupation des personnes participantes.

La Figure 6.2 présente les données d'une étude hypothétique sur la relation entre le fait de réaliser des listes quotidiennes de tâches à faire (une « liste de tâches ») et le stress. Notons qu'il n'est pas clair s'il s'agit d'une expérience ou d'une étude corrélationnelle, car on ne sait pas si la variable indépendante a été manipulée. Si l'équipe de recherche

assignait au hasard certaines personnes participantes à la réalisation de listes quotidiennes de tâches à faire et d'autres, à ne pas en faire, il s'agirait d'une expérience. Si en revanche, l'équipe de recherche se contentait de demander aux personnes participantes si elles font des listes quotidiennes de choses à faire, il s'agirait alors d'une étude corrélationnelle. La distinction est importante, car si l'étude était une expérience, on pourrait conclure que la réalisation de listes quotidiennes de tâches à faire réduit le stress des personnes participantes. Mais si c'était une étude corrélationnelle, on pourrait seulement conclure que ces variables sont statistiquement liées. Il est possible que le stress ait un effet négatif sur la capacité des gens à planifier à l'avance (le problème de directionnalité). Ou peut-être que les personnes plus consciencieuses sont à la fois plus susceptibles de réaliser des listes de tâches à faire et moins susceptibles d'être stressées (le problème de la troisième variable). Le point crucial est que ce qui définit une étude comme expérimentale ou corrélationnelle n'est pas les variables étudiées, ni le fait que les variables soient quantitatives ou catégorielles, ni le type de graphique ou de statistiques utilisés pour analyser les données. Ce qui définit une étude, c'est la *manière* dont elle est menée.

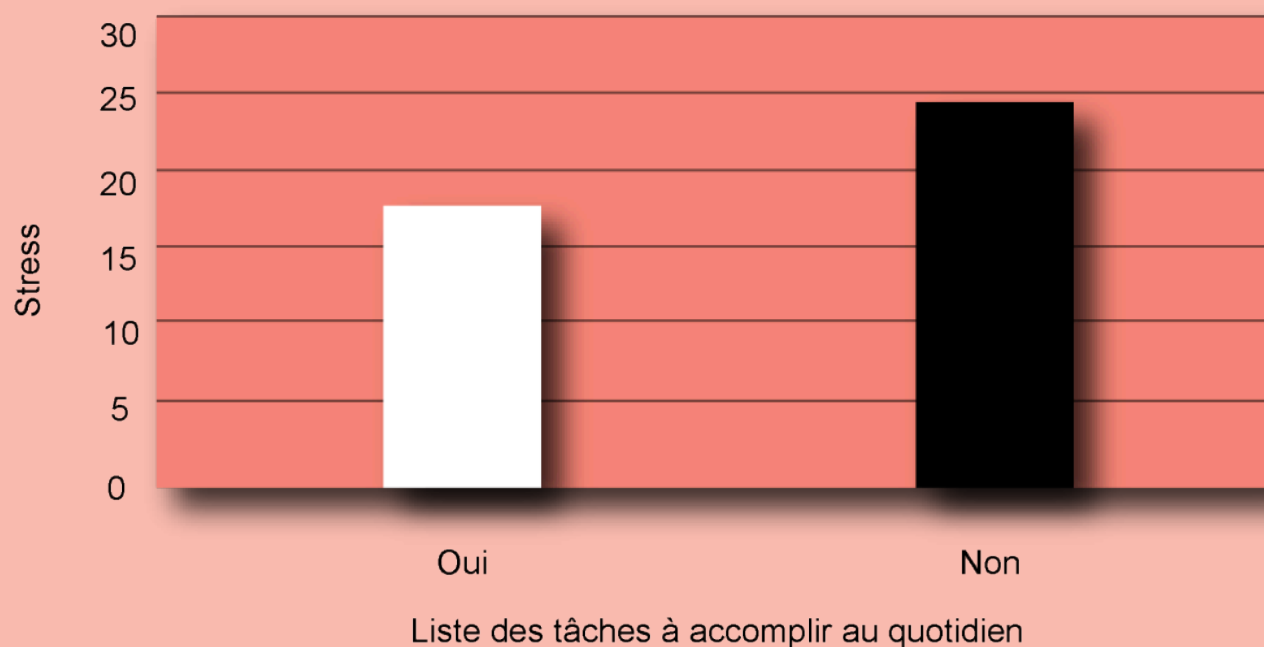


Figure 6.2. Résultats d'une étude hypothétique visant à déterminer si les personnes qui réalisent quotidiennement des listes de tâches à faire sont moins stressées que celles qui ne réalisent pas de telles listes.

Collecte de données dans la recherche corrélationnelle

Là encore la caractéristique principale de la recherche corrélationnelle est qu'aucune des deux variables n'est manipulée. La manière dont les variables sont mesurées et l'endroit où elles le sont n'ont pas d'importance. Des scientifiques pourraient demander à des personnes participantes de se rendre dans un laboratoire pour effectuer une tâche informatisée d'empan numérique à rebours et une tâche informatisée de prise de décision risquée, puis évaluer la relation entre les scores des personnes participantes aux deux tâches. Ou encore des scientifiques pourraient se rendre dans un centre commercial pour interroger les gens sur leurs attitudes à l'égard de l'environnement et leurs habitudes d'achat, puis évaluer la relation entre ces deux variables. Ces deux études seraient corrélationnelles, car aucune variable indépendante n'est manipulée.

Corrélations entre variables quantitatives

Les corrélations entre les variables quantitatives sont souvent présentées à l'aide de **diagrammes de dispersion**. La Figure 6.3 présente des données hypothétiques sur la relation entre le niveau de stress d'une personne et le nombre de symptômes physiques qu'elle présente. Chaque point du diagramme de dispersion représente le score d'une personne pour les deux variables. Par exemple le point encerclé dans la Figure 6.3 représente une personne dont le score de stress était de 10 et qui présentait trois symptômes physiques. Si l'on tient compte de tous les points, on constate que les personnes plus stressées ont tendance à présenter davantage de symptômes physiques. Il s'agit là d'un bon exemple de **relation positive**, dans laquelle les scores d'une variable tendent à être associés à des scores plus élevés pour l'autre variable. En d'autres termes, ils se déplacent dans la même direction, soit vers le haut, soit vers le bas. Une **relation négative** est une relation dans laquelle des résultats plus élevés pour une variable tendent à être associés à des résultats plus faibles pour l'autre variable. En d'autres termes, ils se déplacent dans des directions opposées. Il existe une relation négative entre le stress et le fonctionnement du système immunitaire par exemple, parce que le stress élevé est associé à un fonctionnement plus faible du système immunitaire.

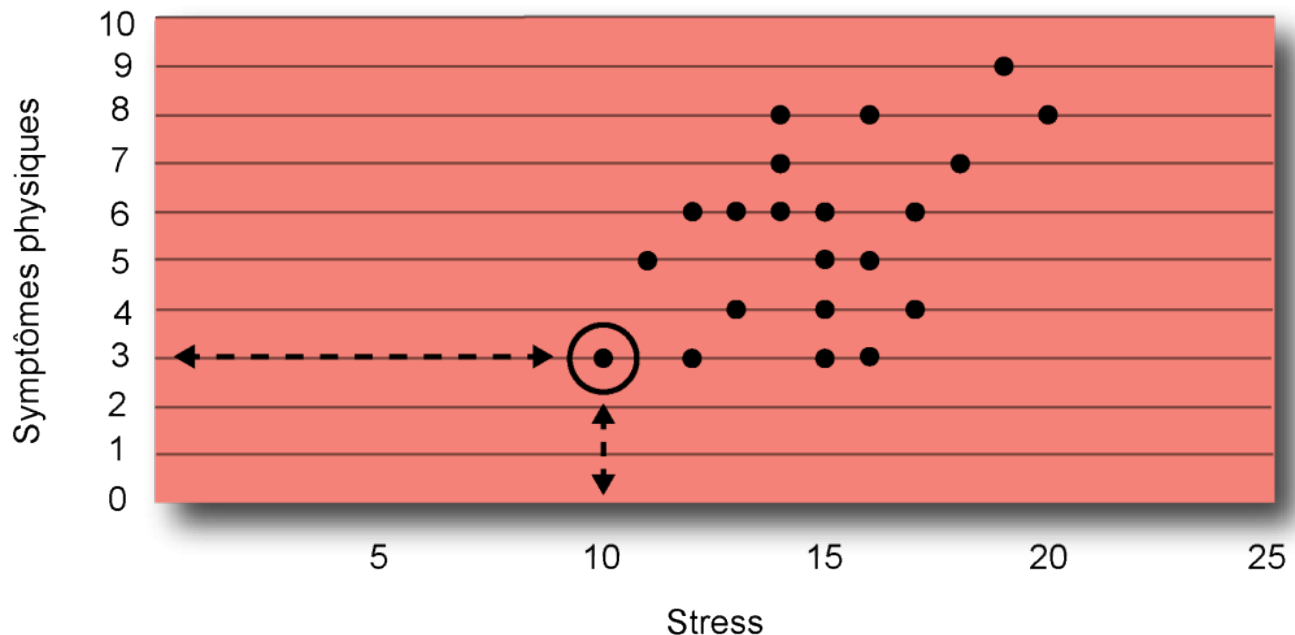


Figure 6.3. Diagramme de dispersion montrant une relation positive hypothétique entre le stress et le nombre de symptômes physiques. Le point encerclé représente une personne dont le score de stress est de 10 et qui présente trois symptômes physiques. Le r de Pearson pour ces données est de + 0,51.

La force d'une corrélation entre des variables quantitatives est généralement mesurée à l'aide d'une statistique appelée le **coefficient de corrélation de Pearson (ou r de Pearson)**. Comme le montre la Figure 6.4, le r de Pearson varie de -1,00 (la relation négative la plus forte possible) à +1,00 (la relation positive la plus forte possible). Une valeur de 0 signifie qu'il n'y a pas de relation entre les deux variables. Lorsque le r de Pearson est égal à 0, les points d'un diagramme de dispersion forment un « nuage » sans forme. Au fur et à mesure que sa valeur évolue vers -1,00 ou +1,00, les points se rapprochent de plus en plus

d'une seule ligne droite. Les coefficients de corrélation proches de $\pm 0,10$ sont considérés comme petits, les coefficients proches de $\pm 0,30$ sont considérés comme moyens et les valeurs proches de $\pm 0,50$ sont considérées comme grandes. Notez que le signe du r de Pearson n'est pas lié à sa force. Des valeurs de r de Pearson de $+ 0,30$ et de $- 0,30$ par exemple sont tout aussi fortes; c'est simplement que l'une représente une relation positive modérée et l'autre une relation négative modérée. À l'exception des coefficients de fidélité, la plupart des corrélations que nous trouvons en psychologie sont de taille petite ou modérée. Le site web [\[en anglais\]](http://rpsychologist.com/d3/correlation/) <http://rpsychologist.com/d3/correlation/>, créé par Kristoffer Magnusson, offre une excellente visualisation interactive des corrélations qui vous permet d'ajuster la force et la direction d'une corrélation, tout en observant les changements correspondants dans le diagramme de dispersion.

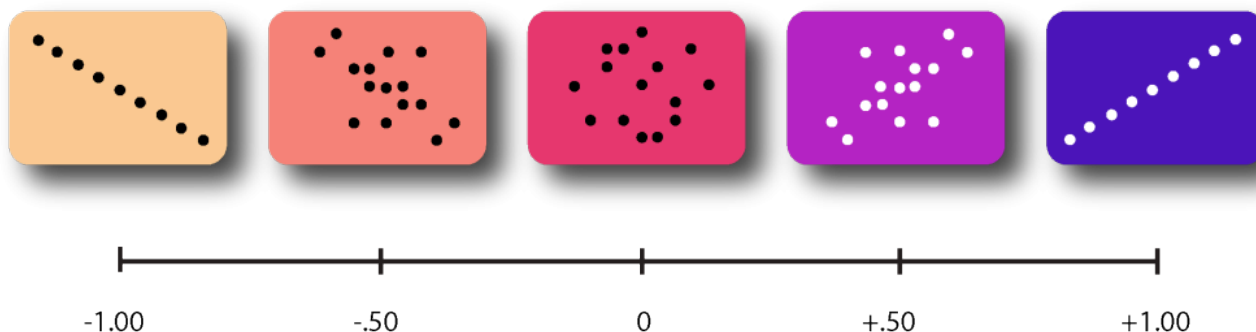


Figure 6.4. Gamme de r de Pearson, de -1,00 (relation négative la plus forte possible) à +1,00 (relation positive la plus forte possible), en passant par 0 (pas de relation).

Il existe deux situations courantes dans lesquelles la valeur du r de Pearson peut être trompeuse. Le r de Pearson est une bonne mesure seulement pour les relations linéaires, dans lesquelles les points sont mieux approximés par une ligne droite. Ce n'est pas une bonne mesure pour les relations non linéaires, dans lesquelles les points sont mieux approximés par une ligne courbe. La Figure 6.5, par exemple, montre une relation hypothétique entre la quantité de sommeil d'une personne par nuit et son niveau de dépression. Dans cet exemple, la ligne qui approxime le mieux les points est une courbe — une sorte de « U » à l'envers — car les personnes qui dorment environ huit heures ont tendance à être les moins déprimées. Les personnes qui dorment peu et celles qui dorment trop ont tendance à être plus déprimées. Même si la Figure 6.5 montre une relation assez forte entre la dépression et le sommeil, le r de Pearson serait proche de zéro parce que les points du diagramme de dispersion ne sont pas bien approximés par une seule ligne droite. Cela signifie qu'il est important de réaliser un diagramme de dispersion et de confirmer qu'une relation est approximativement linéaire avant d'utiliser le r de Pearson. Les relations non linéaires sont assez courantes en psychologie, mais mesurer leur force dépasse le cadre de ce livre.

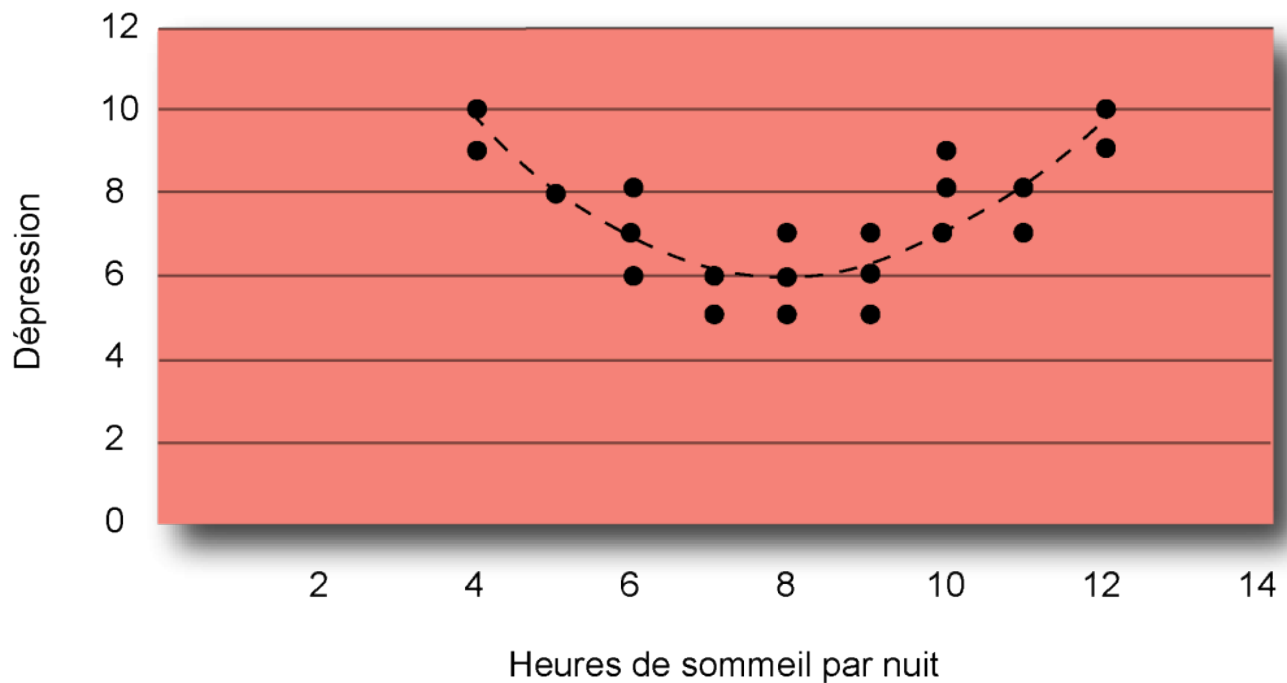


Figure 6.5. Relation hypothétique non linéaire entre le sommeil et la dépression.

Les autres situations courantes dans lesquelles la valeur du r de Pearson peut être trompeuse sont celles où soit une variable, soit les deux variables ont une portée limitée dans l'échantillon par rapport à la population. Ce problème est appelé la **restriction de l'étendue**. Supposons par exemple qu'il existe une forte corrélation négative entre l'âge des personnes et leur appréciation de la musique hip-hop, comme le montre le diagramme de dispersion de la Figure 6.6. Le r de Pearson est ici de $-0,77$. Cependant si nous ne recueillons des données que sur les personnes âgées de 18 à 24 ans — représentées par la zone ombrée de la Figure 6.6 — la relation semblerait alors assez faible. En fait le r de Pearson pour cette étendue restreinte d'âges est de 0. Il est donc judicieux de concevoir les études de manière à éviter toute restriction de l'étendue. Par exemple si l'âge est l'une de vos principales variables, vous pouvez prévoir de collecter des données auprès de personnes d'âges très divers. Étant donné que la restriction de l'étendue n'est pas toujours anticipée ou facilement évitable, c'est une bonne pratique d'examiner vos données pour détecter une éventuelle restriction de l'étendue et d'interpréter le r de Pearson à la lumière de ce problème. (Il existe également des méthodes statistiques permettant de corriger le r de Pearson pour tenir compte de la restriction de l'étendue, mais elles sortent du cadre de ce livre).

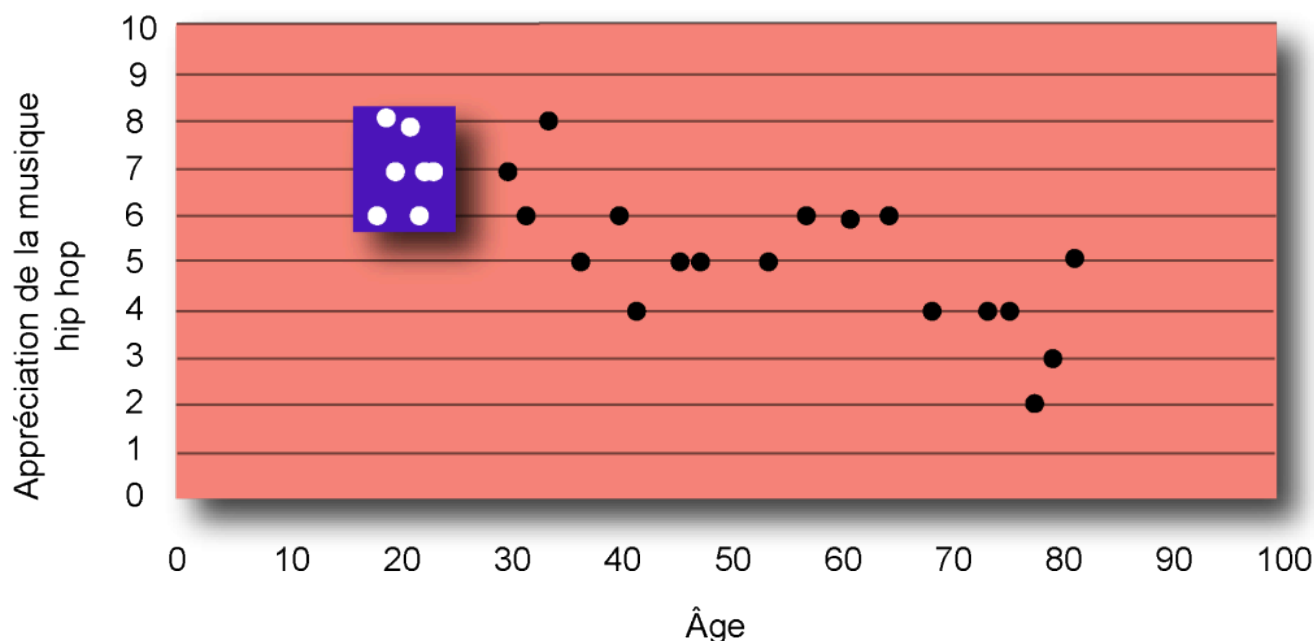


Figure 6.6. Données hypothétiques montrant comment une corrélation globale forte peut sembler faible lorsqu'une variable a un intervalle restreint. La corrélation globale est ici de $-0,77$, mais la corrélation pour les 18-24 ans (dans l'encadré bleu) est de 0.

La corrélation n'implique pas la causalité

Vous avez probablement entendu à plusieurs reprises que « la corrélation n'implique pas la causalité ». Un exemple amusant de ceci vient d'une étude de 2012 qui a montré une corrélation positive (r de Pearson = 0,79) entre la consommation de chocolat par habitant d'une nation et le nombre de prix Nobel décernés aux gens de cette nation (Messerli, 2012). Il semble toutefois évident que cela ne signifie pas que manger du chocolat fait gagner des prix Nobel, et qu'il serait absurde d'essayer d'augmenter le nombre de prix Nobel gagnés en recommandant aux parents de donner plus de chocolat à leurs enfants.

Il y a deux raisons pour lesquelles la corrélation n'implique pas la causalité. La première raison s'appelle le **problème de directionnalité**. Deux variables, X et Y, peuvent être statistiquement liées parce que X cause Y ou parce que Y cause X. Considérons par exemple une étude montrant que le fait de faire de l'exercice ou non est statistiquement lié au degré de bonheur — de sorte que les personnes qui font de l'exercice sont en moyenne plus heureuses que celles qui n'en font pas. Cette relation statistique est compatible avec l'idée que l'exercice physique cause le bonheur, mais elle est également compatible avec l'idée que le bonheur cause l'exercice physique. Peut-être que le fait d'être heureux donne aux gens plus d'énergie ou les incite à chercher des opportunités de socialiser avec d'autres personnes en allant à la salle de sport. La deuxième raison pour laquelle la corrélation n'implique pas la causalité est appelée le **problème de la troisième variable**. Deux variables, X et Y, peuvent être statistiquement liées non pas parce que X cause Y, ou parce que Y cause X, mais parce qu'une troisième variable, Z, cause à la fois X et Y. Par exemple le fait que les nations qui ont remporté le plus de prix Nobel ont tendance à consommer davantage de chocolat reflète probablement la géographie, dans la mesure où les pays européens ont tendance à avoir

des taux de consommation de chocolat par habitant plus élevés, et à investir davantage dans l'éducation et la technologie (encore une fois par habitant) que beaucoup d'autres pays dans le monde. De même la relation statistique entre l'exercice physique et le bonheur pourrait signifier qu'une troisième variable, telle que la santé physique, est à l'origine des deux autres. Le fait d'être en bonne santé physique pourrait inciter les gens à faire de l'exercice et les rendre plus heureux. Les corrélations qui résultent d'une troisième variable sont souvent appelées **corrélations erronées**.

Vous trouverez d'excellents exemples amusants de corrélations erronées sur le site web <http://www.tylervigen.com> (la figure 6.7 en donne un exemple).

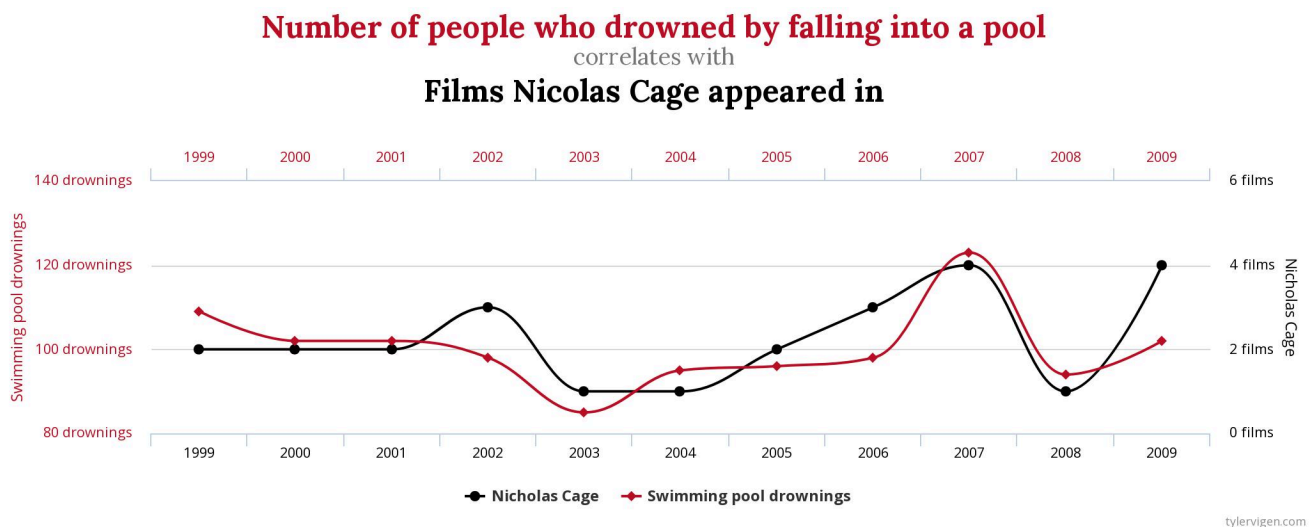


Figure 6.7. Exemple de corrélation fallacieuse. Source: <http://tylervigen.com/spurious-correlations> (CC-BY 4.0)

« Des bonbons en abondance peuvent conduire à la violence »

Bien que les personnes menant des recherches en psychologie savent que la corrélation n'implique pas la causalité, plusieurs journalistes ne le savent pas. Un site web sur la corrélation et la causalité, http://jonathan.mueller.faculty.noctrl.edu/100/correlation_or_causation.htm, renvoie à des dizaines de rapports médiatiques sur de véritables recherches biomédicales et psychologiques. De nombreux titres suggèrent qu'une relation de cause à effet a été démontrée alors qu'une lecture attentive des articles montre que ce n'est pas le cas en raison des problèmes de directionnalité et de troisième variable.

Un de ces articles porte sur une étude montrant que les enfants qui mangent des bonbons tous les jours sont plus susceptibles que les autres d'être arrêtés pour un délit violent plus tard dans leur vie. Mais les bonbons peuvent-ils vraiment « conduire » à la violence, comme le suggère le titre de l'article? Quelles autres explications pouvez-vous donner à cette relation statistique? Comment pourrait-on réécrire le titre pour qu'il ne soit pas trompeur?

Comme vous l'avez appris en lisant ce livre, il existe plusieurs façons d'aborder les problèmes de

directionnalité et de troisième variable. Par exemple, au lieu de simplement mesurer la quantité d'exercice que les gens font, des scientifiques pourraient faire venir des personnes dans un laboratoire et assigner au hasard la moitié d'entre elles à courir sur un tapis roulant pendant 15 minutes et le reste à s'asseoir sur un canapé pendant 15 minutes. Bien que cette modification du devis de recherche semble mineure, elle est extrêmement importante. Maintenant si les personnes qui font de l'exercice sont d'humeur plus positive que celles qui n'en font pas, cela ne peut pas être dû au fait que leur humeur a influencé la quantité d'exercice qu'elles ont fait (car ce sont les scientifiques qui ont aléatoirement réparti les personnes participantes entre les deux conditions; les scientifiques ont déterminé la quantité d'exercice qu'elles ont faite). De même cela ne peut pas être dû au fait qu'une troisième variable (par exemple la santé physique) influence à la fois l'intensité de l'exercice et l'humeur. Les expériences éliminent ainsi les problèmes de directionnalité et de troisième variable, et permettent aux scientifiques de tirer des conclusions solides sur les relations de cause à effet. Pour réviser différemment les notions associées à la corrélation, vous pouvez visionner les vidéos suivantes en français :

- La finance pour tous. (2022, 26 avril). Corrélation n'est pas causalité! — L'instant maths. Institut pour l'éducation financière du public.
<https://www.lafinancepourtous.com/juniors/lyceens/l-instant-maths/correlation-nest-pas-causalite>
- Statistique Canada. (2022, 25 février). Statistique 101 : Corrélation et causalité. Statistique Canada.
<https://www.statcan.gc.ca/fr/afc/litteratie-donnees/catalogue/892000062021002>

Attributions aux médias

- [Nicholas Cage and Pool Drownings](#) © Tyler Viegen is licensed under a [CC BY \(Attribution\)](#) license
- La finance pour tous. (2022, 26 avril). Corrélation n'est pas causalité! – L'instant maths. Institut pour l'éducation financière du public.
<https://www.lafinancepourtous.com/juniors/lyceens/l-instant-maths/correlation-nest-pas-causalite>
- Statistique Canada. (2022, 25 février). Statistique 101 : Corrélation et causalité. Statistique Canada.
<https://www.statcan.gc.ca/fr/afc/litteratie-donnees/catalogue/892000062021002>

35. Corrélation complexe

Objectifs d'apprentissage

1. Expliquer quelques raisons pour lesquelles les personnes en recherche utilisent des modèles corrélationnels complexes.
2. Créer et interpréter une matrice de corrélation.
3. Décrire comment les personnes menant des recherches peuvent utiliser la corrélation partielle et la régression multiple pour contrôler statistiquement les troisièmes variables.

Comme nous l'avons déjà vu, on mène des études corrélationnelles plutôt que des expériences lorsque l'intérêt porte sur des relations non causales ou lorsqu'il porte sur des relations causales, mais que la variable indépendante ne peut pas être manipulée pour des raisons pratiques ou éthiques. Dans cette section, nous examinons certaines approches de la recherche corrélationnelle complexe qui consiste à mesurer plusieurs variables et à évaluer les relations entre elles.

Évaluer les relations entre plusieurs variables

La plupart des recherches corrélationnelles complexes consistent à mesurer plusieurs variables — binaires ou continues — puis à évaluer les relations statistiques entre elles. Par exemple les chercheurs Nathan Radcliffe et William Klein ont étudié un échantillon d'adultes d'âge moyen pour voir comment leur niveau d'optimisme (mesuré à l'aide d'un court questionnaire appelé « *Life Orientation Test* ») est lié à plusieurs autres variables relatives à la survenue d'une crise cardiaque (Radcliffe et Klein, 2002). Ces variables comprenaient leur état de santé, leur connaissance des facteurs de risque de crise cardiaque et leurs croyances quant à leur propre risque de crise cardiaque. Ces scientifiques ont constaté que les personnes participantes les plus optimistes étaient en meilleure santé (par exemple elles faisaient plus d'exercice et avaient une tension artérielle plus basse), connaissaient les facteurs de risque de crise cardiaque et pensaient à juste titre que leur propre risque était inférieur à celui de leurs pairs.

Dans un autre exemple, Ernest Jouriles et ses collègues ont mesuré les expériences d'agressions physiques et psychologiques des adolescentes et adolescents dans leurs relations, et leur détresse psychologique. Étant donné que les mesures de l'agression physique (telles que le *Conflict in Adolescent Dating Relationships Inventory* et le *Relationship Violence Interviews*) ont souvent tendance à produire des distributions très déséquilibrées, les scientifiques ont transformé leurs mesures de l'agression physique en une mesure dichotomique (c'est-à-dire binaire) (0 = n'a pas eu lieu, 1 = a eu lieu). Ceci a aussi été fait pour

les mesures de l'agression psychologique, puis on a mesuré les corrélations entre ces variables, trouvant que les personnes adolescentes ayant subi une agression physique étaient modérément susceptibles d'avoir également subi une agression psychologique, et que le fait d'avoir subi une agression psychologique était lié à des symptômes de détresse psychologique (Jouriles, Garrido, Rosenfield et McDonald, 2009).

Cette approche est souvent utilisée pour évaluer la validité de nouvelles mesures psychologiques. Par exemple lorsque John Cacioppo et Richard Petty ont créé leur Need for Cognition Scale — une mesure du degré auquel les gens aiment penser et apprécient la pensée — ils l'ont utilisée pour mesurer le besoin de cognition d'un large échantillon de personnes étudiant à l'université, avec trois autres variables : l'intelligence, la réponse socialement souhaitable (la tendance à donner ce que l'on pense être la réponse « appropriée ») et le dogmatisme (Caccioppo et Petty, 1982). Les résultats de cette étude sont résumés dans le Tableau 6.1, qui est une **matrice de corrélation** montrant la corrélation (r de Pearson) entre chaque paire possible de variables de l'étude. Par exemple la corrélation entre le besoin de cognition et l'intelligence était de +0,39, la corrélation entre l'intelligence et la réponse socialement souhaitable était de + 0,02, et ainsi de suite. (Seule la moitié de la matrice est remplie car l'autre moitié contiendrait exactement les mêmes informations. En outre comme la corrélation entre une variable et elle-même est toujours de +1,00, ces valeurs sont remplacées par des tirets dans toute la matrice). Dans ce cas, le patron général des corrélations était cohérent avec les idées des personnes menant la recherche, à propos de la manière dont les scores du besoin de cognition devraient être liés à ces autres construits.

Tableau 6.1 Matrice de corrélation montrant les corrélations entre le besoin de cognition et trois autres variables d'après les recherches de Cacioppo et Petty (1982)

	Besoin de cognition	Intelligence	Socialement souhaitable	Dogmatisme
Besoin de cognition	—			
Intelligence	+0,39	—		
Socialement souhaitable	+0,08	+0,02	—	
Dogmatisme	-0,27	-0,23	+0,03	—

Analyse factorielle

Lorsque les scientifiques étudient les relations entre un grand nombre de variables conceptuellement similaires, ces personnes utilisent souvent une technique statistique complexe appelée **analyse factorielle**. Essentiellement, l'analyse factorielle organise les variables en un plus petit nombre de grappes, de sorte qu'elles sont fortement corrélées à l'intérieur de chaque grappe, mais faiblement corrélées entre les grappes. Chaque grappe est alors interprétée comme une mesure multiple du même construit psychologique sous-jacent. Ces construits psychologiques sous-jacents sont aussi appelés facteurs. Par exemple lorsque des personnes effectuent une grande variété de tâches mentales, l'analyse factorielle les organise généralement en deux facteurs principaux — l'un que les scientifiques interprètent comme l'intelligence mathématique (arithmétique, estimation quantitative, raisonnement spatial, etc.) et l'autre que les scientifiques interprètent comme l'intelligence verbale (grammaire, compréhension écrite, vocabulaire, etc.). Les facteurs des inventaires de personnalité inspirés du modèle le Big Five ont été identifiés à partir d'analyses factorielles

des scores obtenus par les personnes sur un grand nombre de traits plus spécifiques. Par exemple les mesures de la gentillesse, de l'esprit grégaire, du niveau d'activité et des émotions positives ont tendance à être fortement corrélées entre elles et sont interprétées comme représentant le construit de l'extraversion. Comme dernier exemple, les chercheurs Peter Rentfrow et Samuel Gosling ont demandé à plus de 1 700 personnes aux études d'évaluer leur degré d'appréciation de 14 genres musicaux populaires différents (Rentfrow et Gosling, 2008). Ils ont ensuite soumis ces 14 variables à une analyse factorielle, qui a permis d'identifier quatre facteurs distincts. Les chercheurs les ont qualifiées de *réfléchies et complexes* (blues, jazz, classique et folk), *intenses et rebelles* (rock, alternatif et heavy métal), *entraînantes et conventionnelles* (country, bande sonore, religieuse, pop) et *énergiques et rythmiques* (rap/hip-hop, soul/funk et musique électronique); voir le Tableau 6.2.

Tableau 6.2 Saturations factorielles des 14 genres musicaux sur quatre composantes principales après rotation Varimax. D'après les travaux de Rentfrow et Gosling (2003)

Genre	Dimension de la préférence musicale			
	<i>Réfléchies et complexes</i>	<i>Intenses et rebelles</i>	<i>Entraînantes et conventionnelles</i>	<i>Énergiques et rythmiques</i>
Blues	.85	.01	-.09	.12
Jazz	.83	.04	.07	.15
Classique	.66	.14	.02	-.13
Folk	.64	.09	.15	-.16
Rock	.17	.85	-.04	-.07
Alternatif	.02	.80	.13	.04
Heavy metal	.07	.75	-.11	.04
Country	-.06	.05	.72	-.03
Bande sonore	.01	.04	.70	.17
Religieuse	.23	-.21	.64	-.01
Pop	-.20	.06	.59	.45
Rap/hip-hop	-.19	-.12	.17	.79
Soul/funk	.39	-.11	.11	.69
Musique électronique	-.02	.15	-.01	.60

Note. N = 1,704. Toutes les saturations de 0,40 ou plus sont en italique ; les saturations les plus élevées pour chaque dimension sont indiquées en caractères gras.

Deux autres points concernant l'analyse factorielle méritent d'être soulignés ici. La première est que les facteurs ne sont pas des catégories. L'analyse factorielle ne nous dit pas que les gens sont extravertis, ou consciencieux, ou qu'ils aiment la musique « réfléchie et complexe », ou la musique « intense et rebelle ». Au contraire les facteurs sont des construits qui fonctionnent indépendamment les uns des autres. Ainsi les personnes qui sont très extraverties peuvent être plus ou moins consciencieuses, et les personnes qui aiment la musique réfléchie et complexe pourraient ou non aimer également la musique intense et rebelle. Deuxièmement, l'analyse factorielle ne révèle que la structure sous-jacente des variables. Il appartient aux scientifiques d'interpréter et d'étiqueter les facteurs, et d'expliquer l'origine de cette structure factorielle

particulière. Par exemple l'une des raisons pour lesquelles l'extraversion et les autres facteurs du Big Five fonctionnent comme des facteurs distincts est qu'ils semblent être contrôlés par des gènes différents (Plomin, DeFries, McClean et McGuffin, 2008).

Explorer les relations causales

Une autre utilisation importante de la recherche corrélacionnelle complexe consiste à explorer les relations causales possibles entre les variables. Cela peut paraître surprenant compte tenu du dicton souvent cité selon lequel « la corrélation n'implique pas la causalité ». Il est vrai que la recherche corrélacionnelle ne peut pas établir sans ambiguïté qu'une variable en cause une autre. Toutefois une recherche corrélacionnelle complexe permet souvent d'écarter d'autres interprétations plausibles. Le principal moyen d'y parvenir est le **contrôle statistique** des tierces variables potentielles. Au lieu de contrôler ces variables par une assignation aléatoire ou en les maintenant constantes comme dans une expérience, les personnes menant la recherche les mesurent et les incluent dans l'analyse statistique appelée **corrélation partielle**. Grâce à cette technique, les scientifiques peuvent examiner la relation entre deux variables, tout en contrôlant statistiquement une ou plusieurs autres variables potentielles.

Par exemple supposons qu'une chercheuse s'intéresse à la relation entre le fait de regarder des émissions de télévision violentes et le comportement agressif, mais qu'elle s'inquiète du fait que le statut socio-économique (SSE) puisse représenter une troisième variable à l'origine de cette relation. Dans ce cas, elle pourrait mener une étude dans laquelle elle mesurerait la quantité de programmes télévisés violents que les personnes participantes regardent dans leur vie quotidienne, le nombre d'actes d'agression qu'elles ont commis et leur statut socio-économique. Elle pourrait d'abord examiner la corrélation entre la consommation de télévision violente et l'agression. Supposons qu'elle trouve une corrélation de $+ 0,35$, ce qui serait considéré comme une corrélation positive de taille modérée. Ensuite elle pourrait utiliser la corrélation partielle pour réexaminer cette relation après avoir contrôlé statistiquement le SSE. Cette technique lui permettrait d'examiner la relation entre la partie de l'écoute de la télévision violente qui est indépendante du SSE et la partie du comportement agressif qui est indépendante du SSE. Si elle constate que la corrélation partielle entre l'écoute de la télévision violente et l'agression, tout en contrôlant le SSE, est de $+ 0,34$, cela suggérerait que la relation entre l'écoute de la télévision violente et l'agression est largement indépendante du SSE (c'est-à-dire que le SSE n'est pas une troisième variable à l'origine de cette relation). En revanche si elle constate qu'après contrôle statistique du statut socio-économique, la corrélation entre la consommation de programmes télévisés violents et l'agression tombe à $+ 0,03$, cela suggérerait que le statut socio-économique est effectivement une troisième variable à l'origine de la relation. Si par contre, elle constate que le contrôle statistique du SSE réduit l'ampleur de la corrélation de $+ 0,35$ à $+ 0,20$, cela suggérerait que le SSE explique une partie, mais pas la totalité de la relation entre la télévision violente et l'agression. Il est important de noter que, si la corrélation partielle constitue un outil important permettant de contrôler statistiquement les variables tierces, sa capacité à tirer des conclusions causales reste limitée, car elle ne règle pas le problème de la directionnalité et il peut exister d'autres variables tierces à l'origine de la relation, que la scientifique n'a pas prises en compte et qu'elle n'a pas contrôlées sur le plan statistique.

Régression

Une fois qu'une relation entre deux variables a été établie, les scientifiques peuvent utiliser cette information pour faire des prédictions sur la valeur d'une variable en fonction de la valeur d'une autre variable. Par exemple une fois que nous avons établi qu'il existe une corrélation entre le QI et la moyenne cumulative générale, nous pouvons utiliser les scores de QI des personnes pour prédire leur moyenne cumulative générale. Ainsi alors que les coefficients de corrélation peuvent être utilisés pour décrire la force et la direction des relations entre les variables, la **régression** est une technique statistique qui permet aux scientifiques de prédire la valeur d'une variable en fonction de la valeur d'une autre variable. La régression peut également être utilisée pour décrire des relations plus complexes entre plus de deux variables. En règle générale, la variable utilisée pour faire la prédiction est appelée la **variable prédictive** et la variable prédite est appelée la **variable résultat ou le critère**. Cette équation de régression a la forme générale suivante :

$$Y = b_1X_1$$

Dans cette formule, Y représente le score prédit de la personne au critère, b_1 représente la pente de la ligne décrivant la relation entre deux variables (ou le coefficient de régression) et X_1 représente le score de la personne à la variable prédictive. Vous pouvez constater que pour prédire le score d'une personne au critère (Y), il suffit de multiplier son score sur la variable prédictive (X) par le coefficient de régression (b_1).

Alors que la **régression simple** consiste à utiliser une variable pour en prédire une autre, la **régression multiple** consiste à mesurer plusieurs variables ($X_1, X_2, X_3, \dots, X_i$) et à les utiliser pour prédire la variable résultat (Y). La régression multiple peut également être utilisée pour décrire simplement la relation entre une seule variable résultat (Y) et un ensemble de variables prédictives ($X_1, X_2, X_3, \dots, X_i$). Le résultat d'une analyse de régression multiple est une équation qui exprime le critère comme une combinaison additive des variables prédictives. Cette équation de régression a la forme générale suivante :

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_iX_i$$

Les coefficients de régression ($b_1, b_2, \dots$) indiquent l'ampleur de la contribution que fait une variable prédictive, en moyenne, par rapport à la prédiction de la variable résultat. Plus précisément, ils indiquent combien la variable de résultat change pour chaque variation d'une unité dans la variable prédictive.

L'avantage de la régression multiple est qu'elle peut montrer si une variable prédictive contribue à une variable résultat au-delà des contributions apportées par d'autres variables prédictives (c'est-à-dire qu'elle peut être utilisée pour montrer si une variable prédictive est liée à une variable résultat, après avoir contrôlé statistiquement d'autres variables prédictives). Supposons qu'une chercheuse veuille savoir comment le revenu et la santé sont liés au bonheur. Ceci est compliqué car le revenu et la santé sont eux-mêmes liés l'un à l'autre. Ainsi si les personnes ayant un revenu plus élevé ont tendance à être plus heureuses, c'est peut-être uniquement parce qu'elles ont tendance à être en meilleure santé. De même si les personnes en meilleure santé ont tendance à être plus heureuses, c'est peut-être uniquement parce qu'elles ont tendance à gagner plus d'argent. Mais une analyse de régression multiple incluant à la fois le revenu et la santé comme variables prédictives montrerait si chacune contribue à la prédiction du bonheur, lorsque l'autre est prise en compte (lorsqu'elle est statistiquement contrôlée). En d'autres termes, la régression multiple permettrait à la chercheuse d'examiner si la partie du revenu, qui n'est pas liée à la santé, prédit ou est liée au bonheur, et si la partie de la santé, qui n'est pas liée au revenu, prédit ou est liée au bonheur. Des recherches de ce type ont d'ailleurs montré que le revenu et la santé ne contribuent que très peu au bonheur, sauf en cas de grande pauvreté ou de maladie (Diener, 2000).

Les exemples présentés dans cette section ne sont qu'un survol de la manière dont les scientifiques utilisent la recherche corrélationnelle complexe pour explorer les relations causales possibles entre les variables. Il est toutefois important de garder à l'esprit que les approches purement corrélationnelles ne peuvent pas établir sans ambiguïté qu'une variable en cause une autre. Le mieux qu'elles peuvent faire est de montrer des modèles de relations qui sont cohérentes avec certaines interprétations causales, et incohérentes avec d'autres.

36. Recherche qualitative

Objectifs d'apprentissage

1. Énumérer plusieurs façons dont la recherche qualitative diffère de la recherche quantitative en psychologie.
2. Décrire les forces et les faiblesses de la recherche qualitative en psychologie comparée à la recherche quantitative.
3. Donner des exemples de recherche qualitative en psychologie.

Qu'est-ce que la recherche qualitative?

Ce livre traite principalement de la **recherche quantitative**, en partie parce que la plupart des études menées en psychologie sont de nature quantitative. Les scientifiques quantitatifs partent généralement d'une question ou d'une hypothèse de recherche ciblée, collectent une petite quantité de données numériques auprès d'un grand nombre de personnes, décrivent les données obtenues à l'aide de techniques statistiques et tirent des conclusions générales sur une grande population. Bien que cette méthode soit largement la plus courante pour mener des recherches empiriques en psychologie, il existe une alternative importante appelée **recherche qualitative**. La recherche qualitative a débuté en anthropologie et en sociologie, mais elle est aujourd'hui également utilisée pour étudier des sujets en psychologie. Les scientifiques qualitatifs commencent généralement par une question de recherche moins ciblée, recueillent de grandes quantités de données relativement « non filtrées » auprès d'un nombre relativement restreint de personnes et décrivent leurs données à l'aide de techniques non statistiques, telles que la théorie ancrée, l'analyse thématique, l'analyse critique du discours ou l'analyse phénoménologique interprétative. Ces personnes sont généralement moins soucieuses de tirer des conclusions générales sur le comportement humain que de comprendre en détail *l'expérience* de leurs personnes participantes à la recherche.

Prenons l'exemple d'une étude menée par les scientifiques Per Lindqvist et ses collègues, qui souhaitaient savoir comment les familles d'adolescentes et adolescents victimes de suicide géraient leur perte (Lindqvist, Johansson et Karlsson, 2008). Cette équipe de recherche n'avait pas de question ou d'hypothèse de recherche spécifique, telle que « Quel est le pourcentage de membres de la famille qui adhèrent à des groupes de soutien aux suicidés? ». Elle souhaitait plutôt comprendre la diversité des réactions des familles, en se concentrant sur *leur* point de vue. Pour répondre à cette question, ces scientifiques ont interrogé les familles de 10 personnes adolescentes mortes par suicide, à leur domicile dans une région rurale de Suède. Les entrevues étaient relativement peu structurées, commençant par une demande générale aux familles de parler de la victime et se terminant par une invitation à parler de toute autre chose qu'elles souhaitaient dire à la personne menant l'entrevue. L'un des thèmes les plus importants qui est ressorti de ces entrevues

est que, même lorsque la vie est redevenue « normale », les familles ont continué à se demander pourquoi leur proche s'était suicidé. Cette lutte s'est avérée particulièrement difficile pour les familles dans lesquelles le suicide était le plus inattendu.

L'objectif de la recherche qualitative

Encore une fois, ce livre porte principalement sur la recherche quantitative en psychologie. La force de la recherche quantitative réside dans sa capacité à fournir des réponses précises à des questions de recherche spécifiques et à tirer des conclusions générales sur le comportement humain. C'est grâce à cette méthode que nous savons que les gens ont une forte tendance à obéir aux figures d'autorité par exemple, et que les étudiantes de premier cycle ne sont pas beaucoup plus bavardes que leurs homologues masculins. Mais si la recherche quantitative permet d'apporter des réponses précises à des questions de recherche spécifiques, elle est loin d'être aussi efficace pour *générer* des questions de recherche nouvelles et intéressantes. De même si la recherche quantitative permet de tirer des conclusions générales sur le comportement humain, elle est loin d'être aussi efficace pour fournir des descriptions détaillées du comportement de groupes particuliers dans des situations particulières. Et la recherche quantitative n'est pas très bonne pour communiquer ce que c'est réellement d'être membre d'un groupe particulier dans une situation particulière.

Mais les faiblesses relatives de la recherche quantitative sont les forces relatives de la recherche qualitative. La recherche qualitative peut aider les scientifiques à générer des questions et des hypothèses de recherche nouvelles et intéressantes. La recherche de Lindqvist et de ses collègues par exemple, suggère qu'il peut y avoir une relation générale entre le degré d'un suicide inattendu et le degré auquel la famille essaie de comprendre pourquoi la personne adolescente s'est suicidée. Cette relation peut maintenant être explorée à l'aide d'une recherche quantitative. Mais il n'est pas certain que cette question aurait été soulevée si les scientifiques n'avaient pas rencontré les familles et écouté ce qu'elles avaient à dire sur leur expérience. La recherche qualitative peut également fournir des descriptions riches et détaillées du comportement humain dans les contextes réels dans lesquels il se produit. Chez les scientifiques qualitatifs, cette profondeur est souvent qualifiée de « description dense » (*thick description*, en anglais) (Geertz, 1973; pour lire une édition plus récente du même auteur, 2017). De même, la recherche qualitative peut donner une idée de ce que c'est que d'être membre d'un groupe particulier ou de se trouver dans une situation particulière — ce que les scientifiques qualitatifs appellent souvent « l'expérience vécue » des personnes participantes à la recherche. Lindqvist et ses collègues décrivent par exemple comment toutes les familles ont spontanément proposé de montrer à la personne menant l'entrevue la chambre de la victime ou l'endroit où le suicide s'est produit — révélant ainsi l'importance de ces lieux physiques pour les familles. Il semble peu probable qu'une étude quantitative aurait permis de découvrir ce détail.

Tableau 6.3 Quelques contrastes entre la recherche qualitative et la recherche quantitative

Qualitative	Quantitative
1. Des informations approfondies sur un nombre relativement restreint de personnes	1. Moins d'informations sur la profondeur avec des échantillons plus grands
2. Les conclusions sont basées sur les interprétations faites par l'enquêteur.	2. Les conclusions sont basées sur des analyses statistiques
3. Globale et exploratoire	3. Spécifique et ciblée

Collecte et analyse des données dans la recherche qualitative

Les méthodes de collecte de données dans la recherche qualitative sont très variées et peuvent impliquer l'observation naturaliste, l'observation des personnes participantes, les données d'archives, les œuvres d'art, et bien d'autres choses encore. Mais l'une des approches les plus courantes, en particulier pour la recherche en psychologie, consiste à mener des **entrevues**. Dans le cadre d'une recherche qualitative, les entrevues peuvent être non structurées — elles consistent en un petit nombre de questions générales ou d'incitations qui permettent aux personnes participantes de parler de ce qui les intéresse — ou structurées, lorsqu'il existe un script strict auquel la personne menant l'entrevue ne déroge pas. La plupart des entrevues se situent entre les deux et sont appelées entrevues semi-structurées, dans lesquelles les scientifiques posent quelques questions cohérentes et peuvent poursuivre en posant des questions plus détaillées sur les sujets abordés. Ces entrevues peuvent être longues et détaillées, mais elles sont généralement menées auprès d'un échantillon relativement petit. L'entrevue non structurée est l'approche utilisée par Lindqvist et ses collègues dans leur recherche sur les familles de victimes de suicide, car ces scientifiques savaient que le degré de divulgation d'un sujet aussi sensible devait être dirigé par les familles, et non par les scientifiques.

Une autre approche utilisée dans la recherche qualitative consiste à faire participer de petits groupes de personnes à des entrevues portant sur un sujet ou une question particulière, appelés **groupes de discussion** (*focus group*, en anglais). L'interaction entre les personnes participantes d'un groupe de discussion permet parfois d'obtenir plus d'informations qu'au cours d'une entrevue individuelle. L'utilisation de groupes de discussion est devenue une technique standard dans le monde des affaires et de l'industrie pour ceux qui veulent comprendre les goûts et les préférences des personnes consommatrices. Le contenu de toutes les entrevues de groupes de discussion est généralement enregistré et transcrit afin de faciliter les analyses ultérieures. Cependant la psychologie sociale nous apprend que la dynamique de groupe est souvent à l'œuvre dans tout groupe, y compris les groupes de discussion, et qu'il est utile d'être conscient de ces possibilités. Par exemple le désir d'être apprécié par les autres peut conduire les personnes participantes à fournir des réponses inexactes qu'elles pensent être perçues favorablement par les autres personnes participantes. On peut en dire de même des caractéristiques de la personnalité. Par exemple les personnes participantes très extraverties peuvent parfois dominer les discussions au sein des groupes de discussion.

L'analyse des données dans la recherche qualitative

Bien que les recherches quantitatives et qualitatives diffèrent généralement sur plusieurs points importants (par exemple la spécificité de la question de recherche, le type de données collectées), c'est la méthode d'analyse des données qui les distingue plus clairement que toute autre chose. Pour illustrer cette idée, imaginons une équipe de scientifiques qui mène une série d'interviews non structurées avec des personnes se rétablissant d'un trouble lié à la consommation d'alcool afin de connaître le rôle de leur foi religieuse dans leur rétablissement. Bien que ce projet semble relever de la recherche qualitative, imaginons qu'une fois les données collectées, elles soient codifiées en fonction de la fréquence à laquelle chaque personne participante mentionne Dieu (ou une « puissance supérieure »), et qu'elles utilisent ensuite des statistiques descriptives et inférentielles pour déterminer si ceux qui mentionnent Dieu plus souvent réussissent mieux à s'abstenir de consommer de l'alcool. Cela ressemble maintenant à de la recherche quantitative. En d'autres termes, la distinction quantitative-qualitative dépend davantage de ce que les scientifiques font avec les données collectées, que de la raison ou de la manière dont ces données ont été collectées.

Mais à quoi ressemble l'analyse des données qualitatives? Tout comme il existe de nombreuses façons de collecter des données dans le cadre d'une recherche qualitative, il existe de nombreuses façons d'analyser les données. Nous nous concentrons ici sur une approche générale appelée la **théorie ancrée** (*grounded theory*, en anglais) (Glaser et Strauss, 1967; pour voir une édition récente, Corbin et Strauss, 2014). Cette approche a été développée dans le domaine de la sociologie dans les années 1960 et a progressivement acquis une popularité en psychologie. Rappelons qu'en recherche quantitative, les scientifiques partent généralement d'une théorie, en déduisent une hypothèse, puis collectent des données pour tester cette hypothèse spécifique. Dans la recherche qualitative utilisant la théorie ancrée, les scientifiques partent des données et développent une théorie ou une interprétation qui est « ancrée » dans ces données. Ces personnes effectuent cette analyse par étapes. Tout d'abord, elles identifient les idées qui se répètent dans les données. Elles organisent ensuite ces idées en un nombre plus restreint de thèmes plus larges. Enfin elles rédigent un **récit théorique** — c'est-à-dire une interprétation des données en fonction des thèmes identifiés. Ce récit théorique se concentre sur l'expérience subjective des personnes participantes, et est généralement étayé par de nombreuses citations directes des personnes participantes elles-mêmes.

Prenons l'exemple d'une étude réalisée par les chercheuses Laura Abrams et Laura Curran, qui ont utilisé l'approche de la théorie ancrée pour étudier l'expérience des symptômes de la dépression post-partum chez les mères à faible revenu (Abrams et Curran, 2009). Leurs données sont le résultat d'entrevues non structurées avec 19 participantes. Le Tableau 6.4 montre les cinq grands thèmes identifiés par ces chercheuses et les idées répétitives plus spécifiques qui composent chacun de ces thèmes. Dans leur rapport de recherche, elles fournissent de nombreuses citations de leurs participantes, comme celle-ci dite par « Destiny » :

Et bien tout récemment, mon appartement a été cambriolé et le fait que son Medicaid, pour une raison quelconque, a été annulé, ce qui fait que beaucoup de choses se sont produites au cours des deux dernières semaines en même temps. Donc en soi je ne veux pas dire que cela m'a presque rendue dingue, mais cela m'a mise dans un mauvais état....Comme si j'étais vraiment déprimée. (p. 357, traduction)

Leur récit théorique se concentre sur l'expérience vécue des participantes face à leurs symptômes, non pas

comme un « trouble affectif » abstrait, mais comme étant étroitement lié à la lutte quotidienne pour élever des enfants seuls souvent dans des conditions difficiles.

Table 6.4 Thèmes et idées récurrentes dans une étude sur la dépression post-partum chez les mères à faible revenu. D'après la recherche d'Abrams et Curran (2009).

Thème	Idées récurrentes
Ambivalence	«Je n'étais pas prête pour ce bébé», «Je ne voulais plus avoir d'enfants».
La surcharge de travail des soignants	«Arrête de pleurer», «J'ai besoin d'une pause», «Je n'en peux plus».
Jonglerie	«Pas le temps de respirer», «Tout le monde dépend de moi», «Naviguer dans le labyrinthe».
Mères seules	«Je n'ai vraiment pas d'aide», «Mon bébé n'a pas de père».
Inquiétude dans la vie réelle	«Je n'ai pas d'argent», «Mon bébé va-t-il s'en sortir ?» «Ce n'est pas sûr ici.»

Le « débat » quantitatif/qualitatif

Compte tenu de leurs différences, il n'est pas surprenant que la recherche quantitative et la recherche qualitative en psychologie et dans les domaines connexes ne coexistent pas en parfaite harmonie. Plusieurs scientifiques en quantitatif critiquent les méthodes qualitatives en raison du fait qu'elles manquent d'objectivité et qu'elles sont difficiles à évaluer en matière de fidélité et de validité, et qu'elles ne permettent pas de généraliser les résultats à d'autres personnes ou situations que celles réellement étudiées. En même temps, certains scientifiques qualitatifs critiquent les méthodes quantitatives en raison du fait qu'elles négligent la richesse du comportement et de l'expérience humaine, et qu'elles répondent plutôt à des questions simples sur des variables facilement quantifiables.

En général cependant, les scientifiques en qualitatif sont bien conscients des questions d'objectivité, de fidélité, de validité et de généralisation. En fait ces personnes ont développé un certain nombre de méthodes pour traiter ces questions (qui dépassent le cadre de notre discussion). Et en général, les scientifiques en quantitatif connaissent le problème de la sursimplification. Ces personnes ne croient pas que l'ensemble du comportement et de l'expérience humaine puisse être décrit de manière adéquate en termes d'un petit nombre de variables et de relations statistiques entre elles. Au contraire, ils utilisent la simplification comme stratégie pour découvrir les principes généraux du comportement humain.

De nombreux scientifiques des deux camps, quantitatif et qualitatif, s'accordent aujourd'hui à dire que les deux approches peuvent et doivent être combinées dans ce que l'on appelle désormais la **recherche sur les méthodes mixtes** (Todd, Nerlich, McKeown et Clarke, 2004). (En fait les études de Lindqvist et de ses collègues et d'Abrams et Curran combinent toutes deux des approches quantitatives et qualitatives). L'une des façons de combiner la recherche quantitative et qualitative consiste à utiliser la recherche qualitative pour générer des hypothèses et la recherche quantitative pour les tester. Encore une fois, alors qu'une étude qualitative pourrait suggérer que les familles qui connaissent un suicide inattendu ont plus de difficulté à résoudre la question du pourquoi, une étude quantitative bien conçue pourrait tester une hypothèse en mesurant ces variables spécifiques avec un large échantillon. Une deuxième approche pour combiner la recherche quantitative et qualitative est appelée la **triangulation**. L'idée est d'utiliser simultanément des

méthodes quantitatives et qualitatives pour étudier les mêmes questions générales et de comparer les résultats. Si les résultats des méthodes quantitatives et qualitatives convergent vers la même conclusion générale, ils se renforcent et s'enrichissent mutuellement. Si les résultats divergent, ils soulèvent une nouvelle question intéressante : pourquoi les résultats divergent-ils et comment peut-on les réconcilier?

L'utilisation de la recherche qualitative peut souvent aider à clarifier les résultats quantitatifs à l'aide de la triangulation. Trenor, Yu, Waight, Zerda et Sha (2008) ont étudié l'expérience des étudiantes en ingénierie dans une université. Lors de la première phase, les étudiantes en ingénierie ont été invitées à répondre à une enquête dans laquelle elles ont évalué un certain nombre de leurs perceptions, notamment leur sentiment d'appartenance. Les résultats ont été comparés en fonction de l'origine ethnique des étudiantes, et statistiquement les différents groupes ethniques n'ont montré aucune différence dans l'évaluation de leur sentiment d'appartenance. On pourrait conclure de ce résultat que l'ethnicité n'a rien à voir avec le sentiment d'appartenance. Cependant au cours de la deuxième phase, l'équipe de recherche a également mené des entrevues avec les étudiantes et, dans ces entrevues, de nombreuses étudiantes issues de minorités ont indiqué que la diversité des cultures à l'université renforçait leur sentiment d'appartenance. Sans la composante qualitative, nous aurions pu tirer des conclusions erronées sur les résultats quantitatifs.

Cet exemple montre comment les recherches qualitatives et quantitatives travaillent ensemble pour nous aider à comprendre le comportement humain. Des scientifiques ont caractérisé la recherche qualitative comme étant la meilleure pour identifier les comportements ou le phénomène, alors que la recherche quantitative est la meilleure pour comprendre le sens ou identifier le mécanisme. Cependant, Bryman (2012, pour une édition canadienne récente, Bell, Bryman et Kleinknecht, 2022) plaide en faveur de l'élimination du fossé entre ces manières arbitrairement différentes d'étudier les mêmes questions.

37. Recherche par observation

Objectifs d'apprentissage

1. Énumérer les différents types de méthodes de recherche par observation et faire la distinction entre chacun d'entre eux.
2. Décrire les forces et les faiblesses de chaque méthode de recherche par observation.

Qu'est-ce que la recherche par observation?

Le terme **recherche par observation** est utilisé pour désigner différents types d'études non expérimentales dans lesquelles le comportement est systématiquement observé et enregistré. L'objectif de la recherche par observation est de décrire une variable ou un ensemble de variables. Plus généralement, l'objectif est d'obtenir un instantané des caractéristiques spécifiques d'un individu, d'un groupe ou d'un environnement. Comme décrit précédemment, la recherche par observation est non expérimentale car rien n'est manipulé ou contrôlé, et de ce fait nous ne pouvons pas tirer de conclusions causales en utilisant cette approche. Les données collectées dans le cadre des études par observation sont souvent de nature qualitative, mais elles peuvent également être quantitatives ou les deux (méthodes mixtes). Il existe plusieurs types de méthodes d'observation qui seront décrites ci-dessous.

Observation naturaliste

L'**observation naturaliste** est une méthode d'observation qui consiste à observer le comportement des personnes dans l'environnement où il se produit habituellement. L'observation naturaliste est donc un type de recherche sur le terrain (par opposition à un type de recherche en laboratoire). Les célèbres recherches de Jane Goodall sur les chimpanzés sont un exemple classique d'observation naturaliste. Dre Goodall a passé trois décennies à observer les chimpanzés dans leur environnement naturel en Afrique de l'Est. Elle a examiné des aspects tels que la structure sociale des chimpanzés, les modes d'accouplement, les rôles des sexes, la structure familiale et les soins apportés à la progéniture, en les observant à l'état naturel. Cependant l'observation naturaliste pourrait plus simplement consister à observer la clientèle dans un magasin d'alimentation, des enfants dans une cour de récréation ou la clientèle hospitalisée dans un service de psychiatrie. Les scientifiques qui pratiquent l'observation naturaliste font généralement leurs observations

le plus discrètement possible, de sorte que les personnes participantes ne se rendent pas compte qu'elles sont en train d'être étudiées. Une telle approche est appelée **observation naturaliste déguisée**. D'un point de vue éthique, cette méthode est considérée comme acceptable si les personnes participantes restent anonymes, et si le comportement a lieu dans un cadre public où les gens ne s'attendent normalement pas à bénéficier d'une vie privée. La clientèle d'une épicerie qui met des articles dans leur panier par exemple, adopte un comportement public facilement observable par le personnel du magasin et les autres personnes dans le magasin. Pour cette raison, la plupart des scientifiques considéreraient qu'il est acceptable, d'un point de vue éthique, de les observer dans une étude. D'un autre côté, l'un des arguments contre l'éthique de l'observation naturaliste du « comportement dans les toilettes » évoqué plus tôt dans le livre est que les gens ont une attente raisonnable en matière d'intimité, même dans des toilettes publiques, et que cette attente a été violée.

Dans les cas où il n'est pas éthique ou pratique de mener une observation naturaliste déguisée, les scientifiques peuvent mener une **observation naturaliste non déguisée** où les personnes participantes sont conscientes de la présence du scientifique et de la surveillance de leur comportement. Toutefois l'observation naturaliste non déguisée pose un problème de réactivité. La **réactivité** est le fait qu'une mesure modifie le comportement des personnes participantes. Dans le cas de l'observation naturaliste non déguisée, le problème de la réactivité est que lorsque les personnes savent qu'elles sont observées et étudiées, elles peuvent agir différemment de ce qu'elles feraient normalement. Ce type de réactivité est connu sous le nom d'**effet Hawthorne**. Par exemple vous pouvez agir très différemment dans un bar si vous savez que quelqu'un vous observe et enregistre vos comportements, ce qui invaliderait l'étude. L'observation déguisée est donc moins réactive et peut donc avoir une plus grande validité parce que les gens ne sont pas conscients que leurs comportements sont observés et enregistrés. Cependant nous savons aujourd'hui que les gens s'habituent souvent à être observés et qu'avec le temps, ces personnes commencent à se comporter naturellement en présence du scientifique. En d'autres termes, avec le temps les gens s'habituent à être observés. Pensez aux émissions de télé-réalité comme Big Brother ou Survivor, où les gens sont constamment observés et enregistrés. Alors qu'ils se comportent de la meilleure façon possible au début, en peu de temps ils flirtent, ont des relations sexuelles, ne portent presque rien, se crient dessus et se comportent parfois de façon embarrassante.

Observation participante

Une autre approche de la collecte de données dans la recherche par observation est l'observation participante. Dans l'**observation participante**, les scientifiques deviennent des personnes qui participent activement au groupe ou à la situation étudiée. L'observation participante est très similaire à l'observation naturaliste en ce sens qu'elle consiste à observer le comportement des personnes dans l'environnement dans lequel il se produit habituellement. Comme pour l'observation naturaliste, les données collectées peuvent inclure des entrevues (généralement non structurées), des notes basées sur leurs observations et interactions, des documents, des photographies et d'autres artefacts. La seule différence entre l'observation naturaliste et l'observation participante est que les personnes ayant pris l'engagement de faire une observation participante deviennent des membres actifs du groupe ou de la situation étudiée. La raison d'être de l'observation participante est qu'il peut y avoir des informations importantes qui ne sont accessibles ou ne

peuvent être interprétées que par une personne participant activement au groupe ou à la situation. Comme l'observation naturaliste, l'observation participante peut être déguisée ou non déguisée. Dans l'**observation participante déguisée**, les personnes menant la recherche se font passer pour des membres du groupe social observé et cachent leur véritable identité de personnes menant une recherche.

Dans un exemple célèbre d'observation participante déguisée, Leon Festinger et ses collègues ont infiltré une secte apocalyptique connue sous le nom de Seekers, dont les membres croyaient que l'apocalypse se produirait le 21 décembre 1954. Voulant étudier comment les membres du groupe réagiraient psychologiquement à l'échec inévitable de la prophétie, les personnes menant la recherche ont soigneusement enregistré les événements et les réactions des membres de la secte dans les jours qui ont précédé et suivi la supposée fin du monde. Sans surprise, les membres de la secte n'ont pas abandonné leur croyance, et sont convaincus que c'était leur foi et leurs efforts qui avaient sauvé le monde de la destruction. Festinger et ses collègues ont ensuite publié un livre sur cette expérience utilisée pour illustrer la théorie de la dissonance cognitive (Festinger, Riecken et Schachter, 1956).

Contrairement à l'**observation participante non déguisée**, les scientifiques font partie du groupe étudié et révèlent leur véritable identité de scientifiques au groupe étudié. Une fois de plus, l'observation déguisée des personnes participantes soulève d'importantes questions éthiques. Premièrement, il n'est pas possible d'obtenir un consentement éclairé et deuxièmement, on a recours à la duperie. Les scientifiques trompent les personnes participantes en leur cachant intentionnellement des informations sur les raisons qui les poussent à faire partie du groupe social étudié. Mais parfois, la participation déguisée est le seul moyen d'accéder à un groupe protecteur (comme une secte). En outre l'observation participante déguisée est moins susceptible de susciter la réactivité que l'observation participante non déguisée.

L'étude de Rosenhan (1973) sur l'expérience des personnes dans un service psychiatrique serait considérée comme une observation participante déguisée parce que Rosenhan et ses pseudo-bénéficiaires ont été admis·es dans des hôpitaux psychiatriques en prétendant être des bénéficiaires afin d'observer comment le personnel soignant traite les bénéficiaires en psychiatrie. Le personnel et les autres personnes dans l'unité ne connaissaient pas leur véritable identité en tant que scientifiques.

Un autre exemple d'observation participante provient d'une étude menée par la sociologue Amy Wilkins sur une organisation religieuse universitaire qui soulignait à quel point ses membres étaient heureux (Wilkins, 2008). Wilkins a passé 12 mois à assister et à participer aux réunions et aux événements sociaux du groupe, et elle a interrogé plusieurs membres du groupe. Dans son étude, Wilkins a identifié plusieurs façons dont le groupe « imposait » le bonheur — par exemple en parlant continuellement du bonheur, en décourageant l'expression d'émotions négatives et en utilisant le bonheur comme un moyen de se distinguer des autres groupes.

L'un des principaux avantages de l'observation participante, c'est que les scientifiques sont beaucoup mieux placés pour comprendre le point de vue et les expériences des personnes qu'ils étudient, lorsqu'ils font partie du groupe social. La principale limite de cette approche est que la simple présence de la personne observatrice peut affecter le comportement des personnes observées. Bien qu'il s'agisse également d'un problème lié à l'observation naturaliste, d'autres problèmes se posent lorsque les scientifiques deviennent des membres actifs du groupe social étudié, car ces scientifiques peuvent modifier la dynamique sociale et/ou influencer le comportement des personnes étudiées. De même si les scientifiques agissent en tant que personnes qui observent les personnes participantes, ces personnes peuvent avoir des inquiétudes quant aux préjugés résultant de l'établissement de relations avec les personnes participantes. Concrètement,

les scientifiques peuvent devenir moins objectifs, ce qui augmente la partialité de la personne chargée de l'expérimentation.

Observation structurée

L'**observation structurée** est une autre méthode d'observation. Dans ce cas, la personne qui enquête observe attentivement un ou plusieurs comportements spécifiques dans un cadre particulier qui est plus structuré que les cadres utilisés dans l'observation naturaliste ou l'observation participante. Souvent le cadre dans lequel les observations sont effectuées n'est pas le cadre naturel. Au lieu de cela, les scientifiques peuvent observer des personnes dans l'environnement du laboratoire. Les scientifiques peuvent également observer des personnes dans un cadre naturel (comme une salle de classe) qui a été structuré d'une certaine manière, par exemple en introduisant une tâche spécifique à laquelle les personnes participantes doivent participer ou en introduisant une situation sociale ou une manipulation spécifique.

L'observation structurée est très similaire à l'observation naturaliste et à l'observation participante dans la mesure où, dans les trois cas, les scientifiques observent un comportement naturel. Toutefois l'observation structurée met l'accent sur la collecte de données quantitatives plutôt que qualitatives. Les scientifiques qui utilisent cette approche s'intéressent à un ensemble limité de comportements. Cela leur permet de quantifier les comportements observés. En d'autres termes, l'observation structurée est moins globale que l'observation naturaliste ou l'observation participante, parce que les scientifiques ayant pris l'engagement d'effectuer une observation structurée s'intéressent à un petit nombre de comportements spécifiques. Par conséquent, plutôt que d'enregistrer tout ce qui se passe, les scientifiques se concentrent uniquement sur des comportements très spécifiques qui les intéressent.

L'équipe de recherche de Robert Levine et Ara Norenzayan a utilisé l'observation structurée pour étudier les différences de « rythme de vie » entre différents pays (Levine et Norenzayan, 1999). L'une de leurs mesures consistait à observer les piétons d'une grande ville pour voir combien de temps il leur fallait pour parcourir 9,144 mètres (à l'origine la longueur du parcours est décrite en pieds). Ces scientifiques ont constaté que les gens de certains pays marchaient de manière systématiquement plus rapide que ceux d'autres pays. Par exemple les gens du Canada et de la Suède ont parcouru 9,144 mètres en un peu moins de 13 secondes en moyenne, tandis que les gens du Brésil et de la Roumanie ont mis près de 17 secondes. Lorsque l'observation structurée a lieu dans le « monde réel » complexe et même chaotique, il est important de se demander quand, où et dans quelles conditions les observations seront effectuées, et qui exactement sera observé. Levine et Norenzayan ont décrit leur processus d'échantillonnage comme suit :

« La vitesse de marche des hommes et des femmes sur une distance de [9,144 mètres] a été mesurée à au moins deux endroits dans les principaux centres-villes de chaque ville. Les mesures ont été prises pendant les heures normales d'ouverture, lors de belles journées d'été. Tous les endroits étaient plats, sans obstacle, avec de larges trottoirs et suffisamment peu encombrés pour permettre aux piétons de se déplacer à des vitesses potentiellement maximales. Pour contrôler les effets de la socialisation, seuls les piétons marchant seuls ont été pris en compte. Les enfants, les personnes souffrant d'un handicap physique évident et les lèche-vitrines n'ont pas été chronométrés. Trente-cinq hommes et 35 femmes ont été chronométrés dans la plupart des villes. » (Levine et Norenzayan, 1999, p. 186, traduction).

Une telle description précise du processus d'échantillonnage rend la collecte de données gérable pour

les personnes chargées de l'observation et permet également un certain contrôle des variables externes importantes. Par exemple en effectuant leurs observations par temps clair en été dans tous les pays, Levine et Norenzayan ont contrôlé les effets de la météo sur la vitesse de marche des gens. Dans l'étude de Levine et Norenzayan, la mesure était relativement directe. Ces scientifiques ont simplement mesuré une distance de 9,144 mètres le long d'un trottoir de la ville et ont ensuite utilisé un chronomètre pour chronométrer les personnes pendant qu'elles marchaient sur cette distance.

Comme autre exemple, les chercheurs Robert Kraut et Robert Johnston ont voulu étudier les réactions des personnes jouant aux quilles à leurs lancers, à la fois lorsqu'elles font face aux quilles et ensuite lorsqu'elles se tournent vers leurs partenaires (Kraut et Johnston, 1979). Mais quelles « réactions » devraient-ils observer? Sur la base de recherches antérieures et de leurs propres essais pilotes, Kraut et Johnston ont dressé une liste de réactions comprenant « sourire fermé », « sourire ouvert », « rire », « visage neutre », « regarder vers le bas », « regarder ailleurs » et « se couvrir le visage » (se couvrir le visage avec les mains). Les responsables de l'observation ont mémorisé cette liste, puis ont pratiqué la codification des réactions avec des enregistrements de personnes jouant aux quilles. Pendant l'étude réelle, ces responsables ont parlé dans un enregistreur audio en décrivant les réactions observées. L'un des résultats les plus intéressants de cette étude était que les personnes jouant aux quilles sourient rarement lorsqu'elles font encore face aux quilles. Elles étaient beaucoup plus enclines à sourire une fois retournées vers leurs partenaires, ce qui suggère que le sourire n'est pas seulement une expression de bonheur, mais aussi une forme de communication sociale.

Dans un autre exemple (en laboratoire cette fois), Dov Cohen et ses collègues avaient des personnes responsables de l'observation qui devaient évaluer les réactions émotionnelles de personnes participantes venant tout juste d'être délibérément bousculées et insultées par une personne complice, alors qu'elles venaient de déposer un questionnaire complété au bout d'un couloir. La personne complice se faisait passer pour une personne travaillant dans le même édifice et qui était frustrée de devoir fermer un tiroir à dossiers à deux reprises pour permettre aux personnes participantes de passer devant (d'abord pour déposer le questionnaire au bout du couloir et ensuite au retour vers la salle qu'on croyait être le lieu où se déroulerait l'étude pour laquelle les personnes participantes s'étaient inscrites). Les deux personnes observatrices étaient placées aux extrémités opposées du couloir afin de pouvoir lire le langage corporel des personnes participantes et d'entendre tout ce qu'elles pourraient dire. Il est intéressant de noter que les scientifiques ont émis l'hypothèse que les personnes participantes du sud des États-Unis, qui est l'un des nombreux endroits au monde où existe une « culture de l'honneur », réagiraient avec plus d'agressivité que les personnes participantes du nord des États-Unis, une prédiction qui a en fait été confirmée par les données d'observation (Cohen, Nisbett, Bowdle et Schwarz, 1996).

Lorsque les observations requièrent un jugement de la part des personnes observatrices — comme dans les études de Kraut et Johnston et de Cohen et ses collègues — un processus appelé **codification** est généralement nécessaire. La codification nécessite généralement de définir clairement un ensemble de comportements cibles. Les personnes observatrices classent ensuite les personnes participantes en fonction du comportement qu'ils ont adopté et du nombre de fois qu'ils ont adopté chaque comportement. Les personnes faisant l'observation peuvent même enregistrer la durée de chaque comportement. Les comportements cibles doivent être définis de manière à guider différentes personnes observatrices pour qu'elles les codifient de la même manière. Cette difficulté de codification illustre la question de la fidélité interjuges, mentionnée au chapitre 4. Les scientifiques doivent démontrer la fidélité interjuges de leur procédure de codification en demandant aux personnes responsables de l'observation de coder les mêmes

comportements de manière indépendante et en montrant ensuite que ces différentes personnes sont en étroite concordance. Kraut et Johnston, par exemple, ont enregistré sur vidéo un sous-ensemble des réactions de leurs personnes jouant aux quilles et ont demandé à deux personnes observatrices de les coder indépendamment. Ces deux personnes ont montré qu'elles étaient d'accord sur les réactions manifestées dans 97 % des cas, ce qui indique une bonne fidélité interjuges.

L'un des principaux avantages de l'observation structurée est qu'elle est beaucoup plus efficace que l'observation naturaliste et l'observation participante. Comme les scientifiques se concentrent sur des comportements spécifiques, cela réduit le temps et les dépenses. De plus l'environnement est souvent structuré de manière à encourager les comportements recherchés, ce qui signifie que les scientifiques n'ont pas besoin de passer autant de temps à attendre que les comportements recherchés se manifestent naturellement. Enfin les scientifiques qui utilisent cette approche peuvent clairement exercer un plus grand contrôle sur l'environnement. Toutefois lorsque les scientifiques exercent plus de contrôle sur l'environnement, cela peut rendre l'environnement moins naturel et ainsi diminuer la validité externe. Il est moins évident par exemple que des observations structurées faites en laboratoire puissent être généralisées à un environnement réel. De plus comme les scientifiques engagés dans l'observation structurée ne sont souvent pas déguisés, il peut y avoir plus de problèmes de réactivité.

Études de cas

Une **étude de cas** est un examen approfondi d'une personne. Parfois des études de cas sont également réalisées sur des unités sociales (par exemple une secte) et des événements (par exemple une catastrophe naturelle). Cependant le plus souvent en psychologie, les études de cas fournissent une description et une analyse détaillées d'une personne. Souvent la personne souffre d'une maladie ou d'un trouble rare ou inhabituel ou d'une lésion d'une région spécifique du cerveau.

Comme de nombreuses méthodes de recherche par observation, les études de cas ont tendance à être de nature plus qualitative. Les méthodes d'étude de cas impliquent un examen approfondi, et souvent longitudinal, d'une personne. En fonction de l'objet de l'étude de cas, les personnes peuvent ou non être observées dans leur environnement naturel. Si le milieu naturel n'est pas l'objet de l'étude, la personne peut être amenée dans le bureau de thérapeutes ou dans le laboratoire de scientifiques pour l'étude. Aussi l'essentiel du rapport de l'étude de cas se concentrera sur des descriptions approfondies de la personne plutôt que sur des analyses statistiques. Cela dit, certaines données quantitatives peuvent également être incluses dans la rédaction d'une étude de cas. Par exemple le score de dépression d'une personne peut être comparé à des scores normatifs ou son score avant et après traitement peut être comparé. Comme pour les autres méthodes qualitatives, une variété de méthodes et d'outils peuvent être utilisés pour collecter des informations sur le cas. Par exemple des entrevues, une observation naturaliste, une observation structurée, des tests psychologiques (par exemple un test de QI) et/ou des mesures physiologiques (par exemple un TACO du cerveau) peuvent être utilisés pour collecter des informations sur la personne.

H.M., maintenant connu sous son vrai nom Henry Molaise, est l'une des études de cas les plus célèbres en psychologie. H.M. souffrait d'une épilepsie intractable et très grave. Le chirurgien a localisé l'épilepsie d'H.M. au niveau de son lobe temporal médian et, en 1953, il a retiré de larges sections de son hippocampe pour tenter d'arrêter les crises. Le traitement a été un succès, en ce sens qu'il a permis de résoudre son épilepsie et que son quotient intellectuel et sa personnalité n'ont pas été affectés. Cependant les médecins se sont rapidement rendu compte qu'H.M. présentait une forme étrange d'amnésie, appelée amnésie antérograde. H.M. était capable de tenir une conversation et de se souvenir de courtes chaînes de lettres, de chiffres et de mots. Essentiellement, sa mémoire à court terme était préservée. Cependant H.M. ne pouvait pas enregistrer de nouveaux événements dans sa mémoire. Il a perdu la capacité de transférer des informations de sa mémoire à court terme à sa mémoire à long terme, ce que les scientifiques en mémoire appellent la consolidation. Ainsi alors qu'il pouvait tenir une conversation avec quelqu'un, il oubliait complètement la conversation une fois qu'elle était terminée. Cette étude de cas a été extrêmement importante pour les scientifiques en mémoire, car elle a suggéré qu'il existe une dissociation entre la mémoire à court terme et la mémoire à long terme, qu'il s'agit de deux capacités différentes sous-tendues par des zones différentes du cerveau. Cela suggère également que les lobes temporaux sont particulièrement importants pour la consolidation des nouvelles informations (c'est-à-dire pour le transfert des informations de la mémoire à court terme à la mémoire à long terme).



Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou se rendre sur le site youtu.be/KkaXNvzE4pk



One or more interactive elements has been excluded from this version of the text. You can view them online here:

<https://ecampusontario.pressbooks.pub/recherchemethodologiepsychologie2023e1/?p=120#oembed-1>

L'histoire de la psychologie est remplie d'études de cas célèbres, telles que la description par Sigmund Freud du cas « Anna O. » (voir l'encadré intitulé Le cas d' « Anna O. ») et la description par John Watson et Rosalie Rayner du petit Albert (Watson et Rayner, 1920), à qui les scientifiques auraient appris à craindre un rat blanc — ainsi que d'autres objets à fourrure — en faisant un bruit fort à chaque fois que le rat s'approchait de lui.

Le cas d' « Anna O. »

Sigmund Freud a utilisé le cas d'une jeune femme qu'il appelait « Anna O. » pour illustrer de nombreux principes de sa théorie de la psychanalyse (Freud, 1961). (Son vrai nom était Bertha Pappenheim, et elle était une des premières féministes qui a par la suite apporté d'importantes contributions au domaine du travail social). Anna s'est présentée à Josef Breuer, un collègue de Freud, vers 1880, avec une série de symptômes physiques et psychologiques étranges. L'un d'entre eux était que, pendant plusieurs semaines, elle n'a pu boire aucun liquide. Selon Freud,

Elle prenait le verre d'eau qu'elle désirait ardemment, mais dès qu'il touchait ses lèvres, elle le repoussait comme une personne souffrant d'hydrophobie... Elle ne se nourrissait que de fruits, tels que des melons, etc., afin de diminuer la soif qui la tourmentait. (p. 9)

Mais selon Freud, une percée se produisit un jour alors qu'Anna était sous hypnose.

[E]lle se plaignit de sa « dame de compagnie » anglaise, qu'elle n'aimait pas, et continua à décrire, avec tous les signes du dégoût, comment elle était entrée une fois dans la chambre de cette dame et comment son petit chien — horrible créature! — avait bu dans un verre. La patiente n'avait rien dit, par politesse. Après avoir exprimé avec énergie la colère qu'elle avait retenue, elle demanda à boire, but une grande quantité d'eau sans aucune difficulté et se réveilla de son hypnose avec le verre à la bouche; la perturbation s'est alors évanouie pour ne plus jamais revenir. (p.9)

L'interprétation de Freud était qu'Anna avait refoulé le souvenir de cet incident, ainsi que l'émotion qu'il avait déclenchée, et que c'était ce qui avait causé son incapacité à boire. De plus il pensait que son souvenir de l'incident ainsi que l'expression de l'émotion qu'elle avait refoulée, avaient fait disparaître le symptôme. En tant qu'illustration de la théorie de Freud, l'étude de cas d'Anna O. est très efficace. Cependant en tant que preuve de la théorie, elle est essentiellement sans valeur. La description ne permet pas de savoir si Anna a réellement refoulé le souvenir du chien buvant dans le verre, si ce refoulement a causé son incapacité à boire, ou si le rappel de ce « traumatisme » a soulagé le symptôme. Cette étude de cas ne permet pas non plus de savoir si l'expérience d'Anna était typique ou atypique.



Figure 6.8. Anna O. "Anna O." est le sujet d'une célèbre étude de cas utilisée par Freud pour illustrer les principes de la psychanalyse. Source : http://en.wikipedia.org/wiki/File:Pappenheim_1882.jpg

Les études de cas sont utiles parce qu'elles offrent un niveau d'analyse détaillé que l'on ne retrouve pas dans de nombreuses autres méthodes de recherche et qu'une analyse plus approfondie peut mener à une meilleure compréhension. Grâce à l'étude de cas, la personne menant la recherche peut acquérir une compréhension affinée de ce qui pourrait être important d'examiner de manière plus approfondie dans le cadre de futures recherches plus contrôlées. Les études de cas sont aussi souvent le seul moyen d'étudier des conditions rares, car il peut être impossible de trouver un échantillon suffisamment large de personnes atteintes de la maladie pour utiliser des méthodes quantitatives. Bien qu'à première vue, une étude de cas portant sur une personne rare puisse sembler ne pas nous apprendre grand-chose sur nous-mêmes, elle

permet souvent de mieux comprendre le comportement normal. Le cas d'H.M. a fourni des informations importantes sur le rôle de l'hippocampe dans la consolidation de la mémoire.



Lecture sur version
imprimée? Scannez ce code
QR avec votre téléphone
cellulaire ou se rendre sur le
site
[https://www.nytimes.com/
2016/08/07/magazine/
the-brain-that-couldnt-rem-
ember.html?_r=0](https://www.nytimes.com/2016/08/07/magazine/the-brain-that-couldnt-remember.html?_r=0)

Toutefois, il est important de noter que même si les études de cas peuvent donner un aperçu de certains domaines et variables à étudier, et peuvent être utiles pour aider à développer des théories, elles ne devraient jamais être utilisées comme preuves pour les théories. En d'autres termes, les études de cas peuvent servir d'inspiration pour formuler des théories et des hypothèses, mais ces hypothèses et théories doivent ensuite être testées formellement à l'aide de méthodes quantitatives plus rigoureuses. La raison pour laquelle les études de cas ne devraient pas être utilisées pour soutenir des théories est qu'elles souffrent de problèmes de validité interne et externe. Les études de cas sont dépourvues des contrôles appropriés que comportent les véritables expériences. À ce titre, elles souffrent de problèmes de validité interne et ne peuvent donc pas être utilisées pour déterminer un lien de causalité. Par exemple lors de l'opération d'H.M., le chirurgien peut avoir accidentellement lésé une autre zone du cerveau d'H.M. (une possibilité suggérée par la dissection du cerveau d'H.M. après son décès) et cette lésion peut avoir contribué à son incapacité à consolider de

nouvelles informations. Le fait est que les études de cas ne permettent pas d'exclure ce type d'explications alternatives. Ainsi comme toutes les méthodes d'observation, les études de cas ne permettent pas de déterminer le lien de causalité. En outre comme les études de cas portent souvent sur une seule personne, et généralement sur une personne anormale, les scientifiques ne peuvent pas généraliser leurs conclusions à d'autres personnes. Rappelons qu'avec la plupart des modèles de recherche, il y a un compromis entre la validité interne et la validité externe. Cependant les études de cas posent des problèmes à la fois de validité interne et de validité externe. Il y a donc des limites à la capacité de déterminer la causalité et de généraliser les résultats. Une dernière limite des études de cas, c'est que les préjugés théoriques du scientifique ont largement la possibilité de colorer ou de biaiser la description du cas. En effet des accusations ont été formulées contre la femme qui a étudié H.M., selon lesquelles elle a détruit un grand nombre de données qui n'ont pas été publiées, et elle a été questionnée pour avoir détruit des données contradictoires, qui ne soutenaient pas sa théorie, sur la manière dont les souvenirs se consolident. Un article fascinant du New York Times décrivant certaines des controverses qui ont suivi la mort d'H.M. et l'analyse de son cerveau est disponible à l'adresse suivante [\[en anglais\]](#) :



Que se passe-t-il lorsque l'on retire l'hippocampe du cerveau ? – Sam Kean [traduit si vous activez les sous-titres [en français](#)]. Lecture sur version imprimée? Scannez ce code QR avec votre téléphone cellulaire ou se rendre sur le site [https://ed.ted.com/lessons/
what-happens-when-you-remove-the-hippocampus-sam-kean](https://ed.ted.com/lessons/what-happens-when-you-remove-the-hippocampus-sam-kean)



Cliquez ici pour lire un retour sur l'histoire d'H.M. ([Gros, 2017](#)).
Lecture sur version imprimée? Scannez ce code QR avec votre
téléphone cellulaire ou se rendre sur le site
[https://lejournel.cnrs.fr/nos-blogs/aux-frontieres-du-cerveau/
limage-de-la-semaine-henry-molaison-lhomme-qui-ne-pouvait-
plus](https://lejournel.cnrs.fr/nos-blogs/aux-frontieres-du-cerveau/limage-de-la-semaine-henry-molaison-lhomme-qui-ne-pouvait-plus)

Recherche archivistique

Une autre approche, souvent considérée comme de la recherche par observation, consiste à analyser des données d'archives qui ont déjà été collectées pour un autre objectif. Un exemple est l'étude de Brett Pelham et de ses collègues sur « l'égoïsme implicite » — c'est-à-dire la tendance des gens à préférer les personnes, les lieux et les choses qui leur ressemblent (Pelham, Carvallo et Jones, 2005). Dans une étude, ils ont examiné les dossiers de sécurité sociale pour montrer que les femmes portant les noms Virginia, Georgia, Louise et Florence étaient particulièrement susceptibles d'avoir déménagé respectivement dans les États de Virginie, de Géorgie, de Louisiane et de Floride.

Comme pour l'observation naturaliste, les mesures peuvent être plus ou moins simples lorsque l'on travaille avec des données d'archives. Par exemple compter le nombre de personnes nommées Virginia qui vivent dans différents États, sur la base des dossiers de sécurité sociale, est relativement simple. Mais prenons l'exemple d'une étude menée par Christopher Peterson et ses collègues sur la relation entre l'optimisme et la santé, à partir de données collectées de nombreuses années auparavant, dans le cadre d'une étude sur le développement de l'adulte (Peterson, Seligman et Vaillant, 1988). Dans les années 1940, des étudiants masculins en bonne santé avaient répondu à un questionnaire à questions ouvertes portant sur des expériences difficiles vécues en temps de guerre. Vers la fin des années 1980, Peterson et ses collègues ont examiné les réponses des hommes au questionnaire afin d'obtenir une mesure du style explicatif — c'est-à-dire de leur façon habituelle d'expliquer les mauvais événements qui leur arrivent. Les personnes plus pessimistes ont tendance à se blâmer eux-mêmes et à s'attendre à des conséquences négatives à long terme qui affectent de nombreux aspects de leur vie, tandis que les personnes plus optimistes ont tendance à blâmer des forces extérieures et à s'attendre à des conséquences négatives limitées. Pour obtenir une mesure du style explicatif de chaque participant, les scientifiques ont utilisé une procédure dans laquelle tous les événements négatifs mentionnés dans les réponses au questionnaire et toutes les explications causales de ces événements ont été identifiés et notés sur des fiches. Celles-ci ont été données à un groupe séparé de personnes responsables de l'évaluation qui ont classé chaque explication en fonction de trois dimensions distinctes de l'optimisme et du pessimisme. La moyenne de ces notes a ensuite été calculée pour obtenir un score de style explicatif pour chaque participant. Les scientifiques ont ensuite évalué la relation statistique entre le style explicatif des hommes lorsqu'ils étaient étudiants de premier cycle et les mesures archivistiques de leur état de santé vers l'âge de 60 ans. Le principal résultat a été que plus les hommes étaient optimistes lorsqu'ils étaient étudiants, plus ils étaient en bonne santé lorsqu'ils étaient plus âgés. Le r de Pearson était de +0,25.

Cette méthode est un exemple d'**analyse de contenu** — une famille d'approches systématiques de la mesure à partir de données d'archives complexes. Tout comme l'observation structurée nécessite de spécifier les comportements d'intérêt et de les noter au fur et à mesure qu'ils se produisent, l'analyse de contenu nécessite de spécifier des mots-clés, des phrases ou des idées et de trouver ensuite toutes les occurrences de ces éléments dans les données. Ces occurrences peuvent ensuite être comptées, chronométrées (par exemple le temps consacré aux sujets de divertissement dans le journal télévisé du soir) ou analysées de diverses autres manières.

Attributions des médias

- Gros, A. (2017, 13 mars). [L'image de la semaine: «Henry Molaison, l'homme qui ne pouvait plus se souvenir»](#). Dans nos blogs aux frontières du cerveau du [CNRS: le journal](#).
- Dwight, D. (2016, 3 août). [The brain that could not remember](#). [New York Times](#).
- [Pappenheim 1882](#) par inconnu se trouve dans la [Public Domain](#).
- [What happens when you remove the hippocampus? – Sam Kean](#) par [TED-Ed](#) sous licence YouTube standard

38. Points clés et exercices

Points clés

- La recherche non expérimentale est une recherche qui ne comporte pas de manipulation d'une variable indépendante.
- Il existe deux grands types de recherche non expérimentale. La recherche corrélacionnelle qui se concentre sur les relations statistiques entre les variables qui sont mesurées mais non manipulées; et la recherche par observation dans laquelle les personnes participantes sont observées et leur comportement est enregistré sans que les personnes menant la recherche n'interfèrent ou ne manipulent aucune variable.
- En général, la recherche expérimentale a une validité interne élevée, la recherche corrélacionnelle une validité interne faible, et la recherche quasi-expérimentale se situe entre les deux.
- La recherche corrélacionnelle consiste à mesurer deux variables et à évaluer la relation entre elles, sans manipulation d'une variable indépendante.
- La corrélation n'implique pas la causalité. Une relation statistique entre deux variables X et Y ne signifie pas nécessairement que X cause Y. Il est également possible que Y cause X, ou qu'une troisième variable Z cause à la fois X et Y.
- Bien que la recherche corrélacionnelle ne puisse pas être utilisée pour établir des relations causales entre les variables, elle permet aux personnes menant les recherches d'atteindre de nombreux autres objectifs importants (établir la fidélité et la validité, fournir des données convergentes, décrire les relations et faire des prédictions).
- Les coefficients de corrélation peuvent aller de -1 à +1. Le signe indique le sens de la relation entre les variables et la valeur numérique indique la force de la relation.
- Les personnes en recherche utilisent souvent la recherche corrélacionnelle complexe pour explorer les relations entre plusieurs variables dans la même étude.
- La recherche corrélacionnelle complexe peut être utilisée pour explorer les relations causales possibles entre les variables à l'aide de techniques, telles que la corrélation partielle et la régression multiple. Ces méthodes peuvent mettre en évidence des patrons de relations qui sont cohérents avec certaines interprétations causales, et incohérents avec d'autres, mais elles ne peuvent pas établir sans ambiguïté qu'une variable est à l'origine d'une autre.
- La recherche qualitative est une alternative importante à la recherche quantitative en psychologie. Elle consiste généralement à poser des questions de recherche plus larges, à collecter des données plus détaillées (par exemple des entretiens) et à utiliser des analyses non statistiques.

- De nombreuses personnes menant des recherches considèrent que la recherche quantitative et la recherche qualitative sont complémentaires et préconisent de les combiner. Par exemple la recherche qualitative peut être utilisée pour générer des hypothèses et la recherche quantitative pour les tester.
- Il existe plusieurs approches différentes de la recherche par observation, notamment l'observation naturaliste, l'observation participante, l'observation structurée, les études de cas et la recherche archivistique.
- L'observation naturaliste permet d'observer des personnes dans leur environnement naturel; l'observation participante consiste à devenir un membre actif du groupe observé; l'observation structurée consiste à codifier un petit nombre de comportements de manière quantitative; les études de cas sont généralement utilisées pour collecter des informations approfondies sur une seule personne; et la recherche archivistique consiste à analyser des données existantes.

Exercices

- **Discussion** : Pour chacune des études suivantes, décidez de quel type de modèle de recherche il s'agit et expliquez pourquoi.
 - Des personnes mènent une recherche dans laquelle il y a des entrevues détaillées avec de jeunes pères, adolescents non mariés, pour savoir ce qu'ils ressentent et ce qu'ils pensent de leur rôle de père, et résument leurs sentiments dans un récit écrit.
 - Des personnes mènent une recherche dans laquelle on mesure l'impulsivité d'un large échantillon d'automobilistes et on examine la relation statistique entre cette variable et le nombre de contraventions reçues par les automobilistes.
 - Des personnes menant une recherche assignent aléatoirement des bénéficiaires souffrant de lombalgie soit à un traitement impliquant l'hypnose, soit à un traitement impliquant l'exercice physique. Après 3 mois, il y a mesure de leur niveau de lombalgie.
- **Discussion** : Pour chacun des éléments suivants, décidez s'il est plus probable que l'étude décrite soit expérimentale ou non expérimentale et expliquez pourquoi.
 - Des psychologues en cognition comparent la capacité des personnes de se rappeler des mots qu'on leur avait demandé de « lire » à celle qu'on leur avait demandé d'« imaginer ».
 - L'équipe en gestion étudie la corrélation entre la moyenne cumulative à l'université par les nouveaux membres du personnel et leurs rapports de performance de la première année.
 - Les spécialistes en ingénierie de l'automobile installent différentes boîtes de vitesses dans un nouveau prototype de voiture, en demandant à chaque fois à plusieurs personnes d'évaluer le confort de la boîte de vitesses.

- Des personnes faisant de la recherche en alimentation étudient la relation entre la température à l'intérieur des réfrigérateurs domestiques et la quantité de bactéries présentes sur les aliments.
- Des personnes menant des recherches en psychologie sociale disent à des personnes participantes de se dépêcher d'aller dans le bâtiment voisin pour compléter une étude. Elles disent aux autres personnes participantes qu'elles peuvent prendre leur temps. On observe ensuite qui s'arrêtent pour aider la personne prétendant être blessée.
- **Pratique** : Pour chacune des relations statistiques suivantes, décidez si le problème de directionnalité est présent et pensez à au moins une troisième variable plausible.
 - Les personnes consommant davantage de homard ont tendance à vivre plus longtemps.
 - Les personnes faisant plus d'exercice ont tendance à peser moins.
 - Les personnes aux études dans un collège et qui boivent plus d'alcool ont tendance à avoir de moins bonnes notes.
- **Pratique** : Élaborez une matrice de corrélation pour une étude hypothétique comprenant les variables de la dépression, de l'anxiété, de l'estime de soi et du bonheur. Incluez les valeurs r de Pearson auxquelles vous vous attendriez.
- **Discussion** : Imaginez une étude corrélationnelle portant sur l'intelligence, le besoin de cognition et les performances des élèves du secondaire dans le cadre d'un cours sur la pensée critique. Une analyse de régression multiple montre que l'intelligence n'est pas liée à la performance dans la classe, mais que le besoin de cognition l'est. Expliquez ce que cette étude a montré en ce qui concerne ce qui est lié à une bonne performance dans le cours de pensée critique.
- **Discussion** : Quelles sont les différences entre une étude qualitative sur les jeunes filles qui jouent au baseball et une étude quantitative sur le même sujet? En quoi les données seraient-elles différentes si l'on interrogeait les filles une à une plutôt que de mener des groupes de discussion ou des enquêtes?
- **Pratique** : Trouvez et lisez une étude de cas publiée en psychologie. (Utilisez "case study" comme terme clé dans une recherche PsycINFO, sélectionnez la langue de votre choix.) Ensuite, faites ce qui suit :
 - Décrivez un problème lié à la validité interne.
 - Décrivez un problème lié à la validité externe.
 - Générez une hypothèse suggérée par l'étude de cas qui pourrait être intéressante à tester dans une étude ultérieure.

39. Références

- Abrams, L. S. et Curran, L. (2009). "And you're telling me not to stress?" A grounded theory study of postpartum depression symptoms among low-income mothers (« Et vous me dites de ne pas stresser ? » Une étude théorique sur les symptômes de la dépression post-partum chez les mères à faible revenu). *Psychology of Women Quarterly*, 33(3), 351–362. <https://doi.org/10.1111/pwqu.2008.33.issue-310.1111/j.1471-6402.2009.01506.x>
- Bryman, A. (2012). *Social Research Methods*, 4th ed. Oxford: OUP.
- Bell, E., Bryman, A. et Kleinknecht, S. (2022). *Social Research Methods (Les méthodes en recherche sociale)* (6e éd. canadienne). Oxford University Press.
- Bushman, B. J. et Huesmann, L. R. (2001). Effects of televised violence on aggression (Effets de la violence télévisée sur l'agressivité). Dans D. Singer et J. Singer (dir.), *Handbook of children and the media* (p. 223–254). Sage.
- Cacioppo, J. T. et Petty, R. E. (1982). The need for cognition (Le besoin de cognition). *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cohen, D., Nisbett, R. E., Bowdle, B. F. et Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An "experimental ethnography" ('L'insulte, l'agression et la culture de l'honneur dans le Sud : Une « ethnographie expérimentale ».). *Journal of Personality and Social Psychology*, 70(5), 945–960. <https://doi.org/10.1037/0022-3514.70.5.945>
- Corbin, J. et Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (Les bases de la recherche qualitative : Techniques et procédures pour développer une théorie ancrée)* 4e éd. Sage Publications.
- Diener, E. (2000). Subjective well-being: The science of happiness, and a proposal for a national index (Le bien-être subjectif : La science du bonheur et une proposition d'indice national). *American Psychologist*, 55(1), 34–43. <https://doi.org/10.1037/0003-066X.55.1.34>
- Festinger, L., Riecken, H. et Schachter, S. (1956). *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world (Quand la prophétie échoue : Une étude sociale et psychologique d'un groupe moderne qui a prédit la destruction du monde.)*. University of Minnesota Press.
- Freud, S. (1966). *Cinq leçons sur la psychanalyse*. Payot.
- Glaser, B. G. et Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research. (La découverte de la théorie ancrée : Stratégies de recherche qualitative)*. Aldine.
- Geertz, C. (1973). *The interpretation of cultures (L'interprétation des cultures)*. Basic Books.
- Geertz, C. (2017). *The interpretation of cultures (L'interprétation des cultures)* 3e éd. Basic Books.

- Jouriles, E. N., Garrido, E., Rosenfield, D. et McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress (Expériences d'agression psychologique et physique dans les relations amoureuses des adolescents : Liens avec la détresse psychologique). *Child Abuse & Neglect*, 33(7), 451–460. <https://doi.org/10.1016/j.chiabu.2008.11.005>
- Kraut, R. E. et Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach (Messages sociaux et émotionnels du sourire : Une approche éthologique). *Journal of Personality and Social Psychology*, 37(9), 1539–1553. <https://doi.org/10.1037/0022-3514.37.9.1539>
- Levine, R. V. et Norenzayan, A. (1999). The pace of life in 31 countries (Le rythme de vie dans 31 pays.). *Journal of Cross-Cultural Psychology*, 30(2), 178–205. <https://doi.org/10.1177/0022022199030002003>
- Lindqvist, P., Johansson, L. et Karlsson, U. (2008). In the aftermath of teenage suicide: A qualitative study of the psychosocial consequences for the surviving family members (Après le suicide d'un adolescent : Une étude qualitative des conséquences psychosociales pour les membres survivants de la famille). *BMC Psychiatry*, 8(26), 1-7. <https://doi.org/10.1186/1471-244X-8-26>
- Loftus, E. F. et Pickrell, J. E. (1995). The formation of false memories (La formation de faux souvenirs). *Psychiatric Annals*, 25(12), 720–725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates (Consommation de chocolat, fonctions cognitives et lauréats du prix Nobel). *New England Journal of Medicine*, 367(16), 1562-1564. <https://doi.org/10.1056/NEJMon1211064>
- Milgram, S. (1963). Behavioral study of obedience (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67(4), 371–378. <https://doi.org/10.1037/h0040525>
- Milgram, S. (1974). Obedience to authority: An experimental view (Soumission à l'autorité : Un point de vue expérimental). Harper & Row.
- Milgram, S. (2017). Soumission à l'autorité: un point de vue expérimental (E. Molinié, Trad.; Nouvelle éd. augmentée). Pluriel.
- Pelham, B. W., Carvallo, M. et Jones, J. T. (2005). Implicit egotism (Égoïsme implicite). *Current Directions in Psychological Science*, 14(2), 106–110. <https://doi.org/10.1111/j.0963-7214.2005.00344.x>
- Peterson, C., Seligman, M. E. P., et Vaillant, G. E. (1988). Pessimistic explanatory style is a risk factor for physical illness: A thirty-five year longitudinal study (Le style explicatif pessimiste est un facteur de risque de maladie physique : Une étude longitudinale sur trente-cinq ans). *Journal of Personality and Social Psychology*, 55(1), 23–27. <https://doi.org/10.1037/0022-3514.55.1.23>
- Plomin, R., DeFries, J. C., McClearn, G. E. et McGuffin, P. (2008). *Behavioral genetics* (5e éd.). Worth.
- Radcliffe, N. M. et Klein, W. M. P. (2002). Dispositional, unrealistic, and comparative optimism: Differential relations with knowledge and processing of risk information and beliefs about personal risk (Optimisme dispositionnel, irréaliste et comparatif : Relations différentielles avec la connaissance et le traitement des

informations sur le risque et les croyances sur le risque personnel). *Personality and Social Psychology Bulletin*, 28(6), 836–846. <https://doi.org/10.1177/0146167202289012>

Rentfrow, P. J. et Gosling, S. D. (2008). The do re mi's of everyday life: The structure and personality correlates of music preferences (The do re mi's of everyday life : The structure and personality correlates of music preferences). *Journal of Personality and Social Psychology*, 84(6), 1236–1256. <https://doi.org/10.1037/0022-3514.84.6.1236>

Rosenhan, D. L. (1973). On being sane in insane places (Être sain d'esprit dans des lieux insensés). *Science*, 179(4070), 250–258. <https://doi.org/10.1126/science.179.4070.25>

Todd, Z., Nerlich, B., McKeown, S. et Clarke, D. D. (dir.) (2004) Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice. (Mélange de méthodes en psychologie : L'intégration des méthodes qualitatives et quantitatives dans la théorie et la pratique). Psychology Press.

Trenor, J.M., Yu, S.L., Waight, C.L., Zerda, K.S et Sha T.-L. (2008). The relations of ethnicity to female engineering students' educational experiences and college and career plans in an ethnically diverse learning environment (Les relations entre l'ethnicité et les expériences éducatives des étudiantes en ingénierie et leurs plans d'études et de carrière dans un environnement d'apprentissage ethniquement diversifié). *Journal of Engineering Education*, 97(4), 449–465. <https://doi.org/10.1002/j.2168-9830.2008.tb00992.x>

Watson, J. B. et Rayner, R. (1920). Conditioned emotional reactions (Réactions émotionnelles conditionnées). *Journal of Experimental Psychology*, 3(1), 1–14. <https://doi.org/10.1037/h0069608>

Wilkins, A. (2008). "Happier than Non-Christians": Collective emotions and symbolic boundaries among evangelical Christians (« Plus heureux que les non-chrétiens » : Émotions collectives et frontières symboliques chez les chrétiens évangéliques.). *Social Psychology Quarterly*, 71(3), 281–301. <https://doi.org/10.1177/019027250807100308>

CHAPITRE 7-LA RECHERCHE PAR ENQUÊTE



Image 7. Enquête par interview

Peu après les attentats terroristes de septembre 2001 à New York et Washington, DC, la chercheuse Jennifer Lerner et ses collègues mènent une enquête sur Internet auprès d'environ 2 000 personnes américaines âgées de 13 à 88 ans (Lerner, Gonzalez, Small et Fischhoff, 2003). Cette équipe de recherche interroge les personnes participantes sur leurs réactions aux attentats et sur leur appréciation de divers risques liés au terrorisme et autres. L'équipe constate notamment que les personnes ont tendance à surestimer la plupart des risques, que les femmes le font davantage que les hommes, et qu'il n'y a pas de différence entre les personnes adolescentes et les adultes. Cependant le résultat le plus intéressant est lié au fait que des personnes participantes ont commencé à ressentir de la colère lorsqu'on leur a demandé ce qui les mettait en colère à propos des attentats, et lorsqu'on leur a présenté une photographie et une capsule audio destinées à susciter la colère. D'autres viennent à ressentir de la peur lorsqu'on leur demande ce qui les amène à s'inquiéter à propos des attentats, et lorsqu'on leur présente une photographie et une capsule audio destinées à susciter la peur. Comme l'équipe de recherche l'avait prévu, les personnes participantes

amenées à ressentir de la colère, ont perçu moins de risques que celles amenées à ressentir de la peur — ce qui montre que la perception des risques est fortement liée à des émotions spécifiques.

L'étude de Lerner et de ses collègues est un exemple de recherche par enquête en psychologie — le sujet de ce chapitre. Nous commencerons par une présentation générale de la recherche par sondage avec sa définition, un peu d'histoire ainsi qu'un aperçu des personnes qui la mènent et des raisons qui les motivent. Nous examinerons ensuite la réponse à une enquête en tant que processus psychologique, et les implications de ce processus pour l'élaboration de bons questionnaires d'enquête. Enfin nous nous pencherons sur certaines questions liées à la réalisation d'une enquête, notamment l'échantillonnage des personnes participantes et la collecte des données.

40. Aperçu de la recherche par enquête

Objectifs d'apprentissage

1. Définir ce qu'est la recherche par enquête, y compris ses deux caractéristiques principales.
2. Décrire les différentes façons dont la recherche par enquête peut être utilisée et donner quelques exemples.

Qu'est-ce que la recherche par enquête ?

La **recherche par enquête** est une méthode quantitative et qualitative qui présente deux caractéristiques importantes. Premièrement, les variables d'intérêt sont mesurées à l'aide d'auto-évaluations (questionnaires ou entretiens). En substance, les scientifiques qui réalisent des enquêtes demandent à leurs personnes participantes (souvent appelés **répondantes** et **répondants** dans les études par enquête) de faire état directement de leurs propres pensées, sentiments et comportements. Deuxièmement, une attention particulière est accordée à la question de l'échantillonnage. En particulier, les personnes qui réalisent des enquêtes ont une préférence marquée pour les grands échantillons aléatoires, car ils fournissent les estimations les plus précises de ce qui est vrai dans la population. En fait la recherche par enquête est peut-être la seule approche en psychologie dans laquelle l'échantillonnage aléatoire est couramment utilisé. Au-delà de ces deux caractéristiques, presque tout est possible dans la recherche par sondage. Les enquêtes peuvent être longues ou courtes. Elles peuvent être menées en personne, par téléphone, par courrier ou par Internet. Elles peuvent porter sur les intentions de vote, les préférences en consommation, les attitudes sociales, la santé ou tout autre sujet sur lesquels il est possible d'interroger les gens et d'obtenir des réponses significatives. Bien que les données d'enquête soient souvent analysées à l'aide de statistiques, de nombreuses questions se prêtent à une analyse plus qualitative.

La plupart des enquêtes sont non expérimentales. Elles sont utilisées pour décrire des variables uniques (par exemple, le pourcentage de gens lors d'une élection qui préfèrent la candidature à la présidence d'une personne plutôt qu'une autre, la prévalence de la schizophrénie dans la population générale, etc.), et pour évaluer les relations statistiques entre les variables (par exemple la relation entre le revenu et la santé). Mais les enquêtes peuvent également être utilisées dans le cadre de la recherche expérimentale. L'étude de Lerner et de ses collègues en est un bon exemple. L'utilisation de mesures d'auto-évaluation et d'un vaste échantillon national permet d'identifier leur travail comme une recherche par sondage. Mais la manipulation d'une variable indépendante (colère ou peur) pour évaluer son effet sur une variable dépendante (jugement du risque) permet également de qualifier leur travail d'expérimental.

Histoire et utilisation de la recherche par sondage

La recherche par enquête trouve ses racines dans les « enquêtes sociales » anglaises et américaines menées au tournant du 20^e siècle par des scientifiques, et des personnes qui préparent des réformes et veulent documenter l'étendue de problèmes sociaux tels que la pauvreté (Converse, 1987). Dans les années 1930, le gouvernement américain commence à mener des enquêtes pour documenter les problèmes sociaux tels que la pauvreté, les conditions économiques et sociales du pays. La nécessité de tirer des conclusions sur l'ensemble de la population contribue à faire progresser les procédures d'échantillonnage. À peu près à la même époque, plusieurs scientifiques qui s'étaient déjà fait un nom dans le domaine des études de marché, en étudiant les préférences de consommation pour les entreprises américaines, s'intéressent aux sondages électoraux. L'élection présidentielle de 1936 entre Alf Landon et Franklin Roosevelt marque un tournant.



Lecture sur version
imprimée? Scannez ce code
QR avec votre téléphone
cellulaire ou se rendre sur le
site
<https://ces-eec.arts.ubc.ca/>

Un magazine appelé *Literary Digest* mène une enquête où on envoie des bulletins de vote (qui sont aussi des demandes d'abonnement) à des millions de personnes américaines. Sur la base de ce « sondage d'opinion », la rédaction prédit que Landon l'emporterait haut la main. Au même moment, les nouvelles enquêtes utilisent des méthodes scientifiques avec des échantillons beaucoup plus petits pour prédire le contraire — à savoir que Roosevelt l'emporterait haut la main. En fait l'un d'entre eux George Gallup, critique publiquement les méthodes du *Literary Digest* avant l'élection et garantit pratiquement que sa prédiction est correcte. Et bien sûr ce fut le cas, démontrant l'efficacité d'une méthodologie d'enquête minutieuse (nous examinerons les raisons pour lesquelles Gallup avait raison plus loin dans ce chapitre). L'intérêt pour les enquêtes en période électorale donne lieu à plusieurs projets à long terme, notamment l'étude canadienne électorale, qui mesure les intentions de vote des personnes vivant au Canada à

l'occasion des élections fédérales depuis 1965. Tout le monde peut accéder aux données et lire les résultats des expériences menées dans le cadre de ces études: [Études Électorales Canadiennes de 2011 et 2015](#)

Depuis les études de marché et les sondages électoraux, la recherche par enquête fait son chemin dans plusieurs domaines académiques, notamment les sciences politiques, la sociologie et la santé publique — où elle continue d'être l'une des principales approches pour collecter de nouvelles données. À partir des années 1930, les psychologues réalisent d'importantes avancées dans la conception des questionnaires, y compris des techniques encore utilisées aujourd'hui, telles que l'échelle de Likert. (Voir « Qu'est-ce qu'une échelle de Likert ? » dans la section 7.2 « Construire des questionnaires d'enquête »). La recherche par enquête a un lien historique fort avec l'étude psychologique sociale des attitudes, des stéréotypes et des préjugés. Les premières personnes menant des recherches sur les attitudes sont également parmi les premières en psychologie à rechercher des échantillons plus larges et plus diversifiés que les échantillons de commodité faciles d'accès comme la population universitaire couramment utilisée en psychologie (et qui l'est toujours).

La recherche par enquête continue d'être importante en psychologie contemporaine. Par exemple les données d'enquête permettent d'estimer la prévalence de divers troubles mentaux et d'identifier les relations statistiques entre ces troubles et divers autres facteurs. La National Comorbidity Survey est une enquête à grande échelle sur la santé mentale menée aux États-Unis (voir <http://www.hcp.med.harvard.edu/ncs>). Dans le cadre d'une partie de cette enquête, près de 10 000 adultes ont reçu une entrevue structurée sur la santé mentale à leur domicile en 2002 et 2003. Le Tableau 7.1 présente les résultats de la prévalence au cours de

la vie de certains troubles de l'anxiété, de l'humeur et de la toxicomanie. (La prévalence au cours de la vie est le pourcentage de la population qui développe le problème à un moment quelconque de sa vie). Il est évident que ce type d'information peut être très utile tant pour les personnes menant des recherches en science fondamentale, qui veulent comprendre les causes et les corrélats des troubles mentaux, que pour les psychologues en clinique et les responsables politiques qui ont besoin de comprendre à quel point ces troubles sont fréquents.

Tableau 7.1 De prévalences au cours de la vie d'après l'étude de comorbidité nationale

Prévalence à vie de troubles*			
	Total	Femme	Homme
Trouble d'anxiété généralisée	5.7	7.1	4.2
Trouble obsessionnel-compulsif	2.3	3.1	1.6
Trouble de dépression caractérisée	16.9	20.2	13.2
Trouble bipolaire	4.4	4.5	4.3
Abus d'alcool	13.2	7.5	19.6
Abus de drogue	8.0	4.8	11.6
*La prévalence à vie d'un trouble est le pourcentage de personnes dans la population développant un trouble à n'importe quel moment dans leur vie.			

Et comme le montre clairement l'exemple présenté dans l'introduction, les enquêtes peuvent même être utilisées comme méthode de collecte de données dans le cadre de la recherche expérimentale afin de tester des hypothèses spécifiques sur les relations de cause à effet entre des variables. De telles études, lorsqu'elles sont menées sur des échantillons importants et diversifiés, peuvent compléter utilement les études en laboratoire menées sur la population universitaire. La recherche par enquête est donc une approche flexible qui peut être utilisée pour étudier une variété de questions de recherche fondamentale et appliquée.

41. Construire des enquêtes

Objectifs d'apprentissage

1. Décrire les processus cognitifs impliqués lorsque vous répondez à une question d'enquête.
2. Expliquer ce qu'est un effet de contexte et donner quelques exemples.
3. Créer un questionnaire d'enquête simple basé sur les principes d'une rédaction et d'une organisation efficaces des questions.

Le cœur de tout projet de recherche par enquête est l'enquête elle-même. Bien qu'il soit facile de penser à des questions intéressantes à poser aux gens, il n'est pas du tout facile de construire une bonne enquête. Le problème vient du fait que les réponses données par les gens peuvent être influencées de manière involontaire par la formulation des questions, l'ordre des questions, les options de réponse proposées et bien d'autres facteurs. Dans le meilleur des cas, ces influences ajoutent du bruit aux données. Dans le pire des cas, elles entraînent des biais systématiques et des résultats trompeurs. Dans cette section, nous examinons donc certains principes de construction des enquêtes afin de minimiser ces effets involontaires et de maximiser ainsi la fidélité et la validité des réponses des personnes interrogées.

Répondre aux enquêtes en tant que processus psychologique

Avant d'examiner les principes spécifiques de la construction d'une enquête, il est utile de considérer la réponse à une enquête comme un processus psychologique.

Un modèle cognitif

La Figure 7.1 présente un modèle des processus cognitifs mobilisés lorsqu'une personne répond à une question d'enquête (Sudman et al., 1996). Les personnes interrogées doivent interpréter la question, récupérer l'information pertinente en mémoire, former un jugement provisoire, convertir ce jugement en l'un des choix de réponse proposés (par exemple une note sur une échelle de 1 à 7) et enfin réviser leur réponse si nécessaire..

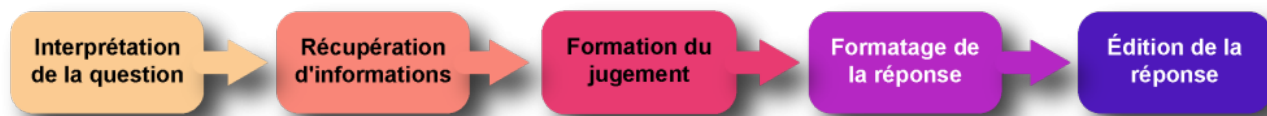


Figure 7.1. Modèle des processus cognitifs impliqués en répondant à un item dans une enquête. [\[Description de la figure\]](#)

Prenons par exemple l'item suivant du questionnaire :

Combien de boissons alcoolisées consommez-vous lors d'une journée typique ?

- _____ beaucoup plus que la moyenne
- _____ un peu plus que la moyenne
- _____ dans la moyenne
- _____ un peu moins que la moyenne
- _____ beaucoup moins que la moyenne

Bien que cet item semble à première vue simple, il pose plusieurs difficultés aux personnes répondantes. Tout d'abord elles doivent interpréter la question. Par exemple elles doivent décider si les « boissons alcoolisées » comprennent la bière et le vin (par opposition aux alcools forts) et si une « journée typique » réfère à un jour de semaine typique, un jour de fin de semaine typique ou les deux. Même si Chang et Krosnick (2003) constatent que les questions sur le comportement « typique » sont plus valides que les questions sur le comportement « passé », leur étude comprend la « semaine typique » à la « semaine passée » (et peut être différente si l'on considère les jours de semaine ou les jours de fin de semaine typiques). Une fois que les personnes interrogées ont interprété la question, elles doivent récupérer les informations pertinentes dans leur mémoire pour y répondre. Mais quelles informations doivent-elles récupérer et comment doivent-elles l'être ? Elles peuvent penser vaguement à quelques occasions récentes au cours desquelles elles ont bu de l'alcool, elles peuvent essayer soigneusement de se rappeler et de compter le nombre de boissons alcoolisées qu'elles ont consommées la semaine dernière, ou elles peuvent retrouver certaines croyances existantes qu'elles ont sur elles-mêmes (par exemple « je ne suis pas un grand buveur »). Elles doivent ensuite utiliser ces informations pour parvenir à un jugement provisoire sur le nombre de boissons alcoolisées qu'elles consomment au cours d'une journée typique. Par exemple ce calcul mental peut consister à diviser le nombre de boissons alcoolisées consommées la semaine dernière par sept pour obtenir un nombre moyen par jour. Elles doivent ensuite formater cette réponse provisoire en fonction des options de réponse effectivement proposées. Dans ce cas, les options posent des problèmes d'interprétation supplémentaires. Par exemple que signifie « dans la moyenne » et qu'est-ce qui est considéré comme « un peu plus » que la moyenne ? Enfin elles doivent décider si elles veulent rapporter la réponse qu'elles ont trouvée ou si elles veulent la modifier d'une manière ou d'une autre. Par exemple si elles pensent qu'elles boivent beaucoup plus que la moyenne, elles ne voudront peut-être pas l'indiquer de peur de mal paraître aux yeux des scientifiques, et choisiront donc plutôt l'option « un peu plus que la moyenne ».

De ce point de vue ce qui, à première vue, semble être une simple question consistant à demander aux personnes combien elles boivent (et à recevoir une réponse directe de leur part) s'avère être beaucoup plus complexe.

Effets du contexte sur les réponses à l'enquête

Là encore cette complexité peut entraîner des influences involontaires sur les réponses des personnes répondantes. Ces **effets** sont souvent appelés **effets de contexte**, car ils ne sont pas liés au contenu de la question, mais au contexte dans lequel la question apparaît (Schwarz et Strack, 1990). Par exemple il y a un **effet de l'ordre des items** lorsque l'ordre, dans lequel les items sont présentés, affecte les réponses des personnes. Un item peut modifier la façon dont les individus répondants interprètent un item ultérieur ou modifier les informations qu'ils récupèrent pour répondre aux items ultérieurs. Par exemple Fritz Strack et ses collègues ont interrogé la population estudiantine sur leur satisfaction générale à l'égard de la vie, et sur la fréquence de leurs fréquentations amoureuses (Strack, Martin et Schwarz, 1988). Lorsque la question portant sur la satisfaction générale à l'égard de la vie était posée en premier, la corrélation entre les deux variables était seulement de -0,12, ce qui suggère que les deux variables ne sont que faiblement liées. En revanche, lorsque la question relative à la fréquence de leurs fréquentations amoureuses est posée en premier, la corrélation entre les deux était de + 0,66, ce qui suggère que les personnes qui ont plus de fréquentations amoureuses ont une forte tendance à être plus satisfaites de leur vie. Le fait de répondre à la fréquence de leurs fréquentations amoureuses en premier rend cette information plus accessible en mémoire, de sorte qu'elles sont plus susceptibles d'évaluer la satisfaction de leur vie à partir de cette information.

Les options de réponse proposées peuvent également avoir des effets inattendus sur les réponses des personnes (Schwarz, 1999). Par exemple lorsqu'on demande aux gens à quelle fréquence ils sont « vraiment irrités » et qu'on leur propose des choix de réponses allant de « moins d'une fois par an » à « plus d'une fois par mois », ils ont tendance à penser à des irritations majeures et à se dire irrités rarement. En revanche, lorsque les choix de réponse vont de « moins d'une fois par jour » à « plusieurs fois par mois », ils ont tendance à penser à des irritations mineures et à se déclarer fréquemment irrités. Les gens ont également tendance à supposer que les choix de réponses intermédiaires représentent ce qui est normal ou typique. Par conséquent s'ils se considèrent comme normaux ou typiques, ils ont tendance à choisir les choix de réponse intermédiaires. Par exemple les gens sont susceptibles de déclarer qu'ils regardent davantage la télévision lorsque les choix de réponse sont centrés sur une option médiane de 4 heures, que lorsqu'ils sont centrés sur une option médiane de 2 heures. Pour atténuer les effets de l'ordre, alternez les questions et les éléments de réponse lorsqu'il n'y a pas d'ordre naturel. Le contrebalancement ou la randomisation de l'ordre de présentation des questions dans les enquêtes en ligne sont de bonnes pratiques pour les questions d'enquête et peuvent réduire les effets de l'ordre de réponse qui montrent que, parmi les personnes indécises lors d'une élection, la personne candidate à une élection, qui est listée en premier dans un bulletin de vote, bénéficie d'une augmentation de 2,5 % simplement parce qu'elle est listée en premier!

Écrire des items pour l'enquête

Types d'items

Les items peuvent être des questions ouvertes ou fermées. Les **questions ouvertes** posent simplement une

question et permettent aux personnes participantes de répondre comme elles l'entendent. Voici quelques exemples de questions ouvertes.

- « Quelle est la chose la plus importante à enseigner aux enfants pour les préparer à la vie ? »
- « Veuillez décrire un moment où vous avez été victime de discrimination en raison de votre âge. »
- « Y a-t-il autre chose dont vous voudriez nous parler ? »

Les questions ouvertes sont utiles lorsque les personnes en recherche ne savent pas comment les individus vont répondre ou lorsqu'on veut éviter d'influencer leurs réponses. Les questions ouvertes sont de nature plus qualitative et sont donc généralement utilisées lorsque les scientifiques ont des questions de recherche plus vaguement définies — souvent dans les premières phases d'un projet de recherche. Les questions ouvertes sont relativement faciles à rédiger, car il n'y a pas de choix de réponse à prendre en compte. Cependant elles demandent plus de temps et d'efforts de la part des personnes participantes et sont plus difficiles à analyser pour la personne menant une recherche, car les réponses doivent être transcrites, codées et soumises à une forme d'analyse qualitative telle que l'analyse de contenu. Un autre inconvénient est que les personnes interrogées sont plus susceptibles de ne pas répondre aux questions ouvertes parce qu'il faut plus de temps pour y répondre. Il est préférable d'utiliser des questions ouvertes lorsque la réponse est incertaine ou lorsque les quantités peuvent facilement être converties en catégories plus tard lors de l'analyse.

Les **questions fermées** posent une question et fournissent une série de choix de réponses parmi lesquelles les personnes participantes peuvent choisir. La question sur l'alcool que nous venons de mentionner en est un exemple tout comme les suivantes :

Quel âge avez-vous ?

- _____ Moins de 18 ans
- _____ 18 à 34
- _____ 35 à 49 ans
- _____ 50 à 70
- _____ Plus de 70 ans

Sur une échelle de 0 (aucune douleur) à 10 (la pire douleur jamais ressentie), quelle est l'intensité de la douleur que vous ressentez en ce moment?

Au cours de votre vie d'adulte, avez-vous déjà été déprimé·e pendant une période de deux semaines ou plus?

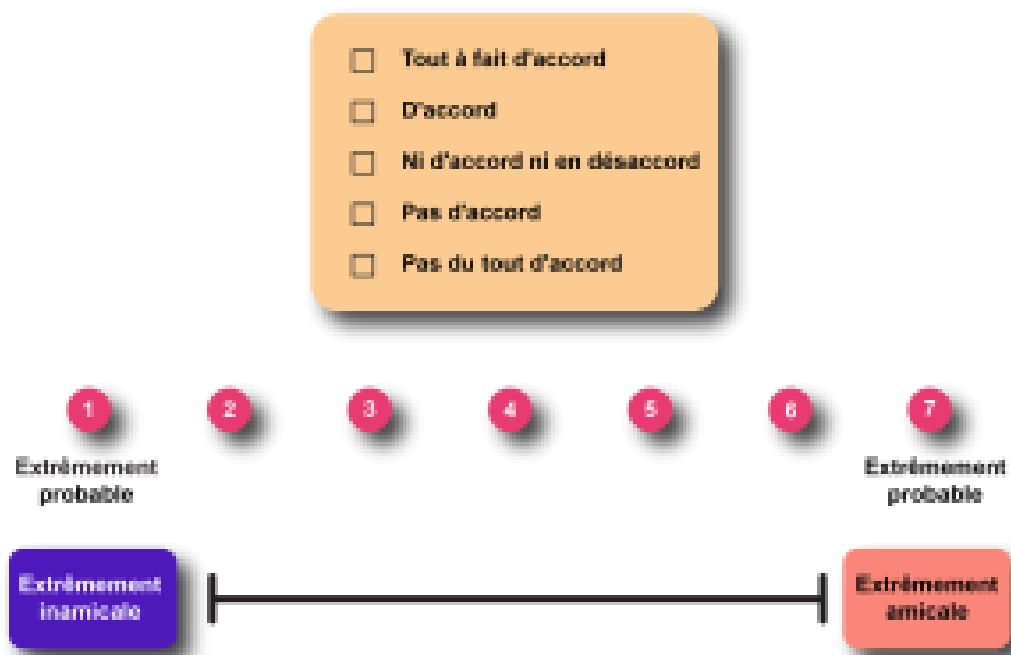
Oui Non

Les questions fermées sont utilisées lorsque les scientifiques ont une bonne idée des différentes réponses qui vont être données. De nature plus quantitative, elles sont également utilisées lorsque les personnes en recherche s'intéressent à une variable ou à un concept bien défini comme le degré d'accord des personnes avec une déclaration, la perception du risque ou la fréquence d'un comportement particulier. Les questions fermées sont plus difficiles à rédiger, car elles doivent inclure un ensemble approprié de choix de réponses. Cependant elles sont relativement rapides et faciles à remplir par les personnes répondantes. Elles sont

également beaucoup plus faciles à analyser pour les scientifiques, car les réponses peuvent être facilement converties en chiffres et saisies dans une feuille de calcul. Pour ces raisons, les questions fermées sont beaucoup plus courantes.

Toutes les questions fermées comportent une série de choix de réponses parmi lesquelles les personnes doivent choisir. Pour les variables catégorielles telles que le sexe, la race ou la préférence pour un parti politique, les catégories sont généralement énumérées et les personnes participantes choisissent celle(s) à laquelle (auxquelles) elles appartiennent. Pour les variables quantitatives, une échelle d'évaluation est généralement fournie. Une **échelle d'évaluation** est un ensemble ordonné de réponses parmi lesquelles les personnes participantes doivent choisir. La Figure 7.2 en donne plusieurs exemples. Le nombre de choix de réponses sur une échelle d'évaluation typique varie de trois à onze, bien que cinq et sept soient probablement les plus courants. Les échelles à cinq points conviennent mieux aux échelles unipolaires où un seul concept est testé, comme la fréquence (jamais, rarement, parfois, souvent, toujours). Les échelles à sept points conviennent mieux aux échelles bipolaires qui comportent un spectre dichotomique, comme l'appréciation (aime beaucoup, aime assez, aime légèrement, ni aime ni n'aime pas, n'aime pas légèrement, n'aime pas assez, n'aime pas du tout). Pour les questions bipolaires, il est utile de proposer une question préalable qui les oriente vers un domaine de l'échelle ; si l'on demande si l'on aime les glaces, il faut d'abord demander « Aimez-vous ou n'aimez-vous pas les glaces de manière générale ? » Une fois que la personne interrogée choisit d'aimer ou de ne pas aimer, précisez-la en lui proposant des choix sur l'échelle à sept points. La ramification améliore à la fois la fidélité et la validité (Krosnick et Berent, 1993). Bien que vous voyiez souvent des échelles avec des étiquettes numériques, il est préférable de ne présenter que des étiquettes verbales aux personnes interrogées, mais de les convertir en valeurs numériques dans les analyses. Évitez les étiquettes partielles, les étiquettes longues ou trop spécifiques. Dans certains cas, les étiquettes verbales peuvent être complétées (ou même remplacées) par des graphiques significatifs. La dernière échelle d'évaluation présentée à la Figure 7.2 est une échelle visuelle analogue, sur laquelle les personnes participantes font une marque quelque part le long de la ligne horizontale pour indiquer la magnitude de leur réponse.

Figure 7.2. Exemples d'échelles d'évaluation pour les questions fermées.
[\[Description de la figure\]](#)



Qu'est-ce qu'une échelle de Likert ?

En lisant des articles sur la recherche en psychologie, vous rencontrerez probablement le terme d'*échelle de Likert*. Bien que ce terme soit parfois utilisé pour désigner presque n'importe quelle échelle d'évaluation (par exemple une échelle de satisfaction de la vie de 0 à 10), ce terme a une signification beaucoup plus précise. Dans les années 1930, le chercheur Rensis Likert (prononcé LICK-ert) crée une nouvelle approche pour mesurer les attitudes des gens (Likert, 1932). Il s'agit de présenter aux gens plusieurs affirmations — y compris des affirmations favorables et défavorables — sur une personne, un groupe ou une idée. Les personnes interrogées expriment ensuite leur accord ou leur désaccord avec chaque affirmation sur une échelle de 5 points : *Tout à fait d'accord*, *d'accord*, *ni d'accord ni en désaccord*, *Pas d'accord*, *Pas du tout d'accord*. Des nombres sont attribués à chaque réponse, puis additionnés pour toutes les questions afin d'obtenir un score représentant l'attitude à l'égard de la personne, du groupe ou de l'idée. Pour les questions formulées en sens inverse (par exemple les questions négatives au lieu des questions positives), un codage inversé est utilisé de manière à ce que la notation numérique des questions aille également dans le sens inverse. L'ensemble des éléments est appelé « échelle de Likert ».

Par conséquent à moins que vous ne mesuriez l'attitude des gens à l'égard de quelque chose en évaluant leur degré d'accord avec plusieurs affirmations à ce sujet, il est préférable d'éviter de parler d'échelle de Likert. Vous utilisez probablement une « échelle d'évaluation ».

Rédiger des items efficaces

Nous pouvons maintenant examiner certains principes de rédaction des items du questionnaire qui minimisent les effets de contexte involontaires, et maximisent la fidélité et la validité des réponses des personnes participantes. Le modèle BRUSO (Peterson, 2000) fournit une ligne directrice approximative pour la rédaction des items du questionnaire. **BRUSO** est un acronyme qui (en anglais) signifie « bref », « pertinent », « non ambigu », « spécifique » et « objectif ». Les items d'un questionnaire efficace sont *brefs* et vont droit au but. Ils évitent les mots longs, trop techniques ou inutiles. Cette brièveté permet aux personnes répondantes de les comprendre plus facilement et de les remplir plus rapidement. Les items d'un questionnaire efficace sont également pertinents (en anglais *relevant*) par rapport à la question de recherche. Si l'orientation sexuelle, la situation matrimoniale ou les revenus d'une personne répondante ne sont pas pertinents, les items les concernant ne devraient probablement pas être inclus. Là encore le questionnaire est plus rapide à remplir mais cela évite d'ennuyer les personnes répondantes avec ce qu'elles percevront, à juste titre, comme des questions non pertinentes, voire « indiscrètes ». Les items efficaces sont également sans ambiguïté (*unambiguous* en anglais); ils ne peuvent être interprétés que d'une seule manière. Le problème de l'item sur l'alcool présenté plus haut dans cette section réside en partie dans le fait que différentes personnes répondantes peuvent avoir des idées différentes sur ce qui constitue « une boisson alcoolisée » ou « une journée typique ». Les items d'un questionnaire efficace sont également *spécifiques*, de sorte que les personnes répondantes savent clairement sur quoi doit porter leur réponse et que les scientifiques savent clairement de quoi il s'agit. Un problème courant est celui des questions fermées qui sont « à double entrée ». Elles posent des questions sur deux sujets conceptuellement distincts, mais n'autorisent qu'une seule réponse. Par exemple « Veuillez évaluer dans quelle mesure vous vous êtes senti-e anxieux/se et déprimé-e ». Cet item devrait probablement être divisé en deux items distincts, l'un sur l'anxiété et l'autre sur la dépression. Enfin les items d'un questionnaire efficace sont *objectifs* en ce sens qu'ils ne révèlent pas les propres opinions de la personne menant la recherche ou amènent les personnes participantes à répondre d'une manière particulière. Le Tableau 7.2 présente quelques exemples d'items de questionnaire pauvres et efficaces sur la base des critères de BRUSO. La meilleure façon de savoir comment les gens interprètent la formulation de l'item est de réaliser un test pilote et de demander à quelques personnes d'expliquer comment elles interprètent l'item.

Tableau 7.2 Modèle BRUSO de rédaction des items d'un questionnaire efficace, plus des exemples

Critère	Pauvre	Efficace
B—Brief	«Êtes-vous actuellement ou avez-vous déjà été en possession d'une arme à feu?»	«Avez-vous déjà possédé une arme à feu?»
R—Relevant	«Quelle est votre orientation sexuelle?»	N'incluez pas cet item à moins qu'il ne soit clairement pertinent pour la recherche.
U—Unambiguous	«Êtes-vous un-e adepte des armes à feu?»	«Possédez-vous actuellement une arme à feu?»
S—Specific	«Qu'avez-vous lu au sujet de la nouvelle mesure de contrôle des armes à feu et de la taxe sur les ventes?»	«Qu'avez-vous lu sur la nouvelle taxe sur les ventes?»
O—Objective	«Qu'avez-vous lu au sujet de la nouvelle mesure de contrôle des armes à feu et de la taxe sur les ventes?»	«Que pensez-vous de la nouvelle mesure de contrôle des armes à feu?»

Pour les questions fermées, il est également important de créer une échelle de réponse appropriée. Pour les variables catégorielles, les catégories présentées doivent généralement être mutuellement exclusives et exhaustives. Les catégories mutuellement exclusives ne se chevauchent pas. Pour une question relative à la religion, par exemple, les catégories *chrétien* et *catholique* ne s'excluent pas mutuellement, mais les catégories *protestant* et *catholique* s'excluent mutuellement. Les catégories exhaustives couvrent toutes les réponses possibles. Bien que les catégories protestante et catholique s'excluent mutuellement, elles ne sont pas exhaustives, car il existe de nombreuses autres catégories religieuses qu'une personne interrogée pourrait sélectionner : *juif*, *hindou*, *bouddhiste*, etc. Dans de nombreux cas, il n'est pas possible d'inclure toutes les catégories possibles, auquel cas une catégorie *autre*, avec un espace permettant à la personne répondante d'indiquer une réponse plus spécifique, est une bonne solution. Si les personnes répondantes peuvent appartenir à plus d'une catégorie (par exemple la race), elles doivent être invitées à choisir toutes les catégories qui s'appliquent.

Pour les échelles d'évaluation, cinq ou sept choix de réponses permettent généralement autant de précision que les personnes répondantes sont capables. Cependant des échelles numériques avec plus de choix peuvent parfois être appropriées. Pour des dimensions telles que l'attrait, la douleur et la probabilité, une échelle de 0 à 10 sera familière à de nombreuses personnes interrogées et facile à utiliser. Quel que soit le nombre de choix de réponses, les plus extrêmes doivent généralement être « équilibrées » autour d'un point médian neutre ou modal. Un exemple d'échelle d'évaluation déséquilibrée mesurant la probabilité perçue pourrait ressembler à ceci :

Peu probable | Un peu probable | Probable | Très probable | Extrêmement probable

Une version équilibrée pourrait ressembler à ceci :

Extrêmement improbable | Assez improbable | Aussi probable que non | Assez probable | Extrêmement probable

Notons toutefois qu'il n'est pas nécessaire d'inclure un choix de réponse intermédiaire ou neutre. Les personnes en recherche choisissent parfois de ne pas l'inclure parce qu'elles veulent encourager les personnes interrogées à réfléchir plus profondément à leur réponse, et de ne pas simplement choisir l'option du milieu par défaut. Cependant l'inclusion d'alternatives intermédiaires sur les dimensions bipolaires peut être utilisée pour permettre aux gens de choisir une option qui n'est ni l'une ni l'autre.

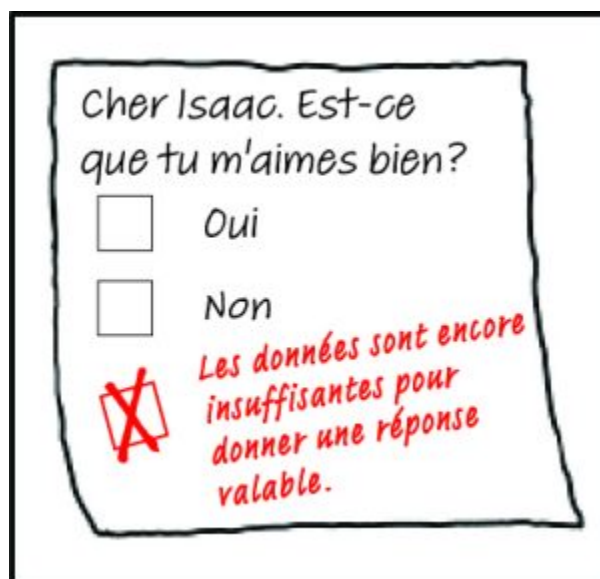


Figure 7.3. Question (traduit en français), extrait de <http://imgs.xkcd.com/comics/question.png> (CC-BY-NC 2.5). [\[Description de la figure\]](#)

Formatage de l'enquête

La rédaction d'items efficaces n'est qu'une partie de la construction d'une enquête. Chaque enquête doit comporter une introduction écrite ou orale qui remplit deux fonctions essentielles (Peterson, 2000). La première consiste à encourager les personnes interrogées à participer à l'enquête. Dans de nombreux types de recherche, un tel encouragement n'est pas nécessaire, soit parce que ces personnes ne savent pas qu'elles participent à une étude (comme dans le cas de l'observation naturaliste), soit parce qu'elles font déjà partie d'une liste de personnes intéressées à participer à des études (*subject pool* en anglais) et qu'ils ont déjà montré leur volonté de participer en s'inscrivant à l'étude et en s'y présentant. Les enquêtes surprennent généralement les personnes interrogées lorsqu'elles répondent au téléphone, consultent leur boîte aux lettres ou leur courrier électronique — et la personne menant l'enquête doit leur expliquer pourquoi elles devraient accepter de participer. L'introduction doit donc expliquer brièvement l'objectif de l'enquête et son importance, fournir des informations sur le commanditaire de l'enquête (les enquêtes menées par des universités ont tendance à générer des taux de réponse plus élevés), reconnaître l'importance de la participation de la personne interrogée et décrire les éventuelles incitations à participer.

La deuxième fonction de l'introduction est d'établir le consentement éclairé. Rappelez-vous que cela implique de décrire aux personnes répondantes tout ce qui pourrait affecter leur décision de participer. Il s'agit notamment des sujets couverts par l'enquête, du temps qu'elle prendra probablement, de la possibilité pour les personnes répondantes de se retirer à tout moment, des questions de confidentialité, etc. Les formulaires de consentement écrits ne sont pas toujours utilisés dans les enquêtes (lorsque l'enquête présente un risque minimal et que le fait de remplir l'instrument d'enquête est souvent accepté par le CÉR comme preuve du consentement à participer), il est donc important que cette partie de l'introduction soit bien documentée, présentée clairement et dans son intégralité à chaque personne interrogée.

L'introduction doit être suivie par les items cruciaux du questionnaire. Mais d'abord il est important de présenter des instructions claires pour remplir le questionnaire, y compris des exemples d'utilisation de toute échelle de réponse inhabituelle. N'oubliez pas que l'introduction est le moment où les personnes répondantes sont généralement le plus intéressé et le moins fatigué. Il est donc judicieux de commencer par les points les plus importants aux fins de la recherche pour ensuite passer aux points les moins importants. Les items doivent également être regroupés par thème ou par type. Par exemple les items utilisant la même échelle d'évaluation (comme une échelle d'accord en 5 points) doivent être regroupés si possible afin de rendre les choses plus rapides et plus faciles pour les personnes répondantes. Les questions démographiques sont souvent présentées en dernier parce qu'elles sont les moins intéressantes pour les personnes participantes, mais aussi parce qu'il est facile d'y répondre si les personnes répondantes sont fatiguées ou s'ennuient. Bien entendu, toute enquête doit se terminer par l'expression d'une appréciation à l'égard de la personne ayant répondu aux questions.

Descriptions de l'image

Figure 7.1 description de l'image

: Organigramme modélisant les processus cognitifs impliqués lors d'une réponse à un item de l'enquête. Dans l'ordre, ces processus sont les suivants:

- Interprétation des questions
- Récupération d'informations
- Formation du jugement
- Formatage des réponses
- Édition des réponses

[\[Retour à la Figure 7.1\]](#)

Figure 7.2 description de l'image : Trois échelles d'évaluation différentes pour des items d'enquête. La première échelle permet de choisir entre « tout à fait d'accord », « d'accord », « ni d'accord ni en désaccord », « en désaccord » et « tout à fait en désaccord ». La seconde est une échelle de 1 à 7, 1 correspondant à « extrêmement improbable » et 7 à « extrêmement probable ». La troisième est une échelle mobile, dont l'une des extrémités est « extrêmement inamicale » et l'autre « extrêmement amicale ». [\[Retour à la Figure 7.2\]](#)

Figure 7.3 description de l'image : Une note indique : « Cher Isaac. Est-ce que tu m'aimes bien ? » avec deux cases à cocher « oui » ou « non ». Quelqu'un a ajouté une troisième case, qu'il a cochée, et qui se lit comme suit : "Les données sont encore insuffisantes pour donner une réponse valable."

[\[Retour à la Figure 7.3\]](#)

Attributions des médias

- [Question](#) par XKCD sous licence [CC BY-NC 2.5](#)

42. Réaliser des enquêtes

Objectifs d'apprentissage

1. Expliquer la différence entre l'échantillonnage probabiliste et l'échantillonnage non probabiliste, et décrire les principaux types d'échantillonnage probabiliste.
2. Définir le biais d'échantillonnage en général et le biais de non-réponse en particulier. Énumérer quelques techniques qui peuvent être utilisées pour augmenter le taux de réponse et réduire le biais de non-réponse.
3. Citer les quatre principales façons de mener une enquête, en indiquant les avantages et les inconvénients de chacune d'entre elles.

Dans cette section, nous examinons comment réaliser une enquête. Nous examinerons d'abord la question de l'échantillonnage, puis les différentes méthodes de collecte des données d'une enquête.

Échantillonnage

Pour l'essentiel toute recherche en psychologie implique un échantillonnage, c'est-à-dire la sélection d'un échantillon à étudier au sein de la population concernée. L'échantillonnage se divise en deux grandes catégories. La première catégorie, l'**échantillonnage probabiliste**, se produit lorsque les scientifiques peuvent spécifier la probabilité que chaque membre de la population soit sélectionné pour l'échantillon. La seconde est l'échantillonnage **non probabiliste**, qui se produit lorsque les scientifiques ne peuvent pas spécifier ces probabilités. La plupart des recherches en psychologie font appel à l'échantillonnage non probabiliste. Par exemple l'**échantillonnage de commodité** — étudier des personnes qui se trouvent à proximité et qui sont disposées à participer — est une forme très courante d'échantillonnage non probabiliste utilisée dans la recherche en psychologie. Parmi les autres formes d'échantillonnage non probabiliste, citons l'**échantillonnage en boule de neige** (dans lequel les actuelles personnes participantes à la recherche aident à recruter de nouvelles personnes participantes pour l'étude), l'**échantillonnage par quotas** (dans lequel les sous-groupes de l'échantillon sont recrutés de manière à être proportionnels aux sous-groupes de la population) et l'**échantillonnage par autosélection** (dans lequel les personnes choisissent de participer à la recherche de leur propre chef, sans être approchées directement par les scientifiques).

Toutefois les personnes qui réalisent des enquêtes sont beaucoup plus susceptibles d'utiliser une forme ou une autre d'échantillonnage probabiliste. Cette tendance s'explique par le fait que l'objectif de la plupart des enquêtes est de faire des estimations précises sur ce qui est vrai dans une population donnée, et que ces estimations sont plus précises lorsqu'elles reposent sur un échantillon représentatif. Par exemple il est important pour les personnes en recherche de fonder leurs estimations des résultats électoraux — qui sont

souvent décidés par quelques points de pourcentage seulement — sur des échantillons représentatifs du corps électoral inscrit.

Par rapport à l'échantillonnage non probabiliste, l'échantillonnage probabiliste exige une spécification très claire de la population, qui dépend bien sûr des questions de recherche auxquelles il faut répondre. La population peut être constituée du corps électoral inscrit en Colombie-Britannique, de toutes personnes canadiennes ayant acheté une voiture au cours de l'année écoulée, des femmes du Québec âgées de plus de 40 ans qui ont subi une mammographie au cours des dix dernières années, ou toutes les anciennes cohortes étudiantes d'une université donnée. Une fois la population spécifiée, l'échantillonnage probabiliste nécessite un **cadre de l'échantillonnage**. Ce cadre d'échantillonnage est essentiellement une liste de tous les membres de la population à partir de laquelle les personnes interrogées sont sélectionnées. Les cadres d'échantillonnage peuvent provenir de diverses sources, notamment d'annuaires téléphoniques, de listes du corps électoral inscrit et de dossiers d'hôpitaux ou d'assurances. Dans certains cas, une carte peut servir de cadre d'échantillonnage, permettant de sélectionner des villes, des rues ou des ménages.

Il existe différentes méthodes d'échantillonnage probabiliste. L'**échantillonnage aléatoire simple** est réalisé de manière à ce que chaque personne de la population ait une probabilité égale d'être sélectionnée pour l'échantillon. Ce type d'échantillonnage peut consister à mettre les noms de toutes les personnes de la base de sondage dans un chapeau, à les mélanger et à en tirer le nombre nécessaire pour l'échantillon. Étant donné que la plupart des bases d'échantillonnage prennent la forme de fichiers informatiques, l'échantillonnage aléatoire est plus susceptible d'impliquer un tri ou une sélection informatisée des personnes répondantes. Une approche courante dans les enquêtes téléphoniques est la composition aléatoire, dans laquelle un ordinateur génère aléatoirement des numéros de téléphone parmi les numéros de téléphone possibles dans une zone géographique donnée.

Une alternative courante à l'échantillonnage aléatoire simple est l'**échantillonnage stratifié**, dans lequel la population est divisée en différents sous-groupes ou « strates » (généralement sur la base de caractéristiques démographiques), puis un échantillon aléatoire est prélevé dans chaque « strate ». L'**échantillonnage stratifié proportionnel** peut être utilisé pour sélectionner un échantillon dans lequel la proportion des personnes répondantes dans chacun des différents sous-groupes correspond à la proportion dans la population. Par exemple étant donné qu'environ 15,3 % de la population canadienne est asiatique, l'échantillonnage aléatoire stratifié peut être utilisé pour s'assurer qu'une enquête menée auprès de 1 000 adultes vivant au Canada comprend environ 153 personnes répondantes asiatiques. L'**échantillonnage stratifié disproportionné** peut également être utilisé pour échantillonner des personnes répondantes supplémentaires issues de sous-groupes particulièrement restreints, ce qui permet de tirer des conclusions valables sur ces sous-groupes. Par exemple comme les communautés noires au Canada représentent un pourcentage relativement faible de la population canadienne (environ 2,9 %), un simple échantillon aléatoire de 1 000 adultes canadiens pourrait inclure trop peu de Canadiens provenant des communautés noires pour que l'on puisse tirer des conclusions à leur sujet, en les distinguant de tout autre sous-groupe. Toutefois si la représentation est importante pour la question de recherche, un échantillonnage stratifié disproportionné pourrait être utilisé pour s'assurer que suffisamment de personnes répondantes canadiennes provenant des communautés noires sont incluses dans l'échantillon, pour tirer des conclusions valables sur l'ensemble des Canadiens provenant des communautés noires.

Un autre type d'échantillonnage probabiliste est l'**échantillonnage en grappes**, dans lequel de grandes grappes de personnes sont échantillonnées de manière aléatoire, puis des personnes au sein de chaque grappe sont échantillonnées de manière aléatoire. Il s'agit de la seule méthode d'échantillonnage

probabiliste qui ne nécessite pas de grille de l'échantillonnage. Par exemple pour sélectionner un échantillon de personnes habitant une petite ville de Saskatchewan, des scientifiques pourraient choisir au hasard plusieurs petites villes, puis sélectionner au hasard plusieurs personnes dans chaque ville. L'échantillonnage en grappes est particulièrement utile pour les enquêtes qui impliquent des entretiens en face-à-face, car il minimise les déplacements des scientifiques. Par exemple au lieu de se rendre dans 200 petites villes pour interroger 200 personnes y habitant, une équipe de recherche peut se rendre dans dix petites villes et interroger 20 personnes qui habitent dans chacune d'entre elles. L'enquête nationale sur la comorbidité (en anglais, *National Comorbidity Survey*) a été réalisée à l'aide d'une forme d'échantillonnage en grappes.

Quelle doit être la taille de l'échantillon d'une enquête ? En général cette estimation dépend de deux facteurs. Le premier est le niveau de confiance dans le résultat souhaité par les scientifiques. Plus l'échantillon est grand, plus toute statistique basée sur cet échantillon a tendance à être proche de la valeur correspondante dans la population. L'autre facteur est une contrainte pratique sous la forme du budget alloué à l'étude. Des échantillons plus grands offrent une plus grande confiance, mais il faut plus de temps, d'efforts et d'argent pour les obtenir. Compte tenu de ces deux facteurs, la plupart des enquêtes utilisent des échantillons dont la taille varie entre 100 et 1 000 personnes. La réalisation d'une analyse de puissance avant de débiter l'enquête aide les scientifiques à faire ce choix.

Taille de l'échantillon et taille de la population

Pourquoi un échantillon d'environ 1 000 personnes est-il considéré comme adéquat pour la plupart des enquêtes, même lorsque la population est beaucoup plus importante ? Considérons par exemple qu'un échantillon de seulement 1 000 personnes est généralement considéré comme un bon échantillon des quelque 25 millions d'adultes de la population canadienne, même s'il ne représente qu'environ 0,00004 % de la population ! La réponse est quelque peu surprenante.

Une partie de la réponse est qu'une statistique basée sur un échantillon plus large aura tendance à être plus proche de la valeur de la population et que cela peut être démontré mathématiquement. Imaginons par exemple que dans un échantillon de personnes inscrites à l'élection, exactement 50 % déclarent avoir l'intention de voter pour la candidate sortante. S'il y a 100 personnes dans cet échantillon, il y a 95 % de chance que le pourcentage réel dans la population se situe entre 40 et 60. Mais si l'échantillon compte 1 000 personnes, il y a 95 % de chance que le pourcentage réel dans la population soit compris entre 47 et 53.

Bien que cet « intervalle de confiance à 95 % » continue de diminuer à mesure que la taille de l'échantillon augmente, il le fait à un rythme plus lent. Par exemple dans le cas d'un échantillon de 2 000 personnes, cette réduction ne ramène l'intervalle de confiance à 95 % qu'à 48-52. Dans de nombreuses situations, la faible augmentation de l'intervalle de confiance au-delà d'un échantillon de 1 000 personnes n'est pas considérée comme valant la peine de consacrer du temps, des efforts et de l'argent supplémentaires.

Une autre partie de la réponse — et peut-être la plus surprenante — est que les intervalles de confiance ne dépendent que de la taille de l'échantillon et non de la taille de la population. Ainsi un échantillon de 1 000 personnes produira un intervalle de confiance à 95 % compris entre 47 et 53, que la taille de la population soit de 100 000 ou d'un million ou 100 millions.

Biais d'échantillonnage

L'échantillonnage probabiliste est développé en grande partie pour résoudre le problème du biais d'échantillonnage. Il y a **biais d'échantillonnage** lorsqu'un échantillon est sélectionné de telle sorte qu'il n'est pas représentatif de l'ensemble de la population et qu'il produit des résultats inexacts. C'est à cause de ce biais que le sondage de *Literary Digest straw* s'est avéré si erroné dans sa prédiction de l'élection présidentielle de 1936. Les listes d'adresses utilisées provenaient en grande partie d'annuaires téléphoniques et de listes de propriétaires d'automobiles enregistrés, qui surreprésentaient les personnes plus aisées, plus susceptibles de voter pour Landon. Gallup a réussi parce qu'il était conscient de ce biais et qu'il a trouvé des moyens d'échantillonner également des personnes moins riches.

Il existe une forme de biais d'échantillonnage à laquelle même un échantillonnage aléatoire minutieux est soumis. Il n'arrive presque jamais que toutes les personnes sélectionnées pour l'échantillon répondent effectivement à l'enquête. Certaines personnes peuvent être décédées ou avoir déménagé, d'autres peuvent refuser de participer parce qu'elles sont trop occupées ou ne sont pas intéressées par l'enquête ou ne participent pas aux enquêtes par principe. Si ces personnes non répondantes diffèrent systématiquement des personnes répondantes, cette différence peut entraîner un **biais de non-réponse**. Par exemple dans une enquête postale sur la consommation d'alcool, la chercheuse Vivienne Lahaut et ses collègues ont constaté qu'environ la moitié seulement de l'échantillon avait répondu après le premier contact et deux rappels de suivi (Lahaut, Jansen, van de Mheen et Garretsen, 2002). Le danger ici est que la moitié des gens ayant répondu pourrait avoir des habitudes de consommation d'alcool différentes de celles de la moitié n'ayant pas répondu, ce qui pourrait conduire à des conclusions inexacts de la part des scientifiques. Pour vérifier l'absence du biais de non-réponse, les scientifiques ont donc effectué des visites inopinées au domicile d'un sous-ensemble de personnes n'ayant pas répondu à l'enquête — revenant jusqu'à cinq fois si l'équipe de recherche ne les trouvait pas chez elles. On a constaté que les personnes non répondantes initiales comprenaient une proportion particulièrement élevée de personnes abstinentes (non-buveuses), ce qui signifiait que leurs estimations de la consommation d'alcool basées uniquement sur les personnes répondantes initiales étaient trop élevées.

Bien qu'il existe des méthodes pour corriger statistiquement le biais de non-réponse, elles sont basées sur des hypothèses concernant les personnes non répondantes — par exemple qu'elles sont plus semblables aux personnes donnant une réponse tardive qu'à celles donnant une réponse précoce — ce qui peut ne pas être correct. C'est pourquoi la meilleure approche pour minimiser le biais de non-réponse est de minimiser le nombre de personnes non répondantes — c'est-à-dire de maximiser le taux de réponse. Il existe de nombreux travaux de recherche sur les facteurs qui influencent les taux de réponse aux enquêtes (Groves, Fowler, Couper, Lepkowski, Singer et Tourangeau, 2004). En général, les entretiens en personne ont les taux de réponse les plus élevés, suivis par les enquêtes téléphoniques, puis par les enquêtes par courrier et par Internet. Parmi les autres facteurs qui augmentent les taux de réponse, citons l'envoi aux personnes répondantes potentielles d'un bref message de prénotification les informant qu'elles seront invitées à participer à une enquête dans un avenir proche, et l'envoi de simples rappels de suivi aux personnes non répondantes au bout de quelques semaines. La longueur et la complexité perçues de l'enquête peuvent également faire la différence, c'est pourquoi il est important que les questionnaires soient aussi courts, simples et pertinents que possible. Enfin l'offre d'une incitation — en particulier en espèces —

est un moyen fiable d'augmenter les taux de réponse. Cependant d'un point de vue éthique, il y a des limites à l'offre d'incitations qui peuvent être si importantes qu'elles sont considérées comme coercitives.

Réaliser l'enquête

Les quatre principaux moyens de réaliser des enquêtes sont les entretiens en personne, le téléphone, le courrier et Internet. Comme pour d'autres aspects de la conception d'une enquête, le choix dépend à la fois des objectifs des scientifiques et du budget. Les entretiens en personne ont les taux de réponse les plus élevés et permettent le contact personnel le plus étroit avec les personnes interrogées. Le contact personnel peut être important par exemple lorsque la personne menant l'enquête doit voir les personnes répondantes et porter un jugement sur celles-ci, comme c'est le cas dans certaines enquêtes sur la santé mentale. Mais l'entrevue en personne est de loin l'approche la plus coûteuse. Les enquêtes téléphoniques ont des taux de réponse plus faibles, mais permettent tout de même un certain contact personnel avec les personnes interrogées. Elles peuvent également être coûteuses, mais le sont généralement moins que les entretiens en personne. Traditionnellement les annuaires téléphoniques ont fourni des bases d'échantillonnage assez complètes. Toutefois cette tendance est moins vraie aujourd'hui, car de plus en plus de personnes choisissent de n'avoir que des téléphones cellulaires et n'installent pas de lignes terrestres qui seraient incluses dans les annuaires téléphoniques. Les enquêtes postales sont encore moins coûteuses, mais elles ont généralement des taux de réponse encore plus faibles — ce qui les rend très sensibles aux biais de non-réponse.

Fait peu surprenant, les enquêtes par Internet sont de plus en plus courantes. Leur construction et leur utilisation ne cessent de se simplifier (voir l'encadré « Création d'enquêtes en ligne »). Bien que le contact initial puisse être établi par courrier avec un lien vers l'enquête en ligne, cette approche ne produit pas nécessairement des taux de réponse plus élevés qu'une enquête par courrier ordinaire. Une meilleure approche consiste à établir un premier contact par courrier électronique avec un lien direct vers l'enquête. Cette approche peut s'avérer efficace lorsque la population est composée de membres d'une organisation qui disposent d'adresses électroniques connues et les utilisent régulièrement (par exemple une communauté universitaire). Pour d'autres populations, il peut être difficile, voire impossible de trouver une liste complète d'adresses électroniques pouvant servir de base d'échantillonnage. Une autre solution consiste à publier une demande de participation à l'enquête avec un lien vers celle-ci sur des sites Internet connus pour être visités par des membres de la population. Mais là encore il est très difficile d'obtenir un échantillon aléatoire de cette manière, car les membres de la population qui visitent les sites Internet sont susceptibles d'être différents de la population dans son ensemble. Néanmoins les méthodes d'enquête par Internet se développent rapidement. Étant donné leur faible coût et le fait qu'il n'y a jamais eu autant de personnes en ligne, les enquêtes par Internet sont susceptibles de devenir l'approche dominante en matière de collecte de données d'enquête dans un avenir proche.

Enfin il est important de noter que certaines des préoccupations que des personnes peuvent avoir au sujet de la collecte de données en ligne (par exemple le fait que les résultats obtenus sur Internet diffèrent de ceux obtenus avec d'autres méthodes) se sont révélées être des mythes. Le Tableau 7.3 (adapté de Gosling, Vazire, Srivastava et John, 2004) aborde trois de ces idées préconçues sur les données collectées dans le cadre d'études réalisées sur Internet :

Tableau 7.3 Quelques idées préconçues et conclusions à propos des études faites sur Internet

Préconception	Recherche
Les échantillons Internet ne sont pas démographiquement diversifiés.	Les échantillons Internet sont plus diversifiés que les échantillons traditionnels dans de nombreux domaines, bien qu'ils ne soient pas complètement représentatifs de la population.
Les échantillons Internet sont inadaptés, isolés socialement ou déprimés.	Les personnes qui utilisent Internet ne diffèrent pas des non-utilisateurs en ce qui concerne les marqueurs d'adaptation et de dépression.
Les résultats obtenus sur Internet diffèrent de ceux obtenus avec d'autres méthodes.	Les données recueillies jusqu'à présent indiquent que les résultats obtenus sur l'Internet sont cohérents avec les résultats obtenus à l'aide de méthodes traditionnelles (par exemple sur l'estime de soi, la personnalité), mais plus de données sont nécessaires.

Création d'enquêtes en ligne

Il existe aujourd'hui plusieurs outils en ligne permettant de créer des questionnaires en ligne. Une fois le questionnaire créé, un lien vers celui-ci peut être envoyé par courrier électronique aux personnes répondantes potentielles ou être intégré dans une page Internet. Les sites Internet suivants font partie de ceux qui proposent des comptes gratuits. Bien que les comptes gratuits limitent le nombre d'éléments du questionnaire et le nombre de personnes répondantes, ils peuvent être utiles pour réaliser des enquêtes à petite échelle, et pour mettre en pratique les principes d'une bonne construction de questionnaire.

Voici quelques outils d'enquête en ligne couramment utilisés :

- SurveyMonkey — <https://surveymonkey.com> version française <https://fr.surveymonkey.com/>
- PsyToolkit — <https://www.psychtoolkit.org/> (gratuit, non commercial, et permet de réaliser de nombreux paradigmes expérimentaux)
- Qualtrics — <https://www.qualtrics.com/> version française <https://www.qualtrics.com/fr/>
- PsycData — <https://www.psychdata.com/>

Petite mise en garde : même lorsque des personnes canadiennes mènent des recherches auprès de la population canadienne, si les données sont conservées sur des serveurs américains, elles peuvent être saisies en vertu du 'Patriot Act'. Pour éviter toute violation des droits, voici une liste de sites d'enquêtes en ligne hébergés au Canada :

- Lime Survey — <https://www.limesurvey.org> en français: <https://www.limesurvey.org/fr>
- Simple Survey — <http://www.simplesurvey.com/>

Il existe également des sites d'enquête hébergés dans d'autres pays que l'Amérique du Nord.

Un autre nouvel outil permettant de réaliser des enquêtes est Mechanical Turk (MTurk) créé par Amazon.com <https://www.mturk.com>. Développé à l'origine pour de simples tests d'utilisabilité, MTurk dispose d'une base de données de plus de 500 000 membres du personnel issus de plus de 190 pays. Vous pouvez mettre en place des tâches simples (par exemple différentes formulations de questions pour tester les items de votre enquête), définir des paramètres en fonction de votre base d'échantillonnage et déployer votre expérience à un coût très

faible. MTurk a été salué comme un moyen peu coûteux de recueillir des données de haute qualité (Buhrmester, Kwang et Gosling, 2011).

43. Points clés et exercices

Points clés

- La recherche par enquête se caractérise par l'utilisation de mesures d'auto-évaluation sur des échantillons soigneusement sélectionnés. Il s'agit d'une approche flexible qui peut être utilisée pour étudier une grande variété de questions de recherche fondamentale et appliquée.
- La recherche par enquête trouve son origine dans la recherche sociale appliquée, les études de marché et les sondages électoraux. Depuis, elle est devenue une approche importante dans de nombreuses disciplines universitaires, notamment les sciences politiques, la sociologie, la santé publique et bien sûr la psychologie.
- La recherche par enquête consiste à demander aux personnes interrogées de rendre compte de leurs propres pensées, sentiments et comportements.
- La plupart des enquêtes sont de nature non expérimentale (elles servent à décrire des variables ou à mesurer des relations statistiques entre des variables), mais elles peuvent également être utilisées pour mesurer des variables dépendantes dans le cadre de véritables expériences.
- Répondre à une question d'enquête est en soi un processus cognitif complexe qui implique l'interprétation de la question, la récupération d'informations, un jugement provisoire, l'expression de ce jugement dans le format de réponse requis et l'édition de la réponse.
- Les réponses aux enquêtes sont soumises à de nombreux effets de contexte dus à la formulation des questions, à l'ordre des items, aux choix des réponses et à d'autres facteurs. Les personnes menant des recherches doivent être sensibles à ces effets lorsqu'elles élaborent des enquêtes et interprètent leurs résultats.
- Les questions de l'enquête sont soit ouvertes, soit fermées. Les questions ouvertes posent simplement une question et permettent aux personnes interrogées de répondre comme elles le souhaitent. Les questions fermées posent une question et proposent plusieurs choix de réponses parmi lesquelles les personnes interrogées doivent choisir.
- Il convient d'utiliser des étiquettes verbales plutôt que des étiquettes numériques, même si les réponses peuvent être converties en données numériques dans les analyses.
- Selon le modèle BRUSO, les items du questionnaire doivent être brefs, pertinents, sans ambiguïté, spécifiques et objectifs.
- La recherche par sondage implique généralement un échantillonnage probabiliste, dans lequel chaque membre de la population a une probabilité connue d'être sélectionné pour faire partie de l'échantillon. Les

types d'échantillonnage probabiliste comprennent l'échantillonnage aléatoire simple, l'échantillonnage aléatoire stratifié et l'échantillonnage en grappes.

- Il y a biais d'échantillonnage lorsqu'un échantillon est sélectionné de telle manière qu'il n'est pas représentatif de la population et qu'il produit des résultats inexacts. La forme la plus répandue de biais d'échantillonnage est le biais de non-réponse, qui se produit lorsque les personnes qui ne répondent pas à l'enquête diffèrent de manière importante de celles qui y répondent. La meilleure façon de minimiser le biais de non-réponse est de maximiser le taux de réponse en prévenant les répondants, en leur envoyant des rappels, en élaborant des questionnaires courts et faciles à remplir et en offrant des incitations.
- Les enquêtes peuvent être menées en personne, par téléphone, par courrier ou sur Internet. Les enquêtes en personne ont les taux de réponse les plus élevés, mais sont les plus coûteuses. Les enquêtes par courrier et par Internet sont moins coûteuses, mais ont des taux de réponse beaucoup plus faibles. Les enquêtes par Internet sont susceptibles de devenir l'approche dominante en raison de leur faible coût.

Exercices

- **Discussion** : Pensez à une question à laquelle on pourrait essayer de répondre en utilisant la recherche par sondage pour chacune des professions suivantes:

- psychologue sociale
- personne menant une recherche en éducation
- personne menant une étude de marché pour une chaîne de supermarchés
- mairie d'une grande ville
- gestionnaire du corps policier à l'université

- **Discussion** : Rédigez une question d'enquête et décrivez brièvement la manière dont une personne pourrait répondre à cette question en vous basant sur le modèle cognitif de réponse à une enquête (ou choisissez n'importe quelle question de l'échelle d'estime de soi de Rosenberg à l'adresse).

Liens :

- [Echelle d'Estime de Soi de Rosenberg](#)
- [Echelle d'Estime de Soi \(ÉES-10\)](#)



*Lecture sur version imprimée?
Scannez ce code QR avec votre
téléphone cellulaire ou se
rendre sur le site
[https://www.irbms.com/
download/documents/
echelle-estime-de-soi-de-rose
nberg.pdf](https://www.irbms.com/download/documents/echelle-estime-de-soi-de-rosenberg.pdf)*

- **Pratique** : Rédigez des questions d'enquête pour chacune des questions générales suivantes. Dans certains cas, une série de questions, plutôt qu'une seule, peut s'avérer nécessaire.

- Dans quelle mesure les personnes répondantes utilisent-elles Facebook ?
- Combien d'exercice les personnes répondantes font-elles ?
- Quelle est la probabilité, selon le type de personnes répondantes, que le premier ministre ou la première ministre sortant.e soit réélue lors de la prochaine élection canadienne ?
- Dans quelle mesure les personnes répondantes éprouvent-elles de la « rage au volant » ?



*Lecture sur version imprimée?
Scannez ce code QR avec votre
téléphone cellulaire ou se
rendre sur le site
[https://www.lrcs.uqam.ca/
wp-content/uploads/2017/
08/ees.pdf](https://www.lrcs.uqam.ca/wp-content/uploads/2017/08/ees.pdf)*

- **Discussion** : Si possible, identifiez un cadre d'échantillonnage approprié pour chacune des populations suivantes. S'il n'y a pas de cadres d'échantillonnage appropriés, expliquez pourquoi.
 - les personnes étudiant dans une université donnée
 - les adultes vivant dans la province d'Alberta
 - les ménages (*households* en anglais) à Montréal, Québec
 - les personnes ayant une faible estime de soi
- **Pratique** : Utilisez un des outils de création d'enquêtes en ligne pour créer un questionnaire de dix questions sur un sujet de votre choix.

44. Références

- Buhrmester, M., Kwang, T. et Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? (Amazon's Mechanical Turk : une nouvelle source de données peu coûteuses mais de grande qualité ?) *Perspectives on Psychological Science*, 6(1), 3-5. <https://doi.org/10.1177/1745691610393980>
- Chang, L. et Krosnick, J.A. (2003). Measuring the frequency of regular behaviors: Comparing the 'typical week' to the 'past week' (Mesurer la fréquence des comportements réguliers : comparer "la semaine typique" et "la semaine dernière"). *Sociological Methodology*, 33(1), 55-80. <https://doi.org/10.1111/j.0081-1750.2003.t01-1-00127.x>
- Converse, J. M. (1987). *Survey research in the United States : Roots and emergence, 1890-1960 (Recherche par sondage aux États-Unis : racines et émergence, 1890-1960)*. University of California Press.
- Gosling, S. D., Vazire, S., Srivastava, S. et John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires (Faut-il faire confiance aux études basées sur le web ? Une analyse comparative de six idées reçues sur les questionnaires en ligne.). *American Psychologist*, 59(2), 93-104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. et Tourangeau, R. (2004). *Survey methodology*. Wiley.
- Krosnick, J.A. et Berent, M.K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format (Comparaisons de l'identification à un parti et des préférences politiques : L'impact du format des questions d'enquête). *American Journal of Political Science*, 27(3), 941-964. <https://doi.org/10.2307/2111580>
- Lahaut, V. M. H. C. J., Jansen, H. A. M., van de Mheen, D. et Garretsen, H. F. L. (2002). Non-response bias in a sample survey on alcohol consumption (Biais de non-réponse dans une enquête par sondage sur la consommation d'alcool). *Alcohol and Alcoholism*, 37(3), 256-260. <https://doi.org/10.1093/alcalc/37.3.256>
- Lerner, J. S., Gonzalez, R. M., Small, D. A. et Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment (Effets de la peur et de la colère sur la perception des risques de terrorisme : Une expérience nationale sur le terrain). *Psychological Science*, 14(2), 144-150.
- Likert, R. (1932). A technique for the measurement of attitudes (Une technique de mesure des attitudes). *Archives of Psychology*, 140, 1-55.
- Miller, J.M. et Krosnick, J.A. (1998). The impact of candidate name order on election outcomes (L'impact de l'ordre des noms des candidats sur les résultats des élections). *Public Opinion Quarterly*, 62(3), 291-330. <https://doi.org/10.1086/297848>
- Natala@aws. (2011, Janvier 26). Re: MTurk CENSUS: About how many workers were on Mechanical Turk

in 2010? Dans *Amazon Web Services Discussion Forums*.
<https://forums.aws.amazon.com/thread.jspa?threadID=58891>

Peterson, R. A. (2000). *Constructing effective questionnaires*. Sage.

Schwarz, N. (1999). Self-reports: How the questions shape the answers (Auto-rapports : comment les questions façonnent les réponses). *American Psychologist*, 54(2), 93–105.
<https://doi.org/10.1037/0003-066X.54.2.93>

Schwarz, N. et Strack, F. (1990). Context effects in attitude surveys: Applying cognitive theory to social research (Effets de contexte dans les sondages des attitudes : application de théorie cognitive à la recherche sociale). Dans W. Stroebe et M. Hewstone (dir.), *European review of social psychology* (Vol. 2, p. 31–50). Wiley.

Strack, F., Martin, L. L. et Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life satisfaction (L'amorçage et la communication : les déterminants sociaux de l'utilisation de l'information dans les jugements de satisfaction de la vie). *European Journal of Social Psychology*, 18(5), 429–442. <https://doi.org/10.1002/ejsp.2420180505>

Sudman, S., Bradburn, N. M. et Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology* (Réfléchir aux réponses : l'application des processus cognitifs aux méthodes des sondages). Jossey-Bass.

CHAPITRE 8-LA RECHERCHE QUASI-EXPÉRIMENTALE



Image 8. Intervention en matière d'éducation

Le préfixe *quasi* signifie « qui ressemble ». La recherche quasi-expérimentale est donc une recherche qui ressemble à une recherche expérimentale, mais qui n'est pas une véritable recherche expérimentale. Rappelons-nous que dans le cas d'une véritable expérience entre groupes, l'assignation aléatoire des conditions est utilisée pour garantir l'équivalence des groupes et que, dans le cas d'une véritable expérience intrasujets, le contrebalancement est utilisé pour se protéger contre les effets de l'ordre. Les quasi-expériences sont dépourvues de l'une de ces caractéristiques fondamentales. Bien qu'une variable indépendante soit manipulée, soit il manque un groupe contrôle ou soit les personnes participantes ne sont pas assignées de manière aléatoire aux conditions (Cook et Campbell, 1979).

Comme la variable indépendante est manipulée avant que la variable dépendante ne soit mesurée, la recherche quasi-expérimentale élimine le problème de directionnalité associé à la recherche non expérimentale. Mais comme les techniques de contrebalancement ne sont pas utilisées, ou que les personnes participantes ne sont pas assignées de façon aléatoire aux conditions — rendant probable

l'existence d'autres différences entre les conditions — la recherche quasi-expérimentale n'élimine pas le problème des variables confondantes. Sur le plan de la validité interne, les quasi-expériences se situent donc généralement entre les études non expérimentales et les véritables expériences.

Les quasi-expériences sont le plus souvent menées dans des contextes de terrain où l'assignation aléatoire est difficile, voire impossible. Elles sont souvent menées pour évaluer l'efficacité d'un traitement — peut-être un type de psychothérapie ou une intervention éducative. Il existe de nombreux types de quasi-expériences, mais nous n'aborderons que quelques-uns des plus courants dans ce chapitre.

45. Devis à un groupe

Objectifs d'apprentissage

1. Expliquer ce qu'est la recherche quasi-expérimentale et la distinguer clairement de la recherche expérimentale et corrélacionnelle.
2. Décrire trois types différents de devis quasi-expérimentaux à un seul groupe.
3. Identifier les menaces à la validité interne associées à chacun de ces devis.

Devis post-test à un seul groupe

Dans un **devis post-test à un seul groupe**, un traitement est mis en œuvre (ou une variable indépendante est manipulée) et une variable dépendante est mesurée une fois après la mise en œuvre du traitement. Imaginons par exemple des scientifiques qui s'intéressent à l'efficacité d'un programme d'éducation antidrogue par rapport aux attitudes des élèves d'une école primaire à l'égard des drogues illicites. Ils pourraient mettre en œuvre le programme antidrogue et, immédiatement après la fin du programme, mesurer l'attitude des élèves à l'égard des drogues illicites.

Il s'agit du type de devis quasi-expérimental le plus faible. L'absence d'un groupe de contrôle ou de comparaison constitue une limite majeure de ce devis. Il n'y a aucun moyen de déterminer quelles auraient été les attitudes de ces élèves s'il n'y avait pas eu de programme antidrogue. Malgré cette limite majeure, les résultats de cette étude sont fréquemment rapportés dans les médias, et sont souvent mal interprétés par la population générale. Par exemple les publicitaires peuvent affirmer que 80 % des femmes ont remarqué que leur peau était plus lumineuse après avoir utilisé le nettoyant de la marque X pendant un mois. S'il n'y a pas de groupe contrôle, cette statistique ne signifie pas grand-chose, voire rien du tout.

Devis prétest-post-test à un seul groupe

Dans un **devis prétest-post-test à un seul groupe**, la variable dépendante est mesurée une fois avant la mise en œuvre du traitement et une fois après. Reprenons l'exemple de scientifiques qui s'intéressent à l'efficacité d'un programme d'éducation antidrogue par rapport aux attitudes des élèves du primaire à l'égard des drogues illicites. Ces personnes pourraient mesurer les attitudes des élèves d'une école primaire particulière pendant une semaine, mettre en œuvre le programme antidrogue pendant la semaine suivante et enfin

mesurer à nouveau leurs attitudes la semaine suivante. Le devis prétest-post-test ressemble beaucoup à une expérience intrasujets dans laquelle chaque personne participante est testée d'abord dans la condition contrôle, puis dans la condition traitement. Il s'en distingue toutefois par le fait que l'ordre des conditions n'est pas contrebalancé parce qu'il n'est généralement pas possible de tester une personne participante d'abord dans la condition traitement, puis dans une condition contrôle « sans traitement ».

Si le score moyen au post-test est meilleur que le score moyen au prétest (par exemple les attitudes envers les drogues illégales sont plus négatives après le programme d'éducation antidrogue), il est logique de conclure que le traitement peut être responsable de l'amélioration. Malheureusement, il n'est souvent pas possible de tirer cette conclusion avec un degré élevé de certitude, car d'autres explications peuvent rendre compte du changement des scores au post-test. Ces explications alternatives menacent la validité interne.

Une explication alternative est celle de l'**histoire**. D'autres choses peuvent s'être produites entre le prétest et le post-test et avoir entraîné un changement entre le prétest et le post-test. Peut-être qu'un programme antidrogue a été diffusé à la télévision et que de nombreux élèves l'ont regardé, ou peut-être qu'une célébrité est morte d'un surdosage et que de nombreux élèves en ont entendu parler.

Une autre explication possible est celle de la **maturation**. Les personnes participantes ont pu changer entre le prétest et le post-test d'une manière qui était prévisible, parce qu'elles grandissent et apprennent. S'il s'agissait d'un programme antidrogue d'un an, les personnes participantes pourraient devenir moins impulsives ou mieux raisonner, ce qui pourrait expliquer leur changement d'attitude à l'égard des drogues illicites.

Une autre menace pour la validité interne des **devis prétest-post-test à un seul groupe** est le **test**, c'est-à-dire le fait que la mesure de la variable dépendante pendant le prétest affecte les réponses des personnes participantes au post-test. Par exemple le fait de remplir le questionnaire sur les attitudes à l'égard des drogues illicites peut avoir eu un effet sur ces attitudes. Le simple fait de remplir ce questionnaire peut avoir suscité des réflexions et des conversations sur les drogues illicites, ce qui a entraîné un changement dans les résultats au post-test.

De même l'**instrumentation** peut constituer une menace pour la validité interne des études utilisant ce devis. On parle d'instrumentation lorsque les caractéristiques de base de l'instrument de mesure changent au fil du temps. Lorsqu'il y a mesure des comportements à l'aide d'observation faite par des êtres humains, ces derniers peuvent avec le temps acquérir des compétences, se fatiguer ou modifier les normes sur lesquelles les observations sont basées. Ainsi les personnes participantes peuvent avoir pris très au sérieux la mesure des attitudes à l'égard des drogues illicites lors du prétest, lorsqu'elle était nouvelle, mais elles peuvent ensuite s'être devenues ennuyées avec la mesure lors du post-test, et avoir été moins prudentes dans l'examen de leurs réponses.

Une autre explication possible d'un changement de la variable dépendante dans un devis prétest-post-test est la **régression vers la moyenne**. Il s'agit du fait statistique qu'une personne qui obtient un score extrêmement élevé ou extrêmement bas sur une variable à une occasion donnée aura tendance à obtenir un score moins extrême à l'occasion suivante. Par exemple un membre de l'équipe de quilles dont la moyenne à long terme est de 150 et qui joue soudainement un 220 obtiendra presque certainement un score inférieur lors de la partie suivante. Son score « régressera » vers sa moyenne de 150. La régression vers la moyenne peut poser un problème lorsque les personnes sont sélectionnées pour une étude plus approfondie en raison de leurs scores extrêmes. Imaginons par exemple que seules les personnes ayant obtenu un score particulièrement élevé au test d'attitudes à l'égard des drogues illicites (celles qui ont des attitudes extrêmement favorables à l'égard des drogues) aient bénéficié du programme antidrogue et aient

été soumises à un nouveau test. La régression vers la moyenne garantit pratiquement que leurs résultats seront inférieurs au post-test, même si le programme de formation n'a aucun effet.

Un concept étroitement lié — et extrêmement important dans la recherche en psychologie — est la **rémission spontanée**. Il s'agit de la tendance de nombreux problèmes médicaux et psychologiques à s'améliorer au fil du temps sans aucune forme de traitement. Le rhume est un bon exemple. Si l'on mesurait aujourd'hui la gravité des symptômes chez 100 personnes souffrant d'un rhume, qu'on leur donnait un bol de soupe au poulet tous les jours, et qu'on mesurait à nouveau la gravité de leurs symptômes en une semaine, on constaterait probablement une nette amélioration. Cela ne signifie pas pour autant que la soupe au poulet soit à l'origine de l'amélioration, car ces personnes se seraient beaucoup améliorées sans aucun traitement du tout. Il en va de même pour de nombreux problèmes psychologiques. Un groupe de personnes gravement déprimées aujourd'hui sera probablement moins déprimé en moyenne dans 6 mois. En examinant les résultats de plusieurs études sur les traitements de la dépression, les chercheurs Michael Posternak et Ivan Miller ont constaté que les personnes participantes placées dans les conditions contrôles avec liste d'attente s'amélioraient en moyenne de 10 à 15 % avant de recevoir le moindre traitement (Posternak et Miller, 2001). Il convient donc d'être très prudent lorsqu'il s'agit d'inférer des liens de causalité à partir de devis prétest-post-test.

Une approche courante pour écarter les menaces à la validité interne décrites ci-dessus consiste à revoir le devis de la recherche afin d'inclure un groupe contrôle, qui ne bénéficie pas de l'effet du traitement. Un groupe contrôle serait soumis aux mêmes menaces liées à l'histoire, à la maturation, aux tests, à l'instrumentation, à la régression vers la moyenne et à la rémission spontanée, et permettrait donc aux scientifiques de mesurer l'effet réel du traitement (s'il existe). Bien entendu, l'inclusion d'un groupe contrôle signifie qu'il ne s'agit plus d'un devis à un seul groupe.

La psychothérapie fonctionne-t-elle?

Les premières études sur l'efficacité de la psychothérapie avaient tendance à utiliser des devis prétest-post-test. Dans un article classique de 1952, le chercheur Hans Eysenck a résumé les résultats de 24 études de ce type, montrant qu'environ deux tiers des patientes et patients s'étaient améliorés entre le prétest et le post-test (Eysenck, 1952). Mais Eysenck a également comparé ces résultats avec des données d'archives provenant d'hôpitaux publics et de compagnies d'assurance, montrant que des personnes patientes similaires se rétablissaient à peu près au même rythme sans avoir reçu de psychothérapie. Ce parallèle a suggéré à Eysenck que l'amélioration constatée auprès des personnes patientes dans les études prétest-post-test pourrait n'être qu'une rémission spontanée. Il convient de noter qu'Eysenck n'a pas conclu à l'inefficacité de la psychothérapie. Il a simplement conclu qu'aucune preuve ne démontrait son efficacité, et il a écrit « la nécessité d'études expérimentales correctement planifiées et exécutées dans ce domaine important » (p. 323, traduction). Vous pouvez lire l'intégralité de l'article ici [en anglais] :

psychclassics.yorku.ca/Eysenck/psychotherapy.htm

Heureusement de nombreux autres scientifiques ont relevé le défi d'Eysenck et en 1980, des centaines d'expériences avaient été menées dans lesquelles les personnes participantes étaient assignées au hasard à des conditions de traitement et de contrôle, et les résultats ont été résumés dans un livre classique de Mary Lee

Smith, Gene Glass et Thomas Miller (Smith, Glass et Miller, 1980). Cette équipe de recherche a constaté que dans l'ensemble la psychothérapie était assez efficace, car environ 80 % des personnes participantes au traitement se sont améliorées davantage que la moyenne des personnes participantes au contrôle. Les recherches ultérieures se sont davantage concentrées sur les conditions dans lesquelles les différents types de psychothérapie sont plus ou moins efficaces.

Devis à série temporelle interrompue

Une variante du devis prétest-post-test est le **devis à série temporelle interrompue** (*interrupted time-series design*, en anglais). Une série temporelle est un ensemble de mesures prises à intervalles réguliers au cours d'une période donnée. Par exemple une entreprise manufacturière peut mesurer la productivité de sa main-d'œuvre chaque semaine pendant un an. Dans un devis comme celui-ci, la série temporelle est « interrompue » par un traitement. Dans un exemple classique, le traitement a consisté à réduire les quarts de travail dans une usine de 10 heures à 8 heures (Cook et Campbell, 1979). Comme la productivité a augmenté assez rapidement après la réduction des quarts de travail dans les équipes et qu'elle est restée élevée pendant de nombreux mois, les scientifiques ont conclu que la réduction des quarts de travail était à l'origine de l'augmentation de la productivité. Il convient de noter que la méthode de série temporelle interrompue ressemble à une méthode prétest-post-test en ce sens qu'elle inclut des mesures de la variable dépendante à la fois avant et après le traitement. Elle diffère toutefois du devis prétest-post-test en ce qu'elle comprend plusieurs mesures prétest et post-test.

La Figure 8.1 présente les données d'une étude hypothétique de série temporelle interrompue. La variable dépendante est le nombre d'absences des personnes aux études par semaine dans un cours de méthodologie de recherche. Le traitement requiert que le personnel enseignant commence à prendre publiquement les présences chaque jour afin que les personnes étudiantes sachent que le personnel enseignant sait qui est présent et qui est absent. Le panneau supérieur de la Figure 8.1 montre à quoi les données pourraient ressembler si ce traitement fonctionnait. Le nombre d'absences est constamment élevé avant le traitement et il y a une baisse immédiate et soutenue des absences après le traitement. Le panneau inférieur de la Figure 8.1 montre à quoi les données pourraient ressembler si ce traitement ne fonctionnait pas. En moyenne, le nombre d'absences après le traitement est à peu près le même qu'avant. Cette figure illustre également un avantage du devis à série temporelle interrompue par rapport à un devis plus simple de type prétest-post-test. S'il n'y avait eu qu'une seule mesure sans traitement avant l'introduction du traitement, à la semaine 7, et une seule après, à la semaine 8, on aurait pu croire que le traitement était responsable de la réduction. Les mesures multiples avant et après le traitement suggèrent que la réduction entre les semaines 7 et 8 n'est rien d'autre qu'une variation normale d'une semaine à l'autre.

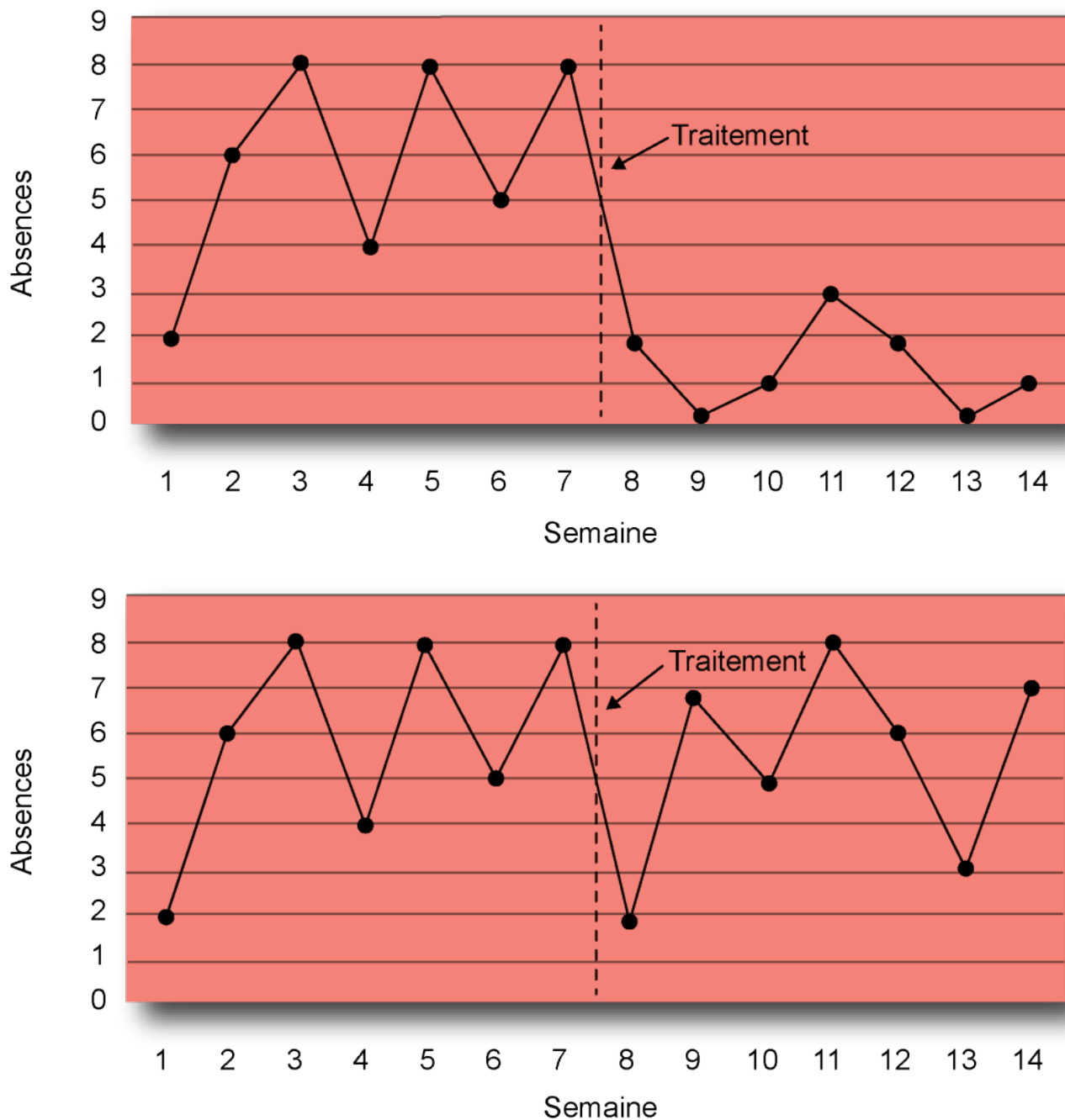


Figure 8.1. Un modèle hypothétique de devis à séries temporelles interrompues. Le panneau supérieur montre des données qui suggèrent que le traitement a entraîné une réduction des absences. Le panneau du bas montre des données qui suggèrent que le traitement n'a pas eu d'effet. [\[Description de la figure\]](#)

Description de l'image

Figure 8.1 – Description de l'image : Deux graphiques linéaires représentant le nombre d'absences par semaine sur 14 semaines. Les sept premières semaines sont sans traitement et les sept dernières semaines

sont avec traitement. Dans le premier graphique, il y a entre quatre et huit absences par semaine. Après le traitement, les absences oscillent entre zéro et trois par semaine, ce qui suggère que le traitement a fonctionné. Dans le deuxième graphique, il n'y a pas de changement notable dans le nombre d'absences par semaine après le traitement, ce qui suggère que le traitement n'a pas fonctionné. [\[Retour à la Figure 8.1\]](#)

46. Devis à groupes non équivalents

Objectifs d'apprentissage

1. Décrire les différents types de devis quasi-expérimentaux à groupes non équivalents
2. Identifier quelques-unes des menaces à la validité interne associées à chacun de ces devis

Rappelons que lorsque les personnes participantes dans une expérience intersujets sont assignées de manière aléatoire à des conditions, les groupes qui en résultent sont susceptibles d'être assez similaires. En fait les scientifiques les considèrent comme équivalents. En revanche lorsque les personnes participantes ne sont pas réparties de manière aléatoire entre les conditions, les groupes résultants sont susceptibles d'être différents à certains égards. Pour cette raison, les scientifiques les considèrent comme non équivalents. Un **devis à groupes non équivalents** est donc un devis intersujets dans lequel les personnes participantes n'ont pas été assignées de manière aléatoire aux conditions. Il existe plusieurs types de devis de groupes non équivalents que nous examinerons.

Devis post-test seulement avec groupes non équivalents

Le premier devis de groupes non équivalents que nous examinerons est le **devis post-test seulement avec groupes non équivalents** (*post-test only nonequivalent groups design*, en anglais). Dans ce devis, les personnes participantes d'un groupe sont exposées à un traitement alors qu'un groupe non équivalent n'est pas exposé au traitement, puis les deux groupes sont comparés. Imaginons par exemple que des scientifiques souhaitent évaluer une nouvelle méthode d'enseignement des fractions aux élèves de troisième année. L'une des solutions consisterait à mener une étude avec un groupe expérimental composé d'une classe de troisième année et un groupe contrôle composé d'une autre classe de troisième année. Il s'agirait d'un devis à groupes non équivalents, car la répartition des élèves entre les classes n'a pas été faite par les scientifiques, au hasard, ce qui signifie qu'il peut y avoir des différences importantes. Par exemple les parents des élèves les plus performants ou les plus motivés auraient pu être plus enclins à demander que leurs enfants soient placés dans la classe de Mme Vachon. Ou encore la direction peut aussi avoir mis les "téméraires" dans la classe de M. Jolicœur parce qu'il exerce une discipline plus stricte. Bien entendu le style du personnel enseignant et même l'environnement de la classe peuvent être très différents, et entraîner des niveaux de réussite ou de motivation différents chez les élèves. Si à la fin de l'étude, il y avait une différence dans la maîtrise des fractions entre les deux classes, elle pourrait avoir été causée par la différence entre les

méthodes d'enseignement— mais elle pourrait aussi avoir été causée par n'importe laquelle de ces variables confondantes.

Bien entendu les scientifiques qui utilisent un devis post-test avec groupes non équivalents peuvent prendre des mesures pour s'assurer que leurs groupes sont aussi similaires que possible. Dans le présent exemple, les scientifiques pourraient essayer de sélectionner deux classes dans la même école, où les élèves des deux classes ont des résultats similaires à un test de mathématiques standardisé et où les enseignants sont du même sexe, ont un âge proche et des styles d'enseignement similaires. Prendre de telles mesures augmenterait la validité interne de l'étude parce qu'elle éliminerait certaines variables confondantes les plus importantes. Mais sans une véritable assignation aléatoire des personnes étudiantes aux conditions, il reste la possibilité d'autres variables confondantes importantes que le scientifique n'était pas en mesure de contrôler.

Devis prétest-post-test avec groupes non équivalents

Une autre façon d'améliorer le devis ayant uniquement un post-test avec groupes non équivalents est d'ajouter un prétest. Dans le **devis prétest-post-test avec groupes non équivalents**, il y a un groupe traitement qui reçoit un prétest, un traitement et ensuite un post-test. Mais en même temps, il y a un groupe contrôle non équivalent qui reçoit un prétest, ne reçoit pas le traitement et reçoit ensuite un post-test. La question n'est donc pas simplement de savoir si les personnes participantes qui reçoivent le traitement s'améliorent, mais si elles s'améliorent davantage que les personnes participantes qui ne reçoivent pas le traitement.

Imaginons par exemple que les élèves d'une école reçoivent un prétest sur leurs attitudes à l'égard des drogues, qu'ils soient ensuite exposés à un programme antidrogue et qu'ils reçoivent enfin un post-test. Les élèves d'une école similaire reçoivent le prétest, ne sont pas exposés à un programme de lutte contre la drogue, et enfin reçoivent un post-test. Encore une fois si les élèves du groupe traité deviennent plus négatifs à l'égard des drogues, ce changement d'attitude pourrait être un effet du traitement mais il pourrait aussi s'agir d'une question d'histoire ou de maturation. S'il s'agit réellement d'un effet du traitement, les élèves du groupe expérimental devraient devenir plus négatifs que les élèves du groupe témoin. Mais s'il s'agit d'une question d'histoire (par exemple la nouvelle de l'overdose d'une célébrité) ou de maturation (par exemple l'amélioration du raisonnement), alors les élèves des deux conditions sont susceptibles de montrer des changements de la même ampleur. Ce type de devis n'élimine cependant pas complètement la possibilité de variables confondantes. Un événement pourrait se produire dans l'une des écoles mais pas dans l'autre (par exemple l'overdose d'une personne) de sorte que les élèves de la première école seraient affectés par cet événement alors que les élèves de l'autre école ne le seraient pas.

Pour revenir à l'exemple de l'évaluation d'une nouvelle mesure de l'enseignement aux élèves de troisième année, cette étude pourrait être améliorée en ajoutant un prétest des connaissances des élèves en matière de fractions. L'évolution des scores entre le prétest et le post-test serait alors évaluée et comparée entre les différents groupes afin de déterminer si l'amélioration des connaissances en matière de fractions a été plus importante dans un groupe que dans l'autre. Bien entendu le style du personnel enseignant et même l'environnement de la classe peuvent être très différents, et entraîner des niveaux de réussite ou de motivation différents chez les élèves indépendamment de l'intervention pédagogique. Une fois de

plus l'histoire différentielle représente également une menace potentielle pour la validité interne. Si l'on découvre de l'amiante dans l'une des écoles entraînant sa fermeture pendant un mois, cette interruption de l'enseignement pourrait produire une différence entre les groupes aux résultats du post-test.

Si les personnes participant à ce type d'étude sont assignées de manière aléatoire aux conditions, il s'agit d'une véritable expérience entre les groupes plutôt que d'une quasi-expérience. En fait c'est le type d'expérience qu'Eysenck réclamait — et qui est maintenant réalisé fréquemment — pour démontrer l'efficacité de la psychothérapie.

Devis à série temporelle interrompue avec groupes non équivalents

L'une des façons d'améliorer la méthode de série temporelle interrompue est d'ajouter un groupe contrôle. Les **devis à série temporelle interrompue avec groupes non équivalents** (*Interrupted Time-Series Design with Nonequivalent Groups* en anglais) consistent à prendre un ensemble de mesures à intervalles réguliers sur une période donnée, avant et après une intervention d'intérêt dans deux groupes non équivalents ou plus. Prenons à nouveau l'exemple d'une entreprise manufacturière qui mesure la productivité de sa main-d'œuvre chaque semaine pendant un an, avant et après avoir réduit les quarts de travail de 10 à 8 heures. Cette méthode pourrait être améliorée en localisant une autre entreprise manufacturière qui n'a pas l'intention de modifier la durée des quarts de travail et en l'utilisant comme groupe contrôle non équivalent. Si la productivité a augmenté assez rapidement après la réduction des quarts de travail dans le groupe expérimental, mais qu'elle est restée constante dans le groupe contrôle, cela constitue une meilleure preuve de l'efficacité du traitement.

De même dans la recherche des effets de la prise de présence sur les absences des personnes inscrites dans un cours de méthodologie de recherche, la méthode pourrait être améliorée en utilisant des personnes inscrites à une autre section du cours de méthodologie de recherche comme groupe contrôle. Si le nombre d'absences était systématiquement plus élevé dans le groupe expérimental avant l'intervention, suivi d'une baisse durable des absences après l'intervention, alors que le groupe contrôle non équivalent présentait des absences systématiquement élevées tout au long du semestre, cela fournirait des preuves supérieures de l'efficacité de l'intervention dans la réduction des absences.

Devis prétest-post-test avec réplication alternée

Certains de ces devis avec groupes non équivalents peuvent être améliorés par l'ajout d'une réplication alternée. En utilisant un **devis prétest-post-test avec réplication alternée**, (*Pretest-post-test Design With Switching Replication*, en anglais), des groupes non équivalents sont soumis à un prétest de la variable dépendante, puis un groupe reçoit un traitement tandis qu'un groupe contrôle non équivalent ne reçoit pas de traitement, la variable dépendante est évaluée à nouveau, puis le traitement est ajouté au groupe contrôle. Enfin la variable dépendante est mesurée une dernière fois.

En guise d'exemple concret, supposons que nous souhaitions introduire une intervention en matière

d'exercice physique pour le traitement de la dépression. Nous recrutons un groupe de patientes et patients souffrant de dépression et un groupe témoin non équivalent d'étudiantes et d'étudiants souffrant de dépression. Nous mesurons d'abord les niveaux de dépression dans les deux groupes, puis nous introduisons l'intervention par l'exercice aux patientes et patients souffrant de dépression, mais nous nous abstenons d'introduire ce traitement aux étudiantes et étudiants. Nous mesurons ensuite les niveaux de dépression dans les deux groupes. Si le traitement est efficace, nous devrions observer une réduction des niveaux de dépression chez les bénéficiaires (qui ont reçu le traitement), mais pas chez les personnes aux études (qui n'ont pas encore reçu le traitement). Enfin pendant que le groupe de bénéficiaires continue à suivre le traitement, nous introduisons le traitement auprès de l'autre groupe, les personnes aux études. C'est maintenant et seulement maintenant que nous devrions voir les niveaux de dépression de ces derniers diminuer.

L'un des points forts de ce devis est qu'il comprend une réplication intégrée. Dans l'exemple donné, nous obtiendrions des preuves de l'efficacité du traitement dans deux échantillons différents (bénéficiaires et personnes aux études). Un autre point fort de ce devis est qu'il permet de mieux contrôler les effets de l'histoire. Il est en effet peu probable qu'un événement extérieur coïncide parfaitement avec l'introduction du traitement dans le premier groupe et qu'il se produise dans le second groupe. Par exemple si un changement de **conditions météorologiques** s'est produit lorsque nous avons introduit le traitement pour la première fois auprès des bénéficiaires, et que cela explique la réduction de la dépression lors de la deuxième mesure de la dépression, nous verrons alors les niveaux de dépression diminuer dans les deux groupes. De la même manière, la **réplication** avec changement de groupes permet de contrôler la maturation et l'instrumentation. Les deux groupes devraient présenter les mêmes taux de rémission spontanée de la dépression et si l'instrument d'évaluation de la dépression changeait à un moment donné de l'étude, le changement serait cohérent dans les deux groupes. Bien entendu les caractéristiques de la demande, les effets placebo et les effets des attentes de la personne menant l'expérimentation peuvent toujours poser problème. Mais ils peuvent être contrôlés à l'aide de certaines des méthodes décrites au chapitre 5.

Devis de réplication alternée avec retrait du traitement

Dans un devis prétest-post-test avec réplication alternée, le premier groupe reçoit un traitement et le second groupe reçoit le même traitement un peu plus tard (tandis que le groupe initial continue à recevoir le traitement). En revanche dans un **devis de réplication alternée avec retrait du traitement**, (*switching replication with treatment removal design*, en anglais), le traitement est retiré du premier groupe lorsqu'il est ajouté au deuxième groupe. Supposons à nouveau que nous mesurons d'abord les niveaux de dépression des bénéficiaires ayant une dépression et des personnes aux études ayant une dépression. Ensuite nous présentons l'intervention d'exercice physique seulement aux bénéficiaires. Après une semaine d'exposition à l'intervention d'exercice pour ce dernier groupe, nous évaluons à nouveau les niveaux de dépression pour les deux groupes. Si l'intervention est efficace, nous devrions voir les niveaux de dépression diminuer pour le groupe des bénéficiaires, mais pas pour le groupe de personnes aux études (parce que ces dernières n'ont pas encore reçu le traitement). Ensuite nous retirerons le traitement au groupe de bénéficiaires souffrant de dépression. Nous leur demanderons donc d'arrêter de faire de l'exercice. Au même moment, nous demanderons au groupe de personnes aux études de commencer à faire de l'exercice. Après une semaine

pendant laquelle ces dernières font de l'exercice alors que les bénéficiaires n'en font pas, nous réévaluerons les niveaux de dépression. Si l'intervention est efficace, nous devrions constater que les niveaux de dépression ont diminué dans le groupe des personnes aux études, mais qu'ils ont augmenté dans le groupe de bénéficiaires (parce que ces personnes ne font plus d'exercice).

La démonstration d'un effet de traitement dans deux groupes, échelonné dans le temps, et l'inversion de l'effet de traitement après le retrait du traitement peuvent fournir des preuves solides de l'efficacité du traitement. En plus de fournir des preuves de la reproductibilité des résultats, ce devis peut également permettre de déterminer si le traitement continue à produire des effets après son retrait. Pour approfondir les connaissances sur l'utilisation de ces devis quasi-expérimentaux dans une situation concrète d'évaluation de l'efficacité d'interventions visant à réduire la consommation d'alcool chez les jeunes consultez:

Benmarhnia, T., & Fuller, D. (2019). Les méthodes quasi-expérimentales: L'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes aux États-Unis. Dans V. Ridde et C. Dagenais (dir.), Évaluation des interventions de santé mondiale: méthodes avancées. Éditions science et bien commun; Institut de recherche pour le développement.
<https://scienceetbiencommun.pressbooks.pub/evalsantemondiale/chapter/quasiexperimentales>

47. Points clés et exercices

Points clés

- La recherche quasi-expérimentale implique la manipulation d'une variable indépendante sans assignation aléatoire des personnes participantes aux conditions ou contrebalancement de l'ordre des conditions.
- Il existe trois types de devis quasi-expérimentaux intrasujets. Il y a le devis post-test à groupe unique, le devis prétest-post-test à groupe unique et le devis à série temporelle interrompue.
- Il existe cinq types de devis quasi-expérimentaux intersujets. Il y a le devis post-test avec groupes non équivalents, le devis prétest-post-test avec groupes non équivalents, le devis à série temporelle interrompue avec groupes non équivalents, le devis prétest-post-test avec réplication alternée et le devis de réplication alternée avec retrait du traitement.
- La recherche quasi-expérimentale élimine le problème de la directionnalité parce qu'elle implique la manipulation de la variable indépendante. Cependant elle n'élimine pas le problème des variables confondantes, car elle n'implique pas d'assignation aléatoire aux conditions ou de contrebalancement. Pour ces raisons, la validité interne de la recherche quasi-expérimentale est généralement supérieure à celle des études non expérimentales, mais inférieure à celle des véritables expérimentations.
- De tous les devis quasi-expérimentaux, ceux qui incluent une réplication alternée sont les plus performants en termes de validité interne.

Exercices

- **Pratique** : Imaginez que deux membres du corps professoral décident de tester l'effet des interrogations quotidiennes sur les performances des personnes inscrites dans un cours de statistiques. Ils décident que la professeure A donnera des interrogations et que la professeure B n'en donnera pas. On compare ensuite les performances des personnes inscrites aux deux sections lors d'un examen final commun. Citez cinq autres variables susceptibles de différer entre les deux sections et d'affecter les résultats.
- **Discussion** : Imaginez qu'un groupe d'enfants obèses soit recruté pour une étude au cours de laquelle leur poids est mesuré, puis que ces enfants participent pendant trois mois à un programme qui les encourage à faire de l'exercice, et enfin que leur poids soit à nouveau mesuré. Expliquez comment chacun des éléments suivants pourrait affecter les résultats :

- - la régression vers la moyenne
 - la rémission spontanée
 - l'histoire
 - la maturation

48. Références

- Cook, T. D., et Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings* (*Quasi-expérimentation : Questions relatives à la conception et à l'analyse sur le terrain*). Houghton Mifflin.
- Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation (Les effets de la psychothérapie : une évaluation). *Journal of Consulting Psychology*, 16(5), 319-324. <https://doi.org/10.1037/h0063633>
- Posternak, M. A. et Miller, I. (2001). Untreated short-term course of major depression : A meta-analysis of studies using outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, 66(2-3), 139-146. [https://doi.org/10.1016/S0165-0327\(00\)00304-9](https://doi.org/10.1016/S0165-0327(00)00304-9)
- Smith, M. L., Glass, G. V. et Miller, T. I. (1980). *The benefits of psychotherapy* (*Les bienfaits de la psychothérapie*). Johns Hopkins University Press.

CHAPITRE 9-LES PLANS FACTORIELS



Dans le chapitre 1, nous avons brièvement décrit une étude menée par Simone Schnall et ses collègues, dans laquelle ces scientifiques ont constaté que le fait de se laver les mains amène les gens à considérer les transgressions morales comme moins mauvaises (Schnall, Benton et Harvey, 2008). Dans une étude différente, mais connexe, Schnall et ses collègues ont cherché à savoir si le fait de se sentir physiquement dégoûté incitait les gens à porter des jugements moraux plus sévères (Schnall, Haidt, Clore et Jordan, 2008). Dans cette expérience, les personnes en recherche ont manipulé les sentiments de dégoût des personnes participantes en les testant soit dans une pièce propre, soit dans une pièce en désordre contenant de la vaisselle sale, une poubelle débordante et un stylo mâchouillé. L'équipe a également utilisé un questionnaire d'auto-évaluation pour mesurer l'attention que les personnes participantes portent à leurs propres sensations corporelles qu'on appelle la « conscience corporelle privée ». La variable dépendante principale, la sévérité des jugements moraux des gens, a été mesurée en décrivant différents comportements (par exemple manger son chien mort, ne pas rendre un portefeuille trouvé) et en demandant aux personnes participantes d'évaluer l'acceptabilité morale de chacun d'entre eux sur une échelle de 1 à 7. Enfin les scientifiques ont demandé aux personnes participantes d'évaluer leur niveau actuel de dégoût ainsi que d'autres émotions. Les principaux résultats de cette étude ont montré que les gens dans la pièce en désordre ressentaient en fait plus de dégoût et portaient des jugements moraux plus sévères que les gens dans la pièce propre —, mais seulement si ces personnes avaient un score relativement élevé à la mesure de la conscience corporelle privée.

Les devis de recherche que nous avons examinés jusqu'à présent étaient simples — centrés sur une question portant sur une variable ou sur une relation entre deux variables. Mais à bien des égards, le devis complexe de cette expérience menée par Schnall et ses collègues est plus typique de la recherche en psychologie. Heureusement nous avons déjà abordé les éléments de base de ces devis dans les chapitres précédents. Dans le présent chapitre, nous examinerons de près comment et pourquoi les scientifiques utilisent des **plans factoriels**, c'est-à-dire des expériences qui comprennent plus d'une variable indépendante.

49. Mettre en place une expérience factorielle

Objectifs d'apprentissage

1. Expliquer pourquoi les personnes menant des recherches incluent souvent plusieurs variables indépendantes dans leurs études.
2. Définir le plan factoriel et utiliser un tableau de plan factoriel pour représenter et interpréter des plans factoriels simples.

Tout comme il est commun pour les études en psychologie d'inclure plusieurs niveaux d'une même variable indépendante (placebo, nouveau médicament, ancien médicament), il est également commun pour elles d'inclure plusieurs variables indépendantes. Schnall et ses collègues ont étudié l'effet du dégoût et de la conscience corporelle privée dans une même étude.

L'inclusion de plusieurs variables indépendantes dans une même expérience est également illustrée par les titres réels suivants, tirés de diverses revues scientifiques:

- « The Effects of Temporal Delay and Orientation on Haptic Object Recognition » (" Les effets du délai temporel et de l'orientation sur la reconnaissance haptique d'objets ")
- « Opening Closed Minds : The Combined Effects of Intergroup Contact and Need for Closure on Prejudice » (" Ouvrir les esprits fermés : les effets combinés du contact intergroupe et du besoin de fermeture sur les préjugés ")
- « Effects of Expectancies and Coping on Pain-Induced Intentions to Smoke » (" Effets des attentes et de l'adaptation sur les intentions de fumer induites par la douleur ")
- « The Effect of Age and Divided Attention on Spontaneous Recognition » (" L'effet de l'âge et de l'attention divisée sur la reconnaissance spontanée ")
- « The Effects of Reduced Food Size and Package Size on the Consumption Behavior of Restrained and Unrestrained Eaters » (" Effets de la réduction de la taille des aliments et de la taille de l'emballage sur le comportement de consommation de personnes qui mangent avec ou sans contraintes ")

Tout comme l'inclusion de plusieurs niveaux d'une même variable indépendante permet de répondre à des questions de recherche plus sophistiquées, l'inclusion de plusieurs variables indépendantes dans une même expérience est également possible. Par exemple au lieu de mener une étude sur l'effet du dégoût sur le jugement moral, et une autre sur l'effet de la conscience corporelle privée sur le jugement moral, Schnall et

ses collègues ont pu mener une étude qui abordait les deux questions. Mais l'inclusion de plusieurs variables indépendantes permet également à l'équipe de répondre à des questions concernant le fait qu'une variable indépendante puisse dépendre du niveau d'une autre. On parle alors d'une interaction entre les variables indépendantes. Schnall et ses collègues ont par exemple observé une interaction entre le dégoût et la conscience corporelle privée, car l'effet du dégoût dépendait du niveau de conscience corporelle privée, faible ou élevé des personnes participantes. Comme nous le verrons, les interactions figurent souvent parmi les résultats les plus intéressants de la recherche en psychologie.

Plans factoriels

Vue d'ensemble

Le plan factoriel est de loin l'approche la plus courante pour inclure plusieurs variables indépendantes (souvent appelées facteurs) dans une expérience. Dans un **plan factoriel**, chaque niveau d'une variable indépendante est combiné avec chaque niveau des autres variables pour produire toutes les combinaisons possibles. Chaque combinaison devient alors une condition de l'expérience. Imaginons par exemple une expérience sur l'effet de l'utilisation d'un téléphone cellulaire (oui ou non) et de l'heure de la journée (jour ou nuit) sur l'aptitude à conduire. Cette expérience est illustrée dans le **tableau des plans factoriels** de la Figure 9.1. Les colonnes du tableau représentent l'utilisation du téléphone cellulaire et les rangées représentent le moment de la journée. Les quatre cellules du tableau représentent les quatre combinaisons ou conditions possibles : utiliser un téléphone cellulaire pendant le jour, ne pas utiliser de téléphone cellulaire pendant le jour, utiliser un téléphone cellulaire pendant la nuit et ne pas utiliser de téléphone cellulaire pendant la nuit. Ce plan particulier est appelé un plan factoriel 2×2 (dit « deux par deux ») parce qu'il combine deux variables, chacune ayant deux niveaux.

Si l'une des variables indépendantes comportait un troisième niveau (par exemple utiliser un téléphone cellulaire tenu dans la main, utiliser un téléphone cellulaire mains libres et ne pas utiliser de téléphone cellulaire), il s'agirait alors d'un plan factoriel 3×2 , et il y aurait six conditions distinctes. Remarquons que le nombre de conditions possibles est le produit des nombres de niveaux. Un plan factoriel 2×2 comporte quatre conditions, un plan factoriel 3×2 en comporte six, un plan factoriel 4×5 en comporte 20, etc. Remarquons également que chaque nombre dans la notation représente un facteur, une variable indépendante. En examinant le nombre de chiffres dans la notation, vous pouvez donc déterminer le nombre de variables indépendantes dans l'expérience. Les plans 2×2 , 3×3 et 2×3 comportent tous deux nombres dans la notation et donc deux variables indépendantes. La valeur numérique de chacun des chiffres représente le nombre de niveaux de chaque variable indépendante. Un « 2 » signifie que la variable indépendante a deux niveaux, un « 3 » signifie que la variable indépendante a trois niveaux, un « 4 » signifie qu'elle a quatre niveaux, etc. Pour illustrer, un plan 3×3 comporte deux variables indépendantes chacune ayant trois niveaux, tandis qu'un plan de $2 \times 2 \times 2$ comporte trois variables indépendantes chacune ayant deux niveaux.

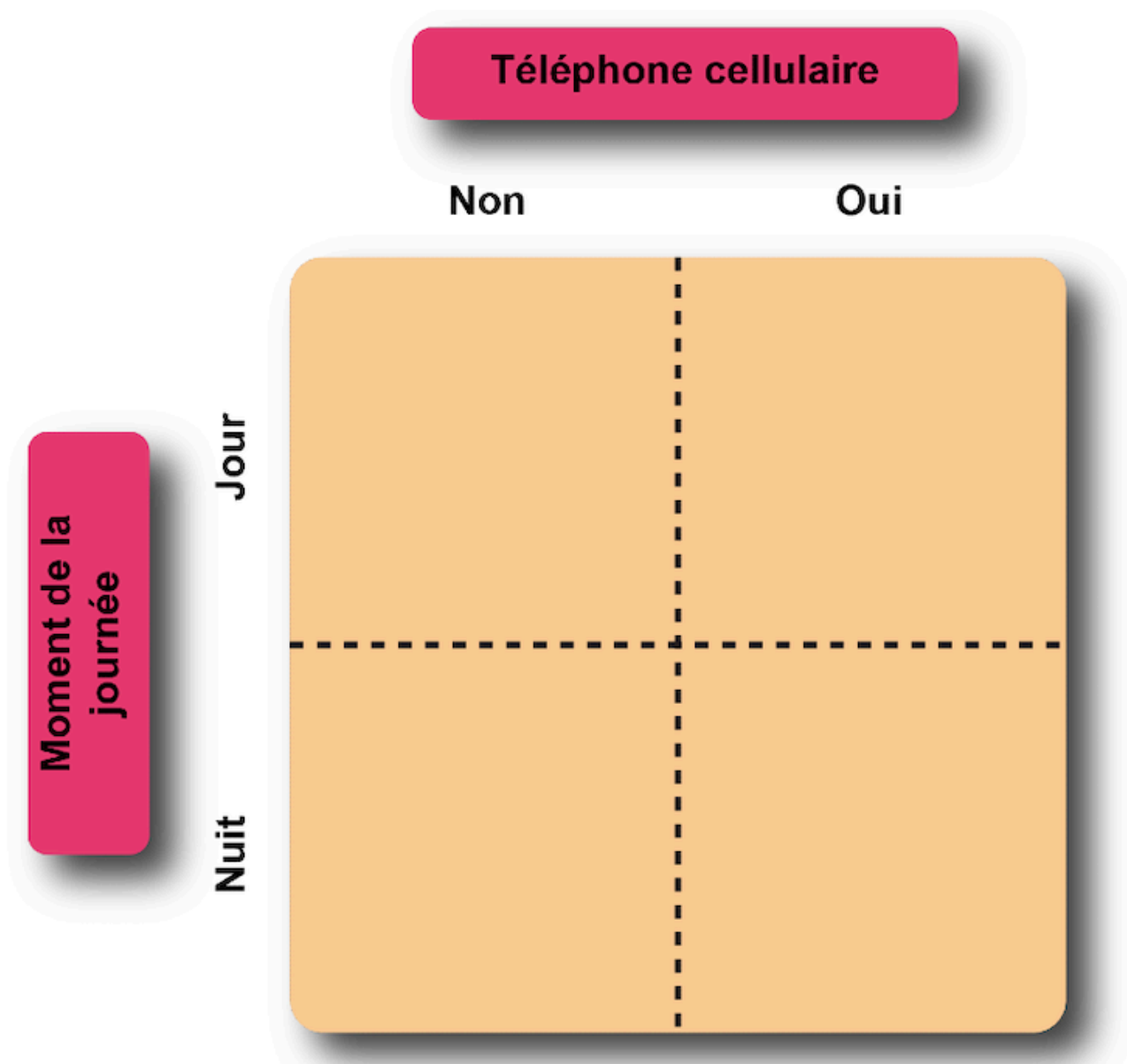


Figure 9.1. Tableau de plan factoriel représentant un plan factoriel 2×2

En principe les plans factoriels peuvent inclure n'importe quel nombre de variables indépendantes avec n'importe quel nombre de niveaux. Par exemple une expérience pourrait inclure le type de psychothérapie (cognitive ou comportementale), la durée de la psychothérapie (2 semaines ou 2 mois) et le sexe du psychothérapeute (femme ou homme). Cela constituerait un plan factoriel $2 \times 2 \times 2$ comportant huit conditions. La Figure 9.2 illustre une façon de représenter ce plan. En réalité il est rare qu'il y ait plus de trois variables indépendantes avec plus de deux ou trois niveaux chacune. Il y a au moins deux raisons à cela. Premièrement le nombre de conditions peut rapidement devenir ingérable. Par exemple l'ajout d'une quatrième variable indépendante à trois niveaux (par exemple l'expérience de la personne thérapeute : faible vs moyenne vs élevée) à l'exemple actuel en ferait un plan factoriel $2 \times 2 \times 2 \times 3$ avec 24 conditions distinctes. Deuxièmement le nombre de personnes participantes nécessaires pour remplir toutes ces conditions (tout

en conservant une capacité raisonnable à détecter un effet sous-jacent réel) peut rendre le plan irréalisable (pour plus d'informations, voir la discussion sur l'importance de la puissance statistique adéquate au chapitre 13). Par conséquent dans la suite de cette section, nous nous concentrerons sur les plans à deux variables indépendantes. Les principes généraux discutés ici s'appliquent directement à des plans factoriels plus complexes.

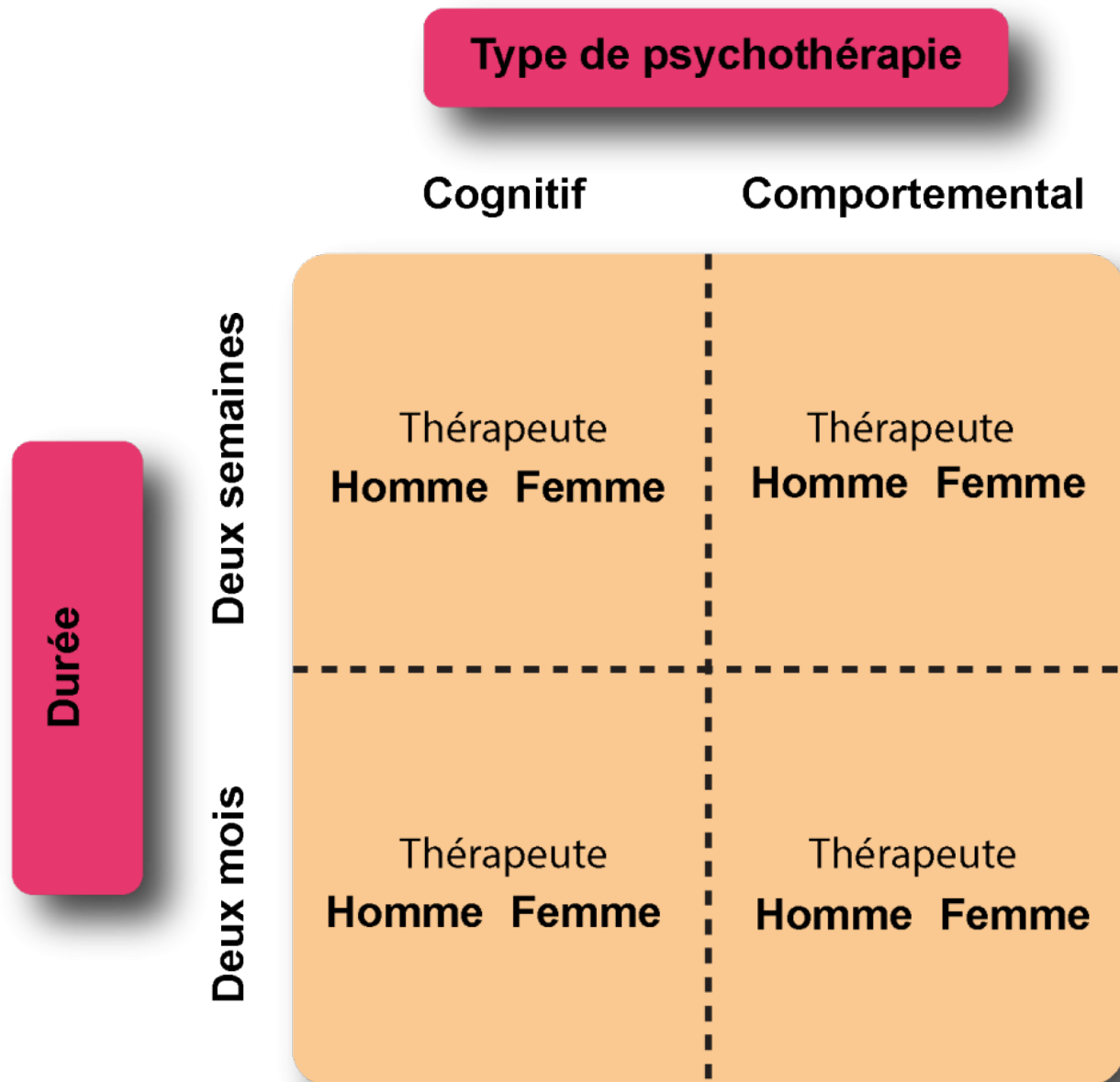


Figure 9.2. Tableau de plan factoriel représentant un plan factoriel $2 \times 2 \times 2$

Assigner les personnes participantes aux conditions

Rappelons que dans un devis intersujets simple, chaque personne participante est testée dans une seule condition. Dans un devis intrasujet simple, chaque personne participante est testée dans toutes les conditions. Dans une expérience factorielle, la décision d'adopter l'approche intersujets ou intrasujets doit être prise séparément pour chaque variable indépendante. Dans un **plan factoriel intersujets**, toutes les variables indépendantes sont manipulées entre les sujets. Par exemple toutes les personnes participantes pourraient être testées en utilisant un téléphone cellulaire *ou* en ne l'utilisant pas, et pendant le jour *ou* pendant la nuit. Cela signifie que chaque personne serait testée dans une et une seule condition. Dans un **plan factoriel intrasujet**, toutes les variables indépendantes sont manipulées à l'intérieur des sujets. Toutes les personnes participantes pourraient être testées à la fois en utilisant un téléphone cellulaire *et* en ne l'utilisant pas, et à la fois pendant le jour *et* pendant la nuit. Cela signifie que chaque personne devrait être testée dans les quatre conditions. Les avantages et les inconvénients de ces deux approches sont les mêmes que ceux évoqués au chapitre 5. Le modèle intersujets est conceptuellement plus simple, évite les effets de l'ordre et du transfert (*order/carryover effects*, en anglais), et minimise le temps et l'effort de chacun des participants. Le modèle intrasujet est plus efficace pour l'équipe de recherche et permet de mieux contrôler les variables externes des personnes participantes.

Puisque les plans factoriels comportent plus d'une variable indépendante, il est également possible de manipuler une variable indépendante entre les sujets et une autre à l'intérieur des sujets. C'est ce qu'on appelle un **plan factoriel mixte**. Par exemple l'équipe de recherche peut choisir de traiter l'utilisation du téléphone cellulaire comme un facteur intrasujets, en testant les mêmes personnes participantes avec et sans téléphone cellulaire (en contrebalançant l'ordre de ces deux conditions). Elle pourrait cependant choisir de traiter l'heure de la journée comme un facteur intersujets en testant chaque participant soit pendant la journée, soit pendant la nuit (peut-être parce que cela ne les oblige à se présenter qu'une seule fois pour le test). Ainsi chaque personne participante dans ce plan mixte serait testée dans deux des quatre conditions.

Que le plan soit intersujets, intrasujets ou mixte, l'assignation des personnes participantes entre les conditions ou les ordres de conditions se fait typiquement de manière aléatoire.

Variables indépendantes non manipulées

Dans de nombreux plans factoriels, l'une des variables indépendantes est une **variable indépendante non manipulée**. L'équipe de recherche la mesure, mais ne la manipule pas. L'étude de Schnall et de ses collègues en est un bon exemple. Une des variables indépendantes était le dégoût, manipulé en testant les personnes participantes dans une pièce propre ou dans une pièce en désordre. L'autre était la conscience corporelle privée, une variable des personnes participantes simplement mesurée. Un autre exemple est celui d'une étude, menée par Halle Brown et ses collègues, dans laquelle les personnes participantes étaient exposées à plusieurs mots qu'elles devaient ensuite rappeler (Brown, Kosslyn, Delamater, Fama et Barsky, 1999). La variable indépendante manipulée était le type de mot. Certains étaient des mots négatifs liés à la santé (par exemple *tumeur*, *coronaire*), tandis que d'autres n'étaient pas liés à la santé (par exemple *élection*, *géométrie*). La variable indépendante non manipulée était le degré faible ou élevé d'hypocondrie (préoccupation excessive à l'égard de symptômes corporels ordinaires). Les résultats de cette étude

montraient que les personnes ayant un niveau d'hypocondrie élevé se rappelaient mieux des mots liés à la santé que celles ayant un niveau d'hypocondrie faible, mais qu'elles n'étaient pas meilleures à se rappeler des mots qui ne sont pas liés à la santé.

De telles études sont extrêmement courantes, et plusieurs points méritent d'être soulignés. Premièrement les variables indépendantes non manipulées sont généralement des variables relatives aux personnes participantes (conscience corporelle privée, hypocondrie, estime de soi, genre, etc.), et sont donc par définition des facteurs intersujets. Par exemple les personnes sont soit peu hypocondriaques, soit très hypocondriaques ; elles ne peuvent pas être testées dans les deux conditions. Deuxièmement ces études sont généralement considérées comme des expériences dès qu'au moins une variable indépendante est manipulée, quel que soit le nombre de variables indépendantes non manipulées incluses. Troisièmement il est important de se rappeler que des conclusions causales peuvent seulement être tirées par rapport à la variable indépendante manipulée. Par exemple Schnall et ses collègues avaient raison de conclure que le dégoût affectait la sévérité des jugements moraux de leurs personnes participantes, car ces scientifiques avaient manipulé cette variable et assigné au hasard les personnes participantes dans la pièce propre ou dans la pièce en désordre. En revanche ces scientifiques n'auraient pas été en mesure de conclure que la conscience corporelle privée des gens influençait la sévérité de leurs jugements moraux, puisque cette variable n'a pas été manipulée. Il se peut par exemple que le fait d'avoir un code moral strict et une conscience corporelle élevée soit causé par une troisième variable (par exemple le névrosisme). Il est donc important de savoir quelles variables d'une étude sont manipulées et lesquelles ne le sont pas.

Études non expérimentales avec plans factoriels

Jusqu'à présent nous avons vu que les expériences factorielles peuvent inclure des variables indépendantes manipulées ou une combinaison de variables indépendantes manipulées et non manipulées. Cependant les plans factoriels peuvent aussi ne comporter que des variables indépendantes non manipulées, auquel cas il ne s'agit plus d'expériences, mais d'études de nature non expérimentale. Prenons l'exemple d'une étude hypothétique dans laquelle des scientifiques se contentent de mesurer l'humeur et l'estime de soi de plusieurs personnes participantes — en les classant comme ayant soit une humeur positive, soit une humeur négative, et comme ayant soit une estime de soi élevée, soit une estime de soi faible — ainsi que leur disposition à avoir des relations sexuelles non-protégées. Il s'agit d'un plan factoriel 2×2 dans lequel l'humeur (positive ou négative) et l'estime de soi (élevée ou faible) sont des facteurs intersujets non manipulés. La disposition à avoir des relations sexuelles non-protégées est la variable dépendante.

Ici encore comme aucune des variables indépendantes n'a été manipulée dans cet exemple, il s'agit d'une étude non expérimentale plutôt que d'une expérience. (L'étude similaire de MacDonald et Martineau [2002] était une expérience parce que ces scientifiques avaient manipulé l'humeur des personnes participantes.) Ceci est important car comme toujours il faut être prudent lorsqu'il s'agit d'inférer une causalité à partir d'études non expérimentales en raison des problèmes de directionnalité et de troisième variable (pour les explications détaillées de problèmes voir chapitre 6). Par exemple un effet de l'humeur des personnes participantes sur leur disposition à avoir des relations sexuelles non protégées pourrait être causé par une autre variable qui se trouve corrélée avec leur humeur.

50. Interpréter les résultats d'une expérience factorielle

Objectifs d'apprentissage

1. Distinguer les effets principaux et les interactions, les reconnaître et donner des exemples de chacun.
2. Dessiner et interpréter des histogrammes et des graphiques linéaires montrant les résultats d'études utilisant des plans factoriels simples.
3. Distinguer les effets principaux des effets simples et reconnaître quand une analyse des effets simples est requise.

Représentation graphique des résultats d'expériences factorielles

Les résultats d'expériences factorielles comportant deux variables indépendantes peuvent être représentés graphiquement en plaçant une variable indépendante sur l'axe des x et l'autre à l'aide de barres ou des lignes de couleurs différentes. (L'axe des y est toujours réservé à la variable dépendante.) La Figure 9.3 présente les résultats de deux expériences factorielles hypothétiques. Le panneau supérieur montre les résultats d'un plan 2×2 . Le moment de la journée (jour vs nuit) est représenté par différentes positions sur l'axe des x , et l'utilisation du téléphone cellulaire (non vs oui) est représentée par des barres de couleurs différentes. (Il serait également possible de représenter l'utilisation du téléphone cellulaire sur l'axe des x et l'heure de la journée avec des barres de couleurs différentes. Le choix dépend de la façon qui permet de communiquer le plus clairement les résultats). Le panneau inférieur de la Figure 9.3 montre les résultats d'un plan 4×2 dans lequel l'une des variables est quantitative. Cette variable, la durée de la psychothérapie, est représentée le long de l'axe des x et l'autre variable, le type de psychothérapie, est représentée par des lignes aux formats différents. Il s'agit d'un graphique linéaire plutôt que d'un diagramme à barres, car la variable sur l'axe des x est quantitative et comporte un petit nombre de niveaux distincts. Les graphiques linéaires conviennent également lorsqu'on représente des mesures effectuées sur un intervalle de temps (également appelées informations de séries temporelles) sur l'axe des x .

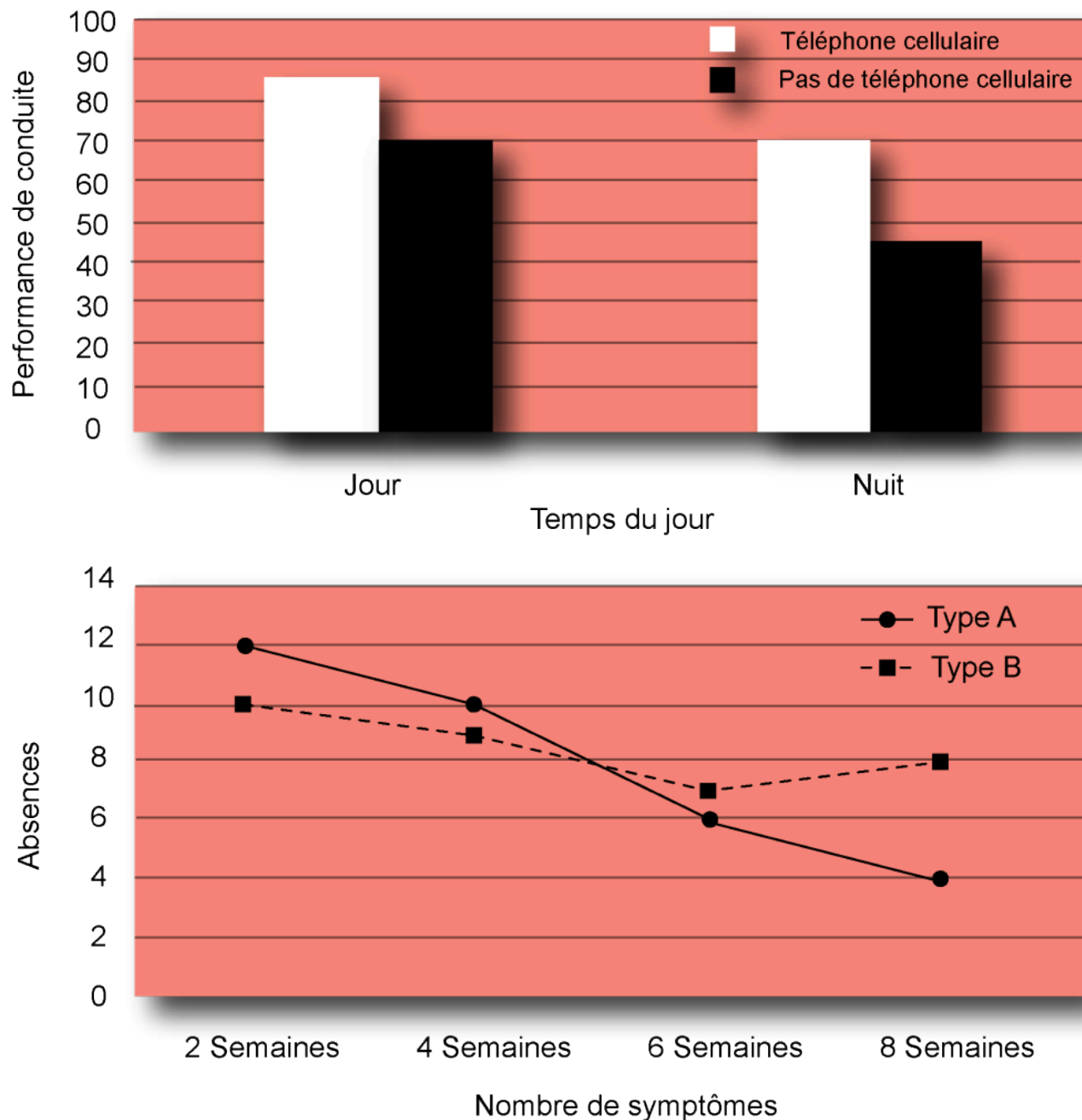


Figure 9.3. Deux façons de représenter les résultats d'une expérience factorielle avec deux variables indépendantes

Effets principaux

Dans les plans factoriels, trois types de résultats sont intéressants : les effets principaux, les effets d'interaction et les effets simples. Un **effet principal** est l'effet d'une variable indépendante sur la variable dépendante — en faisant la moyenne de tous les niveaux de l'autre variable indépendante. Il y a donc un effet principal à considérer pour chaque variable indépendante de l'étude. Le panneau supérieur de la

Figure 9.3 montre un effet principal de l'utilisation du téléphone cellulaire, car la performance de conduite était meilleure en moyenne lorsque les personnes participantes n'utilisaient pas de téléphone cellulaire, que lorsqu'elles en utilisaient un. Les barres blanches sont en moyenne plus élevées que les barres noires. L'étude montre également un effet principal de l'heure de la journée, car la performance de conduite était meilleure le jour que la nuit — que les personnes participantes utilisent ou non un téléphone cellulaire. Les effets principaux sont indépendants les uns des autres, en ce sens que l'existence ou non d'un effet principal d'une variable indépendante ne dit rien sur l'existence ou non d'un effet principal de l'autre variable. Le panneau inférieur de la Figure 9.3 par exemple montre un effet principal clair de la durée de la psychothérapie. Plus la psychothérapie est longue, plus elle est efficace.

Interactions

Il y a un effet d'**interaction** (ou simplement « une interaction ») lorsque l'effet d'une variable indépendante dépend du niveau d'une autre. Bien que cela puisse sembler compliqué, vous avez déjà une compréhension intuitive des interactions. Prenons un exemple de la vie quotidienne : votre camarade vous demande d'aller au cinéma avec une autre personne. Vous lui répondez : « Cela dépend du film que vous allez voir et des autres personnes qui viendront ». Vous voulez vraiment voir le film à succès de l'été, mais vous avez peu d'intérêt pour la comédie romantique légère. En d'autres termes, il y a un effet principal du type de film sur votre décision. Si votre décision d'aller voir l'un ou l'autre de ces films dépend aussi des personnes qui sont emmenées, il y a alors une interaction. Par exemple vous accepteriez d'aller voir la comédie romantique si votre camarade amène une personne que vous aimeriez mieux connaître, mais vous refuseriez si elle vient avec quelqu'un d'autre : voilà une interaction. Les interactions médicamenteuses illustrent aussi bien les interactions de la vie quotidienne. De nombreux hommes âgés prennent du Viagra pour les aider dans la chambre à coucher, et de nombreux hommes prennent des nitrates pour traiter l'angine de poitrine ou les douleurs thoraciques. Chacun de ces médicaments est donc bénéfique en soi (il existe des effets principaux de chacun d'eux sur le bien-être des hommes âgés), mais la combinaison de ces deux médicaments peut être mortelle. En d'autres termes, il existe une interaction très importante entre le Viagra et les médicaments pour le cœur que les hommes âgés doivent connaître pour éviter des conséquences graves.

Examinons maintenant quelques exemples d'interactions tirés de la recherche. Il n'est probablement pas surprenant d'apprendre que l'effet d'une psychothérapie est plus fort chez les personnes très motivées à changer que chez celles qui le sont peu. Il s'agit d'une interaction, car l'effet d'une variable indépendante (suivre ou non une psychothérapie) dépend du niveau d'une autre (la motivation à changer). Schnall et ses collègues ont également mis en évidence une interaction, car l'effet de la propreté ou du désordre de la pièce sur les jugements moraux des personnes participantes dépendait du niveau de conscience corporelle privée, faible ou élevée des personnes. Si elles avaient une conscience corporelle privée élevée, les personnes dans la pièce en désordre portaient des jugements plus sévères. Si elles avaient une faible conscience corporelle, l'état de la pièce propre ou en désordre n'avait pas d'effet.

Dans de nombreuses études, la question principale porte sur une interaction. L'étude de Brown et de ses collègues a été inspirée par l'idée que les personnes très hypocondriaques sont particulièrement attentives à toute information négative liée à la santé. Cela a conduit à l'hypothèse que les personnes ayant une hypocondrie élevée se rappelleraient mieux des mots liés à la santé que les personnes ayant une

hypocondrie faible, mais qu'elles se rappelleraient aussi bien des mots sans lien avec la santé que les personnes ayant un faible niveau d'hypocondrie. Et c'est exactement ce que les résultats ont montré.

Types d'interactions

L'effet d'une variable indépendante peut dépendre du niveau de l'autre variable de plusieurs façons. Le premier type s'appelle des **interactions diffuses**. Des exemples d'interactions diffuses sont présentés dans les deux panneaux supérieurs de la Figure 9.4. Dans le panneau supérieur, la variable indépendante « B » a un effet au niveau 1 de la variable indépendante « A » (il y a une différence dans la hauteur des barres blanche et noire à gauche du graphique), mais aucun effet au niveau 2 de la variable indépendante « A » (il n'y a pas de différence dans la hauteur des barres blanche et noire à droite). Cela ressemble beaucoup à l'étude de Schnall et de ses collègues, qui ont constaté un effet du dégoût chez les personnes ayant une conscience corporelle privée élevée, mais pas chez les personnes ayant une conscience corporelle privée faible. Dans le panneau du milieu, la variable indépendante « B » a un effet plus important au niveau 1 de la variable indépendante « A » qu'au niveau 2 (il y a une plus grande différence dans la hauteur des barres blanche et noire à gauche du graphique et une plus petite différence dans la hauteur des barres blanche et noire à droite). Cela ressemble à l'exemple hypothétique de la conduite automobile où l'utilisation d'un téléphone cellulaire la nuit a un effet important, et l'utilisation d'un téléphone cellulaire le jour a un effet plus faible. En résumé dans les interactions diffuses, il y a un effet d'une variable indépendante à un niveau de l'autre variable indépendante, alors qu'il y a soit un effet faible, soit aucun effet de cette variable indépendante à l'autre niveau de l'autre variable indépendante.

Le deuxième type est l'**interaction croisée**. Une interaction croisée est illustrée dans le panneau inférieur de la Figure 9.4 où la variable indépendante « B » a un effet aux deux niveaux de la variable indépendante « A », mais les effets sont dans des directions opposées. Un autre exemple d'interaction croisée provient d'une étude de Kathy Gilliland portant sur l'effet de la caféine sur les scores aux tests de mémoire verbale des personnes introverties et des personnes extraverties (Gilliland, 1980). Les personnes introverties obtiennent de meilleurs résultats au test de mémoire verbale que les personnes extraverties lorsqu'elles n'ont pas consommé de caféine. Par contre, après avoir ingéré 4 mg de caféine par kilogramme de poids corporel, les personnes extraverties obtiennent de meilleurs résultats que les personnes introverties.

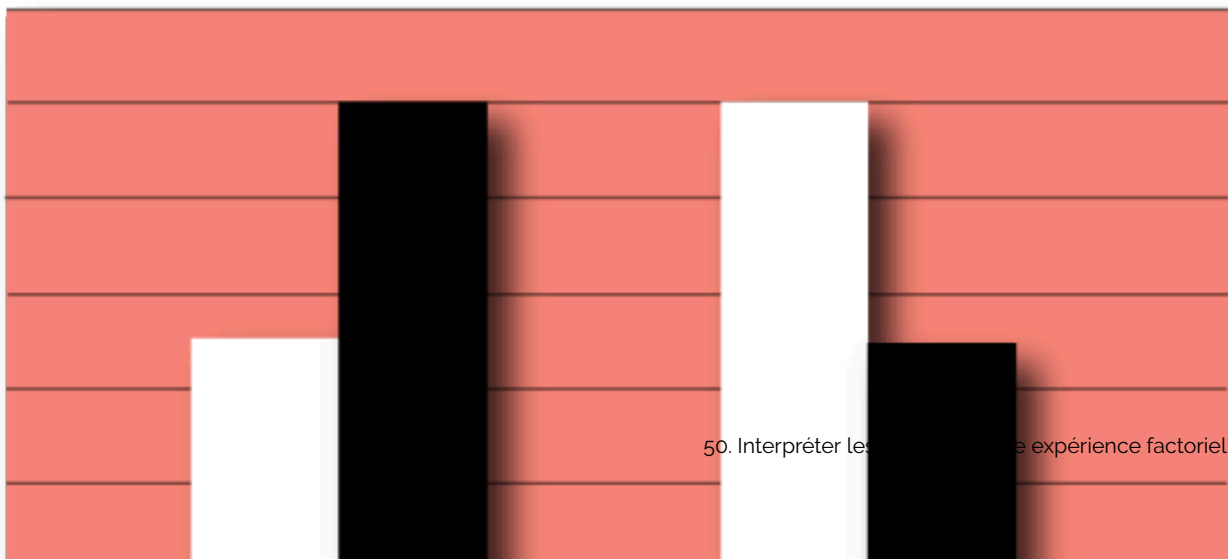
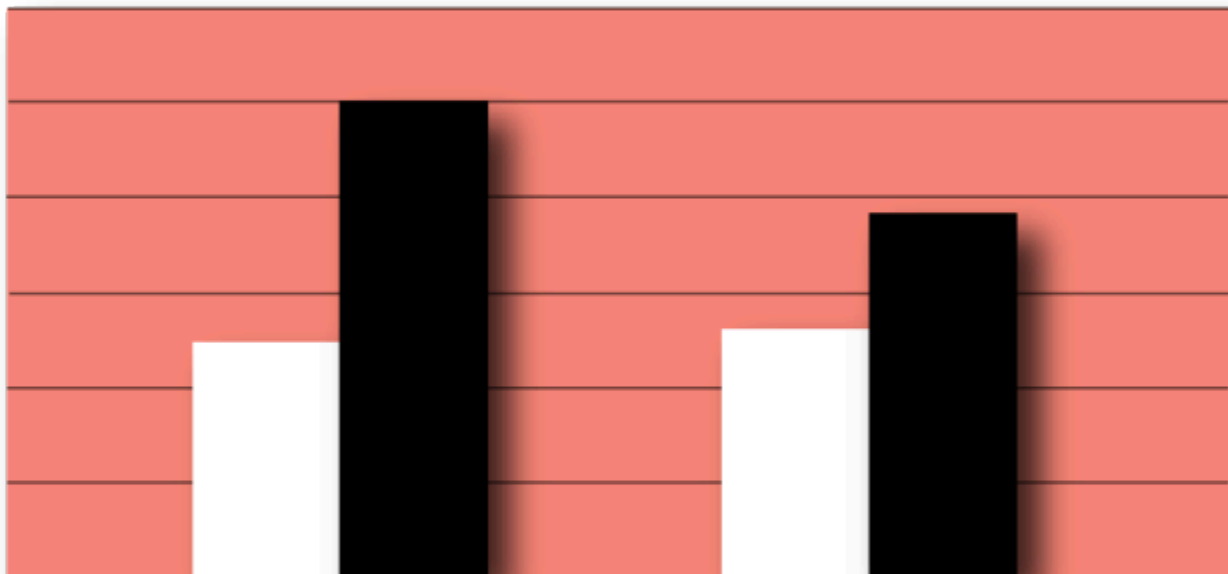
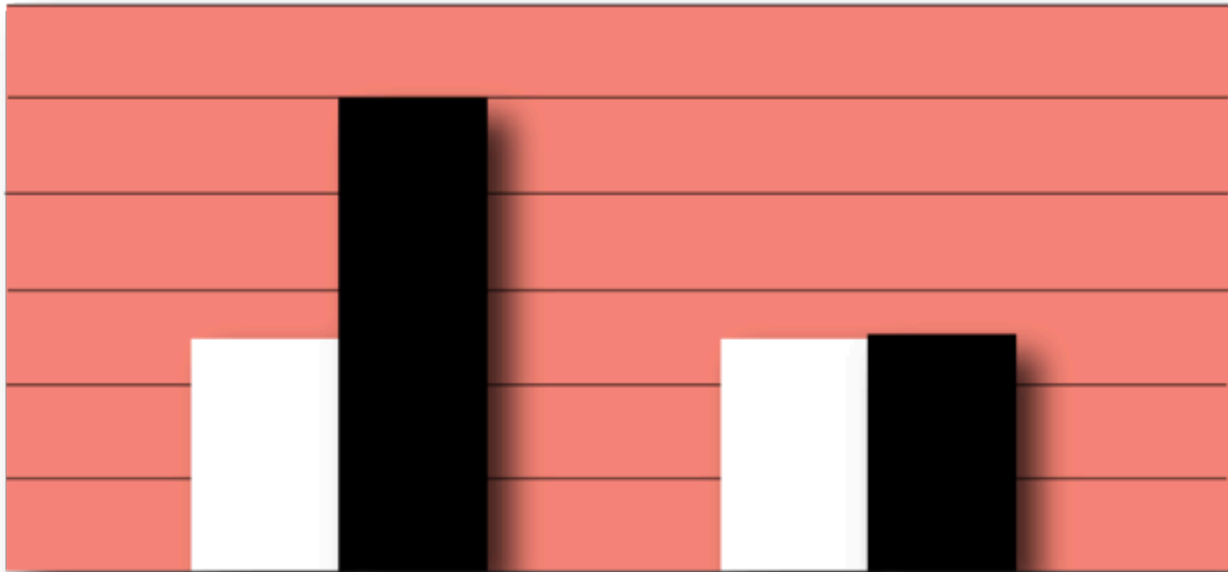


Figure 9.4. Diagrammes à barres montrant trois types d'interactions. Dans le panneau supérieur, une variable indépendante a un effet à un niveau de la deuxième variable indépendante, mais pas à l'autre. Dans le panneau du milieu, une variable indépendante a un effet plus important à un niveau de la deuxième variable indépendante qu'à l'autre. Dans le panneau du bas, une variable indépendante a l'effet inverse à un niveau de la deuxième variable indépendante qu'à l'autre niveau.

La Figure 9.5 présente des exemples de ce même type d'interaction lorsque l'une des variables indépendantes est quantitative et que les résultats sont représentés dans un graphique linéaire. Notons que les deux figures du haut illustrent les deux types d'interaction diffuse que l'on peut trouver, tandis que la figure du bas illustre une interaction croisée (les deux lignes se « croisent » littéralement l'une l'autre).

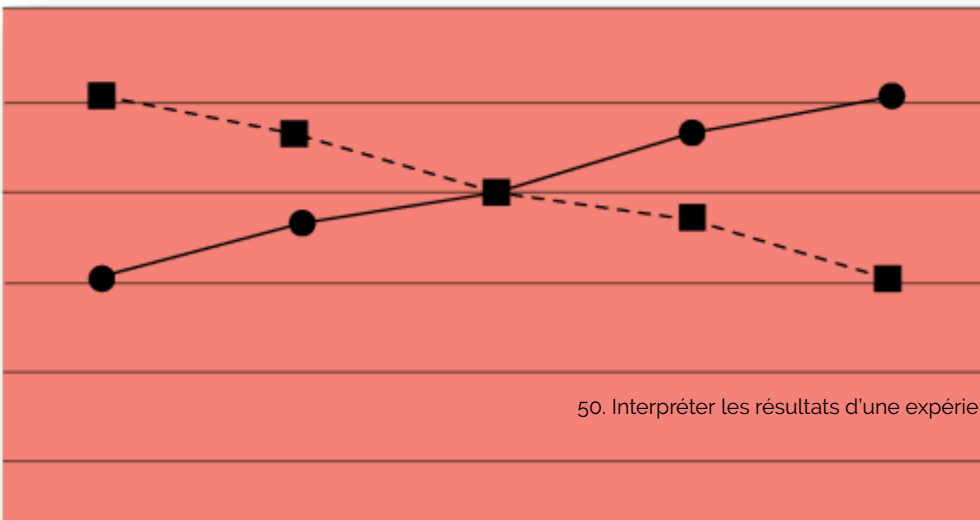
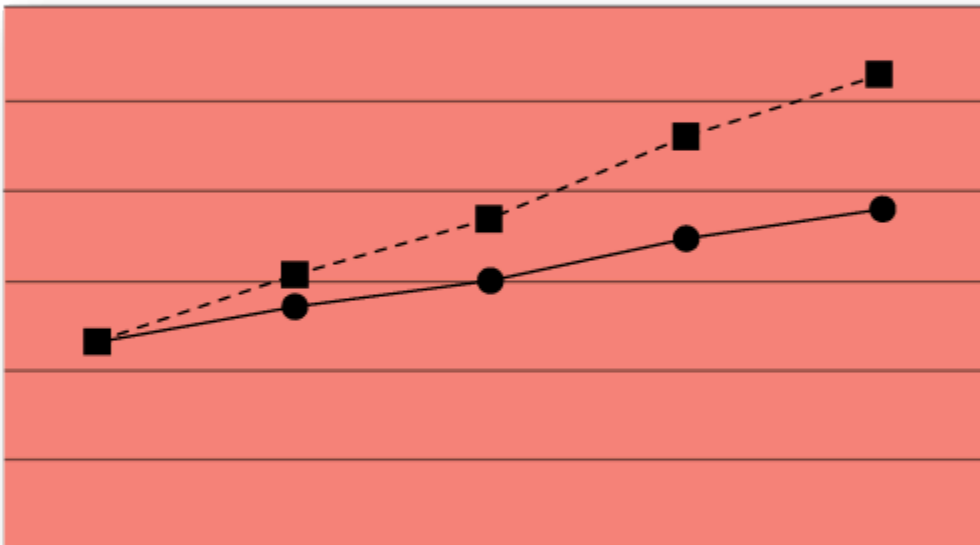
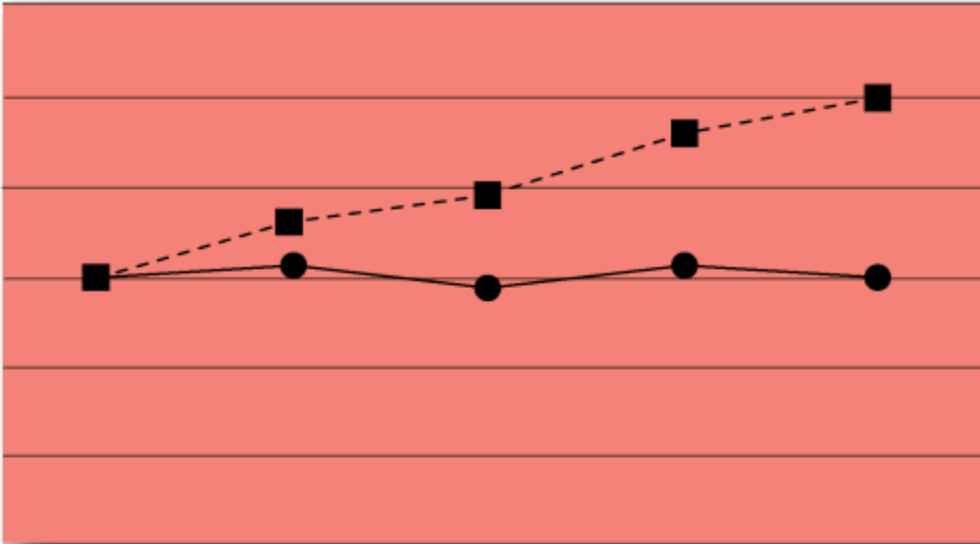


Figure 9.5. Graphiques linéaires illustrant différents types d'interactions. Dans le panneau supérieur, une variable indépendante a un effet à un niveau de la seconde variable indépendante, mais pas à l'autre. Dans le panneau du milieu, une variable indépendante a un effet plus important à un niveau de la deuxième variable indépendante qu'à l'autre. Dans le panneau du bas, une variable indépendante a l'effet inverse à un niveau de la deuxième variable indépendante.
[Description de la figure](#)

Effets simples

Lorsque les scientifiques trouvent une interaction, cela suggère que les effets principaux peuvent être quelque peu trompeurs. Pensons à l'exemple d'une interaction croisée où l'on constate que les personnes introverties obtiennent de meilleurs résultats à un test de mémoire verbale que les personnes extraverties lorsqu'elles n'ont pas ingéré de caféine, mais que les personnes extraverties obtiennent de meilleurs résultats que les personnes introverties lorsqu'elles ont ingéré 4 mg de caféine par kilogramme de poids corporel. Pour examiner l'effet principal de la consommation de caféine, l'équipe de recherche aurait fait la moyenne entre introversion et extraversion et aurait simplement cherché à savoir si dans l'ensemble, les personnes ayant consommé de la caféine étaient mieux ou pires au test de mémoire verbale. Comme l'effet positif de la caféine sur les extraverties serait annulé par les effets négatifs de la caféine sur les introverties, aucun effet principal de la consommation de caféine n'aurait été constaté. De la même façon, pour examiner l'effet principal de la personnalité, l'équipe aurait calculé la moyenne entre les niveaux de la variable caféine afin d'étudier les effets de la personnalité (introversion vs extraversion) indépendamment de la caféine. Dans ce cas, les effets positifs de l'extraversion dans la condition avec caféine seraient annulés par les effets négatifs de l'extraversion dans la condition sans caféine. L'absence d'effets principaux signifie-t-elle qu'il n'y a pas d'effet de la caféine et qu'il n'y a pas d'effet de la personnalité ? Non bien sûr. La présence de l'interaction indique que la situation est plus complexe, que les effets de la caféine sur la performance au test verbal varient selon la personnalité. C'est là que les effets simples entrent en jeu. Les **effets simples** sont un moyen de décomposer l'interaction pour comprendre précisément ce qui se passe. Une interaction nous informe seulement que les effets d'au moins une variable indépendante varient selon le niveau d'une autre variable indépendante. Chaque fois qu'une interaction est détectée, les scientifiques doivent effectuer des analyses supplémentaires pour déterminer d'où provient cette interaction. Bien sûr il est possible de visualiser et d'interpréter l'interaction sur un graphique, mais une analyse des effets simples offre aux scientifiques un moyen plus sophistiqué de décomposer l'interaction. Plus précisément, une analyse des effets simples permet aux scientifiques de déterminer les effets de chaque variable indépendante à chacun des niveaux de l'autre variable indépendante. Ainsi alors qu'une analyse des effets principaux amènerait les scientifiques à faire la moyenne des deux niveaux de la variable de personnalité pour examiner les effets de la caféine sur la performance au test verbal, une analyse des effets simples consisterait à examiner les effets de la caféine chez les personnes introverties, puis les effets de la caféine chez les personnes extraverties. Comme nous l'avons vu précédemment, les personnes en recherche ont également examiné les effets de la personnalité dans la condition sans caféine et ont constaté que, dans cette condition, les introverties obtenaient de meilleurs résultats que les extraverties. Enfin elles ont examiné les effets de la personnalité dans la condition avec caféine et ont constaté que les extraverties obtenaient de meilleurs résultats que les introverties dans

cette condition. Dans le cas d'un plan 2 x 2 comme celui-ci, les scientifiques peuvent étudier deux effets principaux et quatre effets simples.

Schnall et ses collègues ont constaté un effet principal du dégoût sur les jugements moraux (les personnes dans une pièce en désordre portaient des jugements moraux plus sévères). Cependant cette équipe a également découvert une interaction entre la conscience corporelle privée et le dégoût. En d'autres termes, l'effet du dégoût dépendait de la conscience corporelle privée. La présence de cette interaction suggère que l'effet principal peut être quelque peu trompeur. En effet, il n'est pas tout à fait exact de dire que les personnes qui se trouvaient dans une pièce en désordre portaient des jugements moraux plus sévères, car cela n'était vrai que pour la moitié des personnes participantes. En utilisant des analyses des effets simples, Schnall et ses collègues ont pu démontrer que pour les personnes ayant une conscience corporelle privée élevée, il y avait un effet du dégoût sur les jugements moraux. Par contre, pour les personnes ayant une faible conscience corporelle privée, il n'y avait pas d'effet du dégoût sur les jugements moraux. En examinant l'effet du dégoût à chaque niveau de conscience corporelle à l'aide d'analyses des effets simples, Schnall et ses collègues ont été capable de mieux comprendre la nature de l'interaction.

Comme décrit précédemment, Brown et ses collègues ont constaté une interaction entre le type de mots (liés ou non à la santé) et l'hypocondrie (élevée ou faible) sur le rappel des mots. Pour décomposer cette interaction à l'aide d'analyses des effets simples, l'équipe a examiné l'effet de l'hypocondrie à chaque niveau de type de mots. Plus précisément, elle a analysé l'effet de l'hypocondrie sur le rappel de mots liés à la santé, puis elle a examiné l'effet de l'hypocondrie sur le rappel de mots non liés à la santé. Elle a constaté que les personnes ayant une hypocondrie élevée rappelaient plus de mots liés à la santé que les personnes ayant une hypocondrie faible. En revanche, il n'y a pas d'effet de l'hypocondrie sur le rappel de mots non liés à la santé.

Encore une fois, l'examen des effets simples permet de décomposer l'interaction, et il est donc nécessaire d'effectuer ces analyses qu'en présence d'une interaction. Lorsqu'il n'y a pas d'interaction, les effets principaux donnent une image complète et exacte de la situation. En résumé, plutôt que d'établir une moyenne entre les niveaux de l'autre variable indépendante, comme c'est le cas dans une analyse des effets principaux, les analyses des effets simples sont utilisées pour examiner les effets de chaque variable indépendante à chaque niveau de l'autre (ou des autres) variable(s) indépendante(s). Ainsi un scientifique utilisant un plan 2x2 avec quatre conditions devra examiner 2 effets principaux et 4 effets simples. Un scientifique utilisant un plan 2x3 avec six conditions devra étudier 2 effets principaux et 5 effets simples, tandis qu'un scientifique utilisant un plan 3x3 avec neuf conditions devra étudier 2 effets principaux et 6 effets simples. Comme on peut le constater, le nombre d'effets principaux dépend simplement du nombre de variables indépendantes incluses (un effet principal peut être exploré pour chaque variable indépendante), le nombre d'analyses des effets simples dépend du nombre de niveaux des variables indépendantes (parce qu'une analyse distincte de chaque variable indépendante est effectuée à chaque niveau de l'autre variable indépendante).

Pour plus d'informations sur les plans factoriels, les interactions et les tests d'effets simples, consultez les ressources d'unipeds : <https://unipeds.net/ressource/c9-plans-a-plusieurs-facteurs/>

Descriptions de l'image

Description de l'image de la Figure 9.5 : Trois panneaux, chacun montrant un modèle de graphique linéaire différent. Dans le panneau supérieur, une ligne reste constante tandis que l'autre augmente. Dans le panneau du milieu, les deux lignes augmentent, mais à des rythmes différents. Dans le panneau du bas, une ligne descend et l'autre monte, de sorte qu'elles se croisent. [\[Retour à Figure 9.5\]](#)

51. Points clés et exercices

Points clés

- Les personnes en recherche incluent souvent plusieurs variables indépendantes dans leurs expériences. L'approche la plus courante est le plan factoriel, dans lequel chaque niveau d'une variable indépendante est combiné avec chaque niveau des autres variables pour créer toutes les conditions possibles.
- Chaque variable indépendante peut être manipulée entre les sujets ou à l'intérieur des sujets.
- Des variables indépendantes non manipulées (le genre) peuvent être incluses dans les plans factoriels, mais elles limitent les conclusions causales qui peuvent être tirées quant aux effets de la variable non manipulée sur la variable dépendante.
- Dans un plan factoriel, l'effet principal d'une variable indépendante est son effet global moyen sur toutes les autres variables indépendantes. Il y a un effet principal pour chaque variable indépendante.
- Il y a une interaction entre deux variables indépendantes lorsque l'effet de l'une dépend du niveau de l'autre. Certaines des questions de recherche et certains des résultats les plus intéressants en psychologie concernent spécifiquement les interactions.
- Une analyse des effets simples permet aux scientifiques de décomposer les interactions en examinant l'effet de chaque variable indépendante à chaque niveau de l'autre variable indépendante.

Exercises

- **Pratique** : Reprenez les cinq titres d'articles présentés au début de cette section. Pour chacun d'entre eux, identifiez les variables indépendantes et la variable dépendante.
- **Pratique** : Créez un tableau des plans factoriels pour une expérience sur les effets de la température ambiante et du niveau de bruit sur les performances au MCAT. Le MCAT, Medical College Admission Test, est un test de résolution de problèmes, de pensée critique et de connaissances en science naturelle (physique, chimie, biologie), comportementale et sociale. Pour l'exercice pratique, assumez que répondre au test vous donnera un score MCAT. Assurez-vous d'indiquer si chaque variable indépendante sera manipulée entre les sujets ou à l'intérieur des sujets et expliquez pourquoi.
 - [Site internet canadien](#)
 - [Site américain \[en anglais\]](#)
- **Pratique** : Dessinez 8 histogrammes différents pour illustrer chacun des résultats possibles suivants dans une expérience factorielle 2×2 :
 - Pas d'effet principal de A ; pas d'effet principal de B ; pas d'interaction
 - Effet principal de A ; pas d'effet principal de B ; pas d'interaction
 - Pas d'effet principal de A ; effet principal de B ; pas d'interaction
 - Effet principal de A ; effet principal de B ; pas d'interaction
 - Effet principal de A ; effet principal de B ; interaction
 - Effet principal de A ; pas d'effet principal de B ; interaction
 - Pas d'effet principal de A ; effet principal de B ; interaction
 - Pas d'effet principal de A ; pas d'effet principal de B ; interaction



Lecture sur version imprimée?
Scannez ce code QR avec votre
cellulaire ou allez sur
[https://www.afmc.ca/fr/
resources-data/education/
aspirants-applicants/](https://www.afmc.ca/fr/resources-data/education/aspirants-applicants/)



Lecture sur version imprimée?
Scannez ce code QR avec votre
cellulaire ou allez sur
[https://students-residents.aamc
.org/about-mcat-exam/
about-mcat-exam](https://students-residents.aamc.org/about-mcat-exam/about-mcat-exam)

52. Références

- Brown, H. D., Kosslyn, S. M., Delamater, B., Fama, A. et Barsky, A. J. (1999). Perceptual and memory biases for health-related information in hypochondriacal individuals (Biais de perception et de mémoire pour les informations relatives à la santé chez les personnes hypocondriaques). *Journal of Psychosomatic Research*, 47(1), 67-78. [https://doi.org/10.1016/S0022-3999\(99\)00011-2](https://doi.org/10.1016/S0022-3999(99)00011-2)
- MacDonald, T. K. et Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? (L'estime de soi, l'humeur et l'intention d'utiliser des préservatifs : Quand une faible estime de soi conduit-elle à des comportements de santé à risque ?) *Journal of Experimental Social Psychology*, 38(3), 299-306. <https://doi.org/10.1006/jesp.2001.1505>
- Gilliland, K. (1980). The interactive effect of introversion-extraversion with caffeine induced arousal on verbal performance (L'effet interactif de l'introversion-extraversion et de l'excitation induite par la caféine sur la performance verbale). *Journal of Research in Personality*, 14(4), 482-492. [https://doi.org/10.1016/0092-6566\(80\)90006-9](https://doi.org/10.1016/0092-6566(80)90006-9)
- Schnall, S., Benton, J. et Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments (Avec une conscience propre : La propreté réduit la sévérité des jugements moraux). *Psychological Science*, 19(12), 1219-1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>
- Schnall, S., Haidt, J., Clore, G. L. et Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096-1109. <https://doi.org/10.1177/0146167208317771>

CHAPITRE 10-LA RECHERCHE À CAS UNIQUE



Image 10. Groupe d'enfants en classe

Vance Hall et ses collègues ont dû relever le défi d'augmenter le temps de concentration sur le travail scolaire de six élèves du primaire présentant des comportements perturbateurs (Hall, Lund et Jackson, 1968). Pendant plusieurs jours, l'équipe de recherche a soigneusement noté si chaque élève faisait ou non son travail scolaire toutes les 10 secondes pendant une période de 30 minutes. Après avoir établi ce niveau de base, ces personnes ont introduit un traitement. Ce traitement consiste à ce que, lorsque l'élève fait son travail scolaire, le personnel enseignant lui accorde une attention positive sous la forme d'un commentaire tel que « bon travail » ou d'une tape sur l'épaule. Les résultats ont montré que tous les élèves augmentaient le temps consacré à leur travail scolaire et diminuaient leurs comportements perturbateurs pendant la phase de traitement. Par exemple l'élève Robbie consacrait au départ 25 % de son temps au travail scolaire et les 75 % restants « à faire claquer des élastiques, à jouer avec des jouets tirés de sa poche, à parler et à rire avec ses camarades » (Hall et al., 1968, p. 3, traduction). Durant la phase de traitement, l'élève Robbie a consacré 71 % de son temps au travail scolaire et seulement 29 % à d'autres activités. Enfin lorsque l'équipe a

demandé au personnel enseignant de cesser d'offrir cette attention positive, tous les élèves ont diminué leur temps d'études et augmenté leurs comportements perturbateurs. Cela a confirmé que c'était bien l'attention positive qui était responsable de l'augmentation du temps d'étude. Il s'agit d'une des premières études à montrer que le fait de prêter attention à un comportement adéquat — et d'ignorer des comportements inadéquats — peut être un moyen rapide et efficace d'intervenir sur des comportements problématiques dans un contexte appliqué.

La majeure partie de ce livre traite de ce que l'on peut appeler la recherche sur les groupes, qui consiste généralement à étudier un grand échantillon et à combiner leurs données pour en tirer des conclusions générales sur le comportement humain. L'étude de Hall et de ses collègues illustre à l'inverse une recherche utilisant un protocole à cas unique qui examine généralement un petit échantillon et s'intéresse de près à chaque personne. Dans ce chapitre, nous examinons cette approche alternative. Nous commencerons par donner un aperçu de la recherche à cas unique notamment des postulats sur lesquels elle repose, des personnes qui la pratiquent et des raisons pour lesquelles elles la pratiquent. Nous examinerons ensuite quelques protocoles à cas unique connus et la manière dont les données issues de ces devis sont analysées. Enfin nous examinerons certains des points forts et des points faibles de cette recherche à cas unique par rapport à la recherche sur les groupes, et nous verrons comment ces deux approches peuvent se compléter.

53. Aperçu de la recherche à cas unique

Objectifs d'apprentissage

1. Expliquer ce qu'est la recherche à cas unique et en quoi elle diffère des autres types de recherche en psychologie.
2. Expliquer qui utilise la recherche à cas unique et pourquoi.

Qu'est-ce que la recherche à cas unique ?

La **recherche à cas unique** est un type de recherche quantitative qui consiste à étudier en détail le comportement de chaque membre d'un petit nombre de personnes. Notons que le terme « cas unique » ne signifie pas qu'une seule personne est étudiée ; il est plus courant qu'il y ait entre deux et dix personnes. (C'est pourquoi les protocoles de recherche à cas unique sont parfois appelés devis à petit n , où n est le symbole statistique de la taille de l'échantillon.) La recherche à cas unique peut être opposée à la **recherche sur les groupes**, qui implique généralement l'étude d'un grand nombre de personnes et l'examen de leur comportement principalement en matière de moyennes de groupe, d'écart-types, etc. La majeure partie de ce livre est consacrée à la compréhension de la recherche sur les groupes, qui est l'approche la plus courante en psychologie. Toutefois la recherche à cas unique est une alternative importante et constitue l'approche principale dans certains domaines plus appliqués de la psychologie.

Avant de poursuivre, il est important de distinguer la recherche à cas unique des études de cas et d'autres approches plus qualitatives impliquant l'étude détaillée d'un petit échantillon de personnes participantes. Comme décrit au chapitre 6, les études de cas impliquent une analyse et une description approfondies d'une personne, ce qui est généralement de nature essentiellement qualitative. D'une manière plus générale, la recherche qualitative s'attache à comprendre l'expérience subjective des personnes en observant leur comportement et en recueillant des données relativement peu structurées (par exemple des entrevues détaillées), puis en analysant ces données à l'aide de techniques narratives plutôt que quantitatives. En revanche, la recherche à cas unique se concentre sur la compréhension d'un comportement objectif par la manipulation et le contrôle expérimental, la collecte de données hautement structurées et l'analyse de ces données de manière quantitative.

Postulats des protocoles à cas unique

Encore une fois, le protocole à cas unique implique l'étude d'un petit nombre de personnes et l'observation intensive du comportement de chaque personne. Mais pourquoi adopter cette approche plutôt que l'approche par groupes? Le protocole à cas unique repose sur plusieurs postulats importants et il sera utile de les examiner maintenant.

Tout d'abord le **premier postulat** est qu'il est essentiel de se concentrer intensivement sur le comportement de chaque personne. Une des raisons est que la recherche sur les groupes peut masquer les différences individuelles et générer des résultats qui ne représentent le comportement d'aucune personne en particulier. Par exemple un traitement qui a un effet positif pour la moitié des personnes qui y sont exposées, mais un effet négatif pour l'autre moitié semblerait en moyenne n'avoir aucun effet. Cependant la recherche à cas unique révélerait probablement ces différences individuelles. Une deuxième raison de se concentrer intensivement sur chaque personne est que parfois c'est le comportement d'une personne particulière qui est le plus intéressant. Le spécialiste en psychologie scolaire peut par exemple vouloir modifier le comportement d'une personne ayant un comportement perturbateur à l'école. Bien que les recherches publiées (qu'elles soient à cas unique ou sur des groupes) puissent fournir des indications sur la manière d'y parvenir, mener une étude directement sur cet élève serait plus direct et probablement plus efficace.

Un **deuxième postulat** de la recherche à cas unique est qu'il est important de découvrir des relations causales par la manipulation d'une variable indépendante, la mesure minutieuse d'une variable dépendante et le contrôle des variables externes. C'est pourquoi la recherche à cas unique est souvent considérée comme un type de recherche expérimentale présentant une bonne validité interne. Rappelons par exemple que Hall et ses collègues ont mesuré leur variable dépendante (le temps d'étude) à plusieurs reprises, d'abord dans une condition contrôle sans traitement, puis dans une condition de traitement (attention positive du personnel enseignant), puis à nouveau dans une condition contrôle. Étant donné qu'il y a eu une nette augmentation du temps d'étude lorsque le traitement a été introduit, une diminution lorsqu'il a été supprimé et une augmentation lorsqu'il a été réintroduit, il n'y a guère de doute que le traitement était la cause de l'amélioration.

Un **troisième postulat** de la recherche à cas unique est qu'il est important d'étudier des effets robustes et constants qui ont une importance biologique ou sociale. Les personnes investies en recherches appliquées en particulier s'intéressent aux traitements qui ont des effets substantiels sur des comportements importants, et qui peuvent être appliqués de manière fiable dans les contextes réels où ils se produisent. C'est ce que l'on appelle parfois la **validité sociale** (Wolf, 1976). L'étude de Hall et de ses collègues par exemple possédait une bonne validité sociale parce qu'elle montrait des effets robustes et constants de l'attention positive du personnel enseignant sur un comportement d'importance évidente pour le personnel enseignant, les parents et les élèves. De plus le personnel enseignant a trouvé le traitement facile à réaliser, même dans leurs classes du primaire souvent agitées.

Qui utilise la recherche à cas unique ?

La recherche à cas unique existe depuis aussi longtemps que la psychologie elle-même. À la fin des années

1800, Wilhelm Wundt l'un des fondateurs de la psychologie a étudié les sensations et la conscience en se concentrant intensivement sur chacune des personnes participantes de ses recherches. Les recherches de Herman Ebbinghaus sur la mémoire et celles d'Ivan Pavlov sur le conditionnement classique sont d'autres exemples classiques, encore décrits dans presque tous les manuels d'introduction à la psychologie.

Au milieu du 20^e siècle, B. F. Skinner a clarifié plusieurs postulats sous-jacents à la recherche à cas unique et en a perfectionné de nombreuses techniques (Skinner, 1938). Lui et d'autres scientifiques l'ont ensuite utilisée pour décrire comment les récompenses, les punitions et d'autres facteurs externes affectent le comportement au fil du temps. Ces travaux ont été menés principalement sur des sujets non humains — surtout des rats et des pigeons. Cette approche, que Skinner a appelée **l'analyse expérimentale du comportement**, reste un sous-domaine important de la psychologie et continue de s'appuyer presque exclusivement sur la recherche à cas unique. Vous trouverez d'excellents exemples de ces travaux dans n'importe quel numéro de la revue d'analyse expérimentale du comportement (*Journal of the Experimental Analysis of Behavior*). À partir des années 1960, de nombreux scientifiques se sont intéressés à l'utilisation de cette approche pour mener des recherches appliquées principalement auprès d'humains — un sous-domaine désormais appelé **l'analyse appliquée du comportement** (Baer, Wolf, et Risley, 1968). L'analyse appliquée du comportement joue un rôle particulièrement important dans la recherche contemporaine sur les troubles du développement, l'éducation, le comportement organisationnel et la santé, parmi de nombreux autres domaines. D'excellents exemples de ces travaux (y compris l'étude menée par Hall et de ses collègues) peuvent être consultés dans la revue d'analyse appliquée du comportement (*Journal of Applied Behavior Analysis*).

Bien que la plupart des recherches à cas unique contemporaines soient menées dans une perspective comportementale, elles peuvent en principe être utilisées pour répondre à des questions issues de n'importe quelle approche théorique. Par exemple une technique d'apprentissage fondée sur des principes cognitifs de l'apprentissage et de la mémoire pourrait être évaluée en l'appliquant à des élèves du secondaire selon l'approche à cas unique. Cette approche peut également être utilisée par des personnes cliniciennes adoptant n'importe quelle perspective théorique — comportementale, cognitive, psychodynamique ou humaniste — pour étudier les processus de changement thérapeutique de chaque personne faisant partie de leur clientèle, et pour documenter leur amélioration (Kazdin, 1982).

54. Protocole de recherche à cas unique

Objectifs d'apprentissage

1. Décrire les éléments fondamentaux d'un protocole de recherche à cas unique.
2. Concevoir de simples recherches à cas unique en utilisant des protocoles d'inversion et de niveaux de base multiples.
3. Expliquer comment les protocoles de recherche à cas unique abordent la question de la validité interne.
4. Interpréter les résultats de simples recherches à cas unique à l'aide de l'inspection visuelle de données représentées graphiquement.

Caractéristiques générales des protocoles à cas unique

Avant d'examiner des protocoles spécifiques de recherche à cas unique, il est utile de présenter certaines caractéristiques communes à la plupart d'entre eux. Plusieurs de ces caractéristiques communes sont illustrées dans la Figure 10.1, qui présente les résultats d'une étude générique à cas unique. Premièrement la variable dépendante (représentée sur l'axe des y) est mesurée de manière répétée dans le temps (représenté par l'axe des x) à intervalles réguliers. Deuxièmement l'étude est divisée en phases distinctes et la personne participante est testée dans une seule condition par phase. Les conditions sont souvent désignées par des lettres majuscules : A, B, C, etc. Ainsi la Figure 10.1 illustre un protocole dans lequel la personne participante a d'abord été testée dans une condition (A), puis dans une autre condition (B) et enfin à nouveau dans la condition initiale (A). (Il s'agit d'un protocole d'inversion, qui sera décrit plus en détail plus loin.)

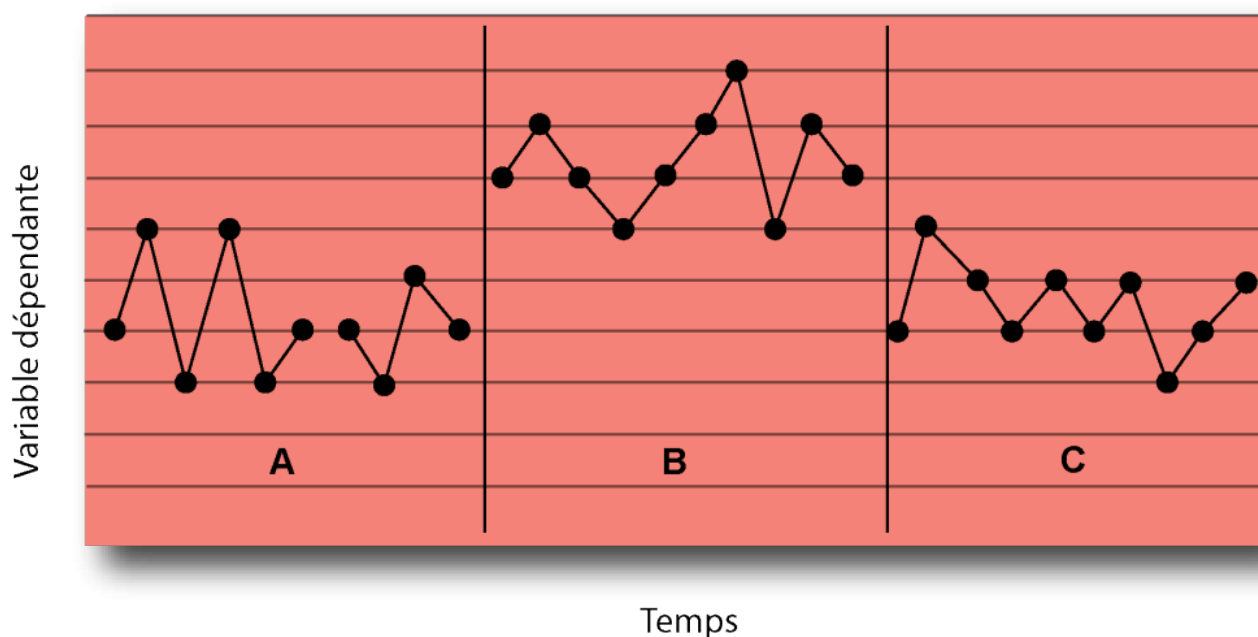


Figure 10.1. Résultats d'une étude générique à cas unique illustrant plusieurs principes de la recherche à cas unique

Un autre aspect important de la recherche à cas unique est que le passage d'une condition à l'autre ne se produit généralement pas après un laps de temps ou un nombre d'observations fixes. Il dépend plutôt du comportement de la personne participante. Plus précisément les scientifiques attendent que le comportement de cette personne dans une condition devienne relativement stable d'une observation à l'autre avant de changer de condition. C'est ce qu'on appelle parfois la **stratégie de l'état stable** (Sidman, 1960). L'idée est que lorsque la variable dépendante a atteint un état stable, tout changement d'une condition à l'autre sera relativement facile à détecter. Rappelons que nous avons rencontré ce même principe lorsque nous avons abordé la recherche expérimentale en général. L'effet d'une variable indépendante est plus facile à détecter lorsque le « bruit » des données est minimisé.

Protocoles d'inversion

Le protocole de recherche à cas unique le plus simple est le **protocole d'inversion**, également appelé «**protocole ABA**». Pendant la première phase (A), un **niveau de base** est établi pour la variable dépendante. Il s'agit du niveau de réponse avant l'introduction de tout traitement, et la phase du niveau de base joue donc le rôle de condition contrôle. Lorsque l'état stable est atteint, la phase B commence, le traitement est introduit. Il peut y avoir une période d'ajustement au traitement pendant laquelle le comportement étudié varie davantage et commence à augmenter ou diminuer. Là encore les scientifiques attendent que la variable dépendante atteigne un état stable afin de déterminer clairement si elle a changé et dans quelle mesure. Enfin les scientifiques retirent le traitement et attendent à nouveau que la variable dépendante retrouve un

état stable. Ce protocole de base peut également être prolongé avec la réintroduction du traitement (ABAB), un autre retour au niveau de base (ABABA), et ainsi de suite.

L'étude de Hall et de ses collègues a utilisé un protocole d'inversion ABAB. La Figure 10.2 illustre approximativement les données de Robbie. Le pourcentage de temps consacré à l'étude (variable dépendante) était faible pendant le premier niveau de base, a augmenté pendant la première phase de traitement jusqu'à ce qu'il se stabilise, a diminué pendant le deuxième niveau de base, puis a de nouveau augmenté pendant la deuxième phase de traitement.

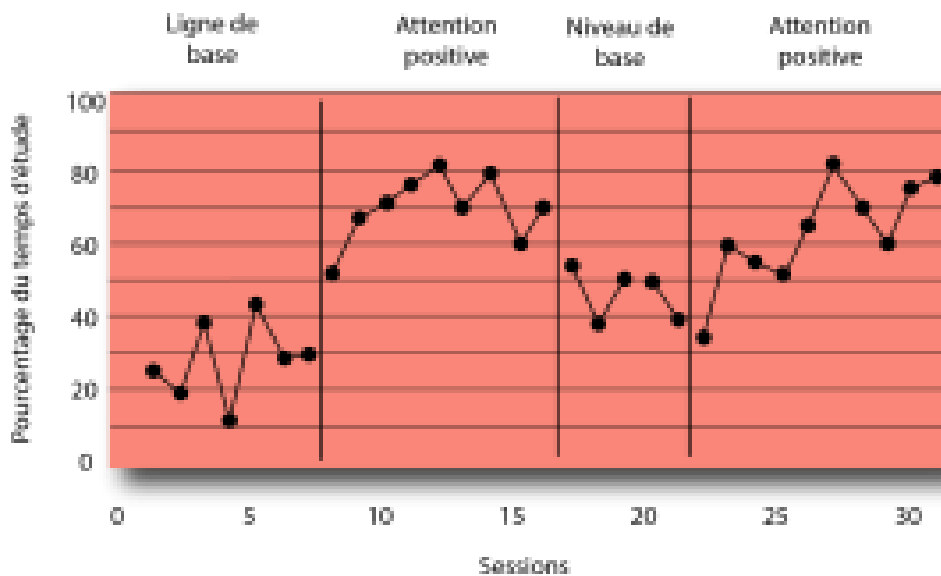


Figure 10.2. Résultats approximatifs de Robbie dans le protocole d'inversion ABAB de Hall et ses collègues
[\[Description de la figure\]](#)

Pourquoi l'inversion — c'est-à-dire le retrait du traitement — est-il considéré comme nécessaire dans ce type de devis ? Pourquoi utiliser un protocole ABA par exemple plutôt qu'un simple protocole AB ? Notons qu'un protocole AB est essentiellement un protocole à série temporelle interrompue appliqué à une personne participante. Rappelons que l'un des problèmes de ce type de protocole est que si la variable dépendante change après l'introduction du traitement, il n'est pas toujours clair que le traitement en soit la cause. Il est possible que quelque chose d'autre ait changé à peu près au même moment, et que cette variable externe soit responsable du changement de la variable dépendante. Mais si la variable dépendante change avec l'introduction du traitement, puis change à nouveau avec le retrait du traitement (en supposant que le traitement ne crée pas d'effet permanent), il est beaucoup plus clair que le traitement (et son retrait) en soit la cause. En d'autres mots, l'inversion augmente considérablement la validité interne de l'étude.

Il existe des variantes proches du protocole de base d'inversion qui permettent d'évaluer plus d'un traitement. Dans un **protocole d'inversion de traitements multiples**, une phase de niveau de base est suivie de phases distinctes au cours desquelles différents traitements sont introduits. Par exemple, des scientifiques peuvent établir un niveau de base du comportement d'étude pour une personne présentant des comportements perturbateurs à l'école (A), puis introduire un traitement impliquant une attention positive de la part du personnel enseignant (B), et ensuite passer à un traitement impliquant une légère punition pour ne pas avoir étudié (C). L'élève pourrait ensuite être revenu à une phase de niveau de base avant de réintroduire

chaque traitement — peut-être dans l'ordre inverse afin de contrôler les effets reportés (*carryover effects*, en anglais). Ce protocole d'inversion de traitements multiples peut également être appelé protocole ABCACB.

Dans un **protocole à traitements alternés**, deux traitements ou plus sont alternés relativement rapidement selon un horaire régulier. Par exemple une attention positive pour l'étude pourrait être utilisée un jour et une légère punition pour ne pas avoir étudié le lendemain, et ainsi de suite. Ou encore un traitement pourrait être appliqué le matin et un autre l'après-midi. Le protocole à traitements alternés peut être un moyen rapide et efficace de comparer des traitements, mais uniquement lorsque les traitements produisent leurs effets rapidement.

Protocoles à niveaux de base multiples

Le protocole d'inversion présente deux problèmes potentiels — liés au retrait du traitement. Le premier est que si un traitement est efficace, il peut être contraire à l'éthique de le retirer. Par exemple si un traitement semble réduire l'incidence de l'automutilation chez un enfant ayant un handicap intellectuel, il serait contraire à l'éthique de retirer ce traitement simplement pour montrer que l'incidence de l'automutilation augmente. Le deuxième problème est que la variable dépendante peut ne pas revenir à son niveau de base après le retrait du traitement. Par exemple lorsque l'attention positive accordée aux études est retirée, l'élève peut continuer à étudier à un rythme plus élevé. Cela pourrait signifier que l'attention positive a eu un effet durable sur le temps d'étude de l'élève, ce qui serait évidemment une bonne chose. Mais cela peut aussi signifier que l'attention positive n'était en réalité pas la cause de l'augmentation du temps d'étude. Il se peut que quelque chose d'autre se soit produit à peu près au même moment que le traitement — par exemple les parents de l'élève ont pu commencer à offrir des récompenses pour ses bonnes notes. Une solution consiste à utiliser un **protocole à niveaux de base multiples**, lequel est représenté à la Figure 10.3. Il existe trois différents types de protocoles à niveaux de base multiples, que nous allons maintenant examiner.

Protocole à niveaux de base multiples parmi les personnes

Dans une version de ce protocole, le niveau de base est établi pour toutes les personnes participantes, puis le traitement est introduit pour chacune d'entre elles. Essentiellement chaque personne est testée dans un protocole AB. L'élément clé de ce protocole est que le traitement est introduit à des *moments* différents pour chaque personne. L'idée est que si la variable dépendante change lors de l'introduction du traitement pour une personne, cela pourrait être une coïncidence. Mais si la variable dépendante change lorsque le traitement est introduit pour plusieurs personnes — en particulier lorsque le traitement est introduit à des moments différents pour les différentes personnes — alors il est peu probable qu'il s'agisse d'une coïncidence.

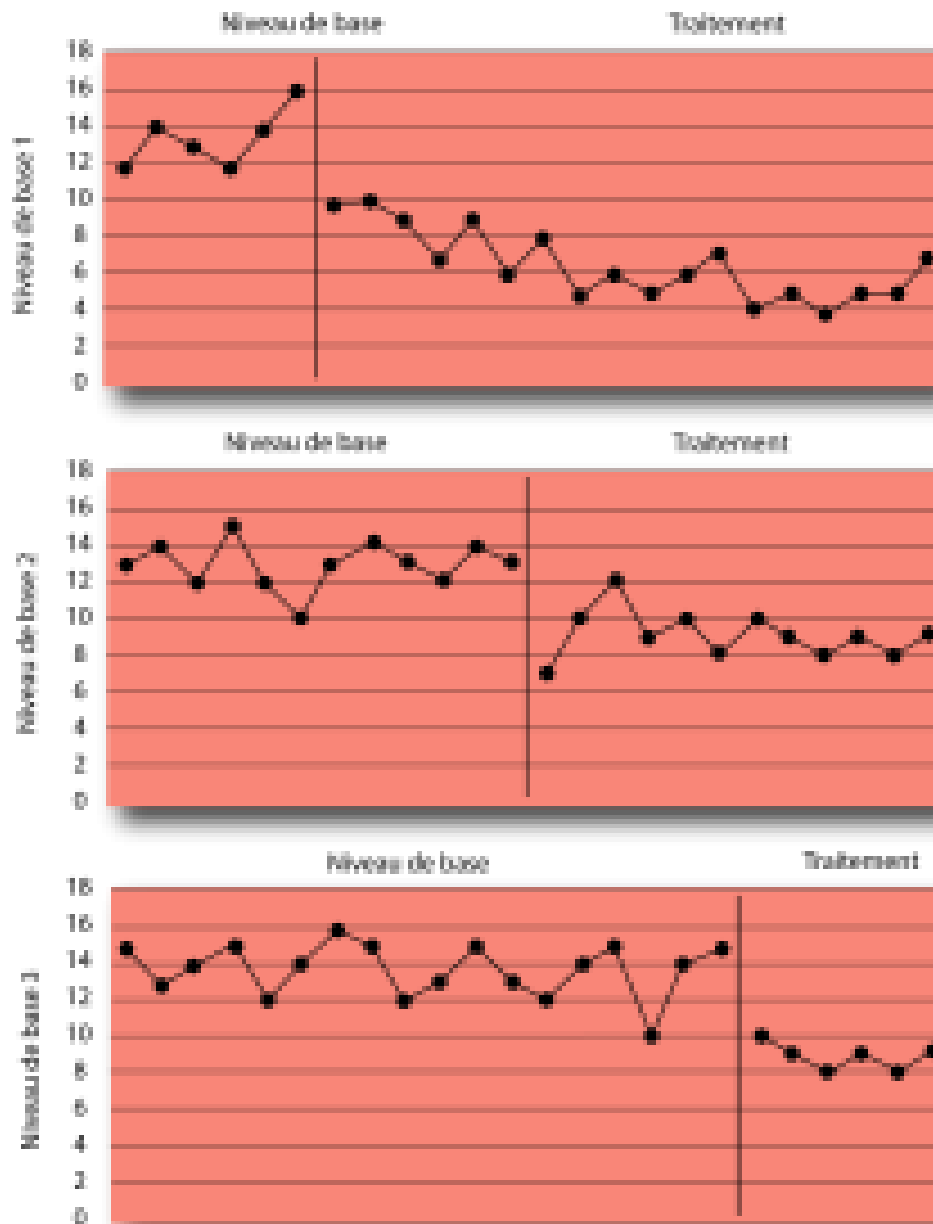


Figure 10.3. Résultats d'une étude générique à bases multiples. Les niveaux de base multiples peuvent concerner des participants, des variables dépendantes ou des contextes différents. Le traitement est introduit à un moment différent pour chaque niveau de base.
[\[Description de la figure\]](#)

Prenons l'exemple d'une étude menée par Scott Ross et Robert Horner (Ross et Horner, 2009). Ces scientifiques s'intéressaient à la manière dont un programme de prévention de l'intimidation dans les écoles affectait les comportements d'intimidation d'élèves à problèmes particuliers. Dans trois écoles différentes, ces scientifiques ont étudié deux élèves qui participaient régulièrement à des comportements d'intimidation. Pendant la phase de niveau de base, l'équipe de recherche a observé ces élèves chaque jour pendant des périodes de 10 minutes durant la récréation de midi, et a compté le nombre de comportements agressifs manifestés envers leurs pairs. Après deux semaines, l'équipe a mis en place le programme dans une première école. Deux semaines plus tard, elle l'a instauré dans la deuxième école. Et deux semaines plus

tard encore, elle l'a instauré dans la troisième école. L'équipe a constaté que le nombre de comportements agressifs manifestés par chaque élève diminuait peu de temps après la mise en œuvre du programme dans l'école de l'élève. Remarquons que si l'équipe de recherche n'avait étudié qu'une seule école ou si elle avait introduit le traitement simultanément dans les trois écoles, il aurait été difficile de déterminer si la réduction des comportements agressifs était due au programme contre l'intimidation ou à quelque chose d'autre s'étant produit à peu près au même moment où il a été introduit (par exemple un congé, une émission de télévision, un changement de météo). Mais grâce au protocole à niveaux de base multiples, ce type de coïncidence aurait dû se produire à trois fois séparément — une occurrence très improbable — pour expliquer les résultats.

Protocole à niveaux de base multiples à travers les comportements

Dans une autre version du protocole à niveaux de base multiples, plusieurs niveaux de base sont établis pour la même personne participante, mais pour des variables dépendantes différentes, et le traitement est introduit à un moment différent pour chaque variable dépendante. Imaginons par exemple une étude visant à mesurer l'effet de la fixation d'objectifs clairs sur la productivité d'un membre du personnel de bureau ayant deux tâches principales: effectuer des visites commerciales et rédiger des rapports. Des niveaux de base pourraient être établis pour les deux tâches. Par exemple les scientifiques pourraient mesurer le nombre de visites commerciales effectuées et de rapports rédigés par le membre du personnel chaque semaine pendant plusieurs semaines. Ensuite le traitement, étant de définir des objectifs clairs, pourrait être introduit pour l'une des tâches à effectuer, et plus tard le même traitement pourrait être introduit pour l'autre tâche. La logique est la même que précédemment. Si la productivité augmente pour l'une des tâches après l'introduction du traitement, il n'est pas certain que le traitement soit à l'origine de cette augmentation. En revanche si la productivité augmente pour les deux tâches après l'introduction du traitement — en particulier lorsque le traitement est introduit à deux moments différents — alors il semble beaucoup plus clair que le traitement en est la cause.

Protocole à niveaux de base multiples dans différents contextes



Lecture sur version
imprimée? Scannez ce code
QR avec votre cellule ou
allez sur
[https://www.erudit.org/fr/
revues/psyedu/
2013-v42-n1-psyedu04724/
1061728ar/](https://www.erudit.org/fr/revues/psyedu/2013-v42-n1-psyedu04724/1061728ar/)

Dans une troisième version du protocole à niveaux de base multiples, plusieurs niveaux de base sont établis pour la même personne, mais dans des contextes différents. Par exemple un niveau de base peut être établi pour le temps qu'un enfant consacre à la lecture pendant son temps libre à l'école, et pendant son temps libre à la maison. Ensuite un traitement tel que l'attention positive peut être introduit d'abord à l'école et ensuite à la maison. Là encore si la variable dépendante change après l'introduction du traitement dans chacun des contextes, cela donne aux scientifiques la confiance que le traitement est effectivement responsable de ce changement.

Pour lire sur diverses utilisations de différents protocoles expérimentaux à cas unique, consultez

[Lanovaz, M.J. \(2013\). L'utilisation de devis expérimentaux à cas unique en psychoéducation. Revue de psychoéducation, 42\(1\), 161-183. DOI:10.7202/1061728ar](#)

Analyse des données dans la recherche à cas unique

En plus du fait qu'elle se concentre sur les personnes, la recherche à cas unique se distingue de la recherche sur les groupes par la manière dont les données sont généralement analysées. Comme nous l'avons vu tout au long de ce livre, la recherche sur les groupes consiste à combiner les données des différentes personnes. Les données de groupe sont décrites à l'aide de statistiques, telles que les moyennes, les écarts-types, les coefficients de corrélation, etc., afin de détecter des patrons généraux. Enfin les statistiques inférentielles sont utilisées pour déterminer si le résultat obtenu avec l'échantillon peut se généraliser à la population. En revanche la recherche à cas unique s'appuie fortement sur une approche très différente, l'**inspection visuelle**. Cela signifie de représenter graphiquement les données de chaque personne participante de manière individuelle comme montré tout au long du présent chapitre, de les examiner attentivement et de porter un jugement afin de déterminer si, et dans quelle mesure, la variable indépendante a eu un effet sur la variable dépendante. Les statistiques inférentielles ne sont généralement pas utilisées.

Lors de l'inspection visuelle de leurs données, les personnes faisant de la recherche à cas unique tiennent compte de plusieurs facteurs dont les changements de **niveau** de la variable dépendante d'une condition à l'autre. Si la variable dépendante est beaucoup plus élevée ou beaucoup plus basse dans une condition que dans une autre, cela suggère que le traitement a eu un effet. Le deuxième facteur est la **tendance**, qui fait référence aux augmentations ou diminutions progressives de la variable dépendante au fil des observations. Si la variable dépendante commence à augmenter ou à diminuer avec un changement de conditions, cela suggère à nouveau que le traitement a eu un effet. Un signe particulièrement révélateur est lorsqu'une tendance change de direction — par exemple lorsqu'un comportement indésirable augmente pendant le niveau de base, mais commence à diminuer après l'introduction du traitement. Le troisième facteur est la

latence, c'est-à-dire le temps nécessaire pour que la variable dépendante commence à changer après une modification des conditions. En général si un changement dans la variable dépendante commence peu de temps après un changement dans les conditions, cela suggère que le traitement en est responsable.

Dans le panneau supérieur de la Figure 10.4, on observe des changements assez évidents dans le niveau et la tendance de la variable dépendante d'une condition à l'autre. En outre les temps de latence pour ces changements sont courts ; le changement se produit immédiatement. Ce patron de résultats suggère fortement que le traitement est responsable des changements de la variable dépendante. Cependant dans le panneau inférieur de la Figure 10.4, les changements de niveau sont assez faibles. Et bien qu'une tendance à la hausse semble apparaître dans la condition de traitement, il semble qu'elle soit en fait la continuation d'une tendance commencée pendant le niveau de base. Ce patron de résultats suggère fortement que le traitement n'est responsable d'aucun changement dans la variable dépendante — du moins pas autant que les personnes pratiquant la recherche à cas unique le souhaiteraient.

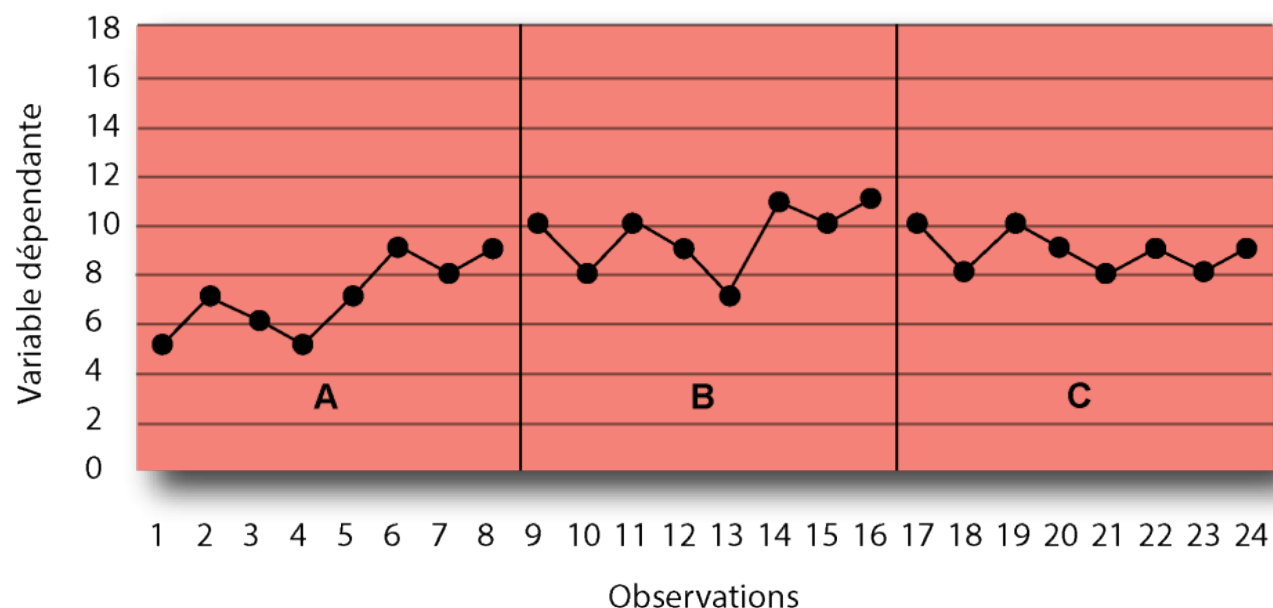
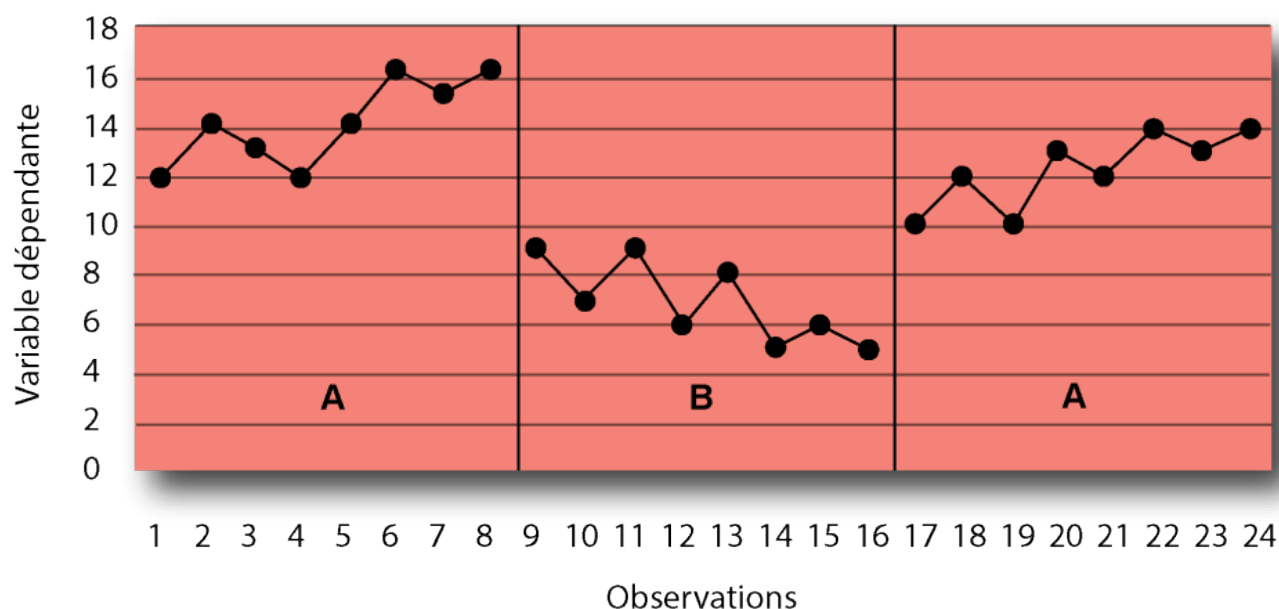


Figure 10.4. Résultats d'une étude générique sur un seul sujet illustrant le niveau, la tendance et la latence. L'inspection visuelle des données suggère un traitement efficace dans le panneau supérieur, mais un traitement inefficace dans le panneau inférieur. [\[Description de la figure\]](#)

Les résultats de la recherche à cas unique peuvent également être analysés à l'aide de procédures statistiques — ce qui devient de plus en plus fréquent. Il existe de nombreuses approches différentes, et les personnes faisant de la recherche à cas unique continuent de débattre de celles qui sont les plus utiles. Une approche s'apparente à ce qui se fait généralement dans les recherches portant sur des groupes.

La moyenne et l'écart-type des réponses de chaque personne dans chaque condition sont calculés et comparés, et des tests statistiques inférentiels, tels que le test *t* ou l'analyse de la variance sont appliqués (Fisch, 2001). (Notons qu'il est moins fréquent de calculer des moyennes *entre* les personnes participantes.) Une autre approche consiste à calculer le **pourcentage de données qui ne se chevauchent pas** (en anglais, *percentage of non-overlapping data*) (PND) pour chaque personne participante (Scruggs et Mastropieri, 2001). Il s'agit du pourcentage de réponses dans la condition traitement qui sont plus extrêmes que la réponse la plus extrême dans une condition contrôle pertinente. Dans l'étude de Hall et de ses collègues par exemple, toutes les mesures du temps d'étude de Robbie lors de la première condition traitement étaient supérieures à la mesure la plus élevée lors du premier niveau de base, pour un PND de 100 %. Plus le pourcentage de données qui ne se chevauchent pas est élevé, plus l'effet du traitement est fort. Néanmoins les approches statistiques formelles d'analyse des données dans la recherche à cas unique sont généralement considérées comme un complément à l'inspection visuelle, et non comme un remplacement.

Description de l'image

Figure 10.2 description longue : Graphiques linéaires montrant les résultats d'une étude avec un protocole d'inversion ABAB. La variable dépendante est faible pendant la première phase du niveau de base ; elle augmente pendant la première phase de traitement ; elle diminue pendant la deuxième phase du niveau de base, mais elle demeure plus élevée que pendant la première phase du niveau de base, et elle atteint son plus haut au niveau pendant la deuxième phase de traitement. [\[Retour à la Figure 10.2\]](#)

Figure 10.3 description longue : Trois graphiques linéaires montrant les résultats d'une étude générique à niveaux de base multiples, dans laquelle différents niveaux de base sont établis, et le traitement est introduit auprès des personnes participantes à différents moments.

Pour le niveau de base 1, le traitement est introduit environ au quart de la durée totale de l'étude. Pendant la phase de niveau de base, la variable dépendante varie généralement entre 12 et 16 unités. Au moment où le traitement est introduit, la variable dépendante chute à 10 unités puis continue de diminuer graduellement au fil des observations jusqu'à la fin de l'étude, variant entre 4 et 10 unités.

Pour le niveau de base 2, le traitement est introduit à mi-parcours de l'étude. Pendant la phase de niveau de base, la variable dépendante varie généralement entre 10 et 15 unités. Au moment où le traitement est introduit, la variable dépendante chute brusquement à 7 unités. Cependant la variable dépendante remonte ensuite à 12 unités peu après sa baisse et varie entre 8 et 10 unités jusqu'à la fin de l'étude.

Pour le niveau de base 3, le traitement est introduit aux trois quarts de la durée totale de l'étude. Pendant la phase de niveau de base, la variable dépendante varie la plupart du temps entre 12 et 16 unités, avec une baisse ponctuelle à 10 unités. Au moment où le traitement est introduit, la variable dépendante chute à 10 unités, puis varie entre 8 et 9 unités jusqu'à la fin de l'étude. [\[Retour à la Figure 10.3\]](#)

Figure 10.4 description longue : Deux graphiques illustrant les résultats d'une étude générique à cas unique avec un protocole ABA. Au premier graphique, dans la condition A, le niveau est élevé et la tendance est croissante. Dans la condition B, le niveau est beaucoup plus bas que dans la condition A et la tendance est décroissante. Dans la deuxième phase de la condition A, le niveau est à nouveau aussi élevé que la première phase et la tendance est croissante. Pour chaque changement, le temps de réponse est court, ce qui suggère que le traitement est la cause du changement.

Au deuxième graphique, dans la condition A, le niveau est relativement bas et la tendance est croissante. Dans la condition B, le niveau est un peu plus élevé que dans la condition A et la tendance est légèrement croissante. Dans la deuxième phase de la condition A, le niveau est à nouveau un peu plus bas que dans la condition B et la tendance est légèrement décroissante. Il est difficile de déterminer le temps de latence de ces changements, car chaque changement est assez minime, ce qui suggère que le traitement est inefficace.

[\[Retour à la Figure 10.4\]](#)

55. Le « débat » entre cas unique et groupe

Objectifs d'apprentissage

1. Expliquer quelques-uns des points de désaccord entre les personnes militantes pour la recherche à cas unique et celles pour la recherche sur les groupes.
2. Identifier plusieurs situations dans lesquelles la recherche à cas unique serait appropriée, et plusieurs autres dans lesquelles la recherche sur les groupes serait appropriée.

La recherche à cas unique est similaire à la recherche sur les groupes — en particulier la recherche expérimentale sur les groupes — à bien des égards. Dans les deux cas, elles utilisent les approches quantitatives dans le but d'établir des relations de cause à effet en manipulant une variable indépendante, en mesurant une variable dépendante et en contrôlant les variables externes. Mais il existe également des différences importantes entre ces approches, et ces différences conduisent parfois à des désaccords. Il convient d'examiner les points de désaccord les plus courants entre les personnes militantes pour la recherche à cas unique et celles militantes pour la recherche sur les groupes, ainsi que sur la manière dont ces désaccords peuvent être résolus. Comme nous le verrons, la recherche à cas unique et la recherche sur les groupes sont probablement mieux conceptualisées en tant qu'approches complémentaires.

Analyse des données

Une série de désaccords porte sur la question de l'analyse des données. Des personnes qui militent en faveur de la recherche sur les groupes craignent que l'inspection visuelle soit inadéquate pour déterminer si, et dans quelle mesure, un traitement a un effet sur une variable dépendante. Une première préoccupation spécifique est que l'inspection visuelle ne soit pas assez sensible pour détecter des effets faibles. Une deuxième préoccupation est que l'inspection visuelle puisse ne pas être fidèle, puisque des scientifiques parviennent à des conclusions différentes à partir du même ensemble de données (Danov et Symons, 2008). Une troisième est que les résultats de l'inspection visuelle — un jugement global sur l'efficacité ou non d'un traitement — ne puissent pas être résumés ou comparés de manière claire et efficace d'une étude à l'autre (contrairement aux mesures de la force de la relation fréquemment utilisées dans la recherche portant sur les groupes).

En général, les personnes menant des recherches à cas unique partagent ces préoccupations. Cependant elles affirment aussi que l'utilisation de la stratégie de l'état stable, combinée à l'accent mis sur des effets

robustes et constants, permet de minimiser la plupart de celles-ci. Si l'effet d'un traitement est difficile à détecter par inspection visuelle parce que l'effet est faible ou que les données ont trop de bruit, les adeptes de la recherche à cas unique cherchent des moyens d'accroître l'intensité de l'effet ou de réduire le bruit dans les données en contrôlant les variables externes (par exemple en administrant le traitement de manière plus consistante). Si l'effet reste difficile à détecter, ces scientifiques pourraient considérer qu'il n'est ni assez fort, ni assez constant pour présenter un intérêt supplémentaire. Par ailleurs plusieurs personnes menant des recherches à cas unique soulignent que l'analyse statistique est de plus en plus courante et beaucoup l'utilisent en complément à l'inspection visuelle — en particulier pour comparer les résultats entre les études (Scruggs et Mastropieri, 2001).

En contrepartie des personnes qui militent en faveur de la recherche à cas unique s'inquiètent de la façon dont les données sont analysées par les personnes menant des recherches sur les groupes. Plus spécifiquement elles soulignent que le fait de se concentrer sur les moyennes de groupe peut être très trompeur. Imaginons à nouveau qu'un traitement ait un effet positif marqué sur la moitié des personnes qui y sont exposées, et un effet négatif tout aussi marqué sur l'autre moitié. Dans une expérience intersujets classique, l'effet positif au traitement sur la moitié de l'échantillon serait statistiquement annulé par l'effet négatif sur l'autre moitié. La moyenne du groupe traitement serait alors la même que celle du groupe contrôle, ce qui donnerait l'impression que le traitement n'a eu aucun effet alors qu'en réalité, il a eu un effet marqué sur tous les membres de l'échantillon!

Mais là encore les personnes menant des recherches sur les groupes partagent cette préoccupation. Bien que ces derniers se concentrent sur les statistiques de groupe, l'importance d'examiner les distributions des scores individuels est aussi considérée comme un élément clé. Par exemple si des personnes ont été positivement affectées par un traitement alors que d'autres l'ont été négativement, cela produirait une distribution bimodale des scores qui pourrait être détectée par l'inspection visuelle d'un histogramme des données. L'utilisation de devis intrasujets est une autre stratégie permettant aux personnes menant des recherches sur les groupes d'observer les effets au niveau individuel, et de même spécifier le pourcentage de personnes présentant des effets forts, moyens, faibles, voire négatifs. Enfin les plans factoriels peuvent être utilisés pour examiner si les effets d'une variable indépendante sur une variable dépendante diffèrent selon les groupes de personnes participantes (introvertis vs extravertis).

Validité externe

Le deuxième point où il y a parfois désaccord entre les personnes menant des recherches à cas unique et celles menant des recherches sur les groupes concerne la validité externe — c'est-à-dire la capacité à généraliser les résultats d'une étude au-delà des personnes et de la situation spécifique actuellement étudiées. En particulier les personnes militantes pour la recherche sur les groupes soulignent la difficulté de savoir si les résultats obtenus avec quelques personnes participantes seulement peuvent être généralisés à d'autres membres de la population. Imaginons par exemple qu'une étude à cas unique démontre qu'un traitement réduit l'automutilation chez deux enfants ayant une déficience intellectuelle. Même si l'effet est fort pour ces deux enfants, comment peut-on savoir si ce traitement fonctionnerait pour d'autres enfants souffrant d'une déficience intellectuelle?

Là encore les personnes menant des recherches à cas unique partagent cette inquiétude. En guise

de réponse, elles notent que les effets forts et consistants qui les intéressent typiquement — même lorsqu'observés dans de petits échantillons — sont susceptibles de se généraliser à d'autres dans la population. Ces personnes menant des recherches à cas unique soulignent également qu'elles accordent une grande importance à répliquer les résultats. Lorsqu'un effet sur un petit échantillon de personnes est observé, typiquement ces scientifiques vont tenter de le répliquer sur un autre petit échantillon, peut-être avec des personnes légèrement différentes ou dans des conditions légèrement différentes. Chaque fois que des résultats similaires sont obtenus, leur confiance dans la généralisation de ces résultats augmente. Les personnes menant des recherches à cas unique soulignent également que les principes du conditionnement classique et opérant — dont la plupart ont été découverts grâce à l'approche à cas unique — ont été généralisés avec succès auprès d'un nombre incroyablement élevé d'espèces et de situations.

En contrepartie encore une fois, les personnes menant des recherches à cas unique expriment leurs propres préoccupations quant à la validité externe des recherches sur les groupes. Un point extrêmement important qu'elles soulignent est que l'étude de grands groupes de personnes participantes ne résout pas entièrement le problème de la généralisation à d'autres personnes. Imaginons par exemple un traitement ayant montré en moyenne un petit effet positif dans une étude menée sur un grand échantillon. Il est probable que quoique plusieurs personnes participantes aient présenté un petit effet positif, d'autres aient montré un grand effet positif et d'autres encore aient présenté un petit effet négatif. Lorsqu'il s'agit d'appliquer ce traitement à un autre grand *groupe*, nous pouvons être relativement certain-e-s qu'il aura en moyenne un petit effet. Mais lorsqu'il s'agit d'appliquer ce traitement à une autre personne, nous ne pouvons pas savoir avec certitude s'il aura un effet faible, important ou même négatif. Un autre point soulevé par ces personnes est que les personnes menant des recherches sur les groupes sont également aux prises avec un problème similaire lorsqu'elles étudient une seule situation, puis généralisent leurs résultats à d'autres situations. Par exemple des scientifiques qui mènent une étude sur l'effet de l'utilisation du téléphone cellulaire chez des automobilistes sur une piste ovale fermée souhaitent probablement appliquer leurs résultats aux automobilistes dans de nombreuses autres situations réelles de conduite. Mais remarquons qu'il faut pour cela généraliser d'une seule situation à une population de situations. La capacité de généralisation ne repose donc pas uniquement sur le nombre de personnes étudiées. Elle nécessite un examen attentif de la similitude entre l'échantillon de personnes et l'échantillon de situations étudiées, et la population de personnes et la population de situations auxquelles on souhaite appliquer la généralisation (Shadish, Cook et Campbell, 2002).

Recherche à cas unique et sur les groupes : des méthodes complémentaires

Comme pour la recherche quantitative et la recherche qualitative, il est probablement préférable de concevoir la recherche à cas unique et la recherche sur les groupes comme des méthodes complémentaires ayant chacune leurs forces et leurs faiblesses, et étant appropriées pour répondre à différents types de questions de recherche (Kazdin, 1982). La recherche à cas unique est particulièrement adaptée pour tester l'efficacité des traitements sur les personnes lorsque l'accent est mis sur des effets forts, consistants et importants d'un point de vue biologique ou social. Elle est également particulièrement utile lorsque le comportement de personnes particulières est l'objet d'intérêt. En contexte clinique, lorsqu'on ne travaille

qu'avec une seule personne à la fois, on peut y voir la seule occasion de réaliser une recherche quantitative systématique.

En revanche la recherche sur les groupes est idéale pour tester l'efficacité des traitements au niveau du groupe. L'un des avantages de cette approche est qu'elle permet aux scientifiques de détecter de faibles effets, qui peuvent être intéressants pour plusieurs raisons. Par exemple la découverte d'un faible effet de traitement peut conduire à des ajustements du traitement qui produiront finalement un effet plus important et plus significatif. La recherche sur les groupes permet également d'étudier les interactions entre les traitements et les caractéristiques des personnes participantes. Par exemple si un traitement est efficace pour celles qui sont très motivées à changer et inefficace pour celles qui sont peu motivées à changer, un devis de groupe peut le détecter beaucoup plus efficacement qu'un protocole à cas unique. La recherche sur les groupes est également nécessaire pour répondre à des questions qui ne peuvent être abordées dans le cadre d'une approche à cas unique, notamment des questions concernant des variables indépendantes qui ne peuvent être manipulées (par exemple le nombre de frères et sœurs, l'extraversion, la culture).

Enfin il est important de comprendre que la recherche à cas unique et la recherche sur les groupes représentent des traditions de recherche différentes. Ce facteur est probablement le plus important pour expliquer l'approche qu'adopte une équipe de recherche. Les équipes œuvrant dans l'analyse expérimentale du comportement et l'analyse appliquée du comportement apprennent à conceptualiser leurs questions de recherche d'une manière qui se prête à l'approche à cas unique. Les scientifiques dans la plupart des autres domaines de la psychologie apprennent à conceptualiser leurs questions de recherche d'une manière qui se prête à l'approche de groupe. En même temps il existe de nombreux sujets en psychologie pour lesquels les recherches des deux traditions se sont enrichies l'une l'autre et ont été intégrées avec succès. Un exemple est la recherche suggérant que les animaux et les humains ont un « sens du nombre » inné — une capacité à percevoir combien d'objets ou d'événements d'un certain type ils ont rencontrés sans avoir besoin de les compter (Dehaene, 2011). Les recherches à cas unique sur les rats et les oiseaux et celles sur les groupes avec des bébés humains ont montré des capacités étonnamment similaires dans ces populations pour discriminer de petits nombres d'objets et d'événements. Ce sens du nombre — qui a probablement évolué bien avant l'espèce humaine — pourrait même être à l'origine des capacités mathématiques avancées de l'humanité.

Le principe des données convergentes

Maintenant que vous avez acquis une connaissance des méthodes de recherche les plus couramment utilisées en psychologie, il devrait être évident qu'aucun devis n'est parfait. Chaque devis de recherche présente des points forts et des points faibles. Les vraies expériences ont généralement une validité interne élevée, mais peuvent avoir des problèmes de validité externe, tandis que la recherche non expérimentale (par exemple la recherche corrélationnelle) a souvent une bonne validité externe, mais une faible validité interne. Chaque étude nous rapproche de la vérité, mais aucune ne peut être considérée comme définitive. C'est l'une des raisons pour lesquelles en science nous disons qu'il n'y a pas de preuve scientifique, mais seulement des données comme évidence.

Alors que les médias ont souvent tendance à tirer des conclusions convaincantes à partir des résultats d'une seule étude, les scientifiques se concentrent plutôt sur l'évaluation d'un ensemble de recherches.

Les scientifiques évaluent les théories non pas en attendant l'expérience parfaite, mais en examinant les tendances générales d'un certain nombre d'études comportant des lacunes. L'idée des **données convergentes** nous incite à examiner le profil des faiblesses qu'on retrouve dans la littérature de recherche, car la nature de ce profil peut soit soutenir, soit compromettre les conclusions que nous souhaitons tirer. Supposons que les résultats d'un certain nombre d'études différentes soient largement cohérents et soutiennent une conclusion particulière. Si toutes ces études présentent des défauts similaires, par exemple si toutes les études sont corrélationnelles et présentent le problème d'une troisième variable et le problème de la directionnalité, la confiance dans les conclusions tirées s'en trouverait ébranlée, car la cohérence des résultats pourrait simplement résulter d'un défaut particulier commun à toutes les études. En revanche si toutes les études présentent des défauts différents et que les faiblesses de certaines d'entre elles constituent la force des autres (la faible validité externe d'une véritable expérience étant contrebalancée par la forte validité externe d'une étude corrélationnelle), nous pouvons alors avoir davantage confiance dans nos conclusions.

Bien qu'il existe des compromis fondamentaux dans les différents devis de recherche, l'ensemble diversifié d'approches utilisées par les scientifiques en psychologie présente des atouts complémentaires qui nous permettent de trouver des données convergentes. Nous pouvons tirer des conclusions significatives et nous rapprocher de la vérité en examinant un grand nombre d'études différentes, chacune présentant des forces et des faiblesses différentes. Si les résultats d'un grand nombre d'études, toutes réalisées à l'aide de devis différents, convergent vers la même conclusion, alors notre confiance dans cette conclusion peut être considérablement accrue. En science, nous visons le progrès pas la perfection.

56. Points clés et exercices

Points clés

- La recherche à cas unique — qui consiste à utiliser un petit nombre de personnes et à se concentrer sur le comportement de chacune d'entre elles — est une alternative importante à la recherche sur les groupes en psychologie.
- Les études à cas unique doivent être distinguées des recherches qualitatives menées auprès d'une seule personne ou d'un petit nombre de personnes. Contrairement à la recherche plus qualitative, la recherche à cas unique se concentre sur la compréhension d'un comportement objectif grâce à la manipulation et au contrôle expérimental, à la collecte de données hautement structurées et à leur analyse quantitative.
- La recherche à cas unique existe depuis les débuts de la psychologie. Aujourd'hui, elle est plus fortement associée à la perspective théorique comportementale, mais elle peut en principe être utilisée pour étudier le comportement dans n'importe quelle perspective.
- Les protocoles de recherche à cas unique consistent généralement à mesurer la variable dépendante de manière répétée au fil du temps et à changer les conditions (p. ex. passer du niveau de base au traitement) lorsque la variable dépendante a atteint un état stable. Cette approche permet à l'équipe de recherche de voir si les changements dans la variable indépendante entraînent des changements dans la variable dépendante.
- Dans un protocole d'inversion, la personne participante est testée d'abord dans une condition de niveau de base, puis dans une condition traitement, avant de revenir à la condition de niveau de base. Si la variable dépendante change lors de l'introduction du traitement puis revient à son niveau de base, cela constitue une preuve solide de l'existence d'un effet du traitement.
- Dans un protocole à niveaux de base multiples, les niveaux de base sont établis pour différentes personnes, différentes variables dépendantes ou différents contextes — et le traitement est introduit à un moment différent pour chaque niveau de base. Si l'introduction du traitement est suivie d'un changement de la variable dépendante pour chacun des niveaux de base, cela constitue une preuve solide de l'existence d'un effet du traitement.
- Les personnes menant des recherches à cas unique analysent généralement leurs données en les représentant graphiquement et en déterminant si la variable indépendante affecte la variable dépendante sur le plan du niveau, de la tendance et de la latence.
- Les différences entre les recherches à cas unique et les recherches sur les groupes conduisent parfois à des désaccords entre les personnes menant des recherches. Ces désaccords portent principalement sur l'analyse des données et la validité externe (notamment la généralisation à d'autres personnes).

- La recherche à cas unique et la recherche sur les groupes doivent sans doute être considérées comme des méthodes complémentaires, avec des forces et des faiblesses différentes, et appropriées pour répondre à différents types de questions de recherche.

Exercices

- **Pratique** : Trouvez et lisez un article publié en psychologie qui fait état d'une nouvelle recherche à cas unique. (Une archive des articles publiés dans le Journal of Applied Behavior Analysis peut être consultée à l'adresse suivante : <http://www.ncbi.nlm.nih.gov/pmc/journals/309/>). Rédigez un bref résumé de l'étude.
- **Pratique** : Concevez une étude à cas unique simple (en utilisant un protocole d'inversion ou à niveaux de base multiples) pour répondre aux questions suivantes. Assurez-vous de spécifier le traitement, de définir opérationnellement la variable dépendante, de décider quand et où les observations seront faites, etc.
 - L'attention positive d'un parent augmente-t-elle le comportement de l'enfant en matière de brossage des dents ?
 - L'auto-évaluation en cours d'étude améliore-t-elle les résultats d'un ou d'une élève aux tests d'orthographe hebdomadaires ?
 - L'exercice physique régulier aide-t-il à soulager la dépression ?
- **Pratique** : Créez un graphique qui affiche les résultats hypothétiques de l'étude que vous avez conçue dans l'exercice 1. Rédigez un paragraphe décrivant ce que montrent les résultats. Assurez-vous de commenter le niveau de base, la tendance et la latence.
- **Discussion** : Imaginez que vous ayez mené une étude à cas unique montrant un effet positif d'un traitement sur le comportement d'un homme souffrant de troubles de l'anxiété sociale. Votre recherche a été critiquée du fait qu'elle ne peut être généralisée à d'autres personnes. Comment pourriez-vous répondre à cette critique ?
- **Discussion** : Imaginez que vous ayez mené une étude de groupe montrant un effet positif d'un traitement sur le comportement d'un groupe de personnes souffrant de troubles d'anxiété sociale, mais que votre recherche ait été critiquée du fait que les effets « moyens » ne peuvent être généralisés à des personnes. Comment pourriez-vous répondre à cette critique ?
- **Pratique** : Reformulez l'étude de Hall et de ses collègues décrite au début de ce chapitre en une étude de groupe, et dressez la liste des forces et des faiblesses de votre nouvelle étude par rapport à l'étude originale.
- **Pratique** : L'effet de génération fait référence au fait que les personnes qui génèrent des informations pendant qu'elles les apprennent (p. ex., en effectuant des auto-évaluations) s'en souviennent mieux par la suite que les personnes qui se contentent de simplement réviser l'information. Concevez une étude à cas

unique sur l'effet de génération appliquée aux personnes aux études universitaires qui apprennent l'anatomie du cerveau.

57. Références

- Baer, D. M., Wolf, M. M. et Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91-97.
- Danov, S. E. et Symons, F. E. (2008). A survey evaluation of the reliability of visual inspection and functional analysis graphs. *Behavior Modification*, 32, 828-839.
- Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (2e éd.). Oxford.
- Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models. *Behavioral Processes*, 54, 137-154.
- Hall, R. V., Lund, D. et Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1, 1-12.
- Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.
- Lanovaz, M. (2013). L'utilisation de devis expérimentaux à cas unique en psychoéducation. *Revue de psychoéducation*, 42(1), 161-183.
<https://umontreal.scholaris.ca/items/4aa6cffa-1976-48a0-9dae-c45faeb63fa2>
- Ross, S. W. et Horner, R. H. (2009). Bully prevention in positive behavior support. *Journal of Applied Behavior Analysis*, 42, 747-759.
- Scruggs, T. E., et Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality*, 9, 227-244.
- Shadish, W. R., Cook, T. D. et Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Authors Cooperative.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.
- Wolf, M. (1976). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

CHAPITRE 11-PRÉSENTER VOTRE RECHERCHE



Image 11. Personne présentant un graphique

« La recherche n'est terminée que lorsque les résultats sont partagés avec la communauté scientifique. »

– American Psychology Association

Imaginez que vous ayez identifié une question de recherche intéressante, analysé la littérature pertinente, conçu et mené une étude empirique, analysé les données et tiré vos conclusions. Il reste encore une étape à franchir dans le processus de recherche scientifique. Il est temps d'ajouter votre recherche à la littérature afin que d'autres puissent en tirer des enseignements et s'en inspirer. N'oubliez pas que la science est un processus social et cumulatif — une immense collaboration entre de nombreuses personnes en recherche réparties dans l'espace et dans le temps. Pour cette raison, on pourrait affirmer que si les recherches ne sont pas rendues publiques sous une forme ou une autre, il n'y a pas vraiment d'engagement scientifique.

Dans ce chapitre, nous verrons comment présenter votre recherche de manière efficace. Nous

commencerons par une discussion sur le style APA (association américaine de psychologie). Ce style est la principale approche de la rédaction adoptée par les personnes menant des recherches en psychologie et dans les domaines connexes. Il est possible que vous trouviez des présentations suivant un autre style plus général, notamment en Europe. Cela inclut au Canada l'[association canadienne en psychologie](#) (ACP) et la [Société Canadienne des Sciences du Cerveau, du Comportement et de la Cognition](#) (SCSCCC). Nous verrons ensuite comment rédiger un article de recherche empirique selon le style APA. Enfin nous examinerons quelques-unes des nombreuses autres façons dont les scientifiques présentent leurs travaux, notamment les articles de synthèse et les articles théoriques, les thèses et autres travaux de personnes aux études ainsi que les communications orales et les communications par affiches lors de congrès.

58. Le style APA (Association américaine de psychologie)

Objectifs d'apprentissage

1. Définir le style APA et énumérer plusieurs de ses caractéristiques les plus importantes.
2. Identifier trois niveaux de style APA et donner des exemples pour chacun d'entre eux.
3. Identifier plusieurs sources d'information sur le style APA.

Qu'est-ce que le style APA ?

Le **style APA** est un ensemble de lignes directrices pour la rédaction en psychologie et dans les domaines connexes. Ces lignes directrices sont énoncées dans le ***Publication Manual of the American Psychological Association*** (APA, 2020). Le manuel de publication a été publié pour la première fois en 1929 sous la forme d'un court article de journal qui fournissait des normes de base pour la préparation des manuscrits devant être soumis pour publication (Bentley, Peerenboom, Hodge, Passano, Warren, et Washburn, 1929).

Il a ensuite été développé et publié sous forme de livre par l'association et en est aujourd'hui à sa septième édition. L'objectif principal du style APA est de faciliter la communication scientifique en favorisant la clarté de l'expression, et en standardisant l'organisation, et le contenu des articles de recherche et des chapitres de livres. Il est plus facile de rédiger au sujet de la recherche lorsque l'on sait quelles informations présenter, dans quel ordre les présenter et même dans quel style les présenter. Enfin il est plus facile de lire un article de recherche lorsqu'il est présenté de manière familière et attendue.

Le style APA est considéré comme un « genre » de rédaction qui convient à la présentation des résultats de la recherche en psychologie — surtout dans les contextes universitaires et professionnels. Il n'est pas synonyme de « bonne rédaction » en général. Vous ne rédigeriez pas dans le style APA une analyse littéraire pour un cours d'anglais, même si elle était basée sur des concepts psychanalytiques. Vous la rédigeriez plutôt dans le style de la *Modern Language Association* (MLA). De même vous ne rédigeriez pas un article de journal dans le style APA même s'il portait sur une nouvelle avancée dans le domaine des neurosciences comportementales. Vous le rédigeriez plutôt dans le style *Associated Press* (AP). De plus vous ne rédigeriez pas un article de recherche empirique en style MLA, en style AP ou dans le style d'un roman d'amour, d'un courriel à un ami ou d'une liste de courses. Vous le rédigeriez dans le style APA. Faire preuve d'une bonne rédaction en général consiste en partie à adopter un style approprié à la tâche de rédaction en cours, et pour écrire au sujet de la recherche en psychologie, c'est le style APA. En plus du manuel, le style APA met à notre

disposition en ligne différentes ressources sur leur site apastyle.apa.org dont des pamphlets, des exemples de documents et des activités.

Les niveaux du style APA

Le style APA étant constitué d'un grand nombre et d'une grande variété de directives — le *Publication Manual* compte près de 430 pages — il peut être utile d'y réfléchir en termes de trois niveaux de base. Le premier niveau est l'organisation et le format général d'un article, le second niveau porte sur le style d'écriture, alors que le troisième niveau vise plutôt les sources citées dans le texte et listées dans la section références.

Le premier niveau porte sur l'organisation générale d'un article (présenté dans le chapitre 2 du manuel de publication) où on décrit notamment que les articles empiriques, en particulier, comportent plusieurs sections distinctes qui apparaissent toujours dans le même ordre :

- **Page titre.** Présente le titre de l'article et les noms et affiliations des auteurs et autrices.
- **Résumé.** Résume la recherche.
- **Introduction.** Décrit les recherches antérieures et la raison d'être de l'étude actuelle.
- **Méthode.** Décrit la manière dont l'étude a été menée.
- **Résultats.** Décrit les résultats de l'étude.
- **Discussion.** Résume l'étude et discute de ses implications.
- **Références.** Liste des références citées dans l'article.

Ce niveau comprend également le format général de l'article : la page titre, les marges, l'alignement du texte général, des titres de section et des sous-sections, l'utilisation de la tabulation et du caractère gras, la présence d'entête, la pagination, la police, la taille des caractères, etc.; il aborde aussi l'accessibilité du style APA. Dans le présent chapitre, un exemple d'article sera présenté à partir de la Figure 11.2.

Le deuxième niveau du style APA peut être qualifié de **style de haut niveau** (abordé aux chapitre 4 « Style d'écriture et grammaire » et chapitre 5 « Directives pour un langage sans biais » du Manuel de publication). Le chapitre 4 inclut des lignes directrices pour exprimer des idées de manière claire et concise. Deux thèmes importants s'y dégagent. Le premier est que le style de rédaction APA est formel plutôt qu'informel. On y adopte un ton approprié pour communiquer avec des collègues professionnels — d'autres scientifiques et membres en pratique clinique — qui partagent un intérêt pour le sujet. Toutefois au-delà de cet intérêt commun, ces collègues ne sont pas nécessairement semblables aux personnes ayant rédigées le texte ou les uns par rapport aux autres. Une personne aux études diplômée de Colombie-Britannique peut rédiger un article qui sera lu par un jeune psychothérapeute de Toronto et un membre émérite du corps professoral en psychologie de Tokyo. L'écriture formelle évite donc l'argot, les contractions, les références à la culture pop, l'humour et d'autres éléments qui seraient acceptables dans le cadre d'une conversation avec un ami ou pour un texte informel. En revanche écrire de manière formelle ne veut pas dire de référer à vous à la troisième personne. En fait selon le style APA, il serait mieux d'utiliser le « je », ou le « nous », si vous êtes plusieurs.

Le deuxième thème du style de haut niveau APA est qu'il est direct. Cela signifie qu'il communique des idées aussi simplement et clairement que possible, en mettant l'accent sur les idées elles-mêmes et non

sur la manière dont elles sont communiquées. Ainsi le style APA minimise les procédés littéraires, tels que la métaphore, l'imagerie, l'ironie, le suspense, etc. Là encore l'humour est réduit au minimum. Les phrases sont courtes et directes. Les termes techniques doivent être utilisés, mais ils le sont pour améliorer la communication, et non pas simplement pour donner à l'écrit une allure plus « scientifique ». Par exemple si les personnes participantes ont plongé leurs mains dans un seau d'eau glacée, il vaut mieux simplement l'écrire ainsi plutôt que d'écrire qu'elles « ont été soumises à un appareil provoquant la douleur ». De plus il n'y a pas de meilleure façon de communiquer qu'un devis intersujets a été utilisé que d'utiliser le terme « devis intersujets ». Pour une mine d'information, consultez dans les lignes directrices du style et de la grammaire de l'APA.

Le style APA et les valeurs de la psychologie

Robert Madigan et ses collègues ont argumenté que le style APA a un objectif qui n'est souvent pas reconnu (Madigan, Johnson, et Linton, 1995). Plus précisément il fait la promotion des valeurs et des hypothèses scientifiques des psychologues. De ce point de vue, plusieurs caractéristiques du style APA, apparaissant à première vue arbitraires, sont en fait tout à fait logiques. Voici plusieurs caractéristiques du style APA et les valeurs ou hypothèses scientifiques qu'elles reflètent.

Caractéristiques du style APA	Valeur ou postulat scientifique
Il y a très peu de citations directes d'autres scientifiques.	Les phénomènes et théories en psychologie sont objectifs et ne dépendent pas des mots spécifiques utilisés par un scientifique particulier pour les décrire.
Les critiques sont dirigées vers le travail des scientifiques, pas vers les personnes personnellement.	L'objectif de la recherche scientifique est de tirer des conclusions générales sur le monde, et non sur la personnalité de quelconque scientifique.
Les références et les citations de références sont nombreuses.	La recherche scientifique est une collaboration à grande échelle entre de nombreux scientifiques.
Les articles empiriques sont organisés avec des sections spécifiques dans un ordre fixe.	Il y a une approche idéale pour mener une recherche empirique en psychologie (même si cet idéal n'est pas toujours atteint dans la recherche actuelle).
Les scientifiques ont tendance à nuancer leurs conclusions, par exemple « Les résultats suggèrent que... ».	Les connaissances scientifiques sont provisoires et toujours sujettes à révision sur la base de nouveaux résultats empiriques.

Un autre élément important du style APA de haut niveau est d'éviter tout langage biaisé envers des groupes particuliers. Il s'agit non seulement d'éviter d'offenser les gens — pourquoi voudriez-vous offenser des personnes intéressées par votre travail ? —, mais aussi d'assurer l'objectivité et l'exactitude scientifiques. Dans la 7e édition,

L'American Psychological Association insiste sur la nécessité de parler de toutes les personnes avec inclusivité et respect. Les personnes qui utilisent le style APA doivent s'efforcer d'utiliser un langage exempt de préjugés et éviter de perpétuer des croyances préjudiciables ou des attitudes dévalorisantes dans leurs écrits. Tout comme vous avez appris à vérifier l'orthographe, la grammaire et la formulation de ce que vous écrivez, entraînez-vous à lire votre travail pour y déceler des préjugés.

Les lignes directrices pour un langage exempt de préjugés contiennent à la fois des lignes directrices générales et des lignes directrices spécifiques qui traitent des caractéristiques individuelles de l'âge, du

handicap, du sexe, de la participation à la recherche, de l'identité raciale et ethnique, de l'orientation sexuelle, du statut socio-économique et de l'intersectionnalité (APA, traduction de la page d'introduction ayant pour titre [Bias free language](#)).

Les principes généraux permettant d'éviter un langage biaisé sont assez simples. Tout d'abord il faut être sensible aux étiquettes en évitant les termes offensants ou ayant des connotations négatives. Il s'agit notamment d'éviter les termes qui identifient les personnes à un trouble ou à un autre problème qu'elles ont. Il faut plutôt parler de la personne, ce que le manuel de publication de l'APA appelle « mettre la personne au premier plan », par exemple il est préférable de parler de *personnes ayant un diagnostic de schizophrénie* plutôt que de *schizophrènes*. Deuxièmement il faut utiliser des termes plus spécifiques plutôt que des termes plus généraux. Par exemple donner le contexte de l'étude, parler en termes de *clients*, de *participants* ou de *patients*, plutôt qu'en termes de *gens*. Par exemple préciser l'origine ethnique/race en disant *Américains chinois* plutôt que de dire Américains d'origine asiatique, si tous les membres du groupe sont en fait des Américains d'origine chinoise. Si on rédige à propos de l'âge, il est préférable de donner l'étendue d'âge exact (p. ex. *entre 65 et 75 ans*) plutôt qu'une indication générale (p. ex. *ayant 65 et plus*). Si on écrit à propos d'un trouble comme la *maladie d'Alzheimer*, il est préférable d'en donner une définition opérationnelle plutôt que d'indiquer s'intéresser à la *démence*. Si on s'intéresse à des personnes ayant un statut socio-économique particulier, il vaut mieux préciser ce statut (p. ex. *sous le seuil de la pauvreté*) plutôt que d'utiliser un terme vague (p. ex. *pauvre*). Troisièmement il faut éviter d'objectiver les personnes participant à la recherche. Au contraire reconnaissez leur contribution active à la recherche. Par exemple « *les étudiantes et étudiants ont rempli le questionnaire* » est préférable à « *On a administré le questionnaire aux sujets* ». Notons que ce principe permet également de rédiger des textes plus clairs et plus attrayants.

Le langage associé à l'identité de genre et l'orientation sexuelle a beaucoup changé dans les dernières années. Dans sa 7e édition, il y a des lignes directrices élaborées pour parler de genre de manière inclusive et respectueuse. On doit notamment utiliser les termes que les personnes emploient pour se décrire elle-même. On doit aussi éviter les termes impliquant la binarité des genres (par exemple dire le sexe opposé). Lorsqu'on rapporte le genre des personnes participant à notre étude, on ne doit pas présumer que tout l'échantillon est cisgenre. Pour référer à tout le monde, on doit inclure les personnes transgenres. Donc lorsqu'on parle d'une personne de manière générique, on ne peut pas utiliser « il » ou « elle », on doit reformuler la phrase pour ne pas avoir à utiliser ces pronoms ou utiliser le pronom singulier « they ». Nous reviendrons sur ce dernier point plus loin dans l'encadré intitulé « L'écriture inclusive en français », puisqu'il n'y a pas de pronom équivalent en français au pluriel comme au singulier. Le Tableau 11.1 présente quelques autres exemples qui respectent les principes généraux de l'APA.

Tableau 11.1 Exemples pour éviter le langage biaisé

Plutôt que...	Utiliser...
homme, hommes	hommes et femmes, gens, personnes
enseignants	corps enseignant
mâles et femelles	hommes et femmes (lorsqu'on parle d'humains), cisgenre homme, femme transgenre
minorité	nommer le groupe spécifique (p. ex., franco-ontariennes et franco-ontariens)
personnes âgées, aînées	personnes plus âgées, personnes entre 65 ans et 80 ans
enfants spéciaux	enfants avec des troubles d'apprentissage

L'écriture inclusive en français

L'écriture inclusive est un style rédactionnel ayant pour but de donner une place à tout le monde, peu importe l'identité de genre. Pour atteindre cet objectif, il est préférable d'adopter ce style dès le début de la rédaction d'un texte. Pour être inclusif, on doit tout d'abord bannir l'utilisation du masculin générique dans le texte. Comme le français est genré de manière binaire (soit c'est masculin, soit c'est féminin), il faut plutôt opter pour des formulations englobantes et des termes épicènes.

Formulation neutres

Les **formulations neutres** ont l'avantage d'inclure les personnes non binaires, les personnes dont on ne connaît pas le genre ainsi que les groupes renfermant des personnes de genres différents. Le tableau suivant présente une liste de termes neutres et d'épicènes facilitant la rédaction inclusive. Un mot est épicène lorsque sa forme est identique au féminin comme au masculin.

Termes		Suggestions
Noms neutres	Nom collectif: auditoire autorité clientèle collectivité communauté corps délégation effectif électorat équipe foule groupe lectorat main-d'œuvre membres milieu de l'éducation pairs personnel public	Le personnel de l'entretien La communauté étudiante Le corps professoral
	Unité administrative: administration bureau conseil décanat département direction faculté ministère présidence rectorat rédaction secrétariat service trésorerie	La direction a rédigé la lettre Le secrétariat a envoyé
	Nom: auxiliaire bénéficiaire bénévole cadre camarade collègue élève enfant gestionnaire guide interprète jeune locataire mentorat partenaire psychologue réceptionniste récipiendaire responsable scientifique spécialiste stagiaire titulaire tuteur	Le mentorat doit remettre Les psychologues enthousiastes Les récipiendaires de la bourse
Adjectifs neutres	Adjectif épïcène: adepte adorable apte brave capable célèbre convenable digne drôle dynamique efficace enthousiaste fiable habile incontournable intègre intraitable noble responsable rude superbe téméraire universitaire utile	L'élève responsable écrit Le personnel enseignant dynamique
Pronoms neutres	Pronom neutre ou épïcène: à qui de qui dont on nous majorité personne plusieurs quelque chose qui quiconque rien tu vous	La majorité de l'électorat Plusieurs personnes participantes

- En plus des termes neutres, il y a différentes techniques de **reformulation**, qui permettent d'éviter les « tournures genrées ». Plusieurs visent notamment à éliminer le besoin d'accorder le participe passé avec le sujet faisant l'action. On peut aussi *enlever des éléments non essentiels* (dont ceux genrés) ou *remplacer le sujet par un nom collectif ou de fonction, un terme neutre, ou un mot épïcène*. Toutes ces options sont très utiles pour la rédaction inclusive. Le tableau suivant donne des exemples de phrases où la formulation genrée est transformée en formulation neutre.

Formulation genrée

Les étudiants ont été réveillés par le feu d'artifice.

Les participantes et participants doivent obligatoirement signer le formulaire de consentement.

Il faut évaluer si le candidat ou la candidate peut être admis.

Nous sommes soulagés d'apprendre cette nouvelle.

Il faut se montrer généreux.

L'examinatrice, Madame Savoie, lit la thèse de Julie.

Formulation neutre

Le feu d'artifice a réveillé le corps étudiant.

Il est obligatoire de signer le formulaire de consentement.

Il faut évaluer l'admissibilité de la candidature.

Quel soulagement d'apprendre cette nouvelle!

Il faut faire preuve de générosité.

Sylvie Savoie examine la thèse de Julie.

Place égale au féminin et au masculin

Dans une langue où le masculin a longtemps été utilisé pour simplifier le texte, la présence du féminin et du masculin peut être marquée en utilisant des doublets complets (p. ex. Les étudiantes et les étudiants doivent suivre une formation en éthique de recherche. Elles et ils devront en fournir la preuve au responsable du cours). Règle générale, les doublets abrégés doivent être utilisés avec parcimonie (p. ex. le point médian: les étudiant-e-s, et la barre oblique pour les mots courts: le/la notaire) puisqu'ils pourraient réduire la lisibilité du texte et posent un problème d'accessibilité. Il est préférable de limiter leurs utilisations dans des contextes bien précis, où il y a notamment manque d'espace.

Sur le portail linguistique du Canada, de nombreuses ressources portent sur l'écriture inclusive, dont notamment une [vidéo de 30 minutes](#). L'écriture inclusive suit 6 principes:

- **Principe 1 :** Combiner les ressources de la langue
- **Principe 2 :** Donner une place égale au féminin et au masculin
- **Principe 3 :** Respecter les préférences des personnes concernées
- **Principe 4 :** Faire des choix représentatifs de la diversité
- **Principe 5 :** Veiller à la clarté de l'information
- **Principe 6 :** Faire preuve de cohérence

Voici quelques ressources portant sur l'écriture inclusive:

Portail linguistique du Canada. [Écriture inclusive](#)

Office québécois de la langue française. [Féminisation et rédaction épiciène](#). Dépannage linguistique.

Institut national de la recherche scientifique. [Inclusivement vôtres! Guide de rédaction inclusive](#)

Université de Montréal. [Inclusivement: Guide d'écriture pour toutes et tous](#)

Université de Sherbrooke. [La rédaction inclusive](#)

Université du Québec (2021). [Guide de communication inclusive](#). Pour des communications qui mobilisent, transforment et ont du style!

Université Laval. [Équité, diversité et inclusion: Outils et formations pertinentes](#)

Le troisième niveau du style APA peut être qualifié de **style de bas niveau** (couvert du chapitre 6 « Les mécanismes du style » au chapitre 10 « Exemples de référence » du Manuel de publication). Le style APA de bas niveau comprend toutes les directives spécifiques relatives à l'orthographe, à la grammaire, aux

références et aux sources citées dans le texte, aux chiffres et aux statistiques, aux figures et aux tableaux, etc. Les directives de bas niveau sont si nombreuses que même les spécialistes d'expérience doivent consulter le Manuel de publication de temps à autre. Le Tableau 11.2 présente les 10 types d'erreurs de style APA les plus fréquentes [en anglais], d'après une analyse des manuscrits soumis à une revue scientifique sur une période de 6 ans (Onwuegbuzie et al., 2010). Ces erreurs ont été commises par des scientifiques professionnels, mais sont probablement similaires à celles que les élèves commettent le plus souvent. Voir également la note 11.8 « Ressources en ligne sur le style APA » dans cette section et bien sûr le manuel de publication lui-même.

Tableau 11.2 Les 10 erreurs du style APA les plus fréquentes [en anglais]

Type d'erreur	Exemple
1. Utilisation de nombres	Ne pas utiliser des chiffres pour 10 et plus
2. Utilisation du trait d'union	Omettre d'unir un adjectif composé, précédé par un nom, avec un trait d'union (p. ex., « enfants premiers nés » devrait être « enfants premiers-nés »)
3. Utilisation de « et al. »	Omettre de l'utiliser après la première citation
4. Les titres	Majuscules et minuscules pas utilisées correctement dans les titres
5. Utilisation du « since »	Utiliser « since » (puisque) pour vouloir dire « because » (parce que)
6. Tableaux et figures	Ne suivent pas le format APA ; informations répétées déjà incluses dans le texte
7. Utilisation de la virgule	Omettre d'utiliser une virgule avant « and » (et) ou « or » (ou) dans une énumération ayant au moins trois éléments
8. Utilisation des abréviations	Omettre d'écrire un terme au complet avant de présenter son abréviation
9. Interligne	Ne pas utiliser les doubles interlignes uniformément
10. Utilisation de « & » dans les références	Utiliser « and » (et) dans le texte ou « & » « and » (et) dans les parenthèses.

N.B. Désormais, selon l'APA 7e édition, l'utilisation de « et al. » dans le texte est toujours requise lorsqu'il y a 3 auteurs ou autrices et plus. Par ailleurs, les mêmes erreurs ne se produisent pas nécessairement en français par exemple le « since ». Une autre erreur fréquente en anglais c'est la règle de mettre les premières lettres des mots de titres en majuscule. En français, la majorité du temps les premières lettres des mots de titres sont en minuscules, sauf celle du premier mot du titre. Selon l'adaptation en français utilisée de l'APA, d'autres différences peuvent exister.

Ressources en ligne sur le style APA

La meilleure source d'information sur le style APA 7e édition est le manuel de publication lui-même. Il existe également de nombreux sites web de qualité sur le style APA, présentant très bien les principes de base pour les personnes novices en recherche. En voici quelques-uns pour les textes rédigés en anglais.

Style APA [en anglais]

<https://apastyle.apa.org/products/publication-manual-7th-edition> QR

Purdue Online Writing Lab [en anglais]

<http://owl.english.purdue.edu/owl/resource/560/01>

En français, il n'y a pas une adaptation officielle des règles de l'APA. Il y a quelques différences d'une adaptation à l'autre. Voici quelques sites de qualité:

Université de Montréal, Québec

<https://bib.umontreal.ca/citer/styles-bibliographiques/apa>

Marc Couture. Adaptation française de l'APA (7e éd.), sous licence [CC BY-NC 4.0](#)

<http://profmcouture.ca/apa/index.htm>

Haute école pédagogique Vaud, Suisse

<https://etudiant.hepl.ch/files/live/sites/files-site/files/filiere-ps/programme-formation/normes-apa-stage-maeps-fps-hep-vaud.pdf>

En plus du style des références, ce dernier document décrit notamment le format des tableaux et des figures, donne des indications sur le format du texte et aborde le langage inclusif. Pour répondre à l'une des particularités de la langue française (qui n'est donc pas discuté dans le Manuel de publication de l'APA), les autres expliquent le langage épique.

En bref, le Tableau 11.3 présente le style APA de l'anglais au français, selon l'adaptation proposée par la bibliothèque de l'Université de Montréal.

Tableau 11.3 Le style APA de l'anglais au français: en bref

Anglais	Français	Explication
" "	« »	
, & ou &	et	
pp.	p.	
2nd ed.	2e éd.	
as cited in	cité dans	citation secondaire dans le texte
n.d.	s.d.	sans date de publication
Retrieved June 24, 2023	Repéré le 24 juin 2023	
Unpublished	Inédit ou inédite	
Ed. ou Eds.	Dir.	Directeur ou directrice scientifique
In	Dans	
et al.	et al.	locution latine
para.	paragr.	paragraphe
manuscript	document	

Sources citées dans le texte et références selon les normes APA

Étant donné que la science est une immense collaboration entre scientifiques, les références aux travaux d'autres membres de la communauté scientifique sont extrêmement importantes. Cette importance se reflète dans l'ensemble des règles détaillées de mise en forme et d'utilisation de ces références.

Références

À la fin d'un article ou d'un chapitre de livre suivant le style APA se trouve une liste contenant les **références** de toutes les sources citées dans le texte (et uniquement les sources citées dans le texte). La liste de références commence sur une page séparée, avec le titre « Références », centré avec une majuscule suivie de minuscules. Les références elles-mêmes sont ensuite classées par ordre alphabétique en fonction du nom de famille du premier auteur ou de la première autrice de chaque citation. (Comme pour le reste du manuscrit, tout doit être double interligne.) De nombreux types d'ouvrages peuvent être cités dans les articles et les chapitres de livres en style APA, notamment des articles de magazines, des sites Internet, des documents gouvernementaux et même des émissions de télévision. Bien entendu vous devez consulter le manuel de publication ou les ressources en ligne sur le style APA pour plus de détails sur la manière de les mettre en page. Nous nous concentrerons ici sur la mise en page des références pour les trois types de travaux les plus courants cités dans le style APA : les articles de périodique, les livres et les chapitres de livres.

Articles de périodique

Pour les articles de périodique, le format générique d'une référence est le suivant :

Autrice, A. A., Auteur, B. B. et Autrice, C. C. (année). Titre de l'article. *Titre du journal*, volume(numéro), p.-p.
<https://doi.org/xx.xxxxxxxx>

Voici un exemple concret:

Martin, P., Speranza, M. et Colombel, F. (2022). Les capacités mnésiques de l'enfant présentant un trouble déficit de l'attention : état des lieux. *Neuropsychiatrie de l'enfance et de l'adolescence*, 70(3), 111-116.
<https://doi.org/10.1016/j.neurenf.2021.12.004>

Il y a plusieurs choses à noter ici. La référence comprend un retrait suspendu (en anglais, *hanging indent*). Cela signifie que la première ligne de la référence n'est pas en retrait, mais que toutes les lignes suivantes le sont. Les noms des auteurs et des autrices apparaissent dans le même ordre que celui de l'article, ce qui reflète leur contribution relative à la recherche. Seuls les noms de famille et les initiales des auteurs et autrices apparaissent, et les noms sont séparés par des virgules avec « et » entre les deux derniers noms. Cette règle s'applique même lorsqu'il n'y a que deux auteurs ou autrices. Toute personne identifiée comme auteurs ou autrices est mentionnée en suivant ce format, mais s'il y a plus de 20 personnes, on écrit les 19 premières, trois points de suspension (...) suivis du nom de la dernière personne. Seul le premier mot du titre de l'article prend une majuscule. Les seules exceptions concernent les mots qui sont des noms propres (p. ex.

« Freud ») ou s'il y a un sous-titre, auquel cas le premier mot du sous-titre prend également une majuscule. Si le titre d'une revue scientifique est en français, il n'y aura que la première lettre du premier mot en majuscule, autrement le reste sera en minuscule (p. ex. *Distances et médiations des savoirs*). En revanche si le titre de la revue est en anglais, tous les mots importants prennent une majuscule (p. ex. *American Psychologist*). Le titre de la revue et le volume du journal sont en italique, mais pas le numéro (le chiffre entre parenthèses). À la toute fin de la référence se trouve l'identifiant de l'objet numérique (DOI), qui fournit un lien permanent vers l'emplacement de l'article sur Internet. Incluez-le s'il est disponible. Il se trouve généralement dans la fiche de l'article sur une base de données électronique (p. ex. PsycINFO) et est habituellement affiché sur la première page de l'article publié.

Livres

Pour un livre, le format générique et un exemple concret sont les suivants :

Autrice, A. A. (année). *Titre du livre* (xe éd., vol. y). Maison d'édition.

Lemaire, P. et Didierjean, P. (2018). *Introduction à la psychologie cognitive* (3e éd.). De Boeck Supérieur.

Chapitres de livre

Pour un chapitre dans un livre dirigé, le format générique et un exemple concret sont les suivants :

Auteur, A. A., Autrice, B. B. et Auteur C. C. (année). Titre du chapitre. Dans A. A. Directrice/Directeur, B. B. Directeur/Directrice et C. C. Directrice/Directeur (dir.), *Titre du livre* (xe éd., vol. x, p. xx-yy). Maison d'édition.

Séguin, C. (2021). La recension des écrits et la recherche documentaire. Dans I. Bourgeois (dir.), *Recherche sociale : De la problématique à la collecte des données* (7e éd., p. 65-84). Presses de l'Université du Québec.

Remarquons que les références des livres et des chapitres de livres sont similaires à celles des articles de périodique, mais qu'il y a aussi des différences. Dans le cas d'un livre dirigé, les noms des responsables de l'édition apparaissent avec les initiales de leur prénom et de leur deuxième prénom suivis de leur nom de famille (et non l'inverse) — avec l'abréviation « dir. » qui apparaît entre parenthèses immédiatement après le nom du dernier directeur ou de la dernière directrice scientifique. Seul le premier mot du titre d'un livre prend une majuscule, et le titre entier est en italique. Pour un chapitre dans un livre dirigé, les numéros de page du chapitre apparaissent entre parenthèses après le titre du livre avec l'abréviation « p. ». Enfin les deux formats se terminent par la maison d'édition.

Sources citées dans le texte

Lorsque vous faites référence à l'idée d'autres personnes en recherche dans le texte, vous devez inclure une **référence dans le texte** à l'ouvrage dans lequel vous avez trouvé cette idée à l'origine, ainsi qu'une référence complète de cette source dans la liste de références. Qu'est-ce qui constitue une idée qui doit être citée ? En général il s'agit de phénomènes découverts par d'autres scientifiques, de théories élaborées

par ces personnes, d'hypothèses dérivées et de méthodes spécifiques utilisées par ces personnes (p. ex. des questionnaires spécifiques ou du matériel de stimulation). Les informations factuelles qui ne sont pas connues de tout le monde doivent également être citées afin que les autres scientifiques puissent aller vérifier ces faits dans les sources. Par exemple dans un article sur l'effet de l'utilisation du téléphone cellulaire sur l'aptitude à conduire, l'autrice ou l'auteur pourrait citer des statistiques officielles sur le nombre d'accidents liés à l'utilisation du téléphone cellulaire se produisant chaque année. Parmi les idées qui n'ont pas besoin d'être citées figurent des concepts méthodologiques et statistiques largement partagés (p. ex. devis intersujets, test *t*) et des affirmations si générales qu'il serait difficile pour quiconque de les contester (p. ex. « La mémoire de travail joue un rôle dans de nombreuses activités quotidiennes »). Attention toutefois, les « connaissances communes » sur le comportement humain sont souvent erronées. Par conséquent en cas de doute, trouvez une référence appropriée à citer ou supprimez l'affirmation douteuse.

Lorsque vous citez une source dans le texte de votre manuscrit, il y a deux façons de le faire. Dans les deux cas, seuls les noms de famille des auteurs ou autrices et l'année de publication sont mentionnés. La première méthode est la citation narrative qui consiste à utiliser les noms de famille des auteurs ou autrices dans la phrase (sans prénoms ni initiales), suivis immédiatement de l'année de publication entre parenthèses. Voici quelques exemples :

- Burger (2008) a reproduit l'étude originale de Milgram (1963) sur l'obéissance.
- Bien que de nombreuses personnes pensent que les femmes sont plus bavardes que les hommes, Mehl et al. (2007) n'ont trouvé pratiquement aucune différence dans le nombre de mots prononcés par les étudiantes et les étudiants.

Il convient de noter plusieurs choses. Premièrement le nom des autrices et auteurs est traité grammaticalement comme des noms de personnes, et non comme des choses. Il est préférable d'écrire « une réplique de l'étude de Milgram (1963) » plutôt que d'écrire « une réplique de Milgram (1963) ». Deuxièmement lorsqu'il y a deux auteurs ou autrices, les noms sont séparés par « et », mais lorsqu'il y a trois auteurs ou plus, le nom de la première autrice ou du premier auteur est suivi par « et al. ». Troisièmement l'année suit immédiatement le nom de la dernière autrice ou du dernier auteur. Pour voir des exemples de citations narratives, courtes, longues et multiples, consultez le site de l'APA [en anglais] : <https://apastyle.apa.org/style-grammar-guidelines/citations/paraphrasing>

La deuxième méthode de citer un article ou un chapitre de livre est de le faire entre parenthèses — en incluant le nom de famille des auteurs ou autrices et l'année de publication entre parenthèses à la suite de l'idée citée. On la nomme la citation entre parenthèses. Voici quelques exemples :

- Les gens peuvent être étonnamment obéissants aux figures d'autorité (Burger, 2008 ; Milgram, 1963).
- Des données récentes suggèrent que les hommes et les femmes sont aussi loquaces (Mehl et al., 2007).

Une chose à noter à propos de ces citations entre parenthèses est qu'elles sont souvent placées à la fin de la phrase, ce qui minimise la perturbation du flux de cette phrase. Notons également que lorsqu'il y a plusieurs citations dans le même ensemble de parenthèses, elles sont classées par ordre alphabétique du nom du premier auteur et séparées par des points-virgules.

Il n'existe pas de règles strictes pour décider lequel des deux styles de citation doit être utilisé. La plupart

des articles et des chapitres de livres contiennent un mélange des deux. En général cependant, la première approche fonctionne bien lorsque vous souhaitez mettre l'accent sur la personne qui a mené la recherche — par exemple si vous comparez les théories de deux personnes célèbres dans ce domaine de recherche. Elle est également efficace lorsque vous décrivez une étude particulière en détail. La deuxième approche fonctionne bien lorsque vous discutez d'une idée générale et surtout lorsque vous souhaitez inclure plusieurs citations pour la même idée.

Dans le Tableau 11.2, la troisième erreur la plus fréquente concerne l'utilisation de « et al. ». Il s'agit de l'abréviation du terme latin « et alia », qui signifie « et autres ». Dans le style APA, si un article ou un chapitre de livre a plus de deux auteurs, vous devez utiliser le nom du premier auteur ou de la première autrice suivi de « et al. ». Si l'article n'a que deux auteurs, les deux doivent être mentionnés dans chaque citation. Voici quelques exemples :

- Rappelons que Mehl et al. (2007) ont constaté que les femmes et les hommes prononçaient en moyenne le même nombre de mots par jour.
- Il existe une forte corrélation positive entre le nombre de tracas quotidiens et le nombre de symptômes ressentis par les personnes (Kanner et al., 1981).

Remarquons qu'il n'y a pas de virgule entre le nom du premier auteur et « et al. ». Notons également qu'il n'y a pas de point après « et », mais qu'il y en a un après « al ». Cela s'explique par le fait que « et » est un mot complet et que « al. » est l'abréviation du mot « alia ».

Enfin même si l'on mentionne la source dans le texte, il faut faire attention aux mots utilisés, autrement on pourrait se faire accuser de plagiat. C'est ce dont il est question dans l'encadré intitulé « Paraphraser, ça s'apprend! ».

Paraphraser, ça s'apprend!

Lorsque vous faites référence à l'idée d'une tierce personne, ou même lorsque vous faites référence à un de vos ouvrages passés, vous devez l'exprimer dans vos propres mots. C'est ce qu'on appelle la **paraphrase**. Il faut mettre dans ses mots l'idée véhiculée par la personne qui a écrit la source d'origine. La paraphrase requiert que l'idée soit reformulée, expliquée ou exprimée dans vos propres mots; ce peut aussi être une synthèse de l'idée d'origine. Bref, écrire une paraphrase ne se résume pas à copier-coller, puis changer quelques mots dans la phrase. Dans ce cas de figure, ce serait du plagiat. Il vaudrait mieux faire une citation directe mot pour mot de la source, mettre le tout entre guillemet ou en retrait, et en donner la source. La paraphrase d'un texte traduit n'est pas une simple traduction mot pour mot de l'information. Encore une fois il faut écrire dans ses mots l'idée du texte.

Écrire des paraphrases est quelque chose qui s'apprend. Complétez les activités interactives suivantes afin de mieux maîtriser les paraphrases.

Quiz interactif sur l'écriture en paraphrases.

<https://integrite.umontreal.ca/boite-a-outils/quiz-generaux/paraphraser/>

Instructions et activités sur la citation et l'écriture en paraphrases [en anglais]

<https://apastyle.apa.org/instructional-aids/paraphrasing-citation-activities.pdf>

59. Rédiger un article de recherche dans le style de l'Association américaine de psychologie (APA)

Objectifs d'apprentissage

1. Identifier les principales sections d'un article de recherche de style APA et le contenu de base de chaque section.
2. Planifier et rédiger un article de recherche efficace suivant le style APA.

Dans cette section, nous verrons comment rédiger un **article de recherche empirique**, c'est-à-dire un article qui présente les résultats d'une ou plusieurs nouvelles études. Rappelons que les sections standards d'un article empirique constituent un genre de plan. Nous examinerons ici chacune de ces sections en détail, y compris les informations qu'elles contiennent, la manière dont ces informations sont mises en page et organisées, ainsi que des conseils pour la rédaction de chaque section. À la fin de cette section, se trouve un exemple d'article de recherche empirique suivant le style APA qui illustre plusieurs de ces principes.

Sections d'un article de recherche

Page titre et résumé

Un article de recherche suivant le style APA commence par une **page titre**. Le titre est centré dans la moitié supérieure de la page. Le titre doit communiquer de manière claire et concise (en 12 mots ou moins) les variables principales et les questions de recherche. Cela nécessite parfois un titre principal suivi d'un sous-titre qui développe le titre principal, auquel cas le titre principal et le sous-titre sont séparés par deux points. Voici quelques titres de numéros récents publiés par la *Revue québécoise de psychologie*.

- Accompagner les personnes étudiantes aux postsecondaires qui vivent de l'anxiété de performance à l'aide de l'approche d'acceptation et d'engagement
- Expression de la souffrance psychique chez la personne adolescente scolarisée en situation de handicap moteur : Une analyse de l'impact de la qualité de l'environnement scolaire en contexte camerounais

- Les effets d'un programme de formation parentale francophone sur les comportements-défis d'enfants autistes
- Stratégies de gestion de conflits et qualité de la relation conjugale

Le titre est mis en gras. Sous le titre figurent les noms des autrices et auteurs, et sur la ligne suivante, leur affiliation institutionnelle, c'est-à-dire l'université ou l'établissement où ces personnes travaillaient au moment de la recherche. Comme nous l'avons déjà vu, les autrices et auteurs sont classés dans un ordre qui reflète leur contribution à la recherche. Lorsque plusieurs ont contribué à parts égales à la recherche, ces personnes sont souvent classées par ordre alphabétique ou dans un ordre déterminé de manière aléatoire.

C'est tellement mignon ! Dans quelle mesure un titre d'article doit-il être informel ?

Dans certains domaines de la psychologie, les titres de nombreux articles empiriques sont informels d'une manière que l'on pourrait qualifier de « mignonne ». Ils prennent généralement la forme d'un jeu de mots ou d'une expression bien connue en rapport avec le sujet étudié. Voici quelques exemples tirés de numéros récents du journal « *Psychological Science* ».

- 'Smells Like Clean Spirit: Nonconscious Effects of Scent on Cognition and Behavior' (« Ça sent l'esprit propre : effets non conscients de l'odeur sur la cognition et le comportement »)
- 'Time Crawls: The Temporal Resolution of Infants' Visual Attention' (« Le temps rampe : la résolution temporelle de l'attention visuelle des nourrissons »)
- 'Scent of a Woman: Men's Testosterone Responses to Olfactory Ovulation Cues' (« Le parfum d'une femme : réponses de la testostérone masculine aux signaux olfactifs de l'ovulation »)
- 'Apocalypse Soon?: Dire Messages Reduce Belief in Global Warming by Contradicting Just-World Beliefs' (« L'apocalypse bientôt ? Les messages alarmistes réduisent la croyance au réchauffement climatique en contredisant la croyance en un monde juste »)
- 'Serial vs. Parallel Processing: Sometimes They Look Like Tweedledum and Tweedledee but They Can (and Should) Be Distinguished' (« Traitement sériel vs traitement parallèle : parfois ils se ressemblent comme Tweedledum et Tweedledee, mais ils peuvent (et doivent) être distingués »)
- 'How Do I Love Thee? Let Me Count the Words: The Social Effects of Expressive Writing' (« Comment t'aimer ? Laisse-moi compter les mots : les effets sociaux de l'écriture expressive »)

La préférence des scientifiques pour ce type de titre varie considérablement. Certains les utilisent régulièrement, tandis que d'autres ne les utilisent jamais. Quels sont les avantages et les inconvénients de l'utilisation de jolis titres d'articles ?

Pour les articles soumis en vue de publication, la page titre comprend également une note de l'autrice ou l'auteur qui énumère toutes les affiliations institutionnelles des autrices et auteurs, les remerciements adressés aux agences ayant financé la recherche ou aux collègues ayant commenté l'article, ainsi que les coordonnées des autrices et auteurs. Par ailleurs comme entête de toutes les pages, il y a un titre abrégé mis en majuscule. Pour les travaux des personnes étudiantes qui ne sont pas soumis en vue de publication — y compris les thèses — les notes d'autrice ou d'auteur ne sont généralement pas nécessaires.

Le **résumé** est une synthèse de l'étude. Il constitue la deuxième page du document et est intitulé *Résumé*. La première ligne n'est pas en retrait. Le résumé présente la question de recherche, un résumé de la méthode, les résultats principaux et les conclusions les plus importantes. Comme le résumé est généralement limité à environ 200 mots, il peut être difficile d'en rédiger un de qualité.

Introduction

L'**introduction** commence à la troisième page du document. Le titre en haut de cette page est le titre complet du document. L'introduction comprend trois sous-sections distinctes, bien qu'elles ne soient généralement pas identifiées par des titres séparés. L'ouverture présente la question de recherche et explique pourquoi elle est intéressante, l'analyse documentaire examine les recherches antérieures pertinentes, et la fermeture reformule la question de recherche et commente la méthode utilisée pour y répondre.

L'ouverture

L'**ouverture**, qui tient généralement en un ou deux paragraphes, présente la question de recherche et explique pourquoi elle est intéressante. Pour capter l'attention du lectorat, le chercheur Daryl Bem recommande de commencer par des observations générales sur le sujet étudié, exprimées dans un langage ordinaire (pas de jargon technique) — des observations qui portent sur les gens et leur comportement (et non sur les scientifiques ou leurs recherches ; Bem, 2003). Des exemples concrets sont souvent très utiles à cet endroit. Selon Bem, voici une mauvaise façon de commencer un article :

La théorie de la dissonance cognitive de Festinger a suscité beaucoup d'attention au cours de la dernière partie du 20^e siècle (p. 191).

Voici un exemple beaucoup plus adéquat:

Une personne, qui a deux croyances incompatibles l'une avec l'autre, peut ressentir de l'inconfort. Par exemple une personne qui sait qu'elle aime fumer, mais qui croit que cela est mauvais pour la santé, peut éprouver un malaise découlant de l'incompatibilité ou de la disharmonie entre ces deux pensées ou cognitions. Ce sentiment d'inconfort a été nommé la dissonance cognitive par le psychologue social Leon Festinger (1957), qui a suggéré que les personnes seraient motivées à supprimer cette dissonance par tous les moyens possibles (traduction, p. 191).

Après avoir capté l'attention du lectorat, l'introduction doit ensuite présenter la question de recherche et expliquer pourquoi elle est intéressante. La réponse comblera-t-elle une lacune dans la littérature ? Permettra-t-elle de mettre à l'épreuve une théorie importante ? Aura-t-elle des implications pratiques ? Donner au lectorat une idée claire de l'objet de la recherche et de son importance les incitera à poursuivre la lecture de l'analyse documentaire — et les aidera à en saisir le sens.

Briser les règles

Le chercheur Larry Jacoby a rapporté plusieurs études montrant qu'un mot vu ou entendu à plusieurs reprises peut sembler plus familier, même si l'on ne se souvient pas de ces répétitions — et que cette tendance est particulièrement prononcée chez les adultes plus âgés. Il a ouvert son article par l'anecdote humoristique suivante :

Un ami dont la mère présente les symptômes de la maladie d'Alzheimer raconte qu'il a emmené sa mère visiter une maison de retraite en vue de son futur déménagement. Lors d'une réunion d'orientation à la maison de retraite, les règles et règlements ont été expliqués, l'un d'entre eux concernait la salle à manger. La salle à manger a été décrite comme similaire à celle d'un bon restaurant, à l'exception du fait que le pourboire n'était pas requis. L'absence de pourboire était un thème central de la réunion, mentionné fréquemment pour souligner la qualité des soins et les avantages d'avoir payé à l'avance. À la fin de la réunion, on a demandé à la mère de l'ami si elle avait des questions. Elle a répondu qu'elle n'en avait qu'une : « Dois-je laisser un pourboire ? » (Jacoby, 1999, p. 3)

Bien que l'humour et les anecdotes personnelles soient généralement déconseillés dans les écrits de style APA, cet exemple est une façon très efficace de commencer, car il capte l'attention du lectorat et fournit un excellent exemple concret du sujet à l'étude.

L'analyse documentaire de la recherche

Immédiatement après l'ouverture vient l'**analyse documentaire de la recherche**, (en anglais *literature review*) qui décrit les recherches antérieures pertinentes sur le sujet, et dont la longueur peut varier de quelques paragraphes à plusieurs pages. Toutefois l'analyse documentaire n'est pas une simple liste d'études antérieures. Elle constitue plutôt une forme d'argumentaire visant à démontrer pourquoi la question de recherche mérite d'être abordée. À la fin de la lecture de l'analyse documentaire, le lectorat doit être convaincu que la question de recherche est cohérente et que la présente étude est une étape logique dans le processus de recherche en cours.

Comme tout argumentaire efficace, l'analyse documentaire doit avoir une certaine structure. Par exemple elle peut commencer par décrire un phénomène de manière générale, l'accompagner de plusieurs études qui l'illustrent, puis décrire deux ou plusieurs théories concurrentes de ce phénomène, et enfin proposer une hypothèse pour tester une ou plusieurs des théories. Elle peut aussi décrire un premier phénomène, puis un autre phénomène qui semble incompatible avec le premier, proposer une théorie qui résout l'incompatibilité, et enfin formuler une hypothèse pour tester cette théorie. Dans la recherche appliquée, l'analyse documentaire peut décrire un phénomène ou une théorie, puis décrire comment ce phénomène ou cette théorie s'applique à une situation concrète importante, puis proposer un moyen de tester si elle s'applique effectivement à cette situation.

Considérer l'analyse documentaire sous cet angle met en évidence plusieurs points. Premièrement il est extrêmement important de commencer par un plan des idées principales à présenter, organisé dans l'ordre

où vous souhaitez les développer. La structure de base de l'argumentation devrait ainsi être claire dans le plan. Deuxièmement il est essentiel de mettre en évidence cette structure dans votre rédaction. Une façon de le faire est de commencer l'analyse documentaire en résumant l'argumentation, avant même de la développer : « Dans cet article, je décrirai deux phénomènes apparemment contradictoires, présenterai une nouvelle théorie susceptible de résoudre cette contradiction, et formulerai finalement une nouvelle hypothèse pour tester la théorie. » Une autre manière est d'ouvrir chaque paragraphe par une phrase qui résume l'idée principale du paragraphe et la relie aux points précédents. Ces phrases d'introduction constituent les « transitions » qui posent des problèmes à de nombreuses personnes novices en recherche. Au lieu de commencer un paragraphe par une description d'une étude antérieure, telle que « Williams (2004) a trouvé que... », il est préférable de commencer par indiquer pourquoi vous décrivez cette étude en particulier. Voici quelques exemples simples :

- Un autre exemple de ce phénomène provient des travaux de Williams (2004).
- Williams (2004) propose une explication de ce phénomène.
- Une perspective alternative a été fournie par Williams (2004).
- Nous avons utilisé une méthode basée sur celle de Williams (2004).

Enfin rappelez-vous que votre objectif est de construire un argumentaire expliquant pourquoi votre question de recherche est intéressante et mérite d'être étudiée — et pas nécessairement pourquoi votre réponse préférée à cette question est correcte. En d'autres termes, votre analyse documentaire doit être équilibrée. Si vous souhaitez mettre l'accent sur la généralité d'un phénomène, vous devez discuter des différentes études qui l'ont démontrée. Toutefois si d'autres études n'ont pas réussi à le démontrer, vous devez également les mentionner. Ou encore si vous proposez une nouvelle théorie, vous devez bien sûr discuter des résultats qui sont compatibles avec cette théorie. Toutefois si d'autres résultats sont incompatibles avec cette théorie, vous devez également en parler. Il est acceptable d'affirmer que l'ensemble des recherches tend à soutenir l'existence d'un phénomène ou à être compatible avec une théorie (et c'est généralement le mieux qu'on puisse espérer en psychologie), mais il n'est pas acceptable d'ignorer les preuves contradictoires. D'ailleurs une grande partie de l'intérêt d'une question de recherche réside dans l'incertitude quant à sa réponse.

La fermeture

La **fermeture** de l'introduction — généralement le dernier ou les deux derniers paragraphes — comprend généralement deux éléments importants. Le premier est un énoncé clair de la principale question de recherche et de l'hypothèse. Cette déclaration est généralement plus formelle et plus précise que dans l'ouverture, et est souvent exprimée en termes de définitions opérationnelles des variables clés. Le deuxième élément est une brève présentation de la méthode ainsi qu'un commentaire sur sa pertinence. Voici par exemple comment Darley et Latané (1968) ont conclu l'introduction de leur article classique sur l'effet du témoin (ou l'effet du spectateur) :

Ces considérations conduisent à l'hypothèse que plus il y a de témoins lors d'une urgence, moins il est probable, ou plus il sera lent, qu'une de ces personnes intervienne pour fournir de l'aide. Pour tester cette proposition, il est nécessaire de créer une situation où une « urgence » réaliste pourrait vraisemblablement se produire. La communication entre les personnes participantes doit également être empêchée afin de

prévenir l'obtention d'informations sur le comportement des autres durant la situation d'urgence. Enfin la situation expérimentale doit permettre d'évaluer la vitesse et la fréquence de la réaction des personnes à l'urgence. L'expérience rapportée ci-dessous a tenté de remplir ces conditions. (traduction, p. 378)

L'introduction mène directement à la section suivante de l'article, la méthode.

Méthode

La **section méthode** décrit la manière dont vous avez mené votre étude. Un principe important pour la rédaction d'une section sur la méthode est qu'elle doit être suffisamment claire et détaillée pour que d'autres personnes en recherche puissent reproduire l'étude en suivant votre « recette ». Cela signifie qu'elle doit décrire tous les éléments importants de l'étude — les caractéristiques démographiques de base des personnes participantes, la manière dont elles ont été recrutées, si elles ont été assignées aléatoirement aux conditions, comment les variables ont été manipulées ou mesurées, comment le contrebalancement a été réalisé, etc. En même temps la méthode doit éviter les détails non pertinents tels que le fait que l'étude a été menée dans la salle de classe 37B du bâtiment de technologie industrielle ou que le questionnaire était recto verso et rempli à l'aide de crayons.

La section sur la méthode commence immédiatement après l'introduction par le titre « Méthode » (et non « Méthodes ») centré sur la page. Immédiatement après, on trouve le sous-titre « Participants », justifié à gauche et en italique. La sous-section « Participants » indique le nombre de personnes participantes, le nombre pour chaque genre documenté (par exemple masculin, féminin), une indication de leur âge, d'autres données démographiques qui peuvent être pertinentes pour l'étude, et la méthode de recrutement y compris toute incitation à la participation.

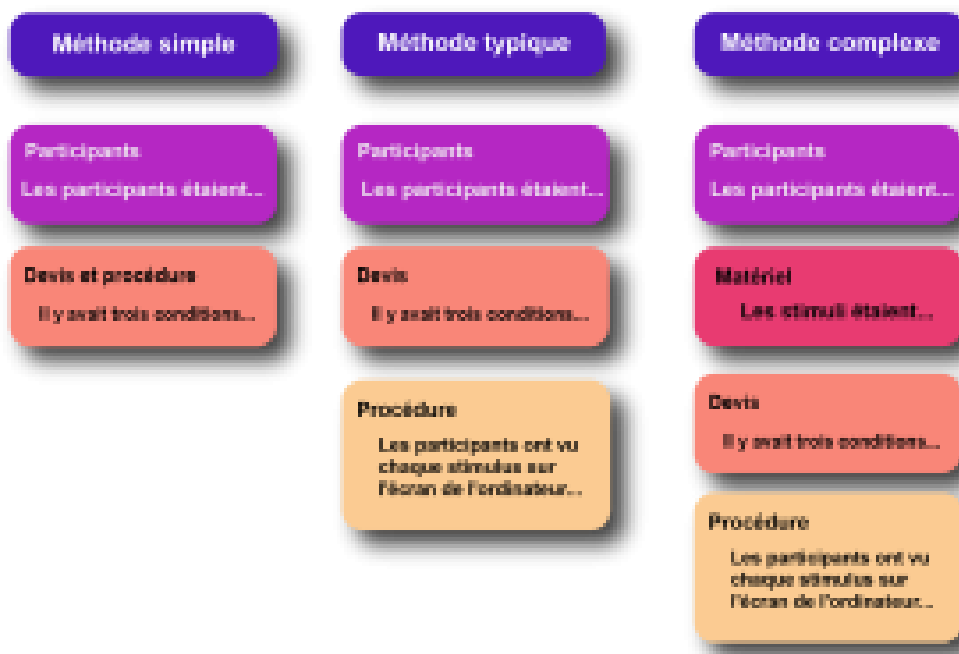


Figure 11.1 Trois façons d'organiser la section méthode.

Après la section participants, la structure peut varier quelque peu. La Figure 11.1 illustre trois approches courantes. Dans la première, la section participants est suivie d'une sous-section sur le devis et la procédure, qui décrit le reste de la méthode. Cette approche fonctionne bien pour les méthodes relativement simples et pouvant être décrites de manière adéquate en quelques paragraphes. Dans la seconde approche, la section participants est suivie de sous-sections distinctes sur le devis et la procédure. Cette approche fonctionne bien lorsque le devis et la procédure sont relativement complexes et nécessitent chacun plusieurs paragraphes.

Quelle est la différence entre le devis et la procédure ? Le devis d'une étude est sa structure générale. Quelles étaient les variables indépendantes et dépendantes ? La variable indépendante a-t-elle été manipulée, et dans l'affirmative a-t-elle été manipulée entre les sujets ou au sein d'un même sujet ? Comment les variables ont-elles été définies sur le plan opérationnel ? La procédure décrit la façon dont l'étude a été menée. Souvent il est plus clair de décrire ce que les personnes participantes ont fait plutôt que de décrire ce que les personnes menant la recherche ont fait. Par exemple les personnes participantes ont donné leur consentement éclairé, lu une série d'instructions, complété un bloc de quatre essais de pratique, complété un bloc de 20 essais expérimentaux, rempli deux questionnaires, ont reçu le débriefage, puis ont été remerciées pour leur participation.

Dans la troisième façon d'organiser une section sur la méthode, la sous-section participants est suivie d'une sous-section sur le matériel, avant les sous-sections sur le devis et la procédure. Cette méthode fonctionne bien lorsque le matériel à décrire est complexe. Il peut s'agir de questionnaires multiples, de vignettes écrites à lire et auxquels réagir, de stimuli perceptifs, etc. Le titre de cette sous-section peut être modifié pour refléter son contenu. Au lieu de « Matériel », il peut s'agir de « Questionnaires », « Stimuli », etc. C'est également dans cette sous-section que l'on réfère à la fidélité et/ou la validité des mesures. C'est ici que vous présenterez les corrélations test-retest, l' α de Cronbach ou d'autres statistiques montrant que les mesures sont cohérentes dans le temps et entre les items, et qu'elles mesurent bien ce qu'elles sont censées mesurer.

Résultats

La **section des résultats** présente les principaux résultats de l'étude, y compris les résultats des analyses statistiques. Bien qu'elle n'inclut pas les données brutes — les réponses ou les scores individuels des personnes participantes — les personnes en recherche doivent conserver leurs données brutes et les mettre à la disposition des autres scientifiques qui en font la demande. De nombreuses revues encouragent le partage ouvert des données brutes en ligne, et certaines exigent désormais la mise à disposition des données et du matériel avant toute publication.

Bien qu'il n'existe pas de sous-sections standardisées, il reste essentiel que la section résultats soit organisée de manière logique. En général elle commence par certains éléments préliminaires. Le premier porte sur l'exclusion éventuelle de certaines personnes participantes ou de certaines réponses, et les raisons de cette exclusion. Les justifications doivent être clairement décrites afin que d'autres scientifiques puissent juger de leur pertinence. Un deuxième élément préliminaire est la manière dont les réponses multiples ont été combinées pour produire les variables principales dans les analyses. Par exemple si les personnes participantes ont évalué l'attrait de 20 stimuli, il faudrait expliquer que, pour obtenir l'indice d'attractivité

pour chaque personne participante, vous avez calculé la moyenne des évaluations. Ou encore si les gens rappelaient autant d'items que possible à partir d'une liste de 20 mots, avez-vous compté le nombre d'items correctement rappelés, calculé le pourcentage d'items correctement rappelés, ou peut-être calculé le nombre d'items corrects moins le nombre d'items incorrects ? Un dernier élément préliminaire est de savoir si la manipulation expérimentale a été réussie. C'est ici que vous indiquerez les résultats des vérifications de toute manipulation.

La section des résultats doit ensuite aborder les questions de recherche principales, une par une. Là encore l'organisation doit être claire. Une approche consisterait à répondre aux questions les plus générales, puis à répondre aux questions plus spécifiques. Une autre approche consisterait à répondre d'abord à la question principale, puis aux questions secondaires. Quoi qu'il en soit, Bem (2003) suggère la structure de base suivante pour discuter de chaque nouveau résultat :

- Rappeler au lectorat la question de recherche.
- Donner la réponse à la question de recherche en mots.
- Présenter les statistiques pertinentes.
- Nuancer la réponse si nécessaire.
- Résumer les résultats.

Notons que seule l'étape 3 (les statistiques) implique nécessairement des chiffres. Les autres étapes consistent à présenter la question de recherche et la réponse à celle-ci en mots. En fait, les résultats de base devraient être compréhensibles même pour le lectorat qui saute les chiffres.

La discussion

La **discussion** est la dernière grande section de l'article de recherche. Elle se compose généralement d'une combinaison des éléments suivants :

- Résumé de la recherche
- Implications théoriques
- Implications pratiques
- Limites de l'étude
- Suggestions pour les recherches futures

La discussion commence généralement par un résumé de l'étude qui apporte une réponse claire à la question de recherche. Dans le cas d'un article court comportant une seule étude, cela peut se limiter à une phrase. Dans un rapport plus long comportant plusieurs études, cela peut nécessiter un paragraphe, voire deux. Le résumé est souvent suivi d'une discussion sur les implications théoriques de la recherche. Les résultats soutiennent-ils des théories existantes ? Si ce n'est pas le cas, comment peut-on les expliquer ? Bien qu'il ne soit pas nécessaire de fournir une explication définitive ou une théorie détaillée pour vos résultats, il est au minimum nécessaire d'esquisser une ou plusieurs explications possibles. Dans la recherche appliquée — et souvent aussi dans la recherche fondamentale —, les implications pratiques de la recherche sont

également abordées. Comment les résultats peuvent-ils être utilisés, et par qui, pour atteindre un objectif concret dans le monde réel ?

Les implications théoriques et pratiques sont souvent suivies d'une discussion sur les limites de l'étude. La validité interne ou externe de l'étude peut poser un problème. La manipulation n'a peut-être pas été très efficace ou les mesures n'ont pas été très fidèles. Il y a peut-être des indices suggérant que la tâche n'a pas été bien comprise par les personnes participantes ou que ces dernières se méfiaient de l'intention des personnes menant la recherche. C'est le moment de discuter de ces enjeux et de leur influence possible sur les résultats. Mais n'en faites pas trop. Toutes les études ont des limites et la majorité du lectorat comprendra qu'un échantillon différent ou des mesures différentes auraient pu produire des résultats différents. Cependant à moins qu'il n'y ait de bonnes raisons de penser que cela aurait été le cas, il n'y a aucune raison de mentionner ces limites routinières. Il est préférable de se concentrer sur deux ou trois limites qui semblent avoir pu influencer les résultats, d'expliquer comment elles auraient pu le faire et de proposer des moyens d'y remédier.

La plupart des discussions se terminent par des suggestions pour des recherches futures. Si l'étude n'a pas répondu de manière satisfaisante à la question de recherche initiale, que faudra-t-il faire pour y parvenir ? Quelles nouvelles questions de recherche l'étude a-t-elle soulevées ? Cette partie de la discussion n'est toutefois pas une simple liste de nouvelles questions. Il s'agit d'une réflexion autour de deux ou trois enjeux non résolus les plus importants. Cela signifie qu'il faut identifier et clarifier chaque question, suggérer des réponses alternatives et même proposer des moyens de les étudier.

Enfin certaines personnes en recherche sont très habiles pour terminer leurs articles par une conclusion radicale ou stimulante. Darley et Latané (1968) par exemple ont terminé leur article sur l'effet du témoin en discutant de l'idée que le fait d'aider les autres dépendrait davantage de la situation que de la personnalité. Leur dernière phrase est la suivante : « Si les gens comprennent les forces situationnelles qui peuvent les faire hésiter à intervenir, ils pourront peut-être mieux les surmonter » (traduction, p. 383). Cependant ce type de conclusion peut être difficile à réussir. Il peut sembler trop exagéré, banal, et finir par diminuer l'impact global de l'article. Il est souvent préférable de conclure simplement en revenant au problème ou à la question présentés dans le paragraphe d'ouverture et en indiquant clairement comment votre recherche a abordé cette question ou ce problème.

Les références

La section des références débute sur une nouvelle page avec le titre « Références » centré en haut de la page. Toutes les références citées dans le texte sont ensuite énumérées selon le format décrit précédemment. Elles sont présentées par ordre alphabétique du nom de famille de la première autrice ou du premier auteur. Si deux sources ont la même première autrice ou le même premier auteur, elles sont classées par ordre alphabétique du nom de famille de la deuxième personne ayant contribué à la rédaction de l'article. Si toutes les personnes identifiées comme ayant contribué sont les mêmes, les références sont classées par ordre chronologique de l'année de publication. Tout ce qui figure dans la liste de références est en double interligne, tant à l'intérieur des références qu'entre elles.

Annexes, tableaux et figures

Les tableaux et les figures sont tous deux utilisés pour présenter des résultats. Les figures peuvent également être utilisées pour présenter des graphiques, illustrer des théories (p. ex. sous la forme d'un organigramme), présenter des stimuli, schématiser des procédures et présenter de nombreux autres types d'informations. Dans la 7^e édition du Style APA, il y a deux emplacements possibles pour les figures et les tableaux. La première option est de les insérer dans le texte après avoir été introduits pour la première fois dans le texte. La deuxième option est d'insérer chaque tableau ou figure, sur une page distincte après la section référence. Si les figures et tableaux sont insérés dans le texte, ils peuvent prendre une page entière ou ils peuvent apparaître sur la même page que le texte. Cependant leur emplacement doit être au haut ou au bas d'une page (pas au centre de la page). Il doit aussi y avoir un double interligne vide entre le tableau ou la figure et le texte. S'ils sont insérés après les références, il y aura d'abord les tableaux, suivis par les figures. Les tableaux sont numérotés dans l'ordre où ils sont mentionnés pour la première fois dans le texte (« Tableau 1 », « Tableau 2 », etc.). Les figures sont numérotées de la même manière (« Figure 1 », « Figure 2 », etc.). Un bref titre explicatif figure au-dessus de chaque tableau, avec la première lettre du premier mot en majuscule. Chaque figure est accompagnée d'une brève légende explicative, où (à l'exception des noms propres) seul le premier mot de chaque phrase est en majuscule. De plus amples détails sur la préparation de tableaux et de figures de style APA sont présentés plus loin dans l'ouvrage.

Les annexes sont la toute dernière section du document. Une annexe est appropriée pour le matériel supplémentaire qui interromprait le flot de l'article s'il était présenté dans l'une des sections principales. Une annexe peut être utilisée pour présenter des listes de mots, des items d'un questionnaire, des descriptions détaillées d'équipements spéciaux ou d'analyses statistiques inhabituelles, ou une liste de références à des études incluses dans une méta-analyse. Chaque annexe commence sur une nouvelle page. S'il n'y en a qu'une, le titre est « Annexe », centré en haut de la page. S'il y en a plusieurs, les titres sont « Annexe A », « Annexe B », et ainsi de suite, et elles apparaissent dans l'ordre où elles ont été mentionnées pour la première fois dans le texte de l'article.

Échantillon d'un article empirique en style APA

Les Figures 11.2, 11.3, 11.4 et 11.5 présentent quelques pages d'une traduction d'un article de recherche empirique de style APA rédigé par Tomoe Suyama, membre de la communauté étudiante au premier cycle à l'Université d'État de Californie à Fresno. L'objectif principal de ces figures est d'illustrer l'organisation et la mise en page de base d'un article de recherche empirique de style APA, bien que de nombreuses conventions de style de haut niveau et de bas niveau puissent également être observées ici.



Figure 11.2-Page de titre et résumé. Pour le travail d'une personne aux études, il n'y a pas de note d'auteur sur la page de titre.

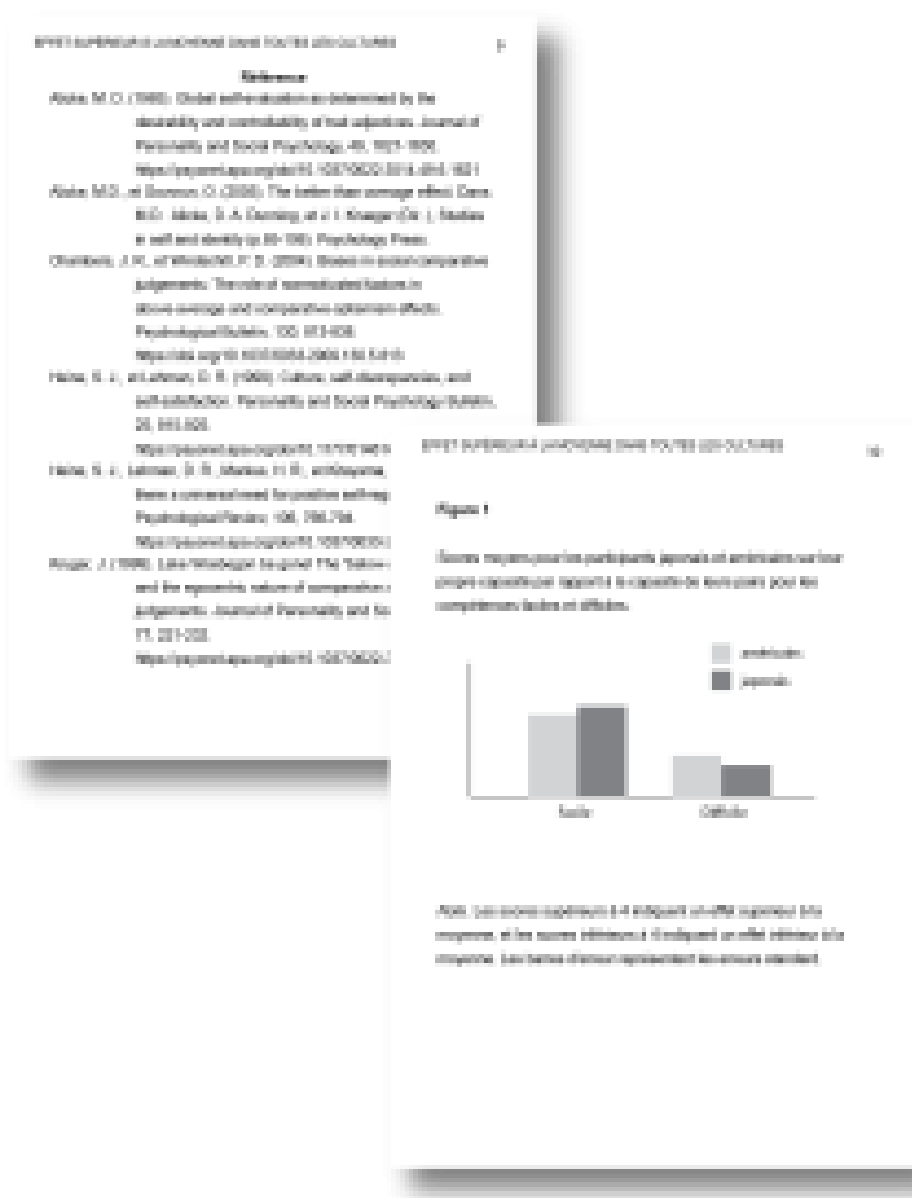


Figure 11.3 – L'introduction et la méthode. Notez que l'introduction est précédée du titre complet et que la section sur la méthode commence immédiatement après la fin de l'introduction.



Figure 11.4-Résultats et discussion La discussion commence immédiatement après la section des résultats.

Figure 11.5-Références et figure. S'il y a des annexes ou des tableaux, ils doivent être placés avant la figure.



Vous pouvez aussi consulter le guide de mise en page pour les travaux des personnes aux études du style APA 7e édition. <https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf> [en anglais]

Description de l'image

Description de l'image de la Figure 11.1 : Tableau montrant trois façons d'organiser une section méthode de style APA.

Dans la méthode simple, il y a deux sous-titres : « Participants » (qui pourrait commencer par « Les participants étaient... ») et « Devis et procédure » (qui pourrait commencer par « Il y avait trois conditions... »).

Dans la méthode classique, il y a trois sous-titres : « Participants » (« Les participants étaient... »), « Devis » (« Il y avait trois conditions... ») et « Procédure » (« Les participants ont vu chaque stimulus sur l'écran de l'ordinateur... »).

Dans la méthode complexe, il y a quatre sous-titres : « Participants » (« Les participants étaient... »), « Matériel » (« Les stimuli étaient... »), « Devis » (« Il y avait trois conditions... »), et « Procédure » (« Les participants ont regardé chaque stimulus sur l'écran de l'ordinateur... »). [\[Retour à la figure 11.1\]](#)

60. Autres formats de présentation

Objectifs d'apprentissage

1. Citer plusieurs façons dont les personnes en psychologie peuvent présenter leur recherche et les situations dans lesquelles elles peuvent les utiliser.
2. Décrire en quoi les documents finaux comme ceux de travaux de personnes aux études diffèrent des documents soumis à des revues scientifiques en vue d'être publiés dans le style APA (Association américaine de psychologie).
3. Décrire l'objectif des communications orales et des communications par affiches lors des congrès académiques.
4. Préparer une courte communication orale de style congrès et une simple communication par affiche.

La rédaction d'un article de recherche empirique dans le style APA n'est qu'un moyen parmi d'autres de présenter une nouvelle recherche en psychologie. Dans cette section, nous examinerons plusieurs autres moyens importants.

Autres types de documents

La section précédente portait sur la rédaction d'articles empiriques destinés à être publiés dans des revues scientifiques. Cependant il existe d'autres types de manuscrits rédigés dans le style APA, dont beaucoup ne seront pas soumis pour publication ailleurs. Nous en examinerons quelques-uns ici.

Articles de synthèse et articles théoriques

Rappelons que les **articles de synthèse** résument la recherche sur un sujet particulier sans présenter de nouveaux résultats empiriques. Lorsque ces articles présentent une nouvelle théorie, ils sont souvent appelés **article théorique**. Les articles de synthèse et les articles théoriques sont structurés de la même manière que les articles empiriques, avec une page titre, un résumé, des références, des annexes, des tableaux et des figures, et ils sont rédigés dans le même style de haut niveau et de bas niveau. Toutefois comme ils ne présentent pas les résultats d'une nouvelle recherche empirique, ils ne comportent pas de section sur les méthodes ou les résultats. Bien entendu le corps du document doit toujours avoir une

organisation logique et comprendre une introduction qui identifie le sujet et explique son importance, une analyse documentaire qui organise les recherches antérieures (en identifiant les relations importantes entre les concepts ou les lacunes dans la recherche), et une conclusion qui résume les principales conclusions et suggère des orientations pour la poursuite de la recherche ou discute des implications théoriques et pratiques. Dans un article théorique, bien entendu, une grande partie du corps du document est consacrée à la présentation de la nouvelle théorie. Les articles théoriques et de synthèse sont généralement divisés en sections, chacune portant un titre approprié. Les sections et leurs titres peuvent varier considérablement d'un article à l'autre (contrairement à un article empirique). Mais quelles qu'elles soient, elles doivent contribuer à organiser le document et à rendre l'argumentation claire.

Documents finaux

Jusqu'à présent nous avons décrit la mise en forme des documents qui seront soumis à une revue scientifique pour publication. En revanche il y a rédaction et préparation d'autres types de documents dans leur forme finale, sans intention de les soumettre pour une publication ailleurs. Il s'agit des **documents finaux** qui comprennent les mémoires, les thèses et autres travaux de personnes aux études. Le style APA 7e édition propose un format spécifique pour les travaux de personnes aux études. Dans ce format, on ajoute sur la page titre le nom et le numéro du cours, le nom du personnel enseignant, ainsi que la date de remise. De plus l'entête de chaque page se limite au numéro de page; aucun court titre n'y figure. Le résumé n'est pas requis non plus. Enfin deux emplacements sont encore possibles pour les tableaux et les figures. Ils peuvent être insérés après la section références ou insérés à proximité de l'endroit où ils sont évoqués. Cette dernière option évite au lectorat d'avoir à feuilleter la fin du document pour les voir. Pour les travaux universitaires, il est important de vérifier les spécifications du format auprès de la personne responsable du cours. Dans un cours sur les méthodes de recherche, les travaux doivent généralement être rédigés comme s'il s'agissait de documents professionnels destinés à être publiés.

Présentations lors de congrès

L'une des façons pour les scientifiques en psychologie de partager leurs recherches est de les présenter lors de **congrès académiques**. (Même si des congrès professionnels en psychologie s'intéressent principalement aux questions associées à la pratique clinique, nous nous intéressons ici aux congrès qui se dédient à la recherche.) Les congrès peuvent aller d'événements à petite échelle impliquant une douzaine de scientifiques qui se réunissent pour un après-midi, à des événements à grande échelle impliquant des milliers de personnes en recherche qui se réunissent pendant plusieurs jours. Bien que les personnes en recherche prenant part à un congrès académique soient susceptibles de discuter de leurs travaux entre eux de manière informelle, il existe deux types de présentations plus formelles : les communications orales (« exposés ») et les communications par affiches. Pour présenter une communication orale ou une affiche lors d'un congrès, il faut généralement soumettre à l'avance un résumé de la recherche à l'équipe organisatrice

du congrès, et obtenir leur approbation – même si le processus d'évaluation par les pairs n'est généralement pas aussi rigoureux que pour les documents soumis à une revue scientifique.

Communications orales

Lors d'une **communication orale**, ou « exposé », la personne invitée à prendre la parole se tient devant un public d'autres scientifiques et leur parle de sa recherche — généralement à l'aide d'un diaporama. Les exposés durent généralement de 10 à 20 minutes, les dernières minutes étant réservées aux questions du public. Dans les grands congrès, les exposés sont généralement regroupés en sessions d'une à deux heures, au cours desquelles tous les exposés portent sur le même thème général.

Lors de la préparation orale, les personnes invitées à présenter doivent garder à l'esprit plusieurs principes généraux. Le premier est que le nombre de diapositives ne doit pas dépasser une par minute de l'exposé. Deuxièmement les exposés sont généralement structurés comme un article de recherche de style APA. Il y a une diapositive avec le titre et le nom des personnes autrices et auteurs, quelques diapositives pour présenter le contexte, quelques autres pour décrire la méthode, quelques diapositives pour les résultats et quelques diapositives pour les conclusions. Troisièmement la personne qui présente doit regarder les membres de l'auditoire et s'adresser à eux sur un ton de conversation moins formel que celui de la rédaction en style APA, mais plus formel qu'une conversation avec une ami-e. Les diapositives ne doivent pas être au centre de la présentation ; elles doivent servir d'appui visuel. À ce titre elles doivent présenter les points principaux sous forme de listes à puces ou de tableaux et figures simples.

Communications par affiches

Une autre façon de présenter la recherche lors d'un congrès est de le faire sous la forme d'une **communication par affiche**. L'affiche est présentée pendant une **session de communications par affiches** d'une à deux heures qui se déroule dans une grande salle sur le site du congrès. Les personnes chargées de la présentation placent leurs affiches sur des tableaux d'affichage disposés dans la salle et se tiennent à proximité. Les autres membres de la conférence circulent ensuite dans la salle, lisent les affiches et discutent avec les personnes présentant leurs affiches. En fait les sessions de communications par affiches sont une version adulte de l'expo-sciences de l'école. Mais elles n'ont rien d'enfantin. Les affiches sont utilisées par les scientifiques d'expériences dans toutes les disciplines, et elles sont de plus en plus courantes. Lors d'un récent congrès de l'American Psychological Association, près de 2 000 affiches ont été présentées au cours de 16 sessions distinctes. L'une des raisons pour lesquelles les affiches sont si populaires est qu'elles encouragent une interaction directe et significative entre les personnes en recherche.

Les affiches sont généralement de grande taille, environ 1.22 mètre de large par 0.91 mètre de haut. Les informations contenues dans l'affiche sont organisées en sections distinctes, comprenant un titre, les noms et affiliations des autrices et auteurs, une introduction, une section sur la méthode, une section sur les résultats, une section sur la discussion ou les conclusions, les références et des remerciements. Bien que les affiches puissent inclure un résumé, celui-ci n'est pas forcément nécessaire car l'affiche elle-même constitue déjà un

bref résumé de la recherche. La Figure 11.6 montre deux façons différentes d'organiser les informations sur une affiche.

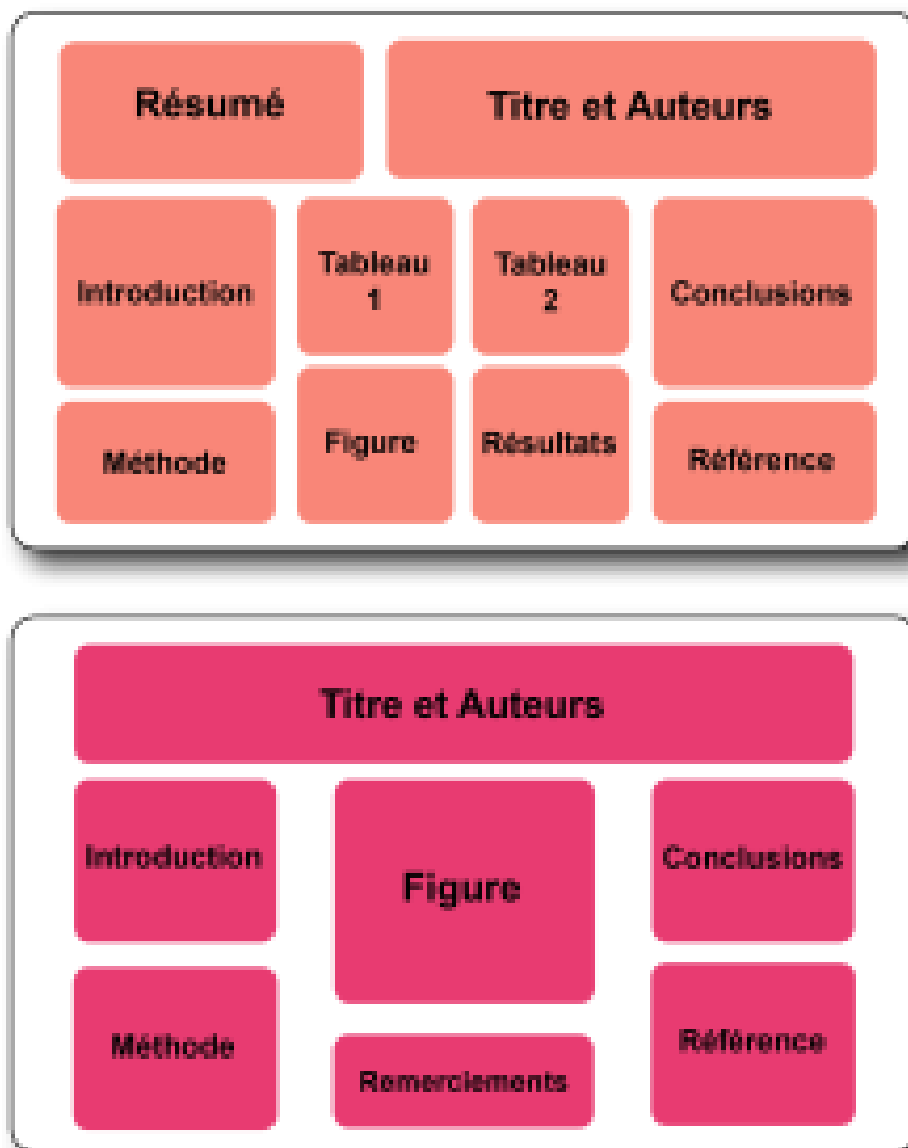


Figure 11.6-Deux façons possibles d'organiser l'information sur une affiche

Étant donné les conditions dans lesquelles les affiches sont souvent présentées — par exemple dans de grandes salles bondées où les gens mangent, boivent et socialisent — elles doivent être conçues de manière à présenter les idées principales de la recherche de la manière la plus simple et la plus claire possible. La taille des caractères sur une affiche doit être grande — par exemple 72 points pour le titre et le nom des personnes responsables de cette communication et 28 points pour le texte principal. Les informations doivent être organisées en sections avec des titres clairs, et le texte doit être divisé en phrases ou en points à puces plutôt qu'en paragraphes. Il est également préférable qu'elles soient organisées en colonnes allant du haut vers le bas plutôt qu'elles soient organisées en rangée allant de gauche à droite sur l'affiche. Il est

ainsi plus facile pour plusieurs personnes de lire en même temps sans risquer de se bousculer. Les affiches incluent souvent des éléments qui ajoutent de l'intérêt visuel. Les figures peuvent être plus colorées que celles d'un document de style APA. Les affiches peuvent également inclure des exemples de stimuli visuels, des photographies de l'appareil ou une simulation des personnes testées. Elles peuvent également inclure des éléments purement décoratifs (comme le logo de l'affiliation), bien qu'il soit préférable de ne pas en abuser.

Encore une fois l'une des principales raisons pour lesquelles les affiches sont devenues un moyen si populaire de présenter des recherches est qu'elles facilitent l'interaction entre les scientifiques. De nombreuses personnes proposent d'emblée de décrire leur recherche aux autres qui les visitent et d'utiliser l'affiche comme support visuel. Au minimum, il est important que les personnes responsables d'une communication se tiennent à côté de leur affiche, saluent les gens qui les visitent, offrent de répondre à leurs questions, soient aptes à répondre aux questions et même aux commentaires critiques occasionnels. Pour les personnes visitant et souhaitant plus d'informations, il est généralement judicieux de mettre à leur disposition un résumé plus détaillé de la recherche, de leur proposer de leur envoyer un résumé détaillé ou de leur fournir les coordonnées (p. ex. une carte d'affaires) leur permettant de demander plus d'informations par la suite.

Pour plus d'informations sur la préparation et les communications orales et les communications par affiches, consultez le site internet de la bibliothèque de l'Université de Montréal à la page intitulée « Participer aux échanges scientifiques: contribuer à des colloques et congrès », boite-outils.bib.umontreal.ca/recherche/participer-echanges-scientifiques et le site internet du carrefour communication scientifique de l'Université du Québec uquebec.libguides.com/carrefour-communication-scientifique.

Congrès académiques

Vous trouverez ci-dessous des liens vers des sites Internet de congrès provinciaux, nationaux et internationaux ainsi que vers un congrès présentant les travaux de personnes étudiant au premier cycle. Pour une liste complète des congrès en psychologie dans le monde entier, voir conferencealerts.com/topic-listing?topic=Psychology

Congrès en français ou avec un volet en français

Congrès de l'ACFAS : www.acfas.ca/evenements/congres

Congrès de la Société québécoise pour la recherche en psychologie (SQRP): www.sqrp.ca/

Congrès de la Société canadienne de psychologie (SCP) convention.cpa.ca/?lang=fr

Congrès de la Société canadienne des sciences du cerveau, du comportement et de la cognition (SCSCCC) www.csbbcs.org/fr/page-daccueil

Congrès international de psychologie clinique fr.clinicalpsychologycongress.com/

Congrès pour la communauté étudiante du premier cycle, exclusivement en anglais

Connecting Minds Undergraduate Research Conference: www.connectingminds.ca

Science Atlantic Psychology Conference: scienceatlantic.ca/conferences/

Grands congrès, exclusivement en anglais

Association for Psychological Science Conference: www.psychologicalscience.org/index.php/convention

Society for Personality and Social Psychology Conference: meeting.spsp.org/

Psychonomic Society Annual Meeting: www.psychonomic.org/annual-meeting

American Psychological Association Convention: www.apa.org/convention

Description de l'image

Figure 11.6 description de l'image:

Deux graphiques illustrant les façons d'organiser l'information sur une affiche.

Dans le premier graphique, le résumé, le titre et le nom des personnes responsables de cette communication (auteurs ou autrices) apparaissent en haut de l'affiche. Sous le résumé et le titre se trouvent quatre colonnes. De haut en bas, de gauche à droite, les colonnes contiennent l'introduction et la méthode, le tableau 1 et une figure, le tableau 2 et les résultats, ainsi que les conclusions et les références.

Dans le deuxième graphique, le titre et les auteurs et autrices apparaissent en haut de l'affiche. Sous le titre se trouvent trois colonnes. De haut en bas, de gauche à droite, les colonnes contiennent l'introduction et la méthode, une figure et les remerciements, et les conclusions et les références. [\[Retour à la Figure 11.6\]](#)

61. Points clés et exercices

Points clés

- Le style APA est un ensemble de lignes directrices pour la rédaction en psychologie. C'est le genre d'écriture que les psychologues utilisent pour communiquer leurs recherches avec d'autres personnes en recherche et psychologues travaillant en clinique.
- Le style APA peut être considéré comme ayant trois niveaux. Il y a l'organisation d'un article de recherche, le style de haut niveau qui comprend l'écriture formelle, directe et utilisant un langage libre de biais, et le style de bas niveau qui consiste en de nombreuses règles spécifiques de grammaire, d'orthographe, de mise en page des références, etc.
- Les références et les citations de références constituent une partie importante du style APA. Il existe des règles spécifiques pour la mise en page des références, leur citation dans le texte d'un article.
- Un article empirique de style APA se compose de plusieurs sections typiques. Les principales sont le résumé, l'introduction, la méthode, les résultats, la discussion et les références.
- L'introduction se compose d'une ouverture qui présente la question de recherche, d'une analyse documentaire qui décrit les recherches antérieures sur le sujet, et d'une fermeture qui réaffirme la question de recherche et commente la méthode. L'analyse documentaire constitue un argument pour justifier l'intérêt de l'étude en cours.
- La section Méthode décrit la méthode de manière suffisamment détaillée pour qu'une autre personne en recherche puisse reproduire l'étude. Elle se compose au minimum d'une sous-section participants, d'une sous-section sur le devis et la procédure.
- La section Résultats décrit les résultats de manière organisée. Chaque résultat principal est présenté en termes statistiques, mais aussi expliqué en mots.
- La Discussion résume généralement l'étude, discute des implications théoriques et pratiques, discute des limites de l'étude, et propose des suggestions pour des recherches futures.
- La recherche en psychologie peut être présentée sous différentes formes. Outre les articles de recherche empirique de style APA, il existe des articles théoriques et de synthèse; des documents finaux comprenant des mémoires, des thèses et des travaux de personnes aux études; des communications orales et des communications par affiches présentées lors de congrès académiques.
- Les communications orales et les communications par affiches présentées lors de congrès suivent certaines directives de style APA, mais elles sont beaucoup moins détaillées que les articles de style APA. Leur fonction est de présenter de nouvelles recherches aux personnes en recherche intéressées et de faciliter l'interaction entre ces personnes.

Exercices

- Pratique : Trouvez une description d'une étude de recherche dans un magazine populaire, un journal, un blogue ou un site internet. Identifiez ensuite cinq différences spécifiques entre la façon dont cette description est rédigée, et la façon dont elle serait rédigée dans le style APA.
- Pratique : Trouvez et corrigez les erreurs dans les références et citations fictives de style APA suivantes.
 - Walters, F. T., et DeLeon, M. (2010). Relation entre la motivation intrinsèque et la précision des auto-évaluations académiques chez les élèves du secondaire. *Revue en Éducation*, 23, 234-256.
 - Moore, Lilia S. (2007). Éthique dans la recherche par sondage. Dans M. Williams & P. L. Lee (eds.), *Ethical Issues in Psychology* (pp. 120-156). Boston, Psychological Research Press.
 - Vang, C., Dumont, L. S., et Prescott, M. P. ont découvert que les gauchers ont une préférence plus forte pour l'art abstrait que les droitiers (2006).
 - Ce résultat a été reproduit à plusieurs reprises (Williamson, 1998 ; Pentecost & Garcia, 2006 ; Armbruster, 2011).
- Pratique : Parcourez le numéro d'une revue scientifique d'intérêt général (p. ex., *Revue québécoise de psychologie*). Lisez l'ouverture des cinq premiers articles et évaluez l'efficacité de chacun d'eux de 1 (très inefficace) à 5 (très efficace). Expliquez chaque note en une ou deux phrases.
- Pratique : Trouvez un article récent dans une revue scientifique et identifiez l'ouverture, l'analyse documentaire et la fermeture de l'introduction.
- Pratique : Trouvez un article récent dans une revue scientifique et mettez en évidence, dans une couleur différente, chacun des éléments suivants de la discussion : résumé, implications théoriques, implications pratiques, limites et suggestions pour des recherches futures.

62. Références

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style (7th ed.)*. American Psychological Association.

American Psychological Association. (s.d.) *Student Paper Setup Guide*.
<https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf>

American Psychological Association. (s.d.). APA Style. <https://apastyle.apa.org>

American Psychological Association. (s.d.). APA Style: *Paraphrasing and citation activities*. Repéré le 23 septembre 2025. <https://apastyle.apa.org/instructional-aids/paraphrasing-citation-activities.pdf>

Bem, D. J. (2003). Writing the empirical journal article. Dans J. M. Darley, M. P. Zanna, et H. R. Roediger III (dir.), *The complete academic: A practical guide for the beginning social scientist* (2e éd.). American Psychological Association.

Bentley, M., Peerenboom, C. A., Hodge, F. W., Passano, E. B., Warren, H. C., et Washburn, M. F. (1929). Instructions in regard to preparation of manuscript. *Psychological Bulletin*, 26, 57–63.

Couture, M. (2025, 25 mars). *Adaptation française de l'APA (7e éd.)*, sous licence CC BY-NC 4.0.
<http://profmcouture.ca/apa/index.htm>

Darley, J. M., et Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 4, 377–383.

Gouvernement du Canada. *Écriture inclusive — les lignes directrices et ressources*.
<https://www.noslangues-ourlanguages.gc.ca/fr/cles-de-la-redaction/ecriture-inclusive-lignes-directrices-ressources>

Girardet, C, et Nuber, N. (s.d.) *Les normes APA (7e éd.)*. Haute école pédagogique Vaud, Suisse
<https://etudiant.hepl.ch/files/live/sites/files-site/files/filiere-ps/programme-formation/normes-apa-stage-maeps-fps-hep-vaud.pdf>

Institut national de la recherche scientifique (s.d.) *Inclusivement vôtres! Guide de rédaction inclusive* Repéré le 22 avril 2026. <https://inrs.ca/linrs/equite-diversite-et-inclusion/guide-de-redaction-inclusive/>

Madigan, R., Johnson, S., et Linton, P. (1995). The language of psychology: APA style as epistemology. *American Psychologist*, 50, 428–436. <https://psycnet.apa.org/doi/10.1037/0003-066X.50.6.428>

Office québécois de la langue française. (s.d.) *Féminisation et rédaction épicienne*, Dépannage linguistique. Repéré le 22 avril 2026.
<https://vitrinelinguistique.oqlf.gouv.qc.ca/banque-de-depannage-linguistique/la-redaction-et-la-communication/feminisation-et-redaction-epicene>

Université Laval. *Outils et formations pertinentes en matière d'équité, diversité et inclusion*. Repéré le 22 avril

2026.

<https://nutriss.ulaval.ca/boite-a-outils/equite-diversite-et-inclusion-edi/outils-et-formations-pertinentes>

Université de Montréal (s.d.). *Citer selon les normes APA, 7e édition*. Repéré le 23 septembre 2025.
<https://bib.umontreal.ca/citer/styles-bibliographiques/apa>

Université de Montréal (s.d.). *Inclusivement: Guide d'écriture pour toutes et tous*. Repéré le 22 avril 2026.
<https://francais.umontreal.ca/soutiller-en-francais/ecriture-inclusive/>

Université de Montréal (s.d.). *Quiz interactif sur l'écriture en paraphrases*. Repéré le 22 avril 2026.
<https://integrite.umontreal.ca/quiz-generaux/paraphraser/>

Université d'Ottawa (s.d.). *Citer et évaluer les sources*. Repéré le 23 septembre 2025.
<https://www.uottawa.ca/bibliotheque/ecriture-citation>

Université de Sherbrooke. *Le français langue commune à UdeS, le français en outil: La rédaction inclusive*. Repéré le 22 avril 2026.
<https://www.usherbrooke.ca/langue/le-francais-en-outils/la-redaction-inclusive>

Université du Québec (2023). *Guide de communication inclusive: pour des communications qui mobilisent, transforment et ont du style!*
<https://reseau.uquebec.ca/system/files/documents/guide-communication-inclusive-universite-du-quebec-2023.pdf>

CHAPITRE 12-LES STATISTIQUES DESCRIPTIVES



Image 12-Graphiques et tableaux

À ce stade nous devons examiner plus en détail les bases de l'analyse des données dans la recherche psychologique. Dans ce chapitre, nous nous concentrons sur les statistiques descriptives — un ensemble de techniques permettant de résumer et de présenter les données de votre échantillon. Nous examinerons tout d'abord certaines des techniques les plus courantes pour décrire des variables uniques, puis certaines des techniques les plus courantes pour décrire les relations statistiques entre les variables. Nous verrons ensuite comment présenter les statistiques descriptives par écrit, et sous la forme de tableaux et de graphiques qui conviendraient à un rapport de recherche suivant le style APA (American Psychological Association). Nous terminerons par quelques conseils pratiques pour l'organisation et la réalisation de vos analyses.

63. Décrire les variables uniques

Objectifs d'apprentissage

1. Utiliser des tableaux de fréquences et des histogrammes pour afficher, et interpréter la distribution d'une variable.
2. Calculer et interpréter la moyenne, la médiane et le mode d'une distribution, et identifier les situations dans lesquelles la moyenne, la médiane ou le mode est la mesure la plus appropriée de la tendance centrale.
3. Calculer et interpréter l'étendue et l'écart-type d'une distribution.
4. Calculer et interpréter les rangs percentiles et les scores z.

Les **statistiques descriptives** désignent un ensemble de techniques permettant de résumer et de présenter des données. Supposons ici que les données soient quantitatives et présentes sous forme de scores sur une ou plusieurs variables pour chaque personne participante à l'étude. Dans la plupart des cas, la question principale de la recherche porte sur une ou plusieurs relations statistiques entre les variables, mais il est également important de décrire chaque variable individuellement. C'est pourquoi nous commencerons par examiner certaines des techniques les plus courantes pour décrire des variables uniques.

La distribution d'une variable

Chaque variable possède une **distribution**, c'est-à-dire la manière dont les scores se répartissent entre les différents niveaux de cette variable. Par exemple dans un échantillon de 100 étudiantes et étudiants universitaires, la distribution de la variable « nombre de frères et sœurs » peut être telle que 10 d'entre eux n'avaient pas de frère et sœur, 30 ont un frère ou une sœur, 40 ont deux frères et sœurs, et ainsi de suite. Dans le même échantillon, la distribution de la variable « genre de la personne » peut être telle que 44 ont un score « masculin », 56 un score « féminin ».

Tableaux de fréquences

La distribution d'une variable peut notamment être représentée dans un **tableau de fréquences**. Le Tableau 12.1 par exemple est un tableau de fréquences montrant une distribution hypothétique des scores

sur l'échelle d'estime de soi de Rosenberg pour un échantillon de 40 personnes aux études collégiales. La première colonne énumère les valeurs de la variable — les scores possibles sur l'échelle de Rosenberg — et la deuxième colonne indique la fréquence de chaque score. Ce tableau montre que trois personnes ont obtenu un score d'estime de soi de 24, cinq un score de 23, et ainsi de suite. Un tel tableau de fréquences permet de visualiser rapidement plusieurs aspects importants d'une distribution tels que l'étendue des scores (de 15 à 24), les scores les plus et les moins fréquents (22 et 17, respectivement), et les scores aberrants qui se démarquent du reste.

Tableau 12.1 *Tableau de fréquences montrant une distribution hypothétique des scores sur l'échelle d'estime de soi de Rosenberg*

Estime de soi	Fréquence
24	3
23	5
22	10
21	8
20	5
19	3
18	3
17	0
16	2
15	1

Quelques autres points méritent d'être soulignés à propos des tableaux de fréquences. Premièrement les niveaux énumérés dans la première colonne vont généralement du plus élevé au plus bas, et ils ne vont généralement pas au-delà des scores les plus élevés et les plus bas des données. Par exemple bien que les scores sur l'échelle d'estime de soi de Rosenberg puissent varier d'un maximum de 30 à un minimum de 0, le Tableau 12.1 n'inclut que les niveaux de 24 à 15 parce que cet intervalle comprend tous les scores de cet ensemble particulier de données. Deuxièmement lorsqu'il existe de nombreux scores différents dans une large étendue de valeurs, il est souvent préférable de créer un tableau de fréquences groupées, dans lequel la première colonne énumère les étendues de valeurs et la deuxième colonne la fréquence des scores dans chaque étendue. Le Tableau 12.2 par exemple est un tableau de fréquences groupées montrant une distribution hypothétique de simples temps de réaction pour un échantillon de 20 personnes participantes. Dans un tableau de fréquences groupées, les étendues doivent toutes être de même largeur et il y en a généralement entre 5 et 15. Enfin les tableaux de fréquences peuvent également être utilisés pour des variables catégorielles auquel cas les niveaux correspondent aux différentes catégories. L'ordre des noms de catégorie est quelque peu arbitraire, mais elles sont souvent énumérées de la plus fréquente en haut à la moins fréquente en bas.

Tableau 12.2 *Tableau de fréquences groupées montrant une distribution hypothétique des temps de réaction*

Temps de réaction (ms)	Fréquence
241–260	1
221–240	2
201–220	2
181–200	9
161–180	4
141–160	2

Histogrammes

Un **histogramme** est une représentation graphique d'une distribution. Il présente les mêmes informations qu'un tableau de fréquence, mais de façon encore plus rapide et facile à comprendre. L'histogramme de la Figure 12.1 présente la distribution des scores d'estime de soi du Tableau 12.1. L'axe des x de l'histogramme représente la variable et l'axe des y représente la fréquence. Au-dessus de chaque niveau de la variable sur l'axe des x se trouve une barre verticale qui représente le nombre de personnes ayant ce score. Lorsque la variable est quantitative, comme dans cet exemple, il n'y a généralement pas d'espace entre les barres. En revanche lorsqu'il s'agit d'une variable catégorielle, il y a généralement un petit écart entre les barres. (L'écart à 17 dans cet histogramme reflète le fait qu'il n'y avait pas de score de 17 dans cet ensemble de données.)

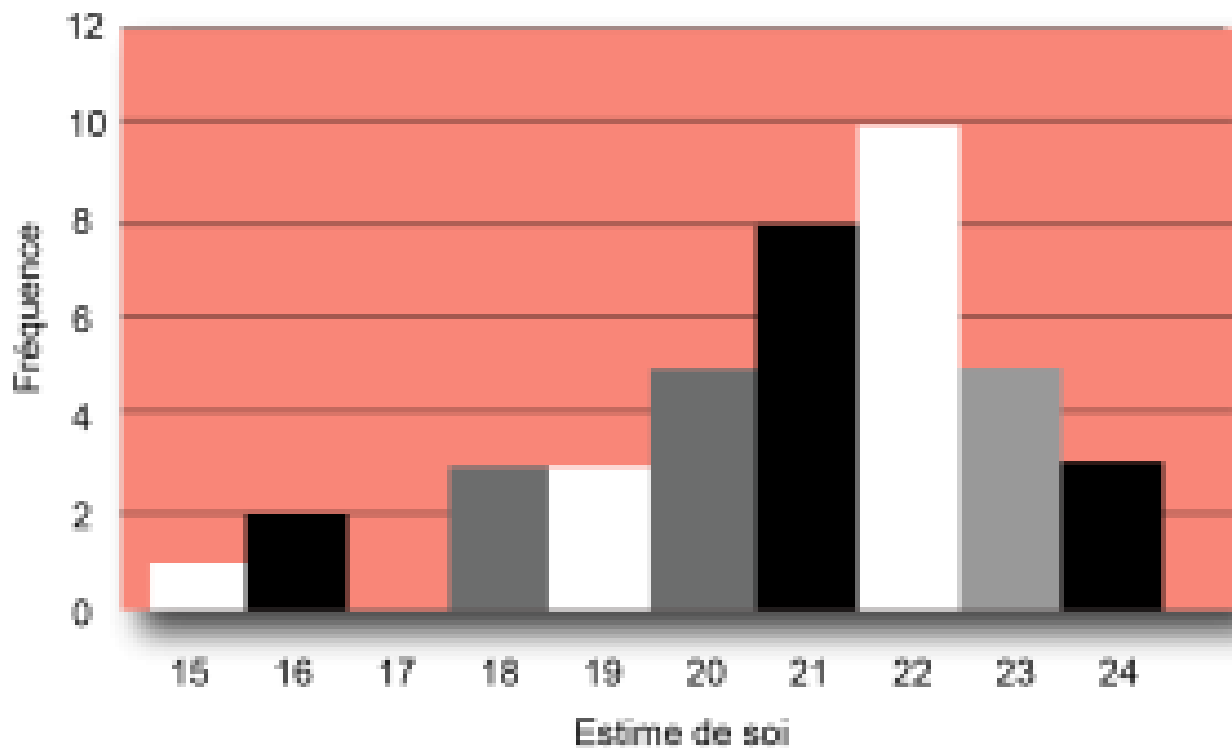


Figure 12.1-Tableau de fréquences montrant une distribution hypothétique des scores sur l'échelle d'estime de soi de Rosenberg

Formes de la distribution

Lorsque la distribution d'une variable quantitative est représentée dans un histogramme, elle a une forme. La forme de la distribution des scores d'estime de soi de la Figure 12.1 est typique. Il y a un pic quelque part vers le milieu de la distribution et des « queues » qui s'effilent dans les deux directions à partir du sommet. La distribution de la Figure 12.1 est unimodale, ce qui signifie qu'elle présente un seul sommet distinct, mais les distributions peuvent également être bimodales, c'est-à-dire qu'elles présentent deux sommets distincts. La Figure 12.2 par exemple montre une distribution hypothétique bimodale des scores de l'inventaire de dépression de Beck. Les distributions peuvent également présenter plus de deux sommets distincts, mais il s'agit de cas relativement rares dans la recherche en psychologie.

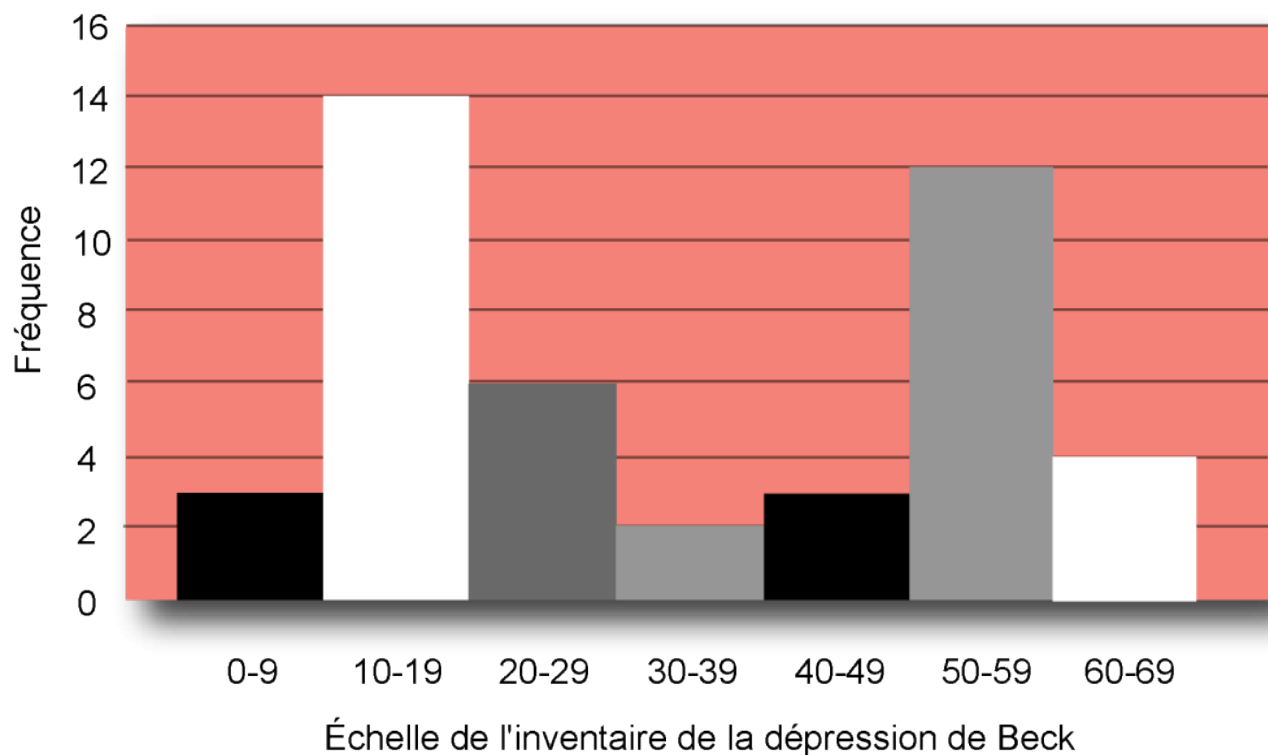


Figure 12.2-Tableau de fréquences groupées montrant une distribution hypothétique de la fréquence de scores à l'échelle de l'inventaire de la dépression de Beck. [Description de l'image](#)

Une autre caractéristique de la forme d'une distribution est qu'elle est symétrique ou asymétrique. La distribution présentée au centre de la Figure 12.3 est **symétrique**. Ses moitiés gauche et droite sont des images miroirs l'une de l'autre. La distribution de gauche est négativement **asymétrique**, avec un sommet décalé vers l'extrémité supérieure de son étendue de scores et une queue négative relativement longue. La distribution de droite est positivement asymétrique, avec un sommet situé vers l'extrémité inférieure de son étendue de scores et une queue positive relativement longue.

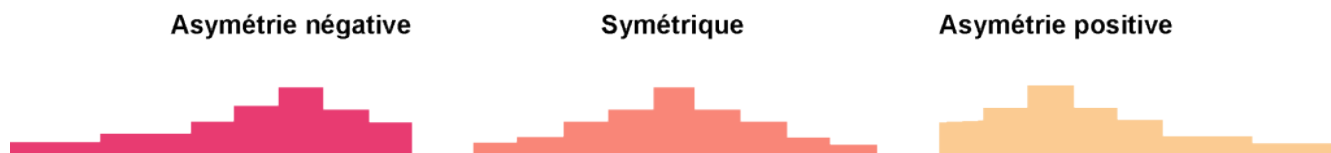


Figure 12.3-Histogrammes montrant la distribution des scores d'estime de soi.

Une **valeur aberrante** est un score extrême qui est beaucoup plus élevé ou beaucoup moins élevé que le reste des scores de la distribution. Parfois les valeurs aberrantes représentent des scores vraiment extrêmes pour la variable étudiée. Par exemple sur l'inventaire de dépression de Beck, une seule personne

cliniquement déprimée peut être une valeur aberrante dans un échantillon de pairs par ailleurs heureux et hautement fonctionnels. Cependant les valeurs aberrantes peuvent également représenter des erreurs ou des malentendus de la part du scientifique ou de la personne participante, des dysfonctionnements de l'équipement ou d'autres problèmes similaires. Nous reviendrons sur la manière d'interpréter les valeurs aberrantes et sur ce qu'il convient de faire à leur sujet plus loin dans ce chapitre.

Mesures de la tendance centrale et de la variabilité

Il est également utile de pouvoir décrire plus précisément les caractéristiques d'une distribution. Nous verrons ici comment le faire en fonction de deux caractéristiques importantes : leur tendance centrale et leur variabilité.

Tendance centrale

La **tendance centrale** d'une distribution est son milieu — le point autour duquel les scores de la distribution tendent à se regrouper. (Un autre terme pour désigner la tendance centrale est celui de moyenne.) Si l'on se reporte à la Figure 12.1 par exemple on constate que les scores d'estime de soi ont tendance à se regrouper autour des valeurs de 20 à 22. Nous examinerons ici les trois mesures les plus communes de la tendance centrale : la moyenne, la médiane et le mode.

La **moyenne** d'une distribution (symbolisée par M) est la somme des scores divisée par le nombre de scores. La formule se présente comme suit :

$$M = \Sigma X / N$$

Dans cette formule, le symbole Σ (la lettre grecque sigma) est le signe de sommation et signifie qu'il faut faire la somme de toutes les valeurs de la variable X . N représente le nombre de scores. La moyenne est de loin la mesure la plus courante de la tendance centrale, et il y a quelques bonnes raisons à cela. Elle donne généralement une bonne indication de la tendance centrale d'une distribution, et elle est facilement comprise par la plupart des gens. En outre la moyenne possède des propriétés statistiques qui la rendent particulièrement utile dans le cadre des statistiques inférentielles.

La **médiane** est une alternative à la moyenne. La médiane est le score central, en ce sens que la moitié des scores de la distribution lui sont inférieurs et l'autre moitié lui sont supérieurs. La façon la plus simple de trouver la médiane est de classer les scores de la plus basse à la plus haute et de localiser le score du milieu. Considérons par exemple l'ensemble suivant de sept scores :

8 4 12 14 3 2 3

Pour trouver la médiane, il suffit de réorganiser les scores du plus faible au plus élevé et de repérer celui qui se trouve au milieu.

2 3 3 **4** 8 12 14

Dans ce cas la médiane est 4, car il y a trois scores inférieurs à 4 et trois scores supérieurs à 4.

Lorsque le nombre de scores est pair, deux scores se trouvent au milieu de la distribution, auquel cas la médiane est la valeur située à mi-chemin entre les deux. Par exemple si l'on ajoute un score de 15 à l'ensemble des données précédentes, il y aura deux scores (4 et 8) au milieu de la distribution, et la médiane sera à mi-chemin entre les deux (6).

Une dernière mesure de la tendance centrale est le mode. Le **mode** est la valeur la plus fréquente dans une distribution. Dans la distribution de l'estime de soi présentée au Tableau 12.1 et à la Figure 12.1 par exemple le mode est 22. Un plus grand nombre de personnes étudiantes a obtenu ce score qu'aucun autre. Le mode est la seule mesure de la tendance centrale qui peut également être utilisée pour les variables catégorielles.

Dans une distribution unimodale et symétrique, la moyenne, la médiane et le mode seront très proches l'un de l'autre au sommet de la distribution. Dans une distribution bimodale ou asymétrique, la moyenne, la médiane et le mode peuvent être très différents. Dans une distribution bimodale, la moyenne et la médiane auront tendance à se situer entre les deux sommets, tandis que le mode se situera au niveau du sommet le plus élevé. Dans une distribution asymétrique, la moyenne diffère de la médiane dans le sens de l'asymétrie (c'est-à-dire dans le sens de la queue la plus longue). Pour les distributions fortement asymétriques, la moyenne peut être tirée si loin dans la direction de l'asymétrie qu'elle ne constitue plus une bonne mesure de la tendance centrale de cette distribution. Imaginons par exemple un ensemble de quatre temps de réaction simples de 200, 250, 280 et 250 millisecondes (ms). La moyenne est de 245 ms. Mais l'ajout d'un score supplémentaire de 5 000 ms — peut-être parce que le participant n'était pas attentif — porterait la moyenne à 1 445 ms. Non seulement cette mesure de la tendance centrale est supérieure à 80 % des scores de la distribution, mais elle ne semble pas non plus très bien représenter le comportement de l'un ou l'autre élément de la distribution. C'est pourquoi les scientifiques préfèrent souvent la médiane pour les distributions fortement asymétriques (telles que les distributions des temps de réaction).

Gardons toutefois à l'esprit qu'il n'est pas obligatoire de choisir une seule mesure de la tendance centrale pour analyser vos données. Chacune d'entre elles fournit des informations légèrement différentes et toutes peuvent être utiles.

Mesures de variabilité

La **variabilité** d'une distribution est la mesure dans laquelle les scores varient autour de leur tendance centrale. Considérons les deux distributions de la Figure 12.4, qui ont toutes les deux la même tendance

centrale. La moyenne, la médiane et le mode de chaque distribution sont 10. Remarquons toutefois que les deux distributions diffèrent en termes de variabilité. Celle du haut présente une variabilité relativement faible, tous les scores étant relativement proches du centre. La distribution du bas présente une variabilité relativement élevée, les scores étant répartis sur une étendue beaucoup plus grande.

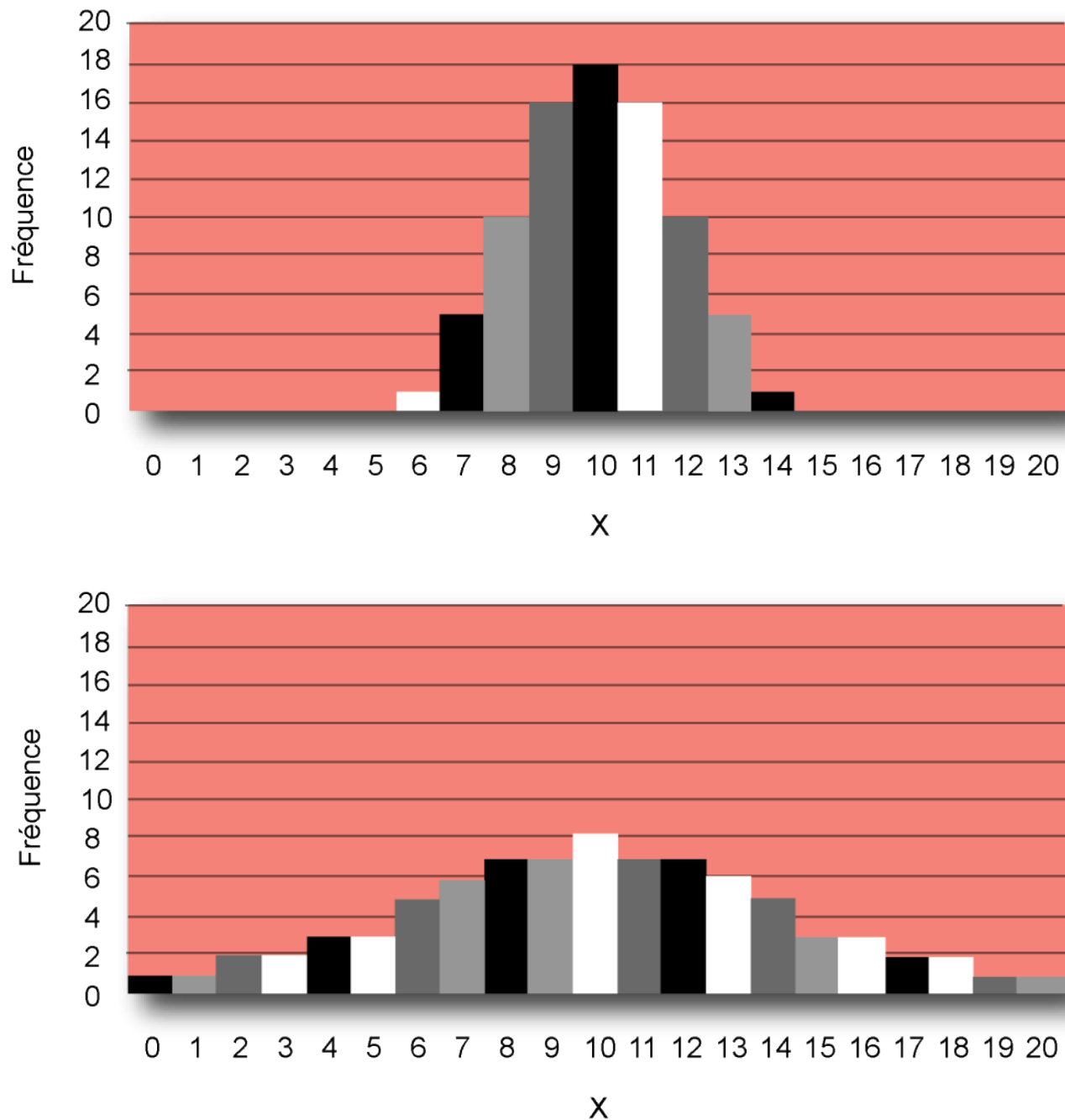


Figure 12.4-Histogramme montrant deux distributions hypothétiques qui ont la même moyenne. [\[Description de la figure\]](#)

Une simple mesure de variabilité est l'**étendue**, qui est simplement la différence entre le score le plus élevé et le score le plus bas de la distribution. L'étendue des scores d'estime de soi du Tableau 12.1 par exemple est la différence entre le score le plus élevé (24) et le score le plus bas (15). En d'autres termes l'étendue est $24 - 15 = 9$. Bien que l'étendue soit facile à calculer et à comprendre, elle peut être trompeuse en cas de valeurs aberrantes. Imaginons par exemple un examen pour lequel tous les élèves ont obtenu un score entre 90 et 100. L'étendue est de 10. Mais s'il y avait une seule personne étudiante ayant obtenu 20 points, l'étendue passerait à 80, ce qui donnerait l'impression que les scores étaient très variables alors qu'en fait, une seule personne se distinguait nettement des autres.

L'écart-type est de loin la mesure de variabilité la plus courante. L'**écart-type** d'une distribution est la distance moyenne entre les scores et la moyenne. Par exemple les écarts-types des distributions de la Figure 12.4 sont de 1,69 pour la distribution du haut et de 4,30 pour celle du bas. En d'autres termes alors que les scores de la distribution supérieure diffèrent de la moyenne d'environ 1,69 unité en moyenne, les scores de la distribution inférieure diffèrent de la moyenne d'environ 4,30 unités en moyenne.

Le calcul de l'écart-type implique une légère complication. Spécifiquement il s'agit de trouver la différence entre chaque score et la moyenne, d'élever au carré chaque différence, de trouver la moyenne de ces différences élevées au carré et enfin de trouver la racine carrée de cette moyenne. La formule se présente comme suit:

$$SD = \sqrt{\frac{\sum (X - M)^2}{N}}$$

Les calculs de l'écart-type sont illustrés pour un petit ensemble de données au Tableau 12.3. La première colonne est un ensemble de huit scores dont la moyenne est de 5. La deuxième colonne est la différence entre chaque score et la moyenne. La troisième colonne est le carré de chacune de ces différences. Remarquons que même si des différences peuvent être négatives, le carré des différences est toujours positif — ce qui signifie que l'écart-type est toujours positif. Au bas de la troisième colonne se trouve la moyenne des différences au carré, également appelée la **variance** (symbolisée par ET^2). Bien que la variance soit elle-même une mesure de la variabilité, elle joue généralement un rôle plus important dans les statistiques inférentielles que dans les statistiques descriptives. Enfin sous la variance se trouve la racine carrée de la variance, qui est l'écart-type.

Tableau 12.3 Calculs de l'écart-type

X	X-M	(X-M) ²
3	-2	4
5	0	0
4	-1	1
2	-3	9
7	2	4
6	1	1
5	0	0
8	3	9
M = 5		$\dot{ET}^2 = 28/8 = 3.50$
		$\dot{ET} = \sqrt{3.50} = 1.87$

N ou N - 1

Si vous avez déjà suivi un cours de statistiques, vous avez peut-être appris à diviser la somme des carrés des différences par $N - 1$ plutôt que par N lorsque vous calculez la variance et l'écart-type. Pourquoi en est-il ainsi?

Par définition l'écart-type est la racine carrée de la moyenne des différences au carré. Cela implique de diviser la somme des différences au carré par N , comme dans la formule qui vient d'être présentée. Le calcul de l'écart-type de cette manière est approprié lorsque votre objectif est simplement de décrire la variabilité d'un échantillon. L'apprentissage de cette méthode met l'accent sur le fait que la variance est en fait la moyenne des différences au carré et que l'écart-type est la racine carrée de cette moyenne.

Toutefois la plupart des calculatrices et logiciels divisent la somme des différences au carré par $N - 1$. Cela est dû au fait que l'écart-type d'un échantillon a tendance à être légèrement inférieur à l'écart-type de la population à partir de laquelle l'échantillon a été sélectionné. En divisant la somme des carrés par $N - 1$, on corrige cette tendance et on obtient une meilleure estimation de l'écart-type de la population. Étant donné que les scientifiques considèrent généralement que leurs données représentent un échantillon sélectionné au sein d'une population plus large — et que ces personnes souhaitent généralement tirer des conclusions sur la population — il est logique d'appliquer systématiquement cette correction.

Rangs percentiles et scores z

Dans de nombreuses situations, il est utile de disposer d'un moyen de décrire l'emplacement d'un score individuel au sein de sa distribution. Le rang percentile est l'une de ces méthodes. Le **rang percentile** d'un score est le pourcentage de scores de la distribution qui sont inférieurs à ce score. Considérons par exemple

la distribution du Tableau 12.1. Pour tout score de la distribution, nous pouvons trouver son rang percentile en comptant le nombre de scores de la distribution qui sont inférieurs à ce score et en convertissant ce nombre en pourcentage du nombre total de scores. Au Tableau 12.1 trois personnes ont obtenu un score d'estime de soi de 24, cinq ont un score d'estime de soi de 23. Dans cette distribution, 32 des 40 scores (80 %) sont inférieurs à 23. Chacune de ces personnes a donc un rang percentile de 80 (on peut également dire qu'elles ont un score « au 80e percentile »). Les rangs percentiles sont souvent utilisés pour rendre compte des résultats de tests standardisés d'aptitudes ou de performances. Si votre rang de percentile à un test d'aptitude verbale était de 40 par exemple cela signifierait que vous avez obtenu des résultats supérieurs à ceux de 40% des personnes qui ont passé le test.

Une autre approche est le score **z**. Le **score z** pour une personne donnée est la différence entre le score de cette personne et la moyenne de la distribution, divisée par l'écart-type de la distribution:

$$z = (X - M) / SD$$

Un score **z** indique dans quelle mesure un score brut est supérieur ou inférieur à la moyenne, mais il l'exprime en écart-type. Par exemple dans une distribution de scores de quotient intellectuel (QI) avec une moyenne de 100 et un écart-type de 15, un score de QI de 110 aurait un score **z** de $(110 - 100) / 15 = +0,67$. En d'autres termes un score de 110 se situe à 0,67 écart-type (environ deux tiers d'un écart-type) au-dessus de la moyenne.

De même un score brut de 85 aurait un score **z** de $(85 - 100) / 15 = -1,00$. En d'autres termes un score de 85 correspond à un écart-type en dessous de la moyenne.

Les scores **z** sont importants pour plusieurs raisons. Encore une fois ils permettent de décrire où se situe le score d'une personne au sein d'une distribution et sont parfois utilisés pour communiquer les résultats de tests standardisés. Ils offrent également une façon d'identifier les valeurs aberrantes. Par exemple les valeurs aberrantes sont parfois définies comme des scores ayant des scores **z** inférieurs à -3,00 ou supérieurs à +3,00. En d'autres termes il s'agit de scores qui s'écartent de plus de trois écarts-types de la moyenne. Enfin les scores **z** jouent un rôle important dans la compréhension et le calcul d'autres statistiques, comme nous le verrons ultérieurement.

Statistiques descriptives en ligne

Bien que plusieurs scientifiques utilisent des logiciels disponibles sur le marché tels que SPSS et Excel pour analyser leurs données, il existe plusieurs outils d'analyse gratuits en ligne qui peuvent également s'avérer extrêmement utiles. Nombre d'entre eux vous permettent de saisir ou de télécharger vos données, puis d'effectuer en un clic plusieurs analyses statistiques descriptives. Parmi ces outils, on peut citer les suivants.

JASP (plusieurs langues, dont plusieurs modules en français)

<https://jasp-stats.org/download/>

Rice Virtual Lab in Statistics (en anglais)

http://onlinestatbook.com/stat_analysis/index.html

VassarStats (en anglais)

<http://faculty.vassar.edu/lowry/VassarStats.html>

Bright Stat (en anglais)

<http://www.brightstat.com>

Pour une liste plus complète, voir <http://statpages.org/index.html>.

Description de l'image

Figure 12.2 description longue : Un histogramme montrant une distribution bimodale des scores à l'inventaire de dépression de Beck. L'axe horizontal est intitulé « Score à l'inventaire de dépression de Beck » et l'axe vertical est intitulé « Fréquence ». Les données sont les suivantes:

- BDI: 0-9, Fréquence: 3
- BDI: 10-19, Fréquence: 14
- BDI: 20-29, Fréquence: 6
- BDI: 30-39, Fréquence: 2
- BDI: 40-49, Fréquence: 3
- BDI: 50-59, Fréquence: 12
- BDI: 60-69, Fréquence: 4

Les deux sommets distincts sont la plage 10-19 et la plage 50-59. [\[Retour à la Figure 12.2\]](#)

Figure 12.4 description longue : Deux histogrammes ayant la même tendance centrale, mais une variabilité différente. Chaque axe horizontal est étiqueté « X » et a des valeurs de 1 à 20, et chaque axe vertical est étiqueté « Fréquence » et a des valeurs de 1 à 20. Chaque histogramme a également une moyenne, une médiane, un mode et une tendance centrale de 10.

Dans le premier histogramme, la variabilité est relativement faible. Les données sont les suivantes:

- X: 6, Fréquence: 1
- X: 7, Fréquence: 5
- X: 8, Fréquence: 10
- X: 9, Fréquence: 16

- X: 10, Fréquence: 18
- X: 11, Fréquence: 16
- X: 12, Fréquence: 10
- X: 13, Fréquence: 5
- X: 14, Fréquence: 1

Dans le second histogramme, la variabilité est relativement élevée. Les données sont les suivantes :

- X: 0, Fréquence: 1
- X: 1, Fréquence: 1
- X: 2, Fréquence: 2
- X: 3, Fréquence: 2
- X: 4, Fréquence: 3
- X: 5, Fréquence: 3
- X: 6, Fréquence: 5
- X: 7, Fréquence: 6
- X: 8, Fréquence: 7
- X: 9, Fréquence: 7
- X: 10, Fréquence: 8
- X: 11, Fréquence: 7
- X: 12, Fréquence: 7
- X: 13, Fréquence: 6
- X: 14, Fréquence: 5
- X: 15, Fréquence: 3
- X: 16, Fréquence: 3
- X: 17, Fréquence: 2
- X: 18, Fréquence: 2
- X: 19, Fréquence: 1
- X: 20, Fréquence: 1

[\[Retour à la Figure 12.4\]](#)

64. Décrire les relations statistiques

Objectifs d'apprentissage

1. Décrire les différences entre les groupes en termes de moyenne et d'écart-type, et en termes de d de Cohen.
2. Décrire les corrélations entre variables quantitatives en termes de r de Pearson.

Comme nous l'avons vu tout au long de ce livre, la plupart des questions de recherche intéressantes en psychologie portent sur les relations statistiques entre les variables. Dans cette section, nous reviendrons sur les deux formes fondamentales de relations statistiques présentées plus tôt dans le livre — les différences entre groupes ou conditions et les relations entre variables quantitatives — et nous les examinerons plus en détail.

Différences entre les groupes ou conditions

Les différences entre les groupes ou les conditions sont généralement décrites à l'aide de moyenne et d'écart-type de chaque groupe ou condition. Par exemple Thomas Ollendick et ses collègues ont mené une étude dans laquelle ces scientifiques ont évalué deux traitements d'une séance pour des phobies simples chez des enfants (Ollendick, Öst, Reuterskiöld, Costa, Cederlund, Sirbu, Davis, Jarrett, 2009). Il y a eu répartition au hasard des enfants ayant une peur intense (par exemple des chiens) à l'une des trois conditions. Dans la condition exposition, les enfants voient l'objet de leur peur sous supervision du personnel soignant. Dans la condition éducation, les enfants apprennent sur les phobies et certaines stratégies pour y faire face. Dans la condition contrôle le liste d'attente, les enfants attendent de recevoir un traitement après la fin de l'étude. La gravité de la phobie de chaque enfant a ensuite été évaluée sur une échelle de 1 à 8 par le personnel soignant qui ne savait pas quel traitement l'enfant avait reçu. Le score de peur moyen dans la condition éducation était de 4,83 avec un écart-type de 1,52, tandis que le score de peur moyen dans la condition exposition était de 3,47 avec un écart-type de 1,77. Le score de peur moyen dans le groupe contrôle était de 5,56 avec un écart-type de 1,21. La gravité de la peur semble inférieure dans les conditions traitements que dans la condition contrôle. En d'autres mots les deux traitements semblent avoir fonctionnés, mais le traitement par exposition semble avoir mieux fonctionné que le traitement par éducation. Comme nous l'avons vu, les différences entre les moyennes des groupes ou des conditions peuvent être présentées dans un graphique à barres, comme celui de la Figure 12.5, où les hauteurs des

barres représentent les moyennes des groupes ou des conditions. Nous verrons bientôt comment créer des graphiques à barres de style APA (American Psychological Association).

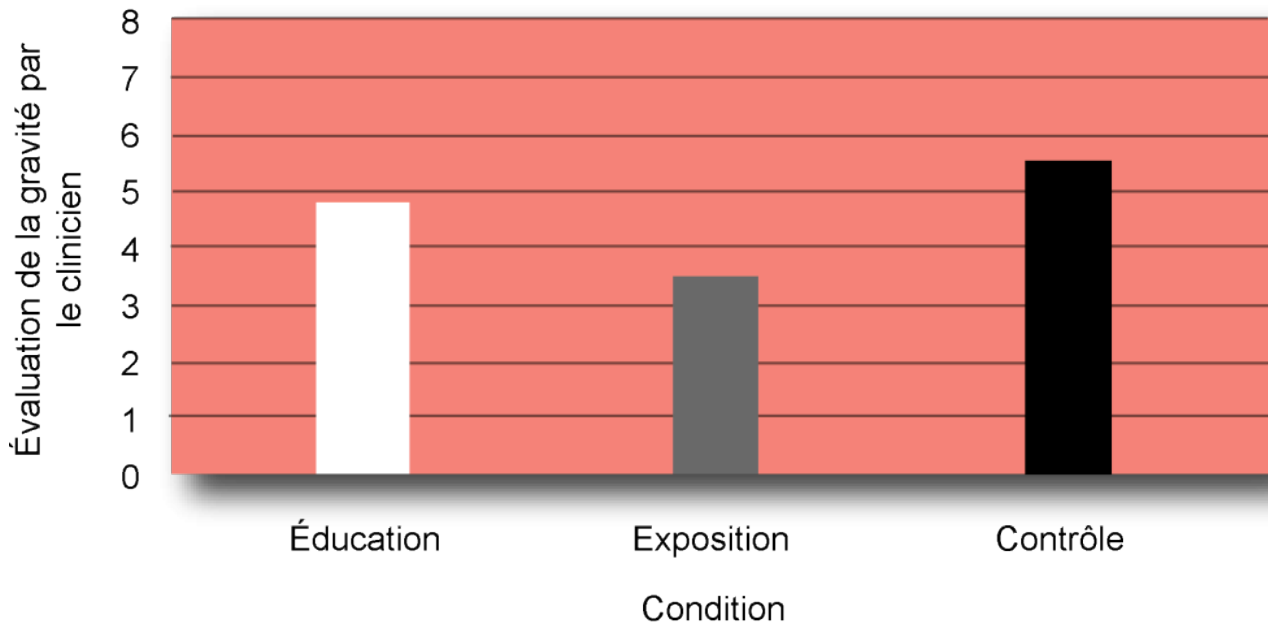


Figure 12.5-Graphique à barres montrant l'évaluation moyenne de la phobie par le personnel soignant pour les enfants dans deux conditions de traitement

Il est également important de pouvoir décrire la force d'une relation statistique, souvent appelée la **taille de l'effet**. La mesure de la taille de l'effet la plus utilisée pour les différences entre les moyennes des groupes ou des conditions s'appelle le **d de Cohen**, qui correspond à la différence entre les deux moyennes divisée par l'écart-type :

$$d = (M_1 - M_2) / \text{ÉT}$$

Dans cette formule, peu importe laquelle des moyennes est M_1 et laquelle est M_2 , si l'une des conditions est le groupe de contrôle, celui-ci est M_2 et l'autre M_1 . Dans le cas contraire la moyenne la plus élevée est généralement M_1 et la moyenne la plus faible M_2 , de sorte que le d de Cohen ait une valeur positive. En effet les valeurs du d de Cohen devraient toujours être positives, de sorte que c'est la différence absolue entre les moyennes qui est prise en compte au numérateur. L'écart-type dans cette formule est généralement une sorte de moyenne des écarts-types des deux groupes, appelée écart-type combiné. Pour calculer l'écart-type combiné, il faut ajouter la somme des carrés des différences pour le groupe 1 à la somme des carrés des différences pour le groupe 2, diviser ce résultat par la somme des tailles des deux échantillons, puis prendre la racine carrée de ce résultat. Toutefois de manière informelle, l'écart-type de l'un ou l'autre groupe peut être utilisé à la place.

Conceptuellement le d de Cohen est la différence entre les deux moyennes exprimées en unités d'écart-type. (remarquez sa similarité avec le score z , qui exprime la différence entre un score individuel et une moyenne en unités d'écart-type). Un d de Cohen de 0.50 signifie que les moyennes des deux groupes diffèrent de 0.50 écart-type (un demi-écart-type). Un d de Cohen de 1.20 signifie qu'elles diffèrent de 1.20 écart-type. Mais comment interpréter ces valeurs en matière de force de la relation ou de taille de l'échantillon? La différence entre les moyennes? Le Tableau 12.4 présente quelques lignes directrices pour l'interprétation des valeurs du d de Cohen dans la recherche en psychologie (Cohen, 1992). Les valeurs proches de 0.20 sont considérées comme petites, les valeurs proches de 0.50 sont considérées comme moyennes et les valeurs proches de 0.80 sont considérées comme grandes. Ainsi une valeur d de Cohen de 0.50 représente une différence de taille moyenne entre deux moyennes, et une valeur d de Cohen de 1.20 représente une très grande différence dans le contexte de la recherche en psychologie. Dans la recherche d'Ollendick et de ses collègues, il y avait une grande différence ($d = 0.82$) entre les conditions d'exposition et d'éducation.

Tableau 12.4 Lignes directrices pour qualifier les valeurs d de Cohen et r de Pearson de « fortes », « moyennes » ou « faibles »

Force de la relation	d de Cohen	r de Pearson
Forte/large	0.80	± 0.50
Moyenne	0.50	± 0.30
Faible/petite	0.20	± 0.10

Le d de Cohen est utile parce qu'il a la même signification, quelle que soit la variable comparée ou l'échelle sur laquelle elle a été mesurée. Un d de Cohen de 0.20 signifie que les moyennes des deux groupes diffèrent de 0.20 écart-type, qu'il s'agisse des scores sur l'échelle d'estime de soi de Rosenberg, du temps de réaction mesuré en millisecondes, du nombre de frères et sœurs ou de la pression artérielle diastolique mesurée en millimètres de mercure. Non seulement cette mesure facilite la communication entre les scientifiques au sujet de leurs résultats, mais cela rend aussi possible de combiner et de comparer les résultats provenant de différentes études utilisant diverses mesures.

Il faut savoir que l'expression « *taille de l'effet* » peut être trompeuse, car elle suggère une relation de cause à effet — c'est-à-dire que la différence entre deux moyennes serait un « effet » du fait d'appartenir à un groupe ou à une condition plutôt qu'à un autre. Imaginons par exemple une étude montrant qu'un groupe de personnes faisant de l'exercice est en moyenne plus heureux qu'un groupe de personnes n'en faisant pas, avec une « taille de l'effet » de $d = 0.35$. Si l'étude était expérimentale — avec des personnes assignées de manière aléatoire aux conditions « exercice » et « sans exercice » — on pourrait conclure que la pratique d'exercice physique a causé une augmentation du bonheur qualifiée de faible à moyenne. En revanche si l'étude était transversale, on pourrait seulement conclure que les personnes faisant de l'exercice sont plus heureuses que celles qui n'en font pas, dans une mesure allant de faible à moyenne. En d'autres termes le simple fait d'appeler la différence une « taille d'effet » ne rend pas la relation causale.

Différences entre les sexes exprimées par le d de Cohen

La chercheuse Janet Shibley Hyde a examiné les résultats de nombreuses études sur les différences psychologiques entre les sexes et les a exprimées en termes de d de Cohen (Hyde, 2007). Voici quelques-unes des valeurs qu'elle a trouvées, en faisant la moyenne de plusieurs études dans chaque cas. (Notons que parce qu'elle considère toujours M_1 , la moyenne pour les hommes et M_2 , celle pour les femmes, les valeurs positives indiquent que les hommes obtiennent des résultats plus élevés et les valeurs négatives indiquent que les femmes obtiennent des résultats plus élevés).

Résolution de problèmes mathématiques	+0.08
Compréhension de lecture	-0.09
Sourire	-0.40
Aggression	+0.50
Attitudes à l'égard des relations sexuelles occasionnelles	+0.81
Efficacité du leadership	-0.02

Hyde souligne que si les hommes et les femmes diffèrent fortement sur certaines variables (par exemple les attitudes à l'égard des relations sexuelles occasionnelles), ils ne diffèrent que faiblement sur la grande majorité d'entre elles. Dans de nombreux cas le d de Cohen est inférieur à 0.10, ce qu'elle qualifie de différence « triviale ». (La différence de bavardage évoquée au chapitre 1 était également triviale: $d = 0.06$.) Bien que les scientifiques et les non-scientifiques mettent souvent l'accent sur les *différences entre les sexes*, Hyde affirme qu'il est au moins aussi logique de considérer les hommes et les femmes comme fondamentalement *similaires*. C'est ce qu'elle appelle l'« hypothèse de similarité entre les genres ».

Corrélations entre les variables quantitatives

Comme nous l'avons vu tout au long de ce livre, de nombreuses relations statistiques intéressantes prennent la forme de corrélations entre des variables quantitatives. Par exemple le chercheur Kurt Carlson et la chercheuse Jacqueline Conard ont mené une étude sur la relation entre la position alphabétique de la première lettre du nom de famille des personnes (de A = 1 à Z = 26) et la rapidité avec laquelle ces personnes répondaient aux appels des consommateurs et consommatrices (Carlson et Conard, 2011). Dans une étude, ces scientifiques ont envoyé des courriels à un grand groupe de personnes inscrites à la maîtrise en administration des affaires en leur offrant des billets de basket-ball gratuits en nombre limité. Les résultats

étaient que plus les noms de famille des personnes étudiantes se situaient vers la fin de l'alphabet, plus ils avaient tendance à répondre rapidement. Ces résultats sont résumés dans la Figure 12.6.

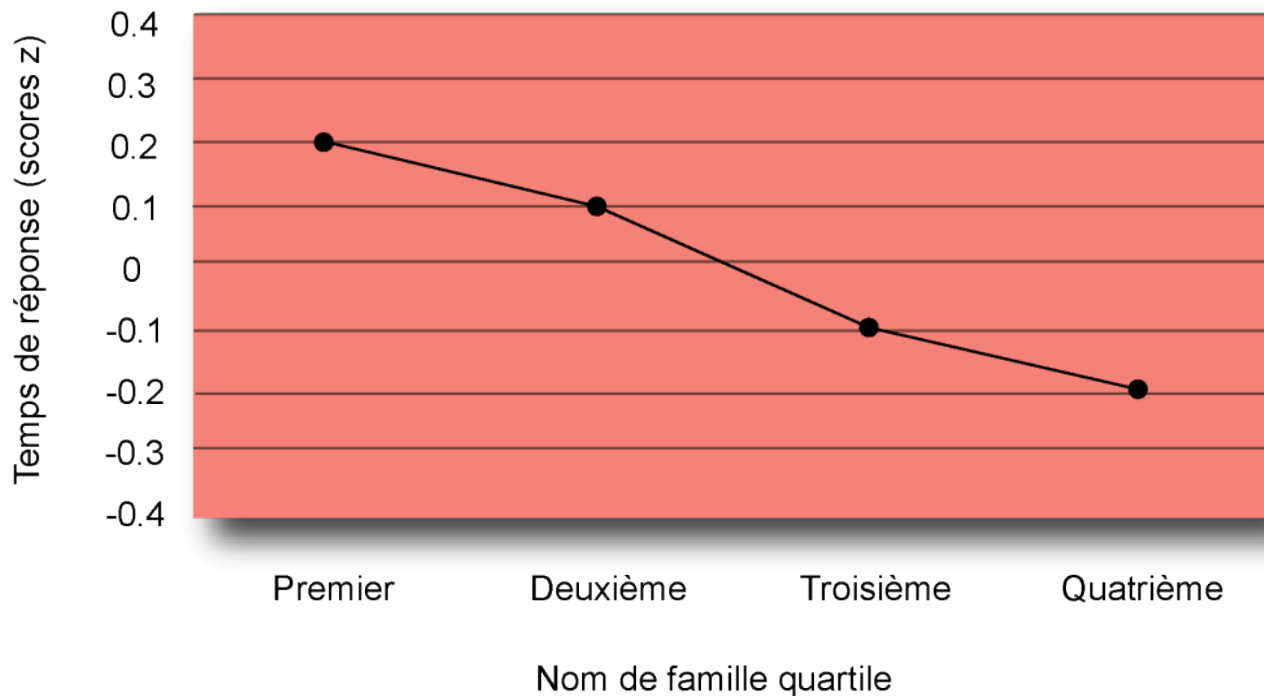


Figure 12.6-Graphique linéaire montrant la relation entre la position alphabétique des noms de famille et la rapidité avec laquelle ces personnes répondent aux offres de biens de consommation

Ces relations sont souvent présentées à l'aide de graphiques linéaires ou de diagrammes de dispersion, qui montrent comment le niveau d'une variable varie en fonction de l'étendue de l'autre. Dans le graphique linéaire de la Figure 12.6 par exemple chaque point représente le temps de réponse moyen des personnes participantes dont le nom de famille se situe dans les premier, deuxième, troisième et quatrième quartiles (ou quarts) de la distribution des noms. Il montre clairement que le temps de réponse tend à diminuer à mesure que les noms de famille se rapprochent de la fin de l'alphabet. Le diagramme de dispersion de la Figure 12.7 illustre la relation entre les scores obtenus par 25 personnes inscrites en méthodologie de recherche sur l'échelle d'estime de soi de Rosenberg à deux occasions, à une semaine d'intervalle. Ici les points représentent des gens et nous pouvons voir que plus les personnes ont obtenu un score élevé la première fois, plus celles-ci ont eu tendance à obtenir un score élevé la deuxième fois. En général les graphiques linéaires sont utilisés lorsque la variable sur l'axe des *x* a (ou est organisée en) un petit nombre de valeurs distinctes, comme les quatre quartiles de la distribution des noms. Les diagrammes de dispersion sont utilisés lorsque la variable sur l'axe des *x* a un grand nombre de valeurs, comme les différents scores possibles d'estime de soi.

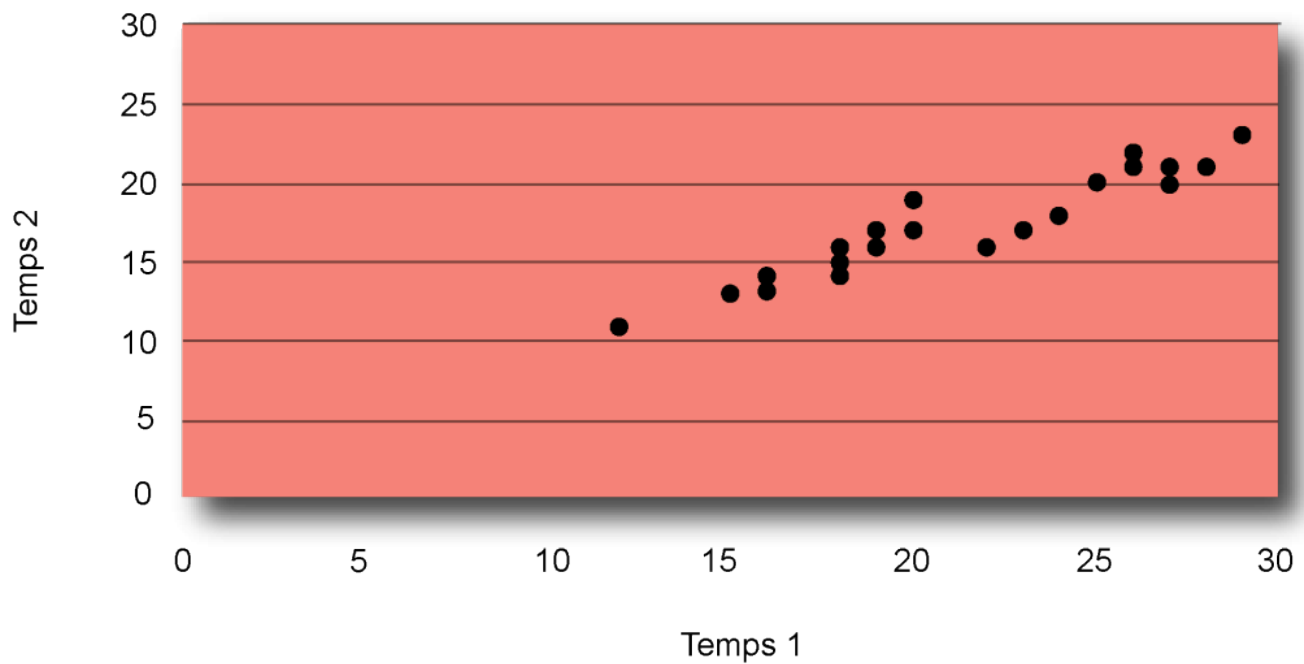


Figure 12.7-Relation statistique entre les scores de personnes aux études à l'échelle d'estime de soi de Rosenberg administrée à deux reprises à une semaine d'intervalle.

La Figure 12.7 illustre une relation positive : des scores plus élevés pour une variable tendent à être associés à des scores plus élevés pour l'autre (formant un nuage de points allant du coin inférieur gauche au coin supérieur droit du graphique.). La Figure 12.6 illustre une relation négative: des scores plus élevés pour une variable tendent à être associés à des scores plus faibles pour l'autre variable (de sorte que les points vont du coin supérieur gauche au coin inférieur droit du graphique).

Ces deux exemples sont également des **relations linéaires**, où les points sont raisonnablement bien ajustés par une seule ligne droite. Les **relations nonlinéaires** sont celles où les points sont mieux ajustés par une ligne courbe. La Figure 12.8 par exemple montre une relation hypothétique entre la quantité de sommeil d'une personne par nuit et son niveau de dépression. Dans cet exemple la ligne qui correspond le mieux aux points est une courbe — une sorte de « U » à l'envers — car les personnes qui dorment environ huit heures ont tendance à être les moins déprimées, tandis que celles qui dorment trop peu et celles qui dorment trop, ont tendance à être plus déprimées. Les relations non linéaires ne sont pas rares en psychologie, mais une discussion détaillée à leur sujet dépasse le cadre de cet ouvrage.

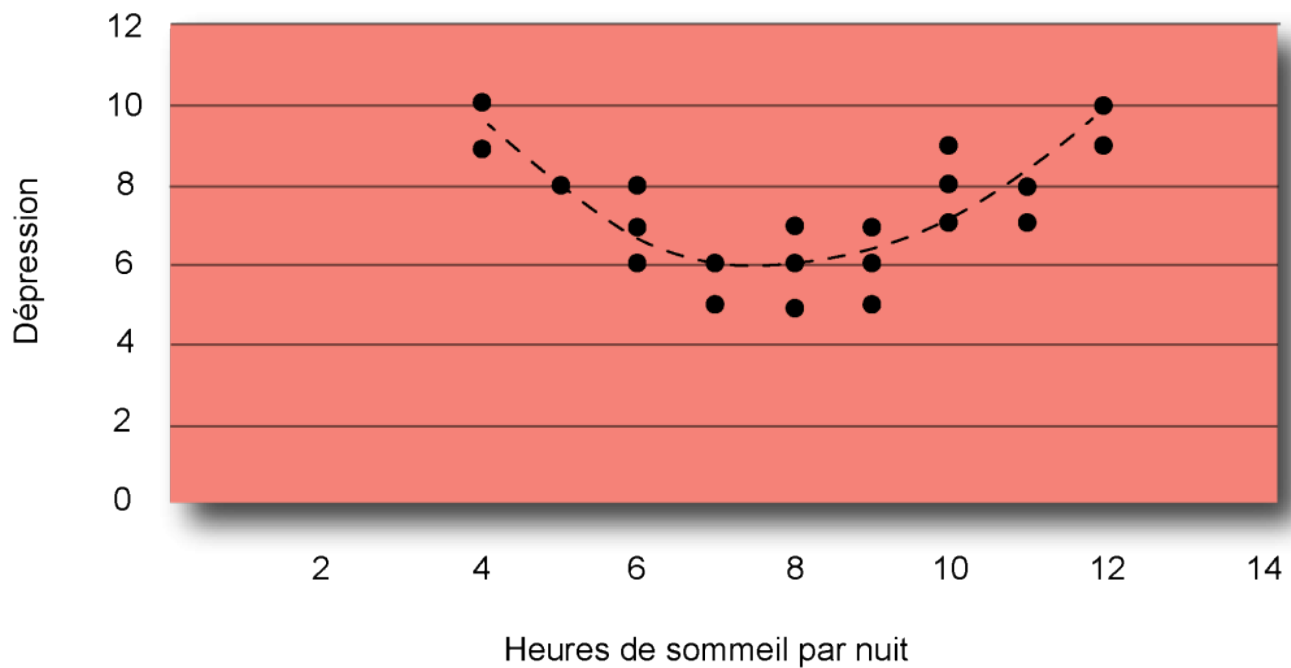


Figure 12.8- Une relation hypothétique non linéaire entre la quantité de sommeil par nuit et le degré de dépression.

Comme nous l'avons vu plus tôt dans le livre, la force d'une corrélation entre des variables quantitatives est généralement mesurée à l'aide d'une statistique appelée r de Pearson. Comme le montre la Figure 12.9, ses valeurs possibles vont de -1.00 à $+1.00$, en passant par zéro. Une valeur de 0 signifie qu'il n'y a pas de relation entre les deux variables. En plus de ses directives pour l'interprétation du d de Cohen, Cohen a proposé des directives pour l'interprétation du r de Pearson dans la recherche en psychologie (voir Tableau 12.4). Les valeurs proches de $\pm 0,10$ sont considérées comme petites, les valeurs proches de $\pm 0,30$ sont considérées comme moyennes et les valeurs proches de $\pm 0,50$ sont considérées comme grandes. Notez que le signe du r de Pearson n'est pas lié à sa force. Des valeurs de r de Pearson de $+ 0,30$ et de $- 0,30$ par exemple sont tout aussi fortes; c'est simplement que l'une représente une relation positive modérée et l'autre une relation négative modérée. Comme le d de Cohen, le r de Pearson est également considéré comme une mesure de « la taille de l'effet », même si la relation n'est pas nécessairement causale.

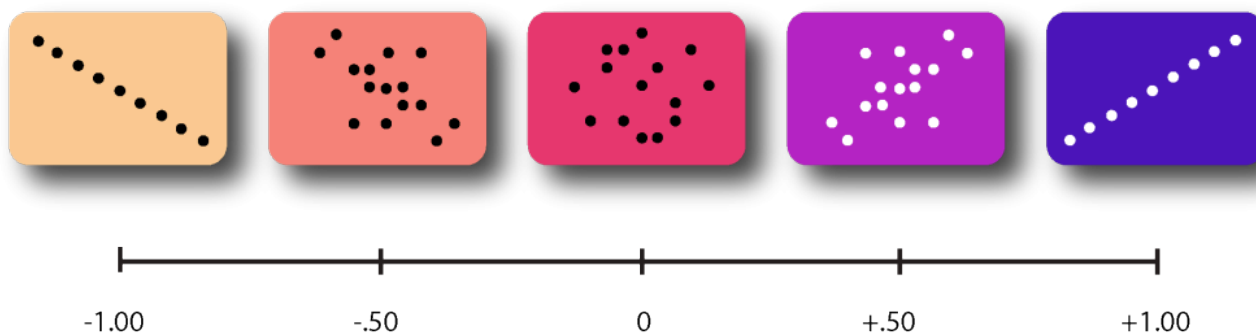


Figure 12.9-Le r de Pearson varie de -1,00 (représentant la relation négative la plus forte possible) à +1,00 (représentant la relation positive la plus forte possible), en passant par 0 (représentant l'absence de relation).

Les calculs du r de Pearson sont plus compliqués que ceux du d de Cohen. Même si vous n'aurez peut-être jamais à les faire à la main, il est toujours instructif de voir comment ils fonctionnent. En termes de calcul le r de Pearson est le « produit croisé moyen des scores z ». Pour le calculer on commence par transformer tous les scores en scores z . Pour la variable X , soustrayez la moyenne de X de chaque score et divisez chaque différence par l'écart-type de X . Pour la variable Y , soustrayez la moyenne de Y de chaque score et divisez chaque différence par l'écart-type de Y . Ensuite pour chaque personne, multipliez les deux scores z ensemble pour former un produit croisé. Enfin prenez la moyenne des produits croisés. La formule se présente comme suit:

$$r = \frac{\sum(z_x z_y)}{N}$$

Le Tableau 12.5 illustre ces calculs pour un petit ensemble de données. La première colonne présente les scores de la variable X , qui a une moyenne de 4.00 et un écart-type de 1.90. La deuxième colonne est le score z pour chacun de ces scores bruts. Les troisième et quatrième colonnes présentent les scores bruts de la variable Y , dont la moyenne est de 40 et l'écart-type de 11.78, ainsi que les scores z correspondants. La cinquième colonne présente les produits croisés. Par exemple le premier est 0.00 multiplié par -0.85, ce qui est égal à 0.00. Le second est 1.58 multiplié par 1.19, ce qui est égal à 1.88. La moyenne de ces produits croisés, indiquée au bas de cette colonne, est le r de Pearson, qui est dans ce cas de +0.53. Il existe d'autres formules pour calculer le r de Pearson à la main, qui peuvent être plus rapides. Cette approche est toutefois beaucoup plus claire pour communiquer conceptuellement ce qu'est le r de Pearson.

Tableau 12.5 Échantillon de calculs pour le r de Pearson

X	z_x	Y	z_y	$z_x z_y$
4	0.00	30	-0.85	0.00
7	1.58	54	1.19	1.88
2	-1.05	23	-1.44	1.52
5	0.53	43	0.26	0.13
2	-1.05	50	0.85	-0.89
$M_x = 4.00$		$M_y = 40.00$		$r = 0.53$
$ÉTx = 1.90$		$ÉTy = 11.78$		

Comme nous l'avons vu précédemment, il existe deux situations communes dans lesquelles la valeur du r de Pearson peut être trompeuse. La première est lorsque la relation étudiée n'est pas linéaire. Même si la Figure 12.8 montre une relation assez forte entre la dépression et le sommeil, le r de Pearson serait proche de zéro parce que les points du diagramme de dispersion ne sont pas bien ajustés par une seule ligne droite. Cela signifie qu'il est important de réaliser un diagramme de dispersion et de confirmer qu'une relation est approximativement linéaire avant d'utiliser le r de Pearson. L'autre cas est celui où l'une ou les deux variables ont une étendue limitée des scores dans l'échantillon par rapport à la population. Ce problème est appelé **restriction de l'étendue de variation**. Supposons par exemple qu'il existe une forte corrélation négative entre l'âge des personnes et leur appréciation de la musique hip-hop, comme le montre le diagramme de dispersion de la Figure 12.10. Le r de Pearson est ici de -.77. Cependant, si nous devons collecter des données uniquement auprès des 18-24 ans — représentés par la zone ombrée de la Figure 12.11 — la relation semble alors assez faible. En fait le r de Pearson pour cette étendue de variation restreinte de l'âge est de 0. Il est donc judicieux de concevoir les études de manière à éviter la restriction de l'étendue de variation. Par exemple si l'âge est l'une de vos principales variables, vous pouvez prévoir de collecter des données auprès de personnes d'âges variés. Étant donné que la restriction de l'étendue de variation n'est pas toujours anticipée ou facilement évitable, il est bon d'examiner vos données pour détecter une éventuelle restriction de l'étendue de variation et d'interpréter le r de Pearson à la lumière de cette restriction. (Il existe également des méthodes statistiques pour corriger le r de Pearson en fonction de la restriction de l'étendue de variation, mais elles sortent du cadre de ce livre).

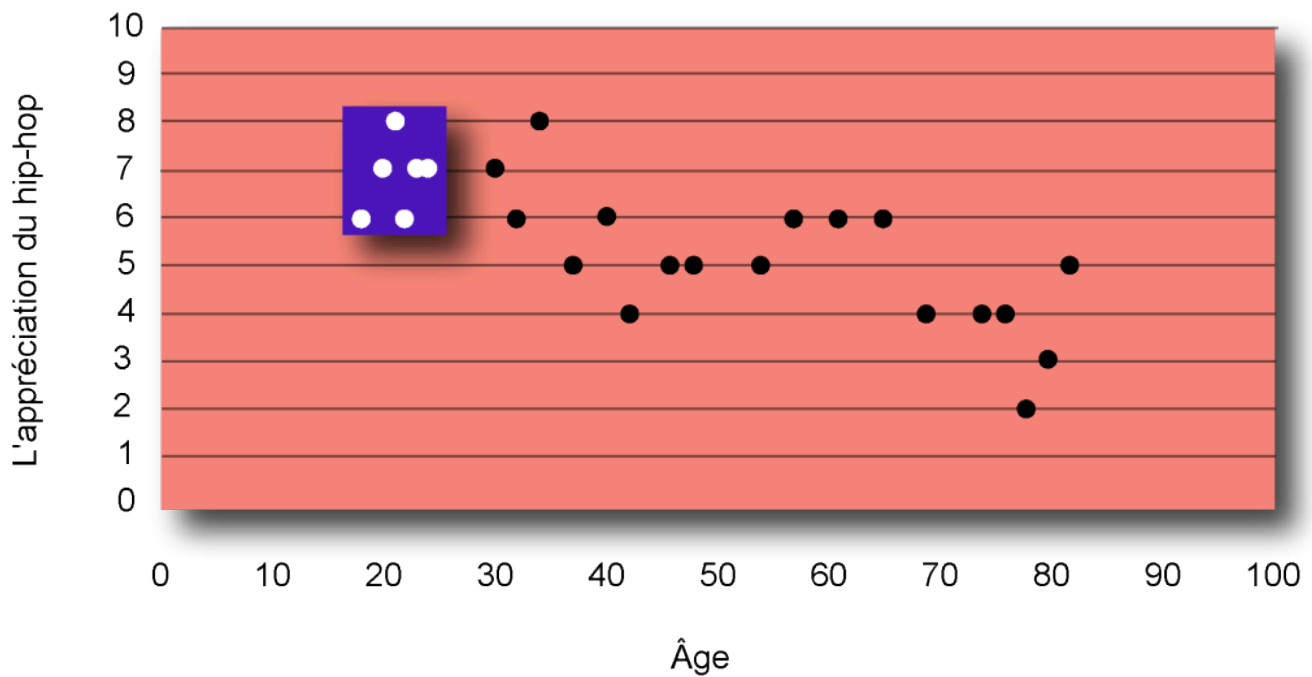


Figure 12.10-Données hypothétiques montrant comment une corrélation globale forte peut sembler faible lorsque l'étendue est restreinte. La corrélation globale est ici de $-0,77$, mais la corrélation pour les 18-24 ans (dans l'encadré bleu) est de 0.

Description des images

Figure 12.5 description longue : Histogrammes. L'axe horizontal est intitulé « Conditions » et l'axe vertical « Évaluation de la gravité par le personnel soignant ». Les données sont les suivantes:

- Conditions : Éducation. Évaluation de la gravité par le personnel soignant : 4.83
- Conditions : Exposition. Évaluation de la gravité par le personnel soignant : 3.47
- Conditions : Contrôle. Évaluation de la gravité par le personnel soignant : 5.56

[\[Retour à la Figure 12.5\]](#)

Figure 12.6 description longue : Graphiques linéaires. L'axe horizontal est intitulé « Quartile du nom de famille » et l'axe vertical est intitulé « Temps de réponse (score z) » et s'étend de -0.4 à 0.4 . Les données sont les suivantes :

- Quartile du nom de famille : Premier. Temps de réponse : 0.2
- Quartile du nom de famille : Deuxième. Temps de réponse : 0.1
- Quartile du nom de famille : Troisième. Temps de réponse : -0.1
- Quartile du nom de famille : Quatrième. Temps de réponse : -0.2

[\[Retour à la Figure 12.6\]](#)

Figure 12.7 description longue: Diagramme de dispersion montrant les scores de personnes étudiantes à l'échelle d'estime de soi de Rosenberg, administrée à deux reprises à une semaine d'intervalle. L'axe horizontal du diagramme de dispersion est intitulé « Temps 1 » et l'axe vertical « Temps 2 ». Chaque point représente les deux scores d'une personne. Par exemple la personne a obtenu 25 la première fois et 20 la deuxième fois. Les points s'échelonnent entre 12, 11 et 28, 23. [\[Retour à la Figure 12.7\]](#)

Figure 12.8 description longue: Diagramme de dispersion montrant la relation hypothétique entre le nombre d'heures de sommeil par nuit et le niveau de dépression. L'axe horizontal est intitulé « Heures de sommeil par nuit » et comporte des valeurs comprises entre 0 et 14, et l'axe vertical est intitulé « Dépression » et comporte des valeurs comprises entre 0 et 12. Une ligne pointillée en forme de U trace la forme approximative des points de données. Deux personnes qui dorment 4 heures par nuit ont obtenu un score de 9 et 10 sur l'échelle de dépression, ce qui correspond au score obtenu par deux personnes qui dorment 12 heures. Les personnes qui dorment 4 et 12 heures ont obtenu les scores les plus élevés sur l'échelle de la dépression, et ces points de données forment les extrémités du U. Trois personnes qui dorment 8 heures ont obtenu des scores de 5, 6 et 7 sur l'échelle de la dépression. Les points de données des personnes qui dorment 8 heures se situent au milieu du U. [\[Retour à la Figure 12.8\]](#)

Figure 12.9 description longue : Cinq diagrammes de dispersion représentant les différentes valeurs du r de Pearson.

Le premier diagramme de dispersion représente le r de Pearson avec une valeur de -1.00. Le diagramme de dispersion montre une ligne diagonale de points qui s'étend du coin supérieur gauche au coin inférieur droit. Plus la valeur de la variable sur l'axe des x est élevée, plus la valeur de la variable sur l'axe des y est faible. Il s'agit de la relation négative la plus forte possible.

Le second diagramme de dispersion représente le r de Pearson avec une valeur de -0.50. Il illustre une relation négative entre les variables sur les axes des x et des y . Les points sont tracés librement autour d'une ligne invisible allant du coin supérieur gauche au coin inférieur droit.

Le troisième diagramme de dispersion représente le r de Pearson avec une valeur de 0. Les points apparaissent de manière aléatoire; il n'y a pas de relation entre les axes x et y .

Le quatrième diagramme de dispersion représente le r de Pearson avec une valeur de +0.50. Il décrit une relation positive entre les variables sur les axes des x et des y . Les points sont vaguement tracés autour d'une ligne invisible allant du coin inférieur gauche au coin supérieur droit.

Le cinquième diagramme de dispersion représente le r de Pearson avec une valeur de +1.00. Le diagramme de dispersion présente une ligne diagonale de points allant du coin inférieur gauche au coin supérieur droit. Plus la valeur de la variable sur l'axe des x est élevée, plus la valeur de la variable sur l'axe des y est élevée. Il s'agit de la relation positive la plus forte possible. [\[Retour à la Figure 12.9\]](#)

Figure 12.10 description longue: Diagramme de dispersion dont l'axe horizontal est intitulé « Âge » et dont les valeurs sont comprises entre 0 et 100, et l'axe vertical est intitulé « Appréciation du hip-hop » et dont les valeurs sont comprises entre 0 et 10. Le r de Pearson dans ce diagramme de dispersion est le suivant : -0.77. Il existe une forte relation négative entre l'âge et l'appréciation du hip-hop, comme le montrent ces paires ordonnées : (20, 8), (40, 6), (69, 4), (80, 3). Mais si l'on restreint l'âge pour n'examiner que les 18-24 ans,

cette relation est beaucoup moins claire. Chacune des sept personnes de cette tranche d'âge évalue son appréciation du hip-hop comme étant soit un 6, un 7 ou un 8. La personne la plus jeune l'évalue à 6 alors que la plus âgée l'évalue à 7, et certaines personnes entre les deux l'évaluent à 8. Comme il n'y a pas de tendance claire, la corrélation pour les 18 à 24 ans est de 0. [\[Retour à la Figure 12.10\]](#)

65. Décrire vos résultats

Objectifs d'apprentissage

1. Rédiger des statistiques descriptives simples dans le style de l'American Psychological Association (APA).
2. Interpréter et créer des figures simples de style APA, y compris des graphiques à barres, des graphiques linéaires et des diagrammes de dispersion.
3. Interpréter et créer des tableaux simples de style APA, y compris des tableaux de moyennes de groupes ou de conditions et des matrices de corrélation.

Une fois vos analyses statistiques descriptives faites, vous aurez besoin de les présenter à d'autres personnes. Dans cette section, nous nous concentrerons sur la présentation des résultats statistiques descriptifs par écrit, sous forme de figures et de tableaux — conformément aux directives de l'American Psychological Association (APA) pour les rapports de recherche écrits. Ces principes peuvent être facilement adaptés à d'autres formats de présentation, tels que les affiches et les diaporamas.

Présenter des statistiques descriptives par écrit

Rappelons que le style APA comprend plusieurs règles pour la présentation des résultats numériques dans le texte (voir 4.31-4.34 dans le *Manuel de publication* de l'APA). Il s'agit notamment d'utiliser des mots uniquement pour les nombres inférieurs à 10 qui ne représentent pas des résultats statistiques précis et d'utiliser des chiffres pour les nombres de 10 et plus. Toutefois les résultats statistiques sont toujours présentés sous forme de chiffres plutôt que de mots, et sont généralement arrondis à deux décimales (par exemple « 2.00 » plutôt que « deux » ou « 2 »). Ils peuvent être présentés soit dans la description narrative des résultats, soit entre parenthèses — comme les citations de référence. Lorsque vous avez un petit nombre de résultats à présenter, il est souvent plus efficace de les écrire. Voici quelques exemples :

L'âge moyen des participantes et participants était de 22.43 ans avec un écart-type de 2.34.

Parmi les participants ayant une faible estime de soi, ceux qui étaient d'humeur négative ont exprimé des intentions plus fortes d'avoir des rapports sexuels non protégés ($M = 4.05$, $ÉT = 2.32$) que ceux qui étaient d'humeur positive ($M = 2.15$, $ÉT = 2.27$).

Le groupe de traitement avait une moyenne de 23.40 (*écart-type* = 9.33), tandis que le groupe de contrôle avait une moyenne de 20.87 (*écart-type* = 8.45).

La corrélation test-retest était de 0.96.

Il existe une corrélation négative modérée entre la position alphabétique des noms de famille des répondants et leur temps de réponse ($r = -0.27$).

Notons que lorsqu'ils sont présentés dans le texte, les termes *moyenne* et *écart-type* sont écrits, mais lorsqu'ils sont présentés entre parenthèses, les symboles *M* et *ÉT* sont utilisés. Notons également qu'il est particulièrement important d'utiliser une construction parallèle pour exprimer des résultats similaires ou comparables de manière similaire. Le troisième exemple est *bien* meilleur que l'alternative non parallèle suivante :

Le groupe de traitement avait une moyenne de 23.40 (*ÉT* = 9.33), tandis que 20.87 était la moyenne du groupe de contrôle, qui avait un écart-type de 8.45.

Présenter des statistiques descriptives dans les figures

Lorsque vous avez un grand nombre de résultats à présenter, vous pouvez souvent le faire de manière plus claire et plus efficace en utilisant une représentation graphique des données telle que des diagrammes circulaires, des graphiques à barres ou des diagrammes de dispersion. Dans un rapport de recherche de style APA, ces graphiques sont présentés sous forme de **figures**. Lorsque vous préparez des figures pour un rapport de recherche de style APA, vous devez garder à l'esprit certaines directives générales. Tout d'abord la figure doit toujours ajouter de l'information importante plutôt que de répéter des informations qui figurent déjà dans le texte ou dans un tableau (si une figure présente des informations de manière plus claire ou plus efficace, vous devez conserver la figure et éliminer le texte ou le tableau). Deuxièmement les figures doivent être aussi simples que possible. Par exemple le manuel de publication décourage l'utilisation de la couleur à moins qu'elle ne soit absolument nécessaire (bien que la couleur puisse toujours être un élément efficace dans les affiches, les présentations de diapositives ou les manuels). Troisièmement les figures doivent pouvoir être interprétées seules. Le lectorat doit être en mesure de comprendre le résultat de base en se référant uniquement à la figure et sa légende, et ne doit pas avoir à se référer au texte pour une explication.

Il existe également plusieurs directives plus techniques pour la présentation des figures, dont les suivantes (voir les sections 5.20 à 5.30 du manuel de publication de l'APA) :

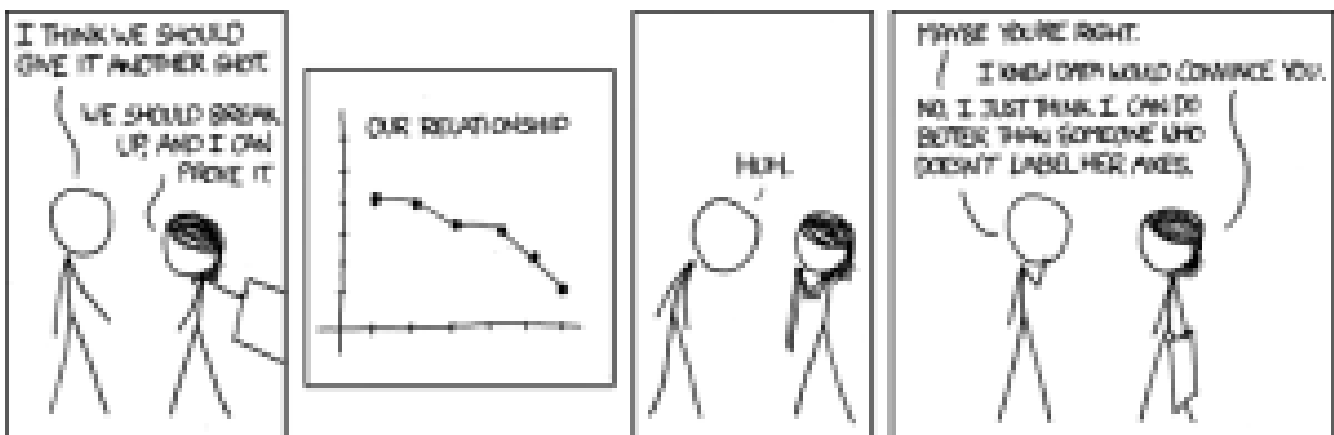
- Mise en page des graphiques
 - En général les diagrammes de dispersion, les graphiques à barres et les graphiques linéaires doivent être légèrement plus larges que hauts.
 - La variable indépendante doit être représentée sur l'axe des *x* et la variable dépendante sur l'axe des *y*.
 - Les valeurs doivent augmenter de gauche à droite sur l'axe des *x* et de bas en haut sur l'axe des *y*.
 - L'axe des *x* et l'axe des *y* doivent commencer par la valeur zéro.
- Étiquettes et légendes des axes
 - Les étiquettes des axes doivent être claires et concises et inclure les unités de mesure si elles ne

figurent pas dans le titre.

- Les étiquettes des axes doivent être parallèles à l'axe.
- Les légendes doivent apparaître à l'intérieur de la figure.
- Le texte doit être rédigé dans la même police de caractères simple et ne doit pas être inférieur à 8 points ni supérieur à 14 points.

- Titres et figures

- Les figures sont dotées du mot « Figure », suivies du numéro de la figure dans l'ordre où elle apparaît dans le texte, et se terminent par un point. Ce titre est en gras.
- Sur la ligne suivante, une brève description de la figure est ajoutée (par exemple « Temps de réaction du groupe témoin par rapport au groupe expérimental »). Ceci en italique.
- Sous la description, la figure est insérée.
- Au bas de la figure, il y a présentation de toutes les informations nécessaires à l'interprétation de la figure telles que les abréviations, les unités de mesure (si elles ne figurent pas dans l'étiquette de l'axe), les unités des barres d'erreur, etc.



"Convincing" repéré à <http://imgs.xkcd.com/comics/convincing.png> (CC-BY-NC 2.5) [Description de l'image](#)

Graphiques à barres

Comme nous l'avons vu tout au long de cet ouvrage, les **graphiques à barres** sont généralement utilisés pour présenter et comparer les scores moyens de deux ou plusieurs groupes ou conditions. Le graphique à barres de la Figure 12.11 est une version en style APA de la Figure 12.4. Remarquons qu'il est conforme à toutes les lignes directrices énumérées. Un nouvel élément de la Figure 12.11 est constitué par les petites barres verticales qui s'étendent à la fois vers le haut et vers le bas à partir du haut de chaque barre principale. Il s'agit de **barres d'erreur**, qui représentent la variabilité dans chaque groupe ou condition. Bien qu'elles s'étendent parfois d'un *écart-type* dans chaque direction, il est plus probable qu'elles s'étendent d'une *erreur standard* dans chaque direction (comme dans la Figure 12.11). L'**erreur standard** est l'écart-type du groupe divisé par la racine carrée de la taille de l'échantillon du groupe. L'erreur standard est utilisée, car

en général une différence entre les moyennes des groupes, qui est supérieure à deux erreurs standards, est statistiquement significative. On peut donc « voir » si une différence est statistiquement significative en se basant sur un graphique à barres avec des barres d'erreur.

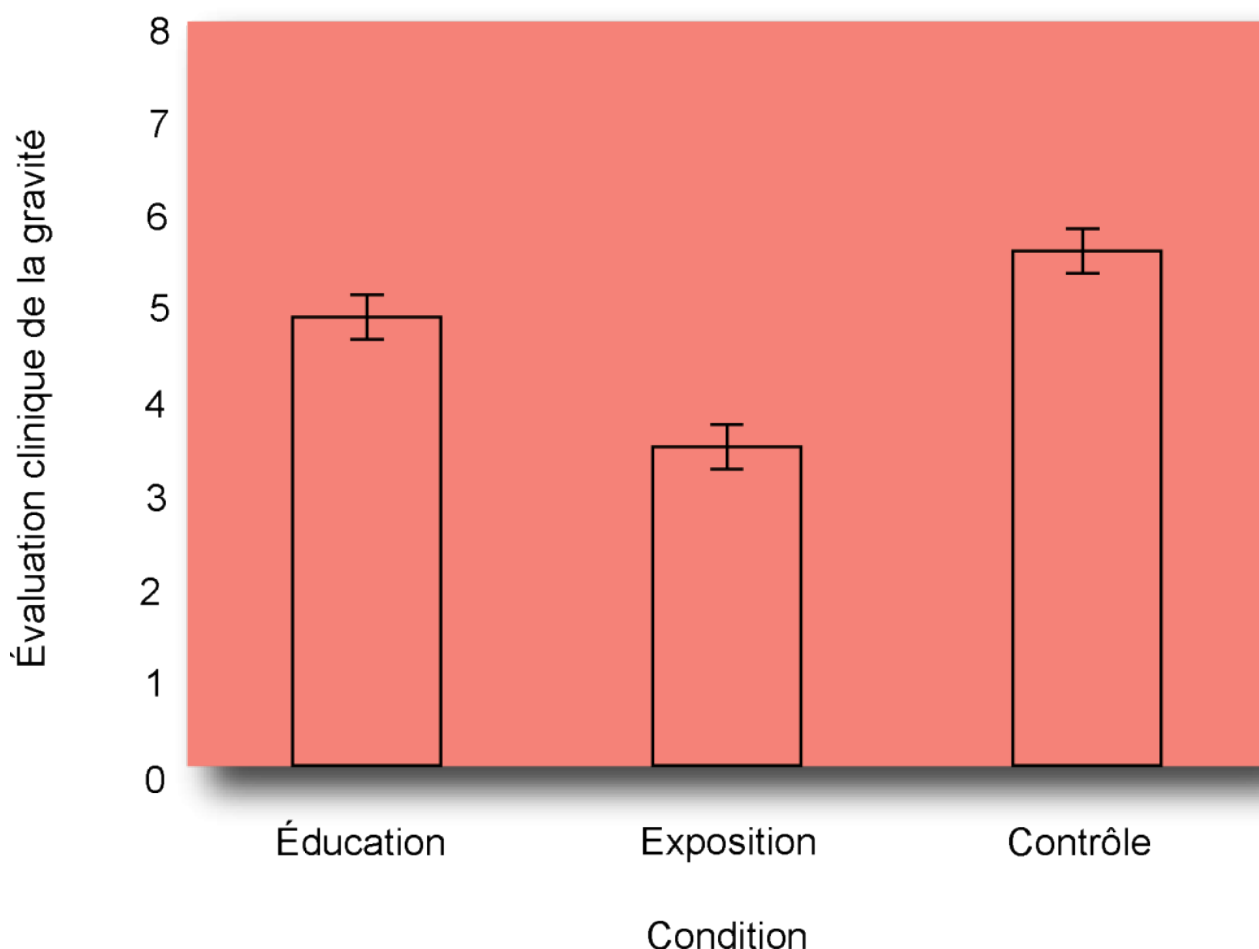


Figure 12.11-Exemple de graphique à barres de type APA, avec des barres d'erreur représentant les erreurs standard, d'après la recherche d'Ollendick et de ses collègues. [\[Description de l'image\]](#)

Graphiques

Les **graphiques linéaires** sont utilisés lorsque la variable indépendante est mesurée de manière plus continue (par exemple le temps) ou pour présenter des corrélations entre des variables quantitatives lorsque la variable indépendante a, ou est organisée en un nombre relativement faible de niveaux distincts. Chaque point d'un graphique linéaire représente le score moyen de la variable dépendante pour les personnes participantes à un niveau de la variable indépendante. La Figure 12.12 est une version en style APA des résultats de Carlson et Conard. Vous remarquerez qu'elle inclut des barres d'erreur représentant l'erreur standard et qu'elle est conforme à toutes les directives énoncées.

Figure x.

Temps de réponse moyen en fonction de la position alphabétique des noms des personnes répondantes dans l'alphabet.

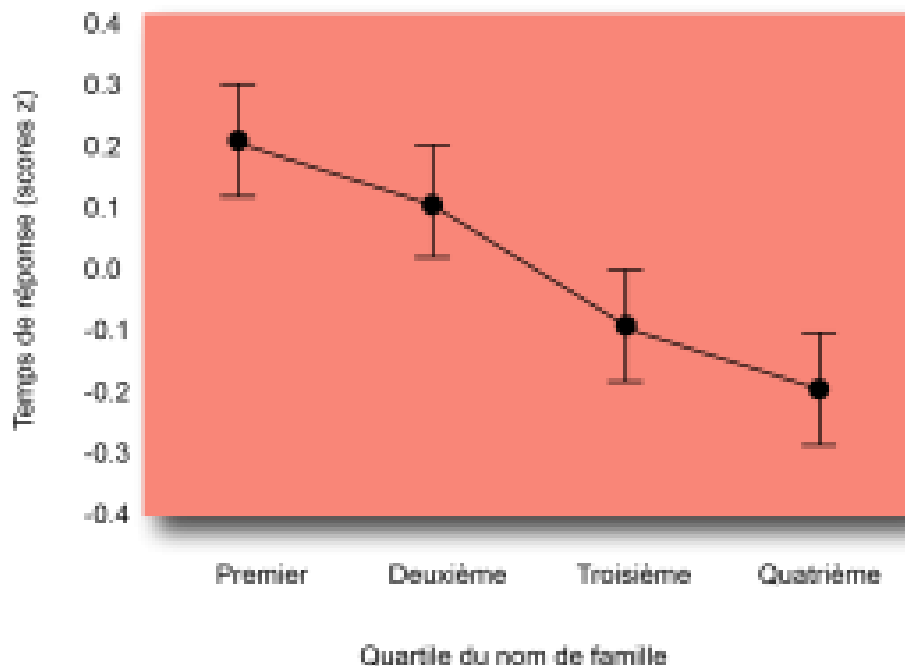


Figure 12.12-Exemple de graphique linéaire de type APA d'après la recherche de Carlson et Conard.
[Description de l'image](#)

Dans la plupart des cas, les informations contenues dans un graphique linéaire pourraient tout aussi bien être présentées dans un graphique à barres. Dans la Figure 12.12 par exemple on pourrait remplacer chaque point par une barre atteignant le même niveau et laisser les barres d'erreur à leur place. Cela souligne la similitude fondamentale des deux types de figures. Dans les deux cas, il s'agit de différences dans le score moyen d'une variable entre les niveaux d'une autre variable. La convention suivie par la plupart des scientifiques consiste toutefois à utiliser un graphique à barres lorsque la variable représentée sur l'axe des x est catégorielle, et un graphique linéaire lorsqu'elle est quantitative.

Diagrammes de dispersion

Les **diagrammes de dispersion** sont utilisés pour présenter les corrélations et les relations entre les variables quantitatives lorsque la variable sur l'axe des x (généralement la variable indépendante) comporte un grand nombre de niveaux. Chaque point d'un diagramme de dispersion représente une personne plutôt que la moyenne d'un groupe de personnes, et il n'y a pas de lignes reliant les points. Le graphique de la Figure 12.13

est une version du style APA de la Figure 12.7, qui illustre quelques points supplémentaires. Premièrement lorsque les variables de l'axe des *x* et de l'axe des *y* sont conceptuellement similaires et mesurées sur la même échelle — comme ici, où il s'agit de mesures de la même variable à deux occasions différentes — cela peut être souligné en faisant en sorte que les axes soient de la même longueur. Deuxièmement lorsque deux personnes ou plus se situent exactement au même point sur le graphique, une façon de l'indiquer consiste à décaler légèrement les points le long de l'axe des *x*. D'autres méthodes consistent à afficher le nombre de personnes entre parenthèses à côté du point ou à agrandir ou assombrir le point proportionnellement au nombre de personnes. Enfin la ligne droite, qui correspond le mieux aux points du diagramme de dispersion appelée ligne de régression, peut également être incluse.

Figure x.

Relation entre les scores sur l'échelle d'estime de soi de Rosenberg obtenus par 25 personnes inscrites en méthodes de recherche à deux occasions, à une semaine d'intervalle. Le *r* de Pearson = 0.96.

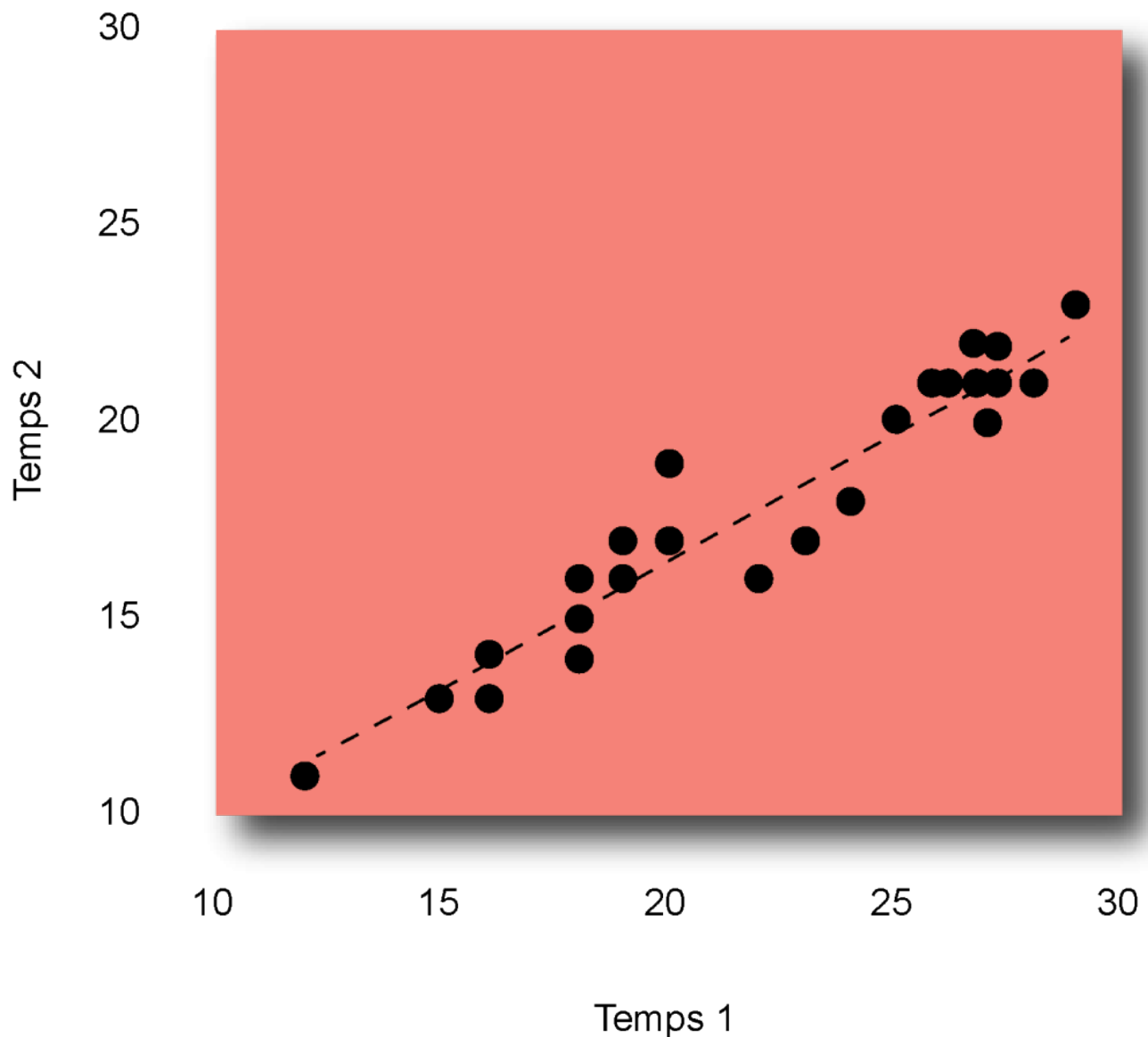


Figure 12.13-Exemple de diagramme de dispersion de type APA. [\[Description de l'image\]](#)

Préserver des statistiques descriptives dans des tableaux

Comme les graphiques, les tableaux peuvent être utilisés pour présenter de grandes quantités d'informations de manière claire et efficace. Les mêmes principes généraux s'appliquent aux tableaux comme aux graphiques. Ils doivent ajouter des informations importantes à la présentation de vos résultats, être aussi simples que possible et pouvoir être interprétés seuls. Une fois de plus nous nous concentrons ici sur les tableaux pour un manuscrit suivant le style APA.

L'utilisation la plus courante des tableaux consiste à présenter plusieurs moyennes et écarts-types,

généralement pour des modèles de recherche complexes comportant plusieurs variables indépendantes et dépendantes. La Figure 12.14 par exemple présente les résultats d'une étude hypothétique similaire à celle de MacDonald et Martineau (2002). (Les moyennes de la Figure 12.14 sont les moyennes rapportées par MacDonald et Martineau, mais les erreurs standard ne le sont pas). Rappelons que ces scientifiques ont catégorisé les personnes participantes comme ayant une faible ou une forte estime de soi, les ont mises dans une humeur négative ou positive, et ont mesuré leurs intentions d'avoir des rapports sexuels non protégés. Ces scientifiques ont également mesuré l'attitude de ces personnes à l'égard des rapports sexuels non protégés. Remarquons que le tableau a des lignes horizontales qui couvrent l'ensemble du tableau en haut et en bas, et juste sous les titres des colonnes. En outre chaque colonne a un titre — y compris la colonne la plus à gauche — et des titres supplémentaires s'étendent sur deux colonnes ou plus, ce qui permet d'organiser les informations et de les présenter de manière plus efficace. Enfin les tableaux de style APA sont numérotés consécutivement à partir de 1 (Tableau 1, Tableau 2, etc.) et portent un titre bref mais clair et descriptif.

Tableau X.
Moyennes et écart-type des intentions d'avoir des rapports sexuels non protégés et des attitudes à l'égard des rapports sexuels non protégés en fonction de l'humeur et de l'estime de soi

Estime de soi	Humeur négative		Humeur positive	
	M	ÉT	M	ÉT
Intentions				
Élevé	2.46	1.97	2.45	2.09
Faible	4.05	2.32	2.15	2.27
Attitudes				
Élevé	1.65	2.23	1.82	2.32
Faible	1.95	2.91	1.23	1.76

Figure 12.14-Exemple de tableau au format APA présentant les moyennes et les écarts types.
[\[Description de l'image\]](#)

Une autre utilisation courante des tableaux consiste à présenter les corrélations — généralement mesurées par le coefficient *r* de Pearson — entre plusieurs variables. Ce type de tableau est appelé **matrice de corrélation**. La Figure 12.15 est une matrice de corrélation basée sur une étude réalisée par David McCabe et ses collègues (McCabe, Roediger, McDaniel, Balota et Hambrick, 2010). Ces scientifiques ont examiné les relations entre la mémoire de travail et plusieurs autres variables. Le tableau indique que la corrélation entre la mémoire de travail et les fonctions exécutives par exemple est extrêmement forte (0.96), que la corrélation entre la mémoire de travail et le vocabulaire est moyenne (0.27) et que toutes les mesures, à l'exception du vocabulaire, ont tendance à diminuer avec l'âge. Notons ici que seule la moitié du tableau

est remplie, car l'autre moitié aurait des valeurs identiques. Par exemple la valeur du r de Pearson dans le coin supérieur droit (mémoire de travail et âge) serait la même que celle dans le coin inférieur gauche (âge et mémoire de travail). La corrélation d'une variable avec elle-même étant toujours de 1.00, ces valeurs sont remplacées par des tirets pour faciliter la lecture du tableau.

Tableau X.
Corrélations entre cinq variables cognitives et l'âge

Mesure	1	2	3	4	5
1. Mémoire de travail	—	.67	.61	.67	
2. Fonctions exécutives	.96	—			
3. Vitesse de traitement	.78	.78	—		
4. Vocabulaire	.27	.45	.08	—	
5. Mémoire épisodique	.73	.75	.52	.38	—
6. Âge	-.59	-.56	-.82	.22	-.41

Figure 12.15 Exemple de tableau au format APA (matrice de corrélation) d'après les recherches de McCabe et ses collègues.
[Description de l'image](#)

Comme pour les graphiques, il n'est pas nécessaire de répéter dans le texte les résultats statistiques précis qui apparaissent dans un tableau. En rédaction, on peut plutôt noter les grandes tendances et attirer l'attention de la lectrice et du lecteur sur des détails (par exemple des corrélations spécifiques) qui présentent un intérêt particulier.

Description de l'image

Description longue « Convaincant » : Bande dessinée en quatre planches. Dans la première planche, un homme dit à une femme : « Je pense que nous devrions essayer encore une fois ». La femme répond : « Nous devrions rompre, et je peux le prouver ».

Dans le deuxième panneau, il y a un graphique linéaire avec une tendance à la baisse intitulé « Notre relation ».

Dans le troisième panneau, l'homme, penché en avant et regardant le graphique dans les mains de la femme, dit « Huh ».

Dans le quatrième panneau, l'homme dit : « Vous avez peut-être raison. » La femme répond : « Je savais que les données te convaincraient. » L'homme répond : « Non, je pense juste que je peux faire mieux que quelqu'un qui n'étiquette pas ses axes. »

[\[Retour à « Convaincant »\]](#)

Figure 12.11 description longue : Un exemple de graphique à barres de style APA, dont l'axe horizontal est intitulé « Conditions » et l'axe vertical « Évaluation de la sévérité par le clinicien ». La légende du graphique indique : « Figures X. Moyenne de l'évaluation de la sévérité de la phobie par le clinicien pour les participants recevant le traitement d'éducation et le traitement d'exposition. Les barres d'erreur représentent les erreurs standard. » En haut de chaque barre de données se trouve une barre d'erreur, qui ressemble à un I majuscule: une ligne verticale avec de courtes lignes horizontales attachées en haut et en bas. La moitié inférieure de chaque barre d'erreur est suspendue au-dessus de la barre de données, tandis que la moitié supérieure dépasse le haut de la barre de données. [\[Retour à la Figure 12.11\]](#)

Figure 12.12 description longue : Un exemple de graphique linéaire, dont l'axe horizontal est intitulé « Quartile du nom de famille » et l'axe vertical « Temps de réponse (scores z) ». Le titre du graphique est : « Figure X. Temps de réponse moyen en fonction de la position alphabétique des noms des personnes répondantes dans l'alphabet. Les temps de réponse sont exprimés sous forme de scores z. Les barres d'erreur représentent les erreurs standard. » Chaque point de données a une barre d'erreur qui dépasse en haut et en bas. [\[Retour à la Figure 12.12\]](#)

Figure 12.13 description longue: Un exemple de diagramme de dispersion de style APA avec un axe horizontal intitulé « Temps 1 » et un axe vertical intitulé « Temps 2 ». Chaque axe a des valeurs comprises entre 10 et 30. Le titre du diagramme de dispersion est « Figure X. Relation entre les scores sur l'échelle d'estime de soi de Rosenberg obtenus par 25 étudiants en méthodes de recherche à deux occasions, à une semaine d'intervalle. Le r de Pearson = 0,96 ». La plupart des points de données sont regroupés autour de la ligne de régression en pointillés, qui s'étend approximativement de (12, 11) à (29, 22). [\[Retour à la Figure 12.13\]](#)

Figure 12.14 description longue : Échantillon de tableau de style APA présentant les moyennes et les écarts-types. Le tableau s'intitule « Tableau X » et a pour titre « Moyennes et écart-type des intentions d'avoir des rapports sexuels non protégés et des attitudes à l'égard des rapports sexuels non protégés en fonction de l'humeur et de l'estime de soi ». Les données sont organisées en fonction de l'humeur négative et de l'humeur positive et détaillent les intentions et les attitudes à l'égard des rapports sexuels non protégés.

Humeur négative :

- Intentions
 - Élevée — Moyenne, 2,46
 - Élevée — Écart-type, 1,97
 - Faible — Moyenne, 4,05
 - Faible — Écart-type, 2,32
- Attitudes
 - Élevée — Moyenne, 1,65
 - Élevée — Écart-type, 2,23

- Faible — Moyenne, 1,95
- Faible — Écart-type, 2,01

Humeur positive :

- Intentions
 - Élevée — Moyenne, 2,45
 - Élevée — Écart-type, 2,00
 - Faible — Moyenne, 2,15
 - Faible — Écart-type, 2,27
- Attitudes
 - Élevée — Moyenne, 1,82
 - Élevée — Écart-type, 2,32
 - Faible — Moyenne, 1,23
 - Faible — Écart-type, 1,75

[\[Retour à la Figure 12.14\]](#)

Figure 12.15 description longue: Échantillon de matrice de corrélation de style APA, intitulé « Tableau X : Corrélations entre cinq variables cognitives et l'âge ». Les cinq variables cognitives sont :

1. Mémoire de travail
2. Fonction exécutive
3. Vitesse de traitement
4. Vocabulaire
5. Mémoire épisodique

Les données sont les suivantes :

Tableau X.
Corrélations entre cinq variables cognitives et l'âge

Mesure	1	2	3	4	5
1. Mémoire de travail	—				
2. Fonctions exécutives	.96	—			
3. Vitesse de traitement	.78	.78	—		
4. Vocabulaire	.27	.45	.08	—	
5. Mémoire épisodique	.73	.75	.52	.38	—
6. Âge	-.59	-.56	-.82	.22	-.41

[\[Retour à la Figure 12.15\]](#)

Attributions des médias

- [Convincing](#) par XKCD [CC BY-NC \(Attribution NonCommercial\)](#)

66. Faire vos analyses

Objectifs d'apprentissage

1. Décrire les étapes de la préparation et de l'analyse d'un ensemble typique de données brutes.
2. Faire la différence entre les analyses planifiées et les analyses exploratoires de données.

Me^ me si vous comprenez les statistiques impliquées, l'analyse des données peut e^ tre un processus compliqué. Il est probable que pour chaque personne participante, il existe des données pour plusieurs variables différentes : des données démographiques, telles que le genre et l'a^ ge, une ou plusieurs variables indépendantes, une ou plusieurs variables dépendantes, et peut-e^ tre une vérification de la manipulation. En outre les données « brutes » (non analysées) peuvent prendre différentes formes — questionnaires papier-crayon remplis, fichiers informatiques remplis de chiffres ou de texte, vidéos ou notes écrites — et il peut e^ tre nécessaire de les organiser, de les codifier ou de les combiner d'une manière ou d'une autre. Il se peut me^ me que des réponses manquantes, incorrectes ou simplement « suspectes » doivent e^ tre traitées. Dans cette section, nous allons considérer quelques conseils pratiques pour rendre ce processus aussi organisé et efficace que possible.

Préparer vos données pour l'analyse

Que vos données brutes soient sur papier ou dans un fichier informatique (ou les deux), il y a quelques précautions à prendre avant de commencer à les analyser. Tout d'abord assurez-vous qu'elles ne contiennent aucune information susceptible d'identifier les personnes, et que vous disposez d'un endroit su^ r où vous pouvez stocker les données et d'un autre endroit su^ r où vous pouvez stocker les formulaires de consentement. À moins que les données ne soient très sensibles, une pièce fermée à clé ou un ordinateur protégé par un mot de passe suffisent généralement. Il est également conseillé de faire des photocopies ou des fichiers de sauvegarde de vos données et de les stocker dans un autre endroit sécurisé — au moins jusqu'à ce que le projet soit terminé. Les scientifiques d'expérience conservent généralement une copie de leurs données brutes et des formulaires de consentement pendant plusieurs années, au cas où des questions concernant la procédure, les données ou le consentement des personnes participantes se poseraient après la fin du projet.

Ensuite vous devez vérifier vos **données brutes** pour vous assurer qu'elles sont complètes et qu'elles

semblent avoir été enregistrées avec précision (que ce soient les personnes participantes, vous-même ou un programme informatique qui avez fait l'enregistrement). À ce stade vous constaterez peut-être que certaines réponses sont illisibles ou manquantes, ou qu'il y a des malentendus évidents (par exemple une réponse de 12 sur une échelle d'évaluation de 1 à 10). Vous devrez décider si ces problèmes sont suffisamment graves pour rendre les données d'une personne inutilisables. S'il manque des informations sur la principale variable indépendante ou dépendante, ou si plusieurs réponses sont manquantes ou suspectes, vous devrez peut-être exclure les données de cette personne des analyses. Si vous décidez d'exclure des données, ne les jetez pas et ne les supprimez pas parce que vous ou un autre scientifique voudriez peut-être les consulter plus tard. Mettez-les plutôt de côté et notez les raisons pour lesquelles vous avez décidé de les exclure, car vous devrez rapporter ces informations.

Vous êtes maintenant fin prêt à saisir vos données dans un tableur ou si elles se trouvent déjà dans un fichier informatique, à les formater en vue de leur analyse. Vous pouvez utiliser un tableur général comme Microsoft Excel ou un programme d'analyse statistique comme SPSS pour créer votre **fichier de données**. (Les fichiers de données créés dans un programme peuvent généralement être convertis pour fonctionner avec d'autres programmes). Le format le plus courant est le suivant: chaque ligne représente un participant et chaque colonne représente une variable (avec le nom de la variable en haut de chaque colonne). Un exemple de fichier de données est présenté dans le Tableau 12.6. La première colonne contient les numéros d'identification de chaque personne participante. Viennent ensuite les colonnes contenant les informations démographiques (genre et âge), les variables indépendantes (humeur, quatre items d'estime de soi et le total des quatre items d'estime de soi) et enfin les variables dépendantes (intentions et attitudes). Les variables catégorielles peuvent généralement être saisies sous forme d'étiquettes de catégorie (par exemple « H » et « F » pour homme et femme) ou de nombres (par exemple « 0 » pour une humeur négative et « 1 » pour une humeur positive). Bien que les étiquettes des catégories soient souvent plus claires, certaines analyses peuvent nécessiter des nombres. Plusieurs programmes statistiques, dont SPSS vous permettent de saisir des nombres, mais aussi d'attacher un nom de catégorie à chaque nombre.

Tableau 12.6 Exemple de fichier de données

ID	SEXE	ÂGE	HUMEUR	SE1	SE2	SE3	SE4	TOTAL	INT	ATT
1	H	20	1	2	3	2	3	10	6	5
2	F	22	1	1	0	2	1	4	4	4
3	F	19	0	2	2	2	2	8	2	3
4	F	24	0	3	3	2	3	11	5	6

Si vous disposez de mesures à réponses multiples — telles que la mesure de l'estime de soi du Tableau 12.6 — vous pouvez combiner les éléments à la main, puis saisir le score total dans votre feuille de calcul. Toutefois il est préférable de saisir chaque réponse comme une variable distincte dans la feuille de calcul — comme pour la mesure de l'estime de soi du Tableau 12.6 — et d'utiliser le logiciel pour les combiner (par exemple à l'aide de la fonction « MOYENNE » dans Excel ou de la fonction « calcul » dans SPSS). Non seulement cette approche est plus précise, mais elle vous permet de détecter et de corriger les

erreurs, d'évaluer la cohérence interne et d'analyser les réponses individuelles si vous décidez de le faire ultérieurement.

Analyses préliminaires

Avant de vous pencher sur vos questions de recherche principales, vous devez souvent effectuer plusieurs analyses préliminaires. Pour les mesures à réponses multiples, vous devez évaluer la cohérence interne de la mesure. Des programmes statistiques tels que SPSS vous permettront de calculer l' α de Cronbach ou le κ de Cohen. Si cela ne vous convient pas, vous pouvez toujours calculer et évaluer une corrélation moitié-moitié.

Ensuite vous devez analyser chaque variable importante séparément. (Cette étape n'est évidemment pas nécessaire pour les variables indépendantes manipulées, car c'est vous, en tant que personne responsable de la recherche, qui avez déterminé la distribution). Réalisez des graphiques à barres pour chacune d'entre elles, notez leur forme et calculez les mesures courantes de la tendance centrale et de la variabilité. Assurez-vous de bien *comprendre* la moyenne de ces statistiques en fonction des variables qui vous intéressent. Par exemple une distribution d'échelles d'évaluation du bonheur sur une échelle de 1 à 10 points peut être unimodale et déséquilibrée, avec une moyenne de 8.25 et un écart-type de 1.14. Mais cela *signifie* que la plupart des personnes participantes se situent à un niveau assez élevé sur l'échelle du bonheur et qu'un petit nombre d'entre eux se situent à un niveau nettement inférieur.

Maintenant le moment est venu d'identifier les valeurs aberrantes, de les examiner de plus près et de décider de ce qu'il convient de faire. Il se peut que vous découvriez que ce qui semble à première vue être une valeur aberrante est le résultat d'une erreur de saisie d'une réponse dans le fichier de données, auquel cas il vous suffit de corriger le fichier de données et de passer à autre chose. Vous pouvez également soupçonner qu'une valeur aberrante représente un autre type d'erreur, d'incompréhension ou de manque d'effort de la part d'une personne participante. Par exemple dans une distribution des temps de réaction dans laquelle la plupart des personnes participantes n'ont pris que quelques secondes à répondre, une personne qui a pris 3 minutes à répondre serait une valeur aberrante. Il est probable que celle-ci n'ait pas compris la tâche (ou du moins que son niveau d'attention n'était pas suffisant). L'inclusion de son temps de réaction aurait un impact important sur la moyenne et l'écart-type de l'échantillon. Dans de telles situations, il peut être justifié d'exclure la réponse ou la personne participante aberrante des analyses. Dans ce cas il convient toutefois de noter quelles réponses ou quelles personnes participantes ont été exclues et pourquoi, et d'appliquer ces mêmes critères de manière cohérente à toutes les réponses et à toutes les personnes participantes. Lorsque vous présentez vos résultats, vous devez indiquer le nombre de réponses ou de personnes ayant participé que vous avez exclu et les critères spécifiques que vous avez utilisés. Encore une fois ne jetez pas littéralement ou ne supprimez pas les données que vous avez choisi d'exclure. Mettez-les simplement de côté parce que vous ou une autre personne responsable d'une recherche pourriez vouloir les consulter plus tard.

Gardons à l'esprit que les valeurs aberrantes ne représentent pas nécessairement une erreur, un malentendu ou un manque d'effort. Elles peuvent représenter des réponses ou des participantes et des participants vraiment extrêmes. Par exemple dans un grand échantillon d'une cohorte étudiante universitaire, la grande majorité a déclaré avoir eu moins de 15 partenaires sexuelles, mais il y avait aussi quelques scores aberrants de 60 ou 70 (Brown et Sinclair, 1999). Bien que ces scores puissent représenter des erreurs, des malentendus ou même des exagérations intentionnelles, il est également plausible qu'ils représentent des estimations honnêtes, et même exactes. Une stratégie consiste à utiliser la médiane et d'autres statistiques qui ne sont pas fortement affectées par les valeurs aberrantes. Une autre stratégie consiste à analyser les données en incluant et en excluant les valeurs aberrantes. Si les résultats sont essentiellement les mêmes, ce qui est souvent le cas, il est logique de laisser les valeurs aberrantes. Si les résultats diffèrent selon que les valeurs aberrantes sont incluses ou exclues, les deux analyses peuvent être rapportées et les différences entre elles discutées.

Analyses planifiées et exploratoires

Enfin il est possible de répondre à vos principales questions de recherche. Lorsque vous avez conçu votre étude, vous avez peut-être émis l'hypothèse qu'une relation particulière pourrait exister dans les données. Dans ce cas vous effectuerez une **analyse planifiée** afin de tester une relation prédite par votre hypothèse. Par exemple si vous vous attendiez à une différence entre les moyennes des groupes ou des conditions, vous pouvez calculer les moyennes et les écarts-types des groupes ou des conditions concernés, créer un graphique à barres pour illustrer les résultats et calculer le *d* de Cohen. Si vous vous attendiez à une corrélation entre des variables quantitatives, vous pouvez créer un graphique linéaire ou un diagramme de dispersion (assurez-vous de vérifier la non-linéarité et la restriction de l'étendue).

Une fois que vous avez réalisé vos analyses planifiées, vous pouvez passer à l'examen de la possibilité qu'il existe, dans vos données, des relations que vous n'aviez pas formulées dans vos hypothèses. Il s'agit alors d'une **analyse exploratoire**, c'est-à-dire d'une analyse que vous entreprenez sans hypothèse préalable. Ces analyses vous permettront d'explorer vos données, à la recherche d'autres résultats intéressants qui pourraient servir de base à de futures recherches (et de matière pour la section discussion de votre document). Daryl Bem (2003) suggère d'

[e]xaminer [vos données] sous tous les angles. Analysez les sexes séparément. Créez de nouveaux indices composites. Si une donnée suggère une nouvelle hypothèse, essayez de trouver ailleurs dans les données des preuves supplémentaires à son appui. Si vous voyez de faibles indices de patrons intéressants, essayez de réorganiser les données pour les mettre davantage en évidence. Si vous n'aimez pas certaines personnes participantes, ou des essais, ou des personnes chargées d'observation ou d'entrevues qui vous ont donné des résultats anormaux, écartez-les (temporairement). Partez pour une expédition de pêche à la recherche de quelque chose — n'importe quoi — d'intéressant (Bem, 2003, p. 186-187, traduction).

Il est important de distinguer les analyses planifiées des analyses exploratoires lors de la rédaction des sections résultats et discussion de votre rapport. En effet les ensembles de données complexes sont susceptibles d'inclure des « patrons » qui se produisent entièrement par hasard, et chaque fois que vous effectuez une autre analyse non planifiée de ces données, vous augmentez la probabilité que ces patrons dus au hasard semblent être de vrais patrons de résultats, ce que l'on appelle une erreur de « type I » (voir le chapitre sur les statistiques inférentielles). Ainsi les résultats découverts dans le cadre d'analyses exploratoires (ce que Bem appelle une « expédition de pêche ») doivent être considérés avec scepticisme et reproduits dans au moins une nouvelle étude avant d'être présentés. Toutefois si vous trouvez des relations intéressantes auxquelles vous ne vous attendiez pas dans les données, expliquez qu'elles pourraient mériter des recherches supplémentaires.

Comprendre les statistiques descriptives

Dans le prochain chapitre, nous examinerons les statistiques inférentielles — un ensemble de techniques permettant de déterminer si les résultats obtenus auprès de votre échantillon sont susceptibles de s'appliquer à la population. Bien que les statistiques inférentielles soient importantes pour des raisons qui seront expliquées sous peu, les personnes débutant en recherche oublient parfois que leurs statistiques descriptives indiquent réellement « ce qui s'est passé » dans leur étude. Par exemple imaginons qu'un groupe traitement de 50 personnes participantes a un score moyen de 34.32 ($ÉT = 10.45$), tandis qu'un groupe témoin de 50 personnes a un score moyen de 21.45 ($ÉT = 9.22$), et que le d de Cohen soit extrêmement élevé (1.31). Bien que la réalisation et la présentation de statistiques inférentielles (comme un test t) constituent sans aucun doute une partie obligatoire dans tout rapport formel sur cette étude, les statistiques descriptives devraient à elles seules indiquer clairement que le traitement a eu un effet. Imaginons également qu'un diagramme de dispersion montre un « nuage » de points indistinct et que le r de Pearson soit négligeable à -0.02. Là encore bien que la réalisation et la présentation de statistiques inférentielles soient requises dans tout rapport formel sur cette étude, les statistiques descriptives devraient à elles seules indiquer clairement que les variables ne sont essentiellement pas liées entre elles. L'essentiel est donc de toujours vous assurer de bien comprendre vos résultats au niveau descriptif d'abord, puis de passer ensuite aux statistiques inférentielles.

67. Points clés et exercices

Points clés

- Chaque variable a une distribution — une façon dont les scores se répartissent entre les différents niveaux. La distribution peut être décrite à l'aide d'un tableau de fréquences et d'un histogramme. Elle peut également être décrite avec des mots en matière de forme, notamment si elle est unimodale ou bimodale, et si elle est symétrique ou asymétrique.
- La tendance centrale, le point milieu d'une distribution, peut être décrite précisément à l'aide de trois statistiques — la moyenne, la médiane et le mode. La moyenne est la somme des scores divisée par le nombre de scores, la médiane est le score du milieu, et le mode est le score le plus fréquent.
- La variabilité ou la dispersion d'une distribution peut être décrite précisément à l'aide de l'étendue et de l'écart-type. L'étendue est la différence entre le score le plus élevé et le plus bas, et l'écart-type est la quantité moyenne par laquelle les scores diffèrent de la moyenne.
- La position d'un score à l'intérieur de sa distribution peut être décrite à l'aide des rangs de percentile ou des scores z . Le rang percentile d'un score est le pourcentage de scores inférieurs à celui-ci, et le score z est la différence entre le score et la moyenne divisée par l'écart-type.
- Les différences entre les groupes ou les conditions sont généralement décrites à l'aide de moyenne et d'écart-type des groupes ou des conditions, ou encore à l'aide du d de Cohen, et présentées dans des graphiques à barres.
- Le d de Cohen est une mesure de la force de la relation (ou de la taille de l'effet) des différences entre les moyennes de deux groupes ou conditions. Il correspond à la différence des moyennes divisée par l'écart-type. En général des valeurs de ± 0.20 , ± 0.50 et ± 0.80 peuvent être considérées comme respectivement faibles, moyennes et grandes.
- Les corrélations entre les variables quantitatives sont généralement décrites à l'aide du r de Pearson et présentées dans des graphiques linéaires ou des diagrammes de dispersion.
- Le r de Pearson est une mesure de la force de la relation (ou de la taille de l'effet) entre des variables quantitatives. Il correspond au produit moyen croisé des deux ensembles de scores z . En général des valeurs de ± 0.10 , ± 0.30 et ± 0.50 peuvent être considérées comme petites, moyennes et fortes, respectivement.
- Dans un article de style APA, les résultats simples sont présentés de la façon la plus efficace dans le texte, tandis que les résultats plus complexes le sont dans des graphiques ou des tableaux.
- Le style APA comprend plusieurs règles pour présenter les résultats numériques dans le texte. Celles-ci incluent d'utiliser des mots uniquement pour les nombres inférieurs à 10 qui ne représentent pas des résultats statistiques précis, d'arrondir les résultats à deux décimales, d'utiliser des mots (par exemple « moyenne ») dans le texte et des symboles (par exemple « M ») entre parenthèses.

- Le style APA comprend aussi plusieurs règles pour présenter les résultats dans les graphiques et tableaux. Les graphiques et les tableaux doivent ajouter des informations plutôt que de les répéter, être aussi simples que possible et pouvoir être interprétés seuls avec une légende descriptive (pour les graphiques) ou un titre descriptif (pour les tableaux).
- Les données brutes doivent être préparées pour l'analyse, en les examinant pour détecter d'éventuelles erreurs, en les organisant et en les entrant dans un tableur.
- Les analyses préliminaires de tout ensemble de données comprennent la vérification de la fidélité des mesures, l'évaluation de l'efficacité des manipulations, l'examen de la distribution des variables individuelles et l'identification des valeurs aberrantes.
- Les valeurs aberrantes qui semblent résulter d'une erreur, d'un malentendu ou d'un manque d'effort peuvent être exclues des analyses. Les critères d'exclusion des réponses ou des personnes participantes doivent être appliqués de la même manière à toutes les données et décrits lors de la présentation des résultats. Les données exclues doivent être mises de côté plutôt que d'être détruites ou supprimées, au cas où elles seraient nécessaires ultérieurement.
- Les statistiques descriptives racontent ce qui s'est produit dans une étude. Bien que les statistiques inférentielles soient aussi importantes, il est essentiel de comprendre d'abord les statistiques descriptives.

Exercices

- Pratique : Faites un tableau de fréquences et un histogramme pour les données suivantes. Rédigez ensuite une brève description de la forme de la distribution en mots.
 - 11, 8, 9, 12, 9, 10, 12, 13, 11, 13, 12, 6, 10, 17, 13, 11, 12, 12, 14, 14
- Pratique : Pour les données de l'exercice 1, calculez la moyenne, la médiane, le mode, l'écart-type et l'étendue.
- Pratique : En utilisant les données des exercices 1 et 2, trouvez
 - les rangs percentile pour les scores de 9 et 14
 - les scores z pour les scores de 8 et 12.
- Pratique : Les données suivantes représentent les scores sur l'échelle d'estime de soi de Rosenberg pour un échantillon de 10 personnes japonaises étudiant à l'université, et de 10 personnes américaines étudiant à l'université. (Bien qu'hypothétiques, ces données sont cohérentes avec les résultats empiriques [Schmitt et Allik, 2005].) Calculez les moyennes et écarts-types des deux groupes, faites un histogramme, calculez le d de Cohen, et décrivez la force de la relation en mots.

Japon	États-Unis
25	27
20	30
24	34
28	37
30	26
32	24
21	28
24	35
20	33
26	36

- Pratique : Les données hypothétiques suivantes représentent les scores d'extraversion et le nombre d'amies et d'amis Facebook de 15 personnes étudiant à l'université. Faites un diagramme de dispersion pour ces données, calculez le r de Pearson et décrivez la relation en mots.

Extraversion	Amis Facebook
8	75
10	315
4	28
6	214
12	176
14	95
10	120
11	150
4	32
13	250
5	99
7	136
8	185
11	88
10	144

- Pratique : Dans une étude classique, des hommes et des femmes ont évalué l'importance de l'attraction physique pour une ou un partenaire à court terme et à long terme (Buss & Schmitt, 1993). Les moyennes et les écarts-types sont les suivants: Hommes / Court terme : $M = 5.67$, $ÉT = 2.34$; Hommes / Long terme : $M = 4.43$, $ÉT = 2.11$; Femmes/ Court terme : $M = 5.67$, $ÉT = 2.48$; Femmes/ Long terme : $M = 4.22$, $ÉT = 1.98$. Présentez ces résultats
 - par écrit
 - dans une figure
 - dans un tableau
- Discussion : Nommez au moins deux façons raisonnables de traiter chacune des valeurs aberrantes suivantes sur la base de la discussion de ce chapitre. (a) Une personne participante estimant la taille de personnes ordinaires indique qu'une femme mesure « 213.36 cm ». (b) Dans une étude sur la mémoire d'objets familiers, une personne participante obtient un score de 0 sur 15. (c) En réponse à une question sur le nombre d'« amies ou amis proches » qu'elle a, une personne participante écrit « 32 ».

68. Références

- American Psychological Association. (2020). Publication manual of the American Psychological Association (7e éd.). <https://doi.org/10.1037/0000165-000>
- American Psychological Association (s.d.). Student Paper Setup: APA style 7e edition. [Guidehttps://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf](https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf)
- Bem, D. J. (2003). Writing the empirical journal article. Dans J. M. Darley, M. P. Zanna et H. L. Roediger III (dir.), *The complete academic: A career guide* (2e éd., p. 185-219). American Psychological Association.
- Brown, N. R. et Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *The Journal of Sex Research*, 36(3), 292-297. <https://doi.org/10.1080/00224499909551999>
- Buss, D. M. et Schmitt, D. P. (1993). Sexual Strategies Theory: An Evolutionary Perspective on Human Mating. *Psychological Review*, 100(2), 204-232. <https://psycnet.apa.org/doi/10.1037/0033-295X.100.2.204>
- Carlson, K. A. et Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. *Journal of Consumer Research*, 38(2), 300-307. <https://doi.org/10.1086/658470>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <https://psycnet.apa.org/doi/10.1037/0033-2909.112.1.155>
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16(5), 259-263. <https://doi.org/10.1111/j.1467-8721.2007.00516.x>
- MacDonald, T. K. et Martineau, A. M. (2002). Self-Esteem, Mood, and Intentions to Use Condoms: When does low self-esteem lead to risky health behaviors ? *Journal of Experimental Social Psychology*, 38(3), 299-306. <https://doi.org/10.1006/jesp.2001.1505>
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A. et Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222-243. <https://doi.org/10.1037/a0017619>
- Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C., Davis 3rd, T. E. et Jarrett, M. A. (2009). One-session treatment of specific phobias in youth: a randomized clinical trial in the United States and Sweden. *Journal of Consulting and Clinical Psychology*, 77(3), 504-516. <https://doi.org/10.1037/a0015158>
- Schmitt, D. P. et Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. *Journal of Personality and Social Psychology*, 89(4), 623-642. <https://doi.org/10.1037/0022-3514.89.4.623>

CHAPITRE 13-LES STATISTIQUES INFÉRENTIELLES



Image 13-Personnes parlant lors d'une célébration

Rappelons que Matthias Mehl et ses collègues, dans leur étude sur les différences de bavardage entre les sexes, ont constaté que les femmes dans leur échantillon parlaient en moyenne 16 215 mots par jour et les hommes, 15 669 mots par jour (Mehl, Vazire, Ramirez-Esparza, Slatcher et Pennebaker, 2007). Malgré cette différence entre les sexes dans leur échantillon, ces scientifiques ont conclu qu'il n'y avait pas de preuve d'une différence de bavardage entre les sexes dans la population. Rappelons également qu'Allen Kanner et ses collègues, dans leur étude sur la relation entre les tracas quotidiens et les symptômes, ont trouvé une corrélation de +0.60 dans leur échantillon (Kanner, Coyne, Schaefer et Lazarus, 1981). Mais ces scientifiques ont conclu que ce résultat signifiait qu'il existait une relation entre les tracas et les symptômes dans la population. Cette affirmation suscite un questionnement sur la manière dont les scientifiques peuvent dire si le résultat de leur échantillon reflète quelque chose qui est vrai pour la population.

La réponse à cette question est que les scientifiques utilisent un ensemble de techniques appelées statistiques inférentielles, qui font l'objet du présent chapitre. Nous nous concentrons en particulier sur

le test de l'hypothèse nulle, l'approche la plus courante des statistiques inférentielles dans la recherche en psychologie. Nous commençons par un aperçu conceptuel du test de l'hypothèse nulle, y compris son objectif et sa logique de base. Nous examinons ensuite plusieurs techniques de test de l'hypothèse nulle permettant de tirer des conclusions sur les différences entre les moyennes et sur les corrélations entre les variables quantitatives. Enfin nous examinerons quelques autres idées importantes liées au test de l'hypothèse nulle, dont certaines peuvent être utiles pour planifier de nouvelles études et interpréter les résultats. Nous nous pencherons également sur certaines critiques formulées à l'encontre des tests de l'hypothèse nulle et sur certaines façons de répondre à ces critiques.

69. Comprendre le test de l'hypothèse nulle

Objectifs d'apprentissage

1. Expliquer l'objectif du test de l'hypothèse nulle, y compris le rôle de l'erreur d'échantillonnage.
2. Décrire la logique fondamentale du test de l'hypothèse nulle.
3. Décrire le rôle de la force de la relation et de la taille de l'échantillon pour déterminer la signification statistique, et porter des jugements raisonnables sur la signification statistique en fonction de ces deux facteurs.

L'objectif du test de l'hypothèse nulle

Comme nous l'avons vu, la recherche en psychologie consiste généralement à mesurer une ou plusieurs variables dans un échantillon et à calculer des données descriptives sommaires (par exemple des moyennes, des coefficients de corrélation) pour ces variables. Ces données descriptives de l'échantillon sont appelées **statistiques**. Toutefois en général l'objectif de la personne conduisant une recherche n'est pas de tirer des conclusions à propos de leur échantillon, mais bien de tirer des conclusions sur la population d'où cet échantillon provient. Ainsi il faut utiliser les statistiques de l'échantillon pour tirer des conclusions sur les valeurs correspondantes dans la population. Ces valeurs correspondantes dans la population sont appelées **paramètres**. Imaginons par exemple qu'une personne responsable d'une recherche mesure le nombre de symptômes dépressifs rapportés par 50 adultes souffrant de dépression clinique et calcule le nombre moyen de symptômes. Cette personne voudra probablement utiliser cette statistique d'échantillon (le nombre moyen de symptômes pour l'échantillon) afin de tirer des conclusions sur le paramètre correspondant dans la population (le nombre moyen de symptômes chez l'ensemble des adultes présentant une dépression clinique).

Malheureusement les statistiques d'échantillon ne sont pas des estimations parfaites des paramètres de population correspondants. En effet toute statistique présente une certaine variabilité aléatoire d'un échantillon à l'autre. Le nombre moyen de symptômes dépressifs peut être de 8.73 dans un échantillon d'adultes présentant une dépression clinique, de 6.45 dans un deuxième échantillon et de 9.44 dans un troisième — et ce, même si tous ces échantillons ont été sélectionnés au hasard dans la même population. De même la corrélation (r de Pearson) entre deux variables peut être de +0.24 dans un échantillon, de -0.04 dans un deuxième échantillon et de +0.15 dans un troisième — là encore bien que ces échantillons aient été

sélectionnés au hasard de la même population. Cette variabilité aléatoire d'une statistique d'échantillon à l'autre est appelée **erreur d'échantillonnage**. (Il est important de noter que le mot *erreur* désigne la variabilité aléatoire entre échantillons et ne signifie pas qu'une faute a été commise. On ne peut pas *faire* une « erreur d'échantillonnage »).

Une des conséquences importantes est que, lorsqu'il existe une relation statistique dans un échantillon, il n'est pas toujours certain qu'il existe une relation statistique dans la population. Une petite différence entre deux moyennes de groupes dans un échantillon peut indiquer qu'il existe une petite différence entre les moyennes de population. Mais il se peut aussi qu'il n'y existe aucune différence dans la population et que la différence observée dans l'échantillon ne soit qu'une question d'erreur d'échantillonnage. De même une valeur de r de Pearson de -0.29 dans un échantillon peut signifier qu'il existe une relation négative dans la population. Mais il se peut aussi qu'il n'y ait pas de relation dans la population et que la relation dans l'échantillon ne soit qu'une question d'erreur d'échantillonnage.

En fait toute relation statistique dans un échantillon peut être interprétée de deux manières :

- Il existe une relation dans la population, et la relation dans l'échantillon en est le reflet.
- Il n'existe pas de relation dans la population, et la relation dans l'échantillon ne reflète que l'erreur d'échantillonnage.

Le but du test de l'hypothèse nulle est simplement d'aider les scientifiques à choisir entre ces deux interprétations.

La logique du test de l'hypothèse nulle

Le **test de l'hypothèse nulle** (souvent appelé test de signification de l'hypothèse nulle ou NHST, pour *Null Hypothesis Significance Testing*) est une démarche formelle qui vise à décider entre deux interprétations possibles d'une relation statistique observée dans un échantillon. L'une des interprétations est appelée **hypothèse nulle** (souvent symbolisée par H_0 et lue comme « H-zéro »). Elle correspond à l'idée qu'il n'existe aucune relation dans la population et que la relation dans l'échantillon reflète uniquement l'erreur d'échantillonnage. De manière informelle l'hypothèse nulle est que la relation dans l'échantillon « est due au hasard ». L'autre interprétation est appelée l'**hypothèse alternative** (souvent symbolisée par H_1). Elle correspond à l'idée qu'il existe une relation dans la population, et que la relation dans l'échantillon reflète cette relation dans la population.

Ainsi toute relation statistique dans un échantillon peut être interprétée de l'une ou l'autre de ces deux manières : elle peut résulter du hasard ou refléter une véritable relation dans la population. Les scientifiques ont donc besoin d'un moyen de trancher entre les deux. Bien qu'il existe de nombreuses techniques spécifiques de test de l'hypothèse nulle, elles reposent toutes sur la même logique générale. Les étapes sont les suivantes :

- Supposer dans un premier temps que l'hypothèse nulle est vraie. Il n'existe pas de relation entre les variables dans la population.
- Déterminer quelle est la probabilité d'obtenir la relation observée entre les variables dans l'échantillon

si l'hypothèse nulle est vraie.

- Si cette relation observée dans l'échantillon est extrêmement improbable, alors il faut **rejeter l'hypothèse nulle** en faveur de l'hypothèse alternative. Si elle n'est pas extrêmement improbable, il faut **retenir l'hypothèse nulle**.

En suivant cette logique, nous pouvons commencer à comprendre pourquoi Mehl et ses collègues ont conclu qu'il n'existait aucune différence de loquacité entre les femmes et les hommes dans la population. Essentiellement ils ont posé la question suivante : « S'il n'y a aucune différence dans la population, quelle est la probabilité d'obtenir une petite différence de $d = 0,06$ dans notre échantillon ? » Leur réponse à cette question était qu'une telle relation observée dans l'échantillon serait assez probable si l'hypothèse nulle était vraie. Par conséquent ils ont retenu l'hypothèse nulle — concluant qu'il n'existe pas de preuve d'une différence entre les sexes dans la population. De la même manière, nous pouvons comprendre pourquoi Kanner et ses collègues ont conclu qu'il existe une corrélation entre les tracasseries et les symptômes dans la population. Ils ont posé la question suivante : « Si l'hypothèse nulle est vraie, quelle est la probabilité que nous trouvions une forte corrélation de $+0,60$ dans notre échantillon ? » Leur réponse : une relation dans l'échantillon serait très improbable si l'hypothèse nulle était vraie. Ces scientifiques ont donc rejeté l'hypothèse nulle en faveur de l'hypothèse alternative — concluant qu'il existe une corrélation positive entre ces variables dans la population.

Une étape cruciale du test de l'hypothèse nulle consiste à déterminer la probabilité d'obtenir le résultat observé dans l'échantillon, ou un résultat encore plus extrême, si l'hypothèse nulle était vraie (Lakens, 2017). Cette probabilité est appelée la **valeur p** . Une valeur p faible signifie que le résultat observé ou un résultat plus extrême seraient improbables si l'hypothèse nulle était vraie, ce qui conduit à rejeter l'hypothèse nulle. Une valeur p qui n'est pas faible signifie que le résultat observé ou un résultat plus extrême seraient vraisemblables si l'hypothèse nulle était vraie, ce qui conduit à retenir l'hypothèse nulle. Mais jusqu'à quel seuil de valeur p peut-on considérer qu'un résultat d'échantillon est suffisamment improbable pour rejeter l'hypothèse nulle ? Dans les tests de l'hypothèse nulle, ce seuil est appelé **α (alpha)** et est presque toujours fixé à 0.05. Autrement dit si la probabilité d'obtenir un résultat au moins aussi extrême que celui observé est de 5 % ou moins sous l'hypothèse nulle, alors on rejette l'hypothèse nulle. Dans ce cas, on dit que le résultat est **statistiquement significatif**.

Si sous l'hypothèse nulle, la probabilité d'obtenir un résultat d'échantillon aussi extrême que celui observé est supérieure à 5 %, l'hypothèse nulle est retenue. Cela ne signifie pas que les scientifiques acceptent l'hypothèse nulle comme vraie, mais simplement qu'il n'existe pas *pour l'instant* suffisamment de preuves pour la rejeter. Les scientifiques utilisent souvent l'expression « ne pas rejeter l'hypothèse nulle » plutôt que « retenir l'hypothèse nulle » et n'emploient jamais l'expression « accepter l'hypothèse nulle ».

La valeur p mal comprise

La valeur p est l'une des quantités les plus mal comprises dans la recherche en psychologie (Cohen, 1994). Même des scientifiques d'expérience l'interprètent parfois de manière erronée, et il n'est pas rare que de telles confusions apparaissent jusque dans les manuels de statistiques !

L'erreur d'interprétation la plus courante consiste à croire que la valeur p représente la probabilité que l'hypothèse nulle soit vraie — c'est-à-dire que le résultat de l'échantillon soit uniquement dû au hasard. Par exemple des personnes responsables de la recherche pourraient affirmer à tort que puisque la valeur p est de 0,02, il n'y a que 2 % de chances que le résultat soit dû au hasard et 98 % de chances qu'il reflète une relation réelle au sein de la population. Mais cette affirmation est incorrecte. En réalité la valeur p correspond à la probabilité d'obtenir un résultat au moins aussi extrême que celui observé dans l'échantillon si l'hypothèse nulle était vraie. Ainsi une valeur p de 0,02 signifie que si l'hypothèse nulle était vraie, un résultat d'échantillon aussi extrême que celui observé ne se produirait que dans 2 % des cas.

Pour éviter ce malentendu, il faut se rappeler que la valeur p n'est pas la probabilité qu'une hypothèse particulière soit vraie ou fausse. Elle correspond plutôt à la probabilité d'obtenir le résultat observé dans l'échantillon si l'hypothèse nulle était vraie.



"Null Hypothesis" repéré à http://imgs.xkcd.com/comics/null_hypothesis.png (CC-BY-NC 2.5).
[\[Description de l'image\]](#)

Rôle de la taille de l'échantillon et de la force de la relation

Rappelons que le test d'hypothèse nulle consiste à répondre à la question suivante : « Si l'hypothèse nulle est vraie, quelle est la probabilité d'obtenir un résultat d'échantillon aussi extrême que celui-ci ? » En d'autres mots, « Quelle est la valeur p ? » Il peut être utile de comprendre que la réponse à cette question dépend uniquement de deux facteurs : la force de la relation et la taille de l'échantillon. Plus la relation observée dans l'échantillon est forte et plus l'échantillon est grand, moins il est probable que ce résultat survienne si l'hypothèse nulle était vraie. En d'autres mots, plus la valeur p est faible. Cela est logique. Imaginons une étude où un échantillon de 500 femmes est comparé à un échantillon de 500 hommes sur une caractéristique psychologique, et où le coefficient d de Cohen est élevé (0,50). S'il n'y avait aucune

différence entre les sexes dans la population, un résultat aussi marqué obtenu avec un si grand échantillon semblerait hautement improbable. Imaginons maintenant une étude similaire avec seulement trois femmes et trois hommes, et où le coefficient d de Cohen est faible (0,10). S'il n'existait aucune différence entre les sexes dans la population, un résultat aussi faible basé sur un si petit échantillon semblerait plausible. C'est précisément pour cette raison que l'hypothèse nulle serait rejetée dans le premier exemple et retenue dans le second.

Bien sûr il arrive parfois que le résultat soit faible, mais que l'échantillon soit grand, ou que le résultat soit fort, mais que l'échantillon soit petit. Dans ces situations, les deux facteurs se compensent de sorte qu'un résultat faible peut être statistiquement significatif si l'échantillon est suffisamment grand et qu'une relation forte peut être statistiquement significative même si l'échantillon est petit. Le Tableau 13.1 montre de manière approximative comment la force de la relation et la taille de l'échantillon se combinent pour déterminer si un résultat d'échantillon est statistiquement significatif. Les colonnes du tableau représentent trois niveaux de force de la relation : faible, moyenne et forte. Les rangées représentent quatre tailles d'échantillon qui peuvent être considérées comme petites, moyennes, grandes et très grandes dans le contexte de la recherche en psychologie. Ainsi chaque cellule du tableau représente une combinaison de la force de la relation et de la taille de l'échantillon. Si une cellule contient le mot « Oui », cette combinaison est statistiquement significative à la fois pour le d de Cohen et le r de Pearson. Si elle contient le mot « Non », elle n'est pas statistiquement significative ni pour l'un ni pour l'autre. Il existe une cellule où la décision diffère entre d et r , ainsi qu'une autre où elle peut varier selon certaines considérations supplémentaires qui sont discutées dans la section 13.2 « Quelques tests de base de l'hypothèse nulle ».

Tableau 13.1 Comment la force de la relation et la taille de l'échantillon se combinent pour déterminer si un résultat est statistiquement significatif

Taille de l'échantillon	Force de la relation		
	Faible	Moyenne	Forte
Petit ($N = 20$)	Non	Non	$d =$ Peut-être $r =$ Oui
Moyen ($N = 50$)	Non	Oui	Oui
Large ($N = 100$)	$d =$ Oui $r =$ Non	Oui	Oui
Très large ($N = 500$)	Oui	Oui	Oui

Bien que le Tableau 13.1 ne fournisse qu'une ligne directrice approximative, il montre très clairement que les relations faibles issues d'échantillons petits ou moyens ne sont jamais statistiquement significatives, et que les relations fortes issues d'échantillons moyens ou grands sont toujours statistiquement significatives.

Si vous gardez cette idée à l'esprit, vous pourrez souvent déterminer si un résultat est statistiquement significatif en vous appuyant uniquement sur les statistiques descriptives. Développer ce type de jugement intuitif est extrêmement utile. D'une part, cela vous permet d'anticiper les résultats de vos tests formels d'hypothèse nulle, et de repérer les problèmes dans vos analyses. Par exemple si la relation observée dans l'échantillon est forte et que l'échantillon est de taille moyenne, on s'attend à ce que l'hypothèse nulle soit rejetée. Si pour une raison quelconque, le test formel de l'hypothèse nulle indique le contraire, vous devrez alors vérifier à nouveau vos calculs et vos interprétations. D'autre part, la capacité à porter ce type de jugement intuitif montre que vous comprenez la logique fondamentale de cette approche, en plus d'être en mesure de réaliser correctement les calculs.

Signification statistique par rapport à la signification pratique

Le Tableau 13.1 illustre un autre point extrêmement important. Un résultat statistiquement significatif n'est pas nécessairement un résultat fort. Même un résultat très faible peut être statistiquement significatif s'il repose sur un échantillon suffisamment grand. Cela est étroitement lié à l'argument de Janet Shibley Hyde concernant les différences entre les sexes (Hyde, 2007). Les différences entre les femmes et les hommes en matière de résolution de problèmes mathématiques et d'aptitude au leadership sont statistiquement significatives. Mais le mot « significatif » peut amener les gens à interpréter ces différences comme fortes et importantes — peut-être même suffisamment importantes pour influencer le choix des cours universitaires ou encore leurs décisions de vote. Or, comme nous l'avons vu, ces différences statistiquement significatives sont en réalité assez faibles, voire « insignifiantes ».

C'est pourquoi il est important de distinguer la signification *statistique* d'un résultat de sa signification *pratique*. La **signification pratique** renvoie à l'importance ou à l'utilité du résultat dans un contexte réel. De nombreuses différences entre les sexes sont statistiquement significatives — et peuvent même présenter un intérêt scientifiques —, mais elles ne sont pas significatives d'un point de vue pratique. En pratique clinique, ce même concept est souvent désigné sous le terme « signification clinique ». Par exemple une étude sur un nouveau traitement de la phobie sociale pourrait montrer qu'il produit un effet positif statistiquement significatif. Pourtant cet effet pourrait ne pas être suffisamment fort pour justifier le temps, les efforts et les autres coûts nécessaires à sa mise en œuvre — surtout si des traitements plus simples et moins coûteux, mais presque aussi efficaces existent déjà. Bien que statistiquement significatif, un tel résultat serait considéré comme dépourvu de signification pratique ou clinique.



THE ANNUAL DEATH RATE AMONG PEOPLE
WHO KNOW THAT STATISTIC IS ONE IN SIX.

"Conditional Risk" repéré au http://imgs.xkcd.com/comics/conditional_risk.png
(CC-BY-NC 2.5). [Description de l'image](#)

Description de l'image

Longue description « Hypothèse nulle » : Une bande dessinée représentant un homme et une femme discutant au premier plan. À l'arrière-plan, un enfant travaille à un bureau. L'homme dit à la femme : « Je ne peux pas croire que les écoles enseignent encore l'hypothèse nulle aux enfants. Je me souviens avoir lu une grande étude qui l'a réfutée de manière concluante il y a des années. » [\[Retour à l'«Hypothèse nulle»\]](#)

Longue description « Risque conditionnel » : Bande dessinée représentant deux randonneurs à côté d'un arbre pendant un orage. Un éclair se déchaîne dans le ciel noir tandis que le tonnerre gronde. L'un des randonneurs dit : « Whoa ! Nous devrions rentrer à l'intérieur ! » L'autre randonneur répond : « Ce n'est pas grave ! La foudre ne tue qu'environ 45 Américains par an, les chances de mourir ne sont donc que d'une sur 7 000 000. Continuons ! » La légende de la bande dessinée indique : « Le taux de mortalité annuel chez les personnes qui connaissent cette statistique est d'un sur six. » [\[Retour au « Risque conditionnel »\]](#)

Attributions des médias

- [Null Hypothesis](#) by XKCD [CC BY-NC \(Attribution NonCommercial\)](#)
- [Conditional Risk](#) by XKCD [CC BY-NC \(Attribution NonCommercial\)](#)

70. Quelques tests de base de l'hypothèse nulle

Objectifs d'apprentissage

1. Réaliser et interpréter des tests t à un échantillon, à échantillons dépendants et à échantillons indépendants.
2. Interpréter les résultats des ANOVAs à un facteur, à mesures répétées et à plan factoriel.
3. Effectuer et interpréter les tests de l'hypothèse nulle du r de Pearson.

Dans cette section, nous examinons plusieurs procédures courantes de test de l'hypothèse nulle. L'accent est mis ici sur la présentation d'informations suffisantes pour vous permettre de réaliser et d'interpréter les versions les plus élémentaires. Dans la plupart des cas, les outils d'analyse statistique en ligne mentionnés au chapitre 12 prendront en charge les calculs — tout comme des programmes tels que Microsoft Excel, JASP et SPSS.

Le test t

Comme nous l'avons vu tout au long de ce livre, de nombreuses études en psychologie portent sur la différence entre deux moyennes. Le test d'hypothèse nulle le plus courant pour ce type de relation statistique est le **test- t** . Dans cette section, nous examinerons trois types de tests t , chacun étant utilisé pour des devis de recherche légèrement différents : le test t à un échantillon, le test t à échantillons dépendants et le test t à échantillons indépendants. Vous avez peut-être déjà suivi un cours de statistiques, mais nous allons réviser ici vos connaissances de base.

Le test t à un échantillon

Le **test t à un échantillon** est utilisé pour comparer une moyenne d'échantillon (M) avec une moyenne hypothétique d'une population (μ_0) qui constitue un point de comparaison pertinent. L'hypothèse nulle stipule que la moyenne de la population (μ) est égale à la moyenne hypothétique de la population : $\mu = \mu_0$. L'hypothèse alternative est que la moyenne de la population est différente de la moyenne hypothétique de la population : $\mu \neq \mu_0$. Pour décider entre ces deux hypothèses, nous devons trouver la probabilité d'obtenir

la moyenne d'échantillon (ou une valeur plus extrême) si l'hypothèse nulle était vraie. Mais pour trouver cette valeur p , il faut d'abord calculer une statistique de test appelé t . (Une **statistique de test** est une statistique qui n'est calculée que pour aider à trouver la valeur p). La formule de t est la suivante :

$$t = \frac{M - \mu_0}{\left(\frac{SD}{\sqrt{N}} \right)}$$

Encore une fois M est la moyenne de l'échantillon et μ_0 est la moyenne hypothétique de la population qui nous intéresse. Le symbole $ÉT$ est l'écart-type de l'échantillon et N est la taille de l'échantillon.

La statistique t (comme toute statistique de test) est utile parce que nous savons comment elle se distribue lorsque l'hypothèse nulle est vraie. Comme le montre la Figure 13.1, cette distribution est unimodale, symétrique, et a une moyenne égale à 0. Sa forme exacte dépend d'un concept statistique appelé degrés de liberté qui, pour un test t à un échantillon, est égal à $N - 1$. (La distribution illustrée à la Figure 13.1 a 24 degrés de liberté.) Le point important est que connaître cette distribution permet de déterminer la valeur p pour n'importe quel score t . Prenons par exemple un score t de 1.50 calculé avec un échantillon de 25 personnes. La probabilité d'obtenir un score t au moins aussi extrême correspond à la proportion des scores t dans la distribution dont la valeur est au moins aussi extrême. Pour l'instant, définissons *extrême* comme éloigné de zéro dans un sens ou dans l'autre. Ainsi la valeur p est la proportion de scores t qui sont égaux ou supérieurs à 1.50 *ou* qui sont égaux ou inférieurs à -1.50 — une valeur qui s'avère être 0.14.

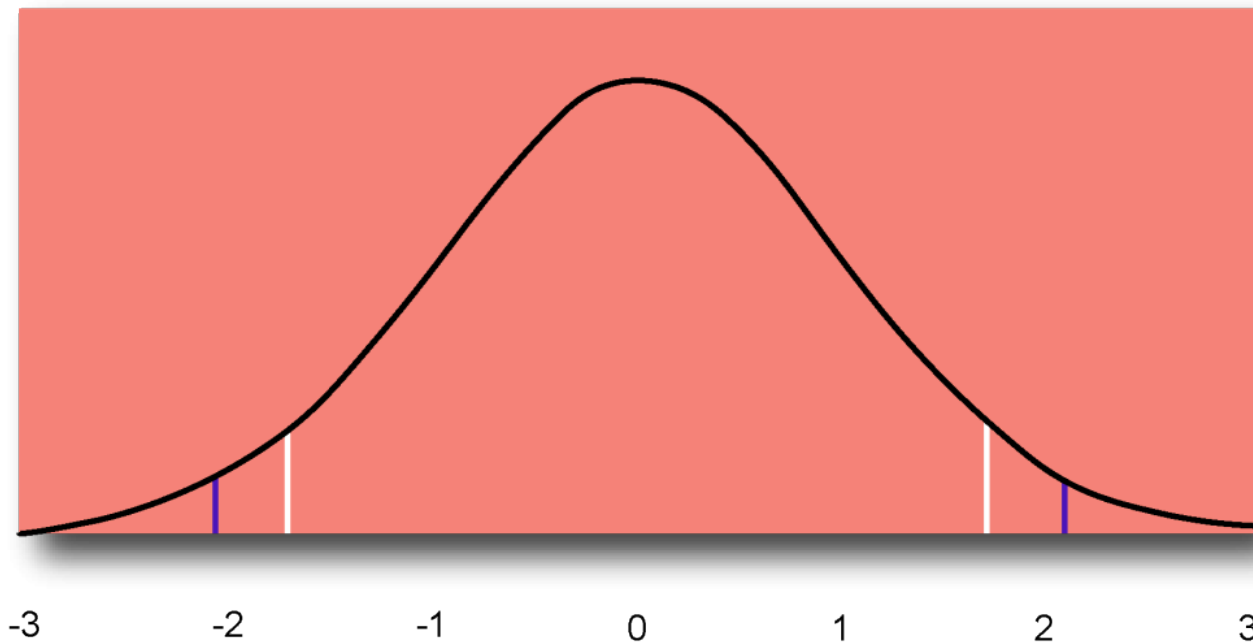


Figure 13.1-Distribution des scores t (avec 24 degrés de liberté) lorsque l'hypothèse nulle est vraie. Les lignes verticales bleues représentent les valeurs critiques bilatérales et les lignes verticales blanches, les valeurs critiques unilatérales avec $\alpha = 0,05$.

Heureusement nous n'avons pas à nous préoccuper directement de la distribution des scores t . Si nous entrons les données de notre échantillon et la moyenne hypothétique d'intérêt dans l'un des outils

statistiques en ligne du chapitre 12 ou dans un programme comme SPSS ou JASP (Excel n'a pas de fonction de test t à un échantillon), le résultat inclura à la fois le score t et la valeur p . À ce stade, le reste de la procédure est simple. Si p est égal ou inférieur à 0.05, nous rejetons l'hypothèse nulle et concluons que la moyenne de la population diffère de la moyenne hypothétique d'intérêt. Si p est supérieur à 0.05, nous retenons l'hypothèse nulle et concluons qu'il n'y a pas suffisamment de preuves pour affirmer que la moyenne de la population diffère de la moyenne hypothétique d'intérêt. (Encore une fois techniquement cela signifie seulement que nous n'avons pas assez de preuves pour conclure qu'il y a une différence.)

Si nous calculons le score t à la main, nous pouvons utiliser un tableau comme le Tableau 13.2 pour prendre la décision. Ce tableau ne fournit pas de valeurs p exactes. Il fournit plutôt les **valeurs critiques** de t pour différents degrés de liberté (df) lorsque α est égal à 0.05. Pour l'instant, concentrons-nous sur les valeurs critiques bilatérales de la dernière colonne du tableau. Chacune de ces valeurs doit être interprétée comme une paire de valeurs : une positive et une négative. Par exemple pour 24 degrés de liberté, les valeurs critiques sont +2.064 et -2.064. Elles sont représentées par les lignes verticales bleues de la Figure 13.1. L'idée est que tout score t inférieur à la valeur critique inférieure (ligne bleue à gauche dans la Figure 13.1) se situe dans les 2.5 % les plus faibles de la distribution, tandis que tout score t supérieur à la valeur critique supérieure (ligne bleue à droite) se situe dans les 2.5 % les plus élevés de la distribution. Par conséquent tout score t au-delà de la valeur critique dans l'une ou l'autre *direction*, se situe dans les 5 % les plus extrêmes des scores t lorsque l'hypothèse nulle est vraie, et correspond à une valeur p inférieure à 0,05. Ainsi si le score t que nous calculons est supérieur à la valeur critique dans l'une ou l'autre direction, nous rejetons l'hypothèse nulle. Si le score t que nous calculons se situe entre les valeurs critiques supérieure et inférieure, nous retenons l'hypothèse nulle.

Tableau 13.2 *Tableau des valeurs critiques de t avec $\alpha = .05$*

<i>dl</i>	Valeur critique	
	Unilatérale	Bilatérale
3	2.353	3.182
4	2.132	2.776
5	2.015	2.571
6	1.943	2.447
7	1.895	2.365
8	1.860	2.306
9	1.833	2.262
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080
22	1.717	2.074
23	1.714	2.069
24	1.711	2.064
25	1.708	2.060
30	1.697	2.042
35	1.690	2.030
40	1.684	2.021
45	1.679	2.014
50	1.676	2.009
60	1.671	2.000
70	1.667	1.994
80	1.664	1.990
90	1.662	1.987
100	1.660	1.984

Jusqu'à présent nous avons examiné ce que l'on appelle un **test bilatéral**, dans lequel nous rejetons l'hypothèse nulle si le score t de l'échantillon est extrême dans l'une ou l'autre direction. Ce test est pertinent lorsque nous pensons que la moyenne de l'échantillon peut différer de la moyenne hypothétique de la population, sans que nous ayons de raison valable de nous attendre à ce que la différence aille dans une direction particulière. Il est toutefois possible d'effectuer un **test unilatéral**, dans lequel nous rejetons l'hypothèse nulle uniquement si le score t de l'échantillon est extrême dans une direction précisée avant la collecte des données. Ce test est pertinent lorsque nous avons de bonnes raisons de penser que la moyenne de l'échantillon différera de la moyenne hypothétique de la population dans une direction particulière.

Voici comment cela fonctionne. Chaque valeur critique unilatérale du Tableau 13.2 peut à nouveau être interprétée comme une paire de valeurs : une positive et une négative. Un score t inférieur à la valeur critique inférieure se situe dans les 5 % inférieurs de la distribution, et un score t supérieur à la valeur critique supérieure se situe dans les 5 % supérieurs de la distribution. Pour 24 degrés de liberté, ces valeurs sont -1.711 et +1.711. (Elles sont représentées par les lignes verticales blanches dans la Figure 13.1.) Dans un test unilatéral, il faut cependant décider avant la collecte des données si l'on s'attend à ce que la moyenne de l'échantillon soit inférieure à la moyenne hypothétique de la population, auquel cas on utilise uniquement la valeur critique inférieure, ou si l'on s'attend à ce que la moyenne de l'échantillon soit supérieure à la moyenne hypothétique de la population, auquel cas on utilise uniquement la valeur critique supérieure. Remarquez que l'on rejette toujours l'hypothèse nulle lorsque le score t de l'échantillon se situe dans les 5 % les plus extrêmes attendus si l'hypothèse nulle est vraie — la valeur α reste donc fixée à 0.05. Nous avons simplement redéfini *extrême* pour ne faire référence qu'à une seule extrémité de la distribution. L'avantage du test unilatéral est que les valeurs critiques sont moins extrêmes. Si la moyenne de l'échantillon diffère de la moyenne hypothétique de la population dans la direction attendue, nous avons plus de chances de rejeter l'hypothèse nulle. L'inconvénient est que si la moyenne de l'échantillon diffère de la moyenne hypothétique de la population dans la direction inattendue, il n'y a aucune possibilité de rejeter l'hypothèse nulle.

Exemple de test t à un échantillon

Imaginons qu'un psychologue de la santé s'intéresse à la précision avec laquelle des personnes étudiant à l'université estiment le nombre de calories contenues dans un biscuit aux pépites de chocolat. Il montre le biscuit à un échantillon de 10 personnes participantes et leur demande individuellement d'estimer le nombre de calories qu'il contient. Le nombre réel de calories dans le biscuit étant de 250, il s'agit de la moyenne hypothétique de la population d'intérêt (μ_0). L'hypothèse nulle stipule que la moyenne des estimations pour la population (μ) est de 250. Comme le chercheur n'a aucune idée précise de la tendance des personnes étudiantes à sous-estimer ou surestimer le nombre de calories, il décide de réaliser un test bilatéral. Imaginons maintenant que les estimations réelles des personnes étudiant à l'université soient les suivantes :

250, 280, 200, 150, 175, 200, 200, 220, 180, 250.

La moyenne des estimations pour l'échantillon (M) est de 212.00 calories et l'écart-type ($ÉT$) est de 39.17. Le psychologue de la santé peut maintenant calculer le score t de son échantillon :

$$t = \frac{212 - 250}{\left(\frac{39.17}{\sqrt{10}} \right)} = -3.07$$

S'il transcrit ses données dans l'un des outils d'analyse en ligne ou utilise SPSS ou JASP, l'outil statistique lui indiquera également que la valeur p bilatérale pour ce score t (avec $10 - 1 = 9$ degrés de liberté) est de 0.013. Comme cette valeur est inférieure à 0.05, le psychologue de la santé rejetterait l'hypothèse nulle et conclurait que les personnes étudiant à l'université ont tendance à sous-estimer le nombre de calories d'un biscuit aux pépites de chocolat. S'il calcule le score t à la main, il pourrait consulter le Tableau 13.2 et constater que la valeur critique de t pour un test bilatéral avec 9 degrés de liberté est de ± 2.262 . Le fait que son score t soit plus extrême que cette valeur critique lui indiquerait que la valeur p est inférieure à 0.05 et qu'il convient de rejeter l'hypothèse nulle. En utilisant le style APA, ces résultats seraient rapportés comme suit : $t(9) = -3.07, p = .01$. Notez que t et p sont en italique, que les degrés de liberté apparaissent entre parenthèses sans décimale et que les valeurs de t et p sont arrondies à deux décimales.

Enfin si ce chercheur avait entrepris cette étude avec de bonnes raisons de s'attendre à ce que les personnes étudiant à l'université sous-estiment le nombre de calories, il aurait pu réaliser un test unilatéral plutôt qu'un test bilatéral. La seule conséquence de cette décision aurait été un changement de la valeur critique, fixée alors à -1,833. Cette valeur légèrement moins extrême aurait rendu un peu plus facile le rejet de l'hypothèse nulle. Toutefois si les personnes aux études avaient en réalité surestimé le nombre de calories — peu importe l'ampleur de cette surestimation — ce chercheur n'aurait pas pu rejeter l'hypothèse nulle.

Le test t à échantillons dépendants

Le **test t à échantillons dépendants** (parfois appelé test t à échantillons appariés) est utilisé pour comparer deux moyennes obtenues à partir du même échantillon, mesuré à deux moments différents ou dans deux conditions différentes. Cette comparaison est appropriée pour les devis prétest-post-test ou les expériences intrasujets. L'hypothèse nulle stipule que les moyennes aux deux moments ou conditions sont identiques dans la population. L'hypothèse alternative stipule qu'elles ne sont pas les mêmes. Ce test peut aussi être unilatéral si la personne chercheuse a de bonnes raisons de s'attendre à ce que la différence aille dans une direction particulière.

On peut considérer le test t à échantillons dépendants comme un cas particulier du test t à un échantillon. La première étape consiste à réduire les deux scores de chaque participant à un seul **score de différence**, en calculant l'écart entre eux. À ce stade, le test t à échantillons dépendants devient un test t à un échantillon sur les scores de différence. La moyenne hypothétique de la population (μ_0) qui nous intéresse est 0, car c'est ce que serait le score de différence moyen, s'il n'y avait pas de différence en moyenne entre les deux temps ou les deux conditions. Nous pouvons maintenant considérer que dans l'hypothèse nulle, le score de différence moyen dans la population est 0 ($\mu_0 = 0$) et que dans l'hypothèse alternative, le score de différence moyen dans la population n'est pas 0 ($\mu_0 \neq 0$).

Exemple du test t à échantillons dépendants

Imaginons que le psychologue de la santé sache désormais que les gens ont tendance à sous-estimer le nombre de calories contenues dans la malbouffe, et qu'il ait mis au point un court programme de formation destiné à améliorer leurs estimations. Pour tester l'efficacité de ce programme, il réalise une étude pré-test/post-test dans laquelle 10 personnes participantes estiment le nombre de calories contenues dans un biscuit aux pépites de chocolat avant la formation, puis à nouveau après celle-ci. Comme il s'attend à ce que le programme augmente les estimations des personnes participantes, il choisit de réaliser un test unilatéral. Imaginons maintenant que les estimations du pré-test sont

230, 250, 280, 175, 150, 200, 180, 210, 220, 190

et que les estimations post-test (pour les mêmes personnes participantes, dans le même ordre) sont

250, 260, 250, 200, 160, 200, 200, 180, 230, 240

Les scores de différence sont donc :

20, 10, -30, 25, 10, 0, 20, -30, 10, 50

Notons que peu importe si la première série de scores est soustraite de la seconde ou la seconde de la première, du moment que cela est fait de la même manière pour toutes les personnes participantes. Dans cet exemple, il est logique de soustraire les estimations du prétest à celles du post-test, de sorte que des scores de différence positifs signifient que les estimations ont augmenté après la formation, tandis que des scores négatifs indiquent qu'elles ont diminué.

La moyenne des scores de différence est de 8.50 avec un écart-type de 27.27. Le psychologue de la santé peut alors calculer le score t pour son échantillon comme suit :

$$t = \frac{8.5 - 0}{\left(\frac{27.27}{\sqrt{10}} \right)} = 1.11$$

En transcrivant ces données dans un outil d'analyse en ligne ou dans Excel, JASP ou SPSS, il obtient une valeur p unilatérale de 0.148 (avec $10 - 1 = 9$ degrés de liberté). Comme cette valeur est supérieure à 0.05, il retient l'hypothèse nulle et conclut que le programme de formation n'augmente pas de manière significative les estimations de calories des personnes. S'il calcule le score t à la main, il peut consulter le Tableau 13.2 et constater que la valeur critique de t pour un test unilatéral avec 9 degrés de liberté est de 1.833. (Le résultat est positif cette fois-ci, car il s'attendait à un score de différence moyenne positif.) Le fait que son score t soit inférieur à cette valeur critique indique que la valeur p est supérieure à 0.05, et qu'il ne peut pas rejeter l'hypothèse nulle.

Le test t à échantillons indépendants

Le **Test t à échantillons indépendants** est utilisé pour comparer les moyennes de deux échantillons distincts (M_1 et M_2). Les deux échantillons peuvent avoir été testés dans des conditions différentes dans le cadre d'une expérience intersujets, ou encore correspondre à des groupes préexistants dans le cadre d'une étude transversale (par exemple femmes et hommes, extravertis et introvertis). L'hypothèse nulle stipule que les moyennes des deux populations sont identiques : $\mu_1 = \mu_2$. L'hypothèse alternative stipule qu'elles ne sont pas

identiques : $\mu_1 \neq \mu_2$. Là encore le test peut être unilatéral si le psychologue de la santé a de bonnes raisons de s'attendre à une différence dans une direction particulière.

Ici la statistique t est un peu plus compliquée, car elle doit prendre en compte deux moyennes d'échantillon, deux écarts-types et deux tailles d'échantillon. La formule est la suivante :

$$t = \frac{M_1 - M_2}{\sqrt{\frac{ET^2_1}{n_1} + \frac{ET^2_2}{n_2}}}$$

Notons que cette formule inclut les écarts-types au carré (les variances) qui apparaissent sous le symbole de la racine carrée. En outre les lettres minuscules n_1 et n_2 désignent les tailles des échantillons des deux groupes ou conditions (par opposition à la majuscule N , qui désigne généralement la taille totale de l'échantillon). Enfin il faut savoir que le test t à échantillons indépendants repose sur $N-2$ degrés de liberté.

Exemple de test t à échantillons indépendants

Le psychologue de la santé veut maintenant comparer les estimations caloriques des personnes qui mangent régulièrement de la malbouffe avec celles de personnes qui en mangent rarement. Comme il croit que la différence peut aller dans une direction comme dans l'autre, il décide d'effectuer un test bilatéral. Il recueille des données auprès d'un échantillon de huit personnes participantes qui mangent régulièrement de la malbouffe, et de sept personnes participantes qui en mangent rarement. Les données sont les suivantes :

Personnes mangeant de la malbouffe: 180, 220, 150, 85, 200, 170, 150, 190

Personnes ne mangeant pas de malbouffe: 200, 240, 190, 175, 200, 300, 240

La moyenne des personnes ne mangeant pas de malbouffe est de 220.71 avec un écart-type de 41.23. La moyenne des personnes mangeant de la malbouffe est de 168.12 avec un écart-type de 42.66. Il peut maintenant calculer son score t comme suit :

$$t = \frac{220.71 - 168.12}{\sqrt{\frac{41.23^2}{8} + \frac{42.66^2}{7}}} = 2.42$$

En transcrivant les données dans un outil d'analyse en ligne ou en utilisant Excel, JASP ou SPSS, il obtient une valeur p bilatérale de 0,015 (avec $15 - 2 = 13$ degrés de liberté). Comme cette valeur p est inférieure à 0.05, le psychologue de la santé rejette l'hypothèse nulle et conclut que les personnes qui mangent régulièrement de la malbouffe font des estimations caloriques inférieures à celles de personnes qui en mangent rarement. S'il calcule le score t à la main, il peut consulter le Tableau 13.2 et constater que la valeur critique de t pour un test bilatéral à 13 degrés de liberté est de ± 2.160 . Le fait que son score t soit plus extrême que cette valeur critique indique que la valeur p est inférieure à 0.05 et qu'il doit rejeter l'hypothèse nulle.

L'analyse de variance

Les tests t sont utilisés pour comparer deux moyennes (une moyenne d'échantillon avec une moyenne de population, les moyennes de deux conditions ou de deux groupes). Lorsqu'il y a plus de deux groupes ou plus de deux conditions à comparer, le test d'hypothèse nulle le plus courant est l'**analyse de variance (ANOVA)**. Dans cette section, nous examinons principalement l'**ANOVA à un facteur**, qui est utilisée pour les devis intersujets avec une seule variable indépendante. Nous examinerons ensuite brièvement d'autres versions de l'ANOVA utilisées pour les devis de recherche intrasujets et plans factoriels.

ANOVA à un facteur

L'ANOVA à un facteur est utilisée pour comparer les moyennes de plus de deux échantillons ($M_1, M_2, \dots, M_G$) dans un devis de recherche intersujets. L'hypothèse nulle stipule que toutes les moyennes sont égales dans la population : $\mu_1 = \mu_2 = \dots = \mu_G$. L'hypothèse alternative stipule que toutes les moyennes ne sont pas égales dans la population.

La statistique de test pour l'ANOVA est appelée F . Il s'agit d'un rapport entre deux estimations de la variance de la population établies à partir des données de l'échantillon. L'une des estimations de la variance de la population est appelée le **carré moyen intergroupe (CM_{inter})** et est établie à partir des différences entre les moyennes des échantillons. L'autre est appelée **carré moyen intragroupe (CM_{intra})** et est établie à partir des différences entre les scores à l'intérieur de chaque groupe. La statistique F est le rapport entre le CM_{inter} et le CM_{intra} et peut donc s'exprimer comme suit :

$$F = CM_{\text{inter}} / CM_{\text{intra}}$$

La raison pour laquelle F est utile est que nous connaissons sa distribution lorsque l'hypothèse nulle est vraie. Comme le montre la Figure 13.2, cette distribution est unimodale et présente une asymétrie positive avec des valeurs qui se concentrent autour de 1. La forme précise de la distribution dépend à la fois du nombre de groupes et de la taille de l'échantillon, auxquels s'associent des degrés de liberté. Le degré de liberté intergroupe est le nombre de groupes moins un : $dl_{\text{inter}} = (G - 1)$. Le degré de liberté intragroupe est la taille totale de l'échantillon moins le nombre de groupes : $dl_{\text{intra}} = N - G$. Ici encore la connaissance de la distribution de F lorsque l'hypothèse nulle est vraie nous permet de trouver la valeur p .

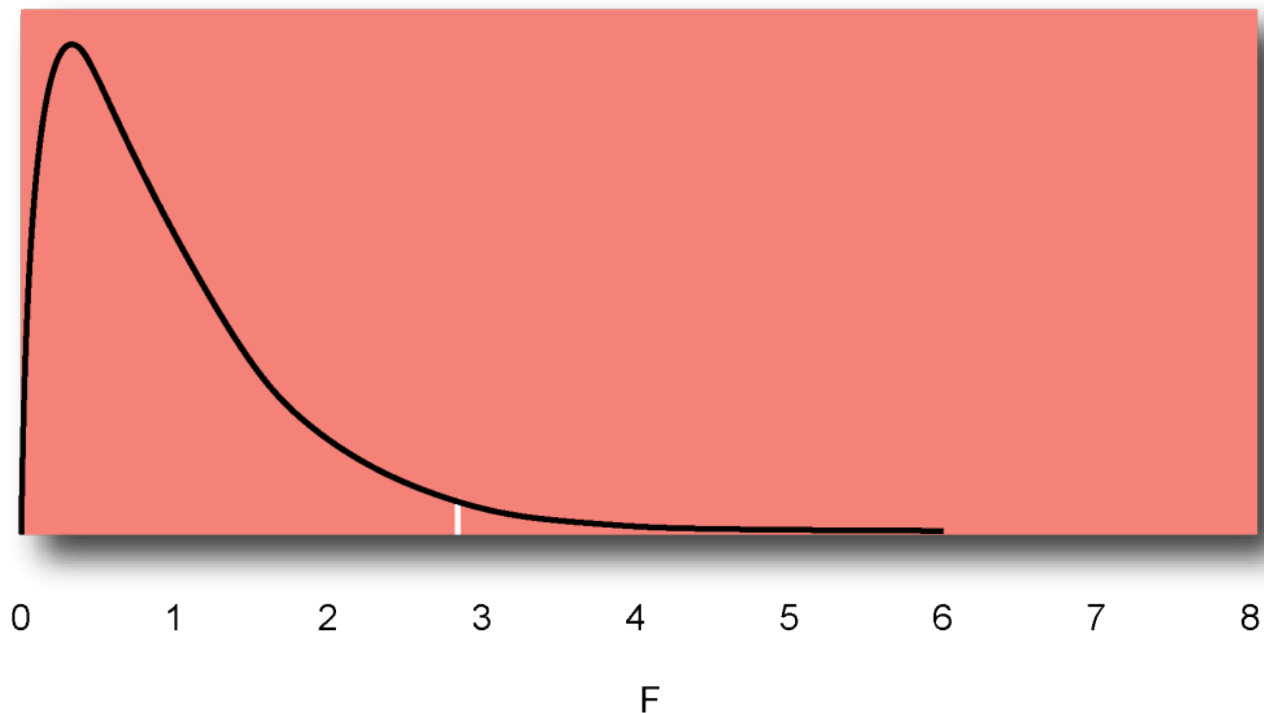


Figure 13.2-Distribution du rapport F avec 2 et 37 degrés de liberté lorsque l'hypothèse nulle est vraie. La ligne verticale blanche représente la valeur critique lorsque α est égal à 0,05.

Les outils en ligne du chapitre portant sur les statistiques descriptives et les logiciels comme Excel, SPSS et JASP calculent directement F et fournissent la valeur p . Si p est inférieur ou égal à 0,05, nous rejetons l'hypothèse nulle et concluons qu'il existe des différences entre les moyennes des groupes dans la population. Si p est supérieur à 0,05, nous retenons l'hypothèse nulle et concluons qu'il n'y a pas suffisamment de preuves pour affirmer qu'il existe des différences. Dans le cas improbable où nous calculons F à la main, nous pouvons utiliser un tableau de valeurs critiques comme le Tableau 13.3 « Tableau de valeurs critiques » pour prendre la décision. L'idée est que tout rapport F supérieur à la valeur critique a une valeur p inférieure à 0,05. Par conséquent si le rapport F calculé est supérieur à la valeur critique, nous rejetons l'hypothèse nulle. Si le rapport F calculé est inférieur à la valeur critique, nous retenons l'hypothèse nulle.

Tableau 13.3 Tableau de
valeurs critiques lorsque $\alpha =$
.05

<i>dl_{intra}</i>	<i>dl_{inter}</i>		
	2	3	4
8	4.459	4.066	3.838
9	4.256	3.863	3.633
10	4.103	3.708	3.478
11	3.982	3.587	3.357
12	3.885	3.490	3.259
13	3.806	3.411	3.179
14	3.739	3.344	3.112
15	3.682	3.287	3.056
16	3.634	3.239	3.007
17	3.592	3.197	2.965
18	3.555	3.160	2.928
19	3.522	3.127	2.895
20	3.493	3.098	2.866
21	3.467	3.072	2.840
22	3.443	3.049	2.817
23	3.422	3.028	2.796
24	3.403	3.009	2.776
25	3.385	2.991	2.759
30	3.316	2.922	2.690
35	3.267	2.874	2.641
40	3.232	2.839	2.606
45	3.204	2.812	2.579
50	3.183	2.790	2.557
55	3.165	2.773	2.540
60	3.150	2.758	2.525
65	3.138	2.746	2.513
70	3.128	2.736	2.503
75	3.119	2.727	2.494
80	3.111	2.719	2.486
85	3.104	2.712	2.479
90	3.098	2.706	2.473
95	3.092	2.700	2.467
100	3.087	2.696	2.463

Exemple d'ANOVA simple

Imaginons que le psychologue de la santé veuille comparer les estimations de calories des personnes étudiant en psychologie, des personnes étudiant en nutrition et des diététistes. Il recueille les données suivantes :

Personnes étudiant en psychologie: 200, 180, 220, 160, 150, 200, 190, 200
Personnes étudiant en nutrition: 190, 220, 200, 230, 160, 150, 200, 210, 195
Diététistes: 220, 250, 240, 275, 250, 230, 200, 240

Les moyennes sont respectivement de 187.50 ($\text{ÉT} = 23.14$), 195.00 ($\text{ÉT} = 27.77$) et 238.13 ($\text{ÉT} = 22.35$). Il semble donc que les diététistes aient fait des estimations beaucoup plus précises en moyenne. Le psychologue transcrit presque certainement ces données dans un programme comme Excel, SPSS ou JASP, qui calcule la valeur F et la valeur p . Le Tableau 13.4 présente les résultats de l'ANOVA à un facteur obtenus avec Excel pour ces données. Ce tableau, appelé tableau de l'ANOVA, montre que le CM_{inter} est de 5 971. 88, le CM_{intra} de 602.23 et que leur rapport, F , est de 9.92. La valeur p est de 0.0009. Comme cette valeur est inférieure à 0.05, le psychologue rejette l'hypothèse nulle et conclut que les moyennes des estimations caloriques des trois groupes ne sont pas identiques dans la population. Remarquons que le tableau d'ANOVA inclut également la « somme des carrés » (SC) intergroupe et intragroupe. Ces valeurs servent à calculer le CM_{inter} et le CM_{intra} , mais elles ne sont généralement pas rapportées par les scientifiques. Finalement, si le psychologue calcule le rapport F à la main, il peut consulter le Tableau 13.3 et constater que la valeur critique de F avec 2 et 21 degrés de liberté est de 3.467 (la même valeur que celle indiquée dans le Tableau 13.4 sous F_{critique}). Le fait que son score F soit plus extrême que cette valeur critique indique que sa valeur p est inférieure à 0.05 et qu'il doit rejeter l'hypothèse nulle.

Tableau 13.4 Résultats typiques d'une ANOVA simple avec Excel

ANOVA						
Source of variation	SC	dl	CM	F	valeur p	Fcritique
Entre les groupes	11,943.75	2	5,971.875	9.916234	0.000928	3.4668
À l'intérieur des groupes	12,646.88	21	602.2321			
Total	24,590.63	23				

Élaborations de l'ANOVA

Comparaisons post hoc

Lorsque nous rejetons l'hypothèse nulle dans une ANOVA simple, nous concluons que les moyennes des groupes ne sont pas toutes identiques dans la population. Mais cela peut signifier plusieurs choses. Avec trois groupes, cela peut signifier que les trois moyennes diffèrent significativement les unes des autres. Cela peut aussi indiquer qu'une seule moyenne diffère des deux autres, alors que ces deux dernières ne diffèrent pas

significativement entre elles. Par exemple il se peut que les estimations caloriques moyennes des personnes étudiant en psychologie, des personnes étudiant en nutrition et des diététistes soient toutes significativement différentes les unes des autres. Mais il se peut que la moyenne des diététistes diffère significativement de celles des personnes étudiant en psychologie et en nutrition, alors que ces deux groupes ne diffèrent pas significativement entre eux. C'est pourquoi les résultats statistiquement significatifs d'une ANOVA à un facteur sont généralement suivis de **comparaisons post hoc**, c'est-à-dire des tests supplémentaires portant sur certaines paires de moyennes de groupes afin de déterminer lesquelles diffèrent effectivement.

Une approche des comparaisons post hoc consiste à effectuer une série de tests t à échantillons indépendants qui compare chaque moyenne de groupe à chacune des autres moyennes de groupe. Mais cette approche présente un problème. En général lorsqu'on effectue un test t alors que l'hypothèse nulle est vraie, il y a environ 5 % de chances de la rejeter à tort (erreur de type I, voir la section 13.3 « Considérations supplémentaires » pour plus d'informations). Si l'on effectue plusieurs tests t alors que l'hypothèse nulle est vraie, la probabilité de rejeter par erreur au moins une hypothèse nulle augmente avec chaque test effectué. C'est pourquoi, les scientifiques n'effectuent généralement pas de comparaisons post hoc à l'aide de tests t standard, car le risque de rejeter par erreur au moins une hypothèse nulle est trop élevé. Ils recourent plutôt à des procédures modifiées du test — parmi lesquelles il y a la procédure de Bonferroni, le test de la plus petite différence significative de Fisher (LSD) et le test de Tukey sur les différences honnêtement significatives (HSD). Les détails de ces méthodes dépassent le cadre de ce livre, mais il est important de comprendre leur objectif. Il s'agit de maintenir le risque de rejeter par erreur une hypothèse nulle vraie à un niveau acceptable (proche de 5 %).

ANOVA à mesures répétées

Rappelons que l'ANOVA à un facteur convient aux devis intersujets dans lesquels les moyennes comparées proviennent de groupes distincts de personnes. Elle ne convient pas aux devis intra-sujets, dans lesquels les moyennes comparées proviennent des mêmes personnes testées dans des conditions différentes ou à des moments différents. Ce cas nécessite une approche légèrement différente appelée **ANOVA à mesures répétées**. Les principes de base de l'ANOVA à mesures répétées sont les mêmes que celles de l'ANOVA à un facteur. La différence principale est que le fait de mesurer la variable dépendante à plusieurs reprises pour chaque participant permet d'obtenir une estimation plus précise du CM_{intra} . Imaginons par exemple qu'une étude mesure le temps de réaction. Certaines personnes participantes seront plus rapides ou plus lentes que d'autres, en raison de différences individuelles stables dans leur système nerveux, leurs muscles et d'autres facteurs. Dans un devis intersujets, ces différences individuelles stables s'ajoutent simplement à la variabilité à l'intérieur des groupes et augmentent la valeur du CM_{intra} (ce qui à son tour diminue la valeur de F). Toutefois dans un devis intrasujet, ces différences individuelles stables peuvent être mesurées et soustraites de la valeur du CM_{intra} . Une valeur plus petite du CM_{intra} se traduit par une valeur plus élevée de F et par un test plus sensible.

ANOVA factorielle

Lorsque plus d'une variable indépendante est incluse dans un plan factoriel, l'approche appropriée est l'**ANOVA factorielle**. Là encore les principes de base sont les mêmes que pour l'ANOVA à un facteur et l'ANOVA à mesures répétées. La différence principale est qu'elle produit un rapport F et une valeur p pour chaque effet principal, et pour chaque interaction. Reprenons l'exemple des estimations caloriques: imaginons que le psychologue de la santé examine l'effet du programme d'étude (psychologie vs nutrition) et du type d'aliment (biscuit vs hamburger) dans un plan factoriel. Une ANOVA factorielle produirait des rapports F et des valeurs p distincts pour l'effet principal du programme d'étude, l'effet principal du type d'aliment, ainsi que pour l'interaction entre le programme d'étude et le type d'aliment. Des ajustements appropriés doivent être apportés selon que le devis est intersujets, intrasujets ou mixte.

Test des coefficients de corrélation

Pour les relations entre variables quantitatives, où le r de Pearson (le coefficient de corrélation) sert à décrire la force de la relation, le test d'hypothèse nulle approprié est le test du coefficient de corrélation. La logique de base est exactement la même que pour les autres tests de l'hypothèse nulle. Dans ce cas, l'hypothèse nulle est qu'il n'y a pas de relation dans la population. On utilise la lettre grecque rho (ρ) pour représenter le paramètre correspondant : $\rho = 0$. L'hypothèse alternative est qu'il existe une relation dans la population : $\rho \neq 0$. Comme pour le test t , ce test peut être bilatéral si les scientifiques n'ont pas d'attente particulière quant à la direction de la relation, ou unilatéral si une direction est attendue.

Il est possible d'utiliser le coefficient de corrélation de l'échantillon pour calculer un score t avec $N - 2$ degrés de liberté, puis de procéder comme pour un test t . Toutefois en raison de sa méthode de calcul, le coefficient de corrélation peut également être traité comme une statistique de test à part entière. Les outils statistiques en ligne et les logiciels statistiques tels qu'Excel, JASP et SPSS calculent généralement le coefficient de corrélation et fournissent la valeur p qui lui est associée. Comme toujours, si la valeur p est égale ou inférieure à 0.05, nous rejetons l'hypothèse nulle et concluons qu'il existe une relation entre les variables dans la population. Si la valeur p est supérieure à 0.05, nous retenons l'hypothèse nulle et concluons qu'il n'y a pas suffisamment de preuves pour affirmer qu'il existe une relation dans la population. Si nous calculons le coefficient de corrélation à la main, nous pouvons utiliser un tableau comme le Tableau 13.5, qui présente les valeurs critiques de r pour différentes tailles d'échantillons lorsque α est égal à 0.05. Une valeur du coefficient de corrélation observée dans l'échantillon, qui est plus extrême que la valeur critique, est statistiquement significative.

Tableau 13.5 Tableau des
valeurs critiques de r lorsque α
= 0.05

N	Valeur critique de r	
	Unilatérale	Bilatérale
5	.805	.878
10	.549	.632
15	.441	.514
20	.378	.444
25	.337	.396
30	.306	.361
35	.283	.334
40	.264	.312
45	.248	.294
50	.235	.279
55	.224	.266
60	.214	.254
65	.206	.244
70	.198	.235
75	.191	.227
80	.185	.220
85	.180	.213
90	.174	.207
95	.170	.202
100	.165	.197

Exemple de test d'un coefficient de corrélation

Imaginons que la psychologue de la santé s'intéresse à la corrélation entre les estimations caloriques des personnes et leur poids. Comme elle n'a pas d'attente quant à la direction de la relation, elle décide d'effectuer un test bilatéral. Elle calcule le coefficient de corrélation pour un échantillon de 22 personnes étudiant à l'université et trouve que le r de Pearson est de -0.21. Le logiciel statistique qu'elle utilise lui indique que la valeur p est de 0.348. Comme cette valeur est supérieure à 0.05, elle retient l'hypothèse nulle et conclut qu'il n'y a pas de relation entre les estimations caloriques des personnes et leur poids. Si elle calculait le coefficient de corrélation à la main, elle pourrait consulter le Tableau 13.5 et constater que la valeur critique pour $22 - 2 = 20$ degrés de liberté est de 0.444. Le fait que le coefficient de corrélation pour son échantillon soit moins extrême que cette valeur critique lui indique que la valeur p est supérieure à 0.05, et qu'elle doit retenir l'hypothèse nulle.

71. Considérations supplémentaires

Objectifs d'apprentissage

1. Définir les erreurs de type I et de type II, expliquer pourquoi elles se produisent et identifier certaines mesures qui peuvent être prises pour minimiser leur probabilité.
2. Définir la puissance statistique, expliquer son rôle dans la planification de nouvelles études, et utiliser des outils en ligne pour calculer la puissance statistique de devis de recherche simples.
3. Énumérer certaines critiques des tests conventionnels d'hypothèse nulle, ainsi que des moyens d'y répondre.

Dans cette section, nous examinons quelques questions additionnelles liées au test de l'hypothèse nulle, dont certaines sont utiles pour planifier les études et interpréter les résultats. Nous considérons également certaines critiques formulées depuis longtemps à l'encontre des tests de l'hypothèse nulle, ainsi que des mesures prises par les scientifiques en psychologie pour y répondre.

Erreurs dans les tests de l'hypothèse nulle

Lors du test de l'hypothèse nulle, les scientifiques tentent de tirer une conclusion raisonnable au sujet de la population à partir de l'échantillon. Malheureusement l'exactitude de cette conclusion n'est pas garantie. Cette divergence est illustrée par la Figure 13.3. Les rangées de ce tableau représentent les deux décisions possibles que les scientifiques peuvent prendre dans le cadre des tests de l'hypothèse nulle : rejeter ou retenir l'hypothèse nulle. Les colonnes représentent les deux états possibles du monde : l'hypothèse nulle est fausse ou elle est vraie. Les quatre cellules du tableau représentent donc les quatre résultats distincts d'un test de l'hypothèse nulle. Deux de ces résultats — rejeter l'hypothèse nulle lorsqu'elle est fausse et la retenir lorsqu'elle est vraie — sont des décisions correctes. Les deux autres — rejeter l'hypothèse nulle lorsqu'elle est vraie et la retenir lorsqu'elle est fausse — constituent des erreurs.

	L'état réel du monde	
Décision	H_0 faux	H_0 vrai
Rejeter H_0	Décision correcte	Erreur de type I
Retenir H_0	Erreur de type II	Décision correcte

Figure 13.3 – Deux types de décisions correctes et deux types d'erreurs dans le test de l'hypothèse nulle
[\[Description de l'image\]](#)

Rejeter l'hypothèse nulle lorsqu'elle est vraie s'appelle une **erreur de type I**. Cette erreur signifie que nous concluons qu'il existe une relation dans la population alors qu'en réalité il n'y en a pas. Les erreurs de type I se produisent parce que, même lorsqu'il n'existe aucune relation dans la population, l'erreur d'échantillonnage à elle seule peut parfois produire un résultat extrême. En fait lorsque l'hypothèse nulle est vraie et que α est fixé à 0,05, nous rejetons l'hypothèse nulle par erreur dans 5 % des cas. (C'est pour cette raison qu' α est parfois appelé le « taux d'erreur de type I ».) Retenir l'hypothèse nulle alors qu'elle est fausse s'appelle une **erreur de type II**. Cette erreur signifie que nous concluons qu'il n'existe pas de relation dans la population alors qu'en réalité il y en a une. En pratique les erreurs de type II se produisent principalement parce que le devis de recherche manque de puissance statistique suffisante pour détecter la relation (par exemple l'échantillon est trop petit). Nous reviendrons prochainement sur la question de la puissance statistique.

En principe il est possible de réduire le risque d'erreur de type I avec un α inférieur à 0,05. Par exemple avec un α de 0,01, si l'hypothèse nulle est vraie, il n'y a que 1 % de chances de la rejeter par erreur. Mais rendre plus difficile le rejet des vraies hypothèses nulles rend aussi plus difficile le rejet des fausses hypothèses nulles, ce qui augmente le risque d'erreur de type II. Inversement, il est possible de réduire le risque d'erreur de type II avec un α supérieur à 0,05 (par exemple 0,10). Mais faciliter le rejet des fausses hypothèses nulles facilite également le rejet des vraies hypothèses nulles, et augmente donc le risque d'erreur de type I. Cela explique en partie pourquoi la convention est de fixer α à 0,05. Il existe un certain consensus parmi les scientifiques pour dire que ce seuil maintient les taux d'erreurs de type I et de type II à des niveaux acceptables.

La possibilité de commettre des erreurs de type I et de type II a plusieurs implications importantes pour l'interprétation des résultats de nos propres recherches ainsi que celles des autres. L'une d'elles est que nous devons faire preuve de prudence lorsque nous interprétons les résultats de toute étude prise isolément, puisqu'il existe toujours la possibilité qu'ils reflètent une erreur de type I ou de type II. C'est pour cette raison

que les personnes en recherche considèrent la réplication de leurs études comme essentielle. Chaque fois qu'une étude est répliquée avec un résultat similaire, les personnes en recherche peuvent à juste titre avoir davantage confiance que ce résultat reflète un phénomène réel et pas seulement une erreur de type I ou de type II.

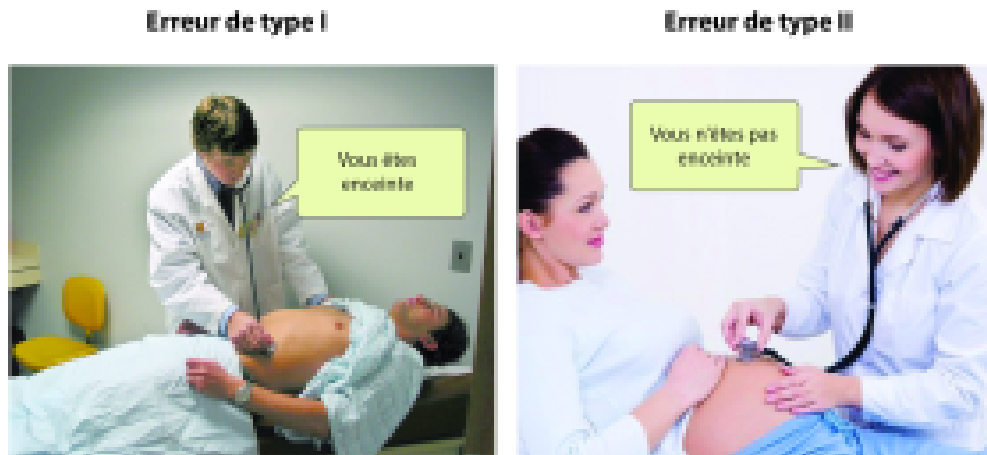


Figure 13.4 Un exemple humoristique illustrant comment les erreurs de type I et de type II peuvent se produire lors d'exams de grossesse. [\[Description de l'image adaptée\]](#)

Un autre problème lié aux erreurs de type I est le **problème du tiroir à dossiers** (Rosenthal, 1979) Rosenthal, R. (1979). L'idée est que lorsque les scientifiques obtiennent des résultats statistiquement significatifs, ces personnes ont tendance à les soumettre pour publication, et les responsables de la rédaction et de l'évaluation ont tendance à les accepter. En revanche lorsque les scientifiques obtiennent des résultats non significatifs, ces personnes ont tendance à ne pas les soumettre pour publication ou s'ils sont soumis, les responsables de la rédaction et de l'évaluation ont tendance à ne pas les accepter. Les scientifiques finissent par ranger ces résultats non significatifs dans un tiroir (ou de nos jours dans un dossier de leur disque dur). Cette tendance a notamment pour effet que la littérature publiée contient probablement une proportion plus élevée d'erreurs de type I que ce à quoi on pourrait s'attendre seulement sur la base des considérations statistiques. Même lorsqu'il existe une relation entre deux variables dans la population, la littérature de recherche publiée est susceptible d'exagérer la force de cette relation. Imaginons par exemple que la relation entre deux variables dans la population soit positive, mais faible (par exemple $\rho = +0.10$). Si plusieurs scientifiques mènent des études sur cette relation, l'erreur d'échantillonnage est susceptible de produire des résultats allant de relations négatives faibles (par exemple $r = -0.10$) à des relations positives modérément fortes (par exemple $r = +0.40$). Mais en raison du problème du tiroir à dossiers, il est probable que seules les études produisant des relations positives modérées à fortes soient publiées. Il en résulte que l'effet rapporté dans la littérature publiée paraît plus fort qu'il ne l'est réellement dans la population.

Le problème du tiroir à dossiers est difficile à résoudre, car il découle des pratiques traditionnelles de la recherche et de la publication scientifique. Une solution proposée est celle des rapports enregistrés où les responsables de la rédaction et de la révision de revues évaluent les recherches soumises pour publication sans connaître leurs résultats (voir cos.io/rr/). L'idée est que si la question de recherche est jugée intéressante et la méthodologie rigoureuse, alors un résultat non significatif devrait être considéré tout aussi important et digne d'être publié qu'un résultat significatif. En l'absence d'un changement aussi radical dans

la manière dont la recherche est évaluée en vue de sa publication, les scientifiques peuvent toujours veiller à conserver leurs résultats non significatifs et de les partager aussi largement que possible (par exemple dans des archives accessibles au public et lors de congrès professionnels). De nombreuses disciplines scientifiques disposent désormais de revues consacrées spécifiquement à la publication de résultats non significatifs. En psychologie par exemple il existe le « Journal of Articles in Support of the Null Hypothesis » (www.jasnh.com).

En 2014, Uri Simonsohn, Leif Nelson et Joseph Simmons ont publié un article (Simonsohn, Nelson, et Simmons, 2014) accusant la communauté scientifique en psychologie de produire trop d'erreurs de type I en adoptant des pratiques de recherche qu'ils ont appelées **p-hacking**. Les scientifiques pratiquant le p-hacking prennent diverses décisions au cours du processus de recherche pour augmenter leurs chances d'obtenir un résultat statistiquement significatif (et une erreur de type I) en supprimant arbitrairement les valeurs aberrantes, en choisissant de manière sélective les variables dépendantes à rapporter, en ne présentant que les résultats significatifs, etc., jusqu'à ce que leurs résultats donnent une valeur p souhaitable. Cet article révolutionnaire a contribué à un débat majeur dans le domaine sur les normes de publication et les moyens d'améliorer la fidélité de nos résultats, qui se poursuit encore aujourd'hui.

Puissance statistique

La **puissance statistique** d'un devis de recherche est la probabilité de rejeter l'hypothèse nulle compte tenu de la taille de l'échantillon et de la force de la relation attendue. Par exemple la puissance statistique d'une étude avec 50 personnes et un r de Pearson attendu de +0.30 dans la population est de 0.59. En d'autres mots, il y a 59 % de chances de rejeter l'hypothèse nulle si en réalité la corrélation dans la population est bien de +0.30. La puissance statistique est le complément de la probabilité de commettre une erreur de type II. Ainsi dans cet exemple la probabilité de commettre une erreur de type II serait de $1 - 0.59 = 0.41$. Il est donc clair que les scientifiques doivent s'intéresser à la puissance de leurs devis de recherche afin d'éviter de commettre des erreurs de type II. En particulier ces personnes doivent s'assurer que leur devis de recherche a une puissance suffisante avant de collecter des données. Une recommandation fréquente est qu'une puissance de 0,80 est adéquate. Cela signifie qu'il y a 80 % de chances de rejeter l'hypothèse nulle pour une force de relation attendue donnée.

La question de savoir comment calculer la puissance pour différents devis de recherche et tests d'hypothèse nulle dépasse la portée de ce livre. Cependant il existe des outils en ligne qui vous permettent de le faire en entrant la taille de votre échantillon, la force de la relation attendue et le niveau α pour divers tests d'hypothèses (voir « Calcul de la puissance en ligne »). En outre le Tableau 13.6 indique la taille de l'échantillon nécessaire pour obtenir une puissance de 0.80 pour les relations faibles, moyennes et fortes à la fois pour un test t bilatéral à échantillons indépendants et pour un test bilatéral du r de Pearson. En particulier les relations faibles nécessitent de très grands échantillons pour offrir une puissance statistique adéquate.

Tableau 13.6 Tailles d'échantillon nécessaires pour obtenir une puissance statistique de 0,80 parmi différentes forces de relation attendues pour un test *t* à échantillons indépendants et un test *r* de Pearson

Test de l'hypothèse nulle		
Force de la relation	Test <i>t</i> à échantillons indépendants	Test <i>r</i> de Pearson
Forte (<i>d</i> = .80, <i>r</i> = .50)	52	28
Moyenne (<i>d</i> = .50, <i>r</i> = .30)	128	84
Faible (<i>d</i> = .20, <i>r</i> = .10)	788	782

Que devez-vous faire si vous découvrez que votre devis de recherche n'a pas une puissance suffisante ? Imaginons par exemple que vous meniez une expérience intersujets avec 20 personnes dans chacune des deux conditions et que vous vous attendiez à une différence moyenne ($d = 0.50$) dans la population. La puissance statistique de ce devis n'est que de 0.34. En d'autres mots, même s'il existe une différence moyenne dans la population, il n'y a qu'une chance sur trois de rejeter l'hypothèse nulle et environ deux chances sur trois de commettre une erreur de type II. Compte tenu du temps et des efforts nécessaires à la réalisation de l'étude, cela peut sembler inacceptable : une chance trop faible de rejeter l'hypothèse nulle, et une chance trop élevée de commettre une erreur de type II.

Étant donné que la puissance statistique dépend principalement de la force de la relation et de la taille de l'échantillon, il y a essentiellement deux mesures à prendre pour augmenter la puissance statistique : augmenter la force de la relation ou augmenter la taille de l'échantillon. Il est parfois possible d'augmenter la force de la relation en utilisant une manipulation plus forte ou en contrôlant plus rigoureusement les variables externes afin de réduire le bruit dans les données (par exemple en utilisant un devis intrasujet plutôt qu'un devis intersujets). La stratégie commune consiste toutefois à augmenter la taille de l'échantillon. Pour toute force de relation attendue, il y aura toujours un échantillon suffisamment grand pour obtenir une puissance adéquate.

Calcul de la puissance en ligne

Les liens suivants renvoient à des outils qui vous permettent de calculer la puissance statistique pour divers devis de recherche et tests d'hypothèse nulle en fournissant des informations sur la force de la relation attendue, la taille de l'échantillon et le niveau α . Ils vous permettent également de calculer la taille de l'échantillon nécessaire pour atteindre le niveau de puissance souhaité (par exemple 0.80). Le premier est un outil en ligne. Le second est un programme gratuit téléchargeable appelé G*Power.

- Page de Russ Lenth sur la puissance et la taille de l'échantillon : github.com/rvlenth/piface
- G*Power : www.gpower.hhu.de

Problèmes liés au test de l'hypothèse nulle et quelques solutions

Encore une fois le test de l'hypothèse nulle constitue l'approche la plus courante des statistiques inférentielles en psychologie. Il n'est cependant pas exempt de critiques. En fait ces dernières années, les critiques sont devenues si importantes que l'American Psychological Association a réuni un groupe de travail chargé de formuler des recommandations sur la façon d'y répondre (Wilkinson et Task Force on Statistical Inference, 1999). Dans cette section, nous examinons certaines de ces critiques et recommandations.

Critiques des tests de l'hypothèse nulle

Certaines critiques du test de l'hypothèse nulle portent sur la mauvaise compréhension qu'en ont les scientifiques. Nous avons déjà vu par exemple que la valeur p est souvent interprétée à tort, comme la probabilité que l'hypothèse nulle soit vraie (alors qu'il s'agit en fait de la probabilité d'obtenir le résultat observé dans l'échantillon *si* l'hypothèse nulle était vraie). Une autre interprétation erronée étroitement liée consiste à croire que $1-p$ équivaut à la probabilité de répliquer un résultat statistiquement significatif. Dans une étude, 60 % d'un échantillon de scientifiques d'expérience pensaient qu'une valeur p de 0.01 pour un test t à échantillons indépendants avec 20 personnes dans chaque échantillon — signifiait qu'il y avait 99 % de chances de répliquer le résultat statistiquement significatif (Oakes, 1986). Comme notre discussion sur la puissance statistique l'a montré, ce chiffre est beaucoup trop optimiste. Le Tableau 13.5 montre que même s'il y a une grande différence entre les moyennes dans la population, il faut 26 personnes par échantillon pour obtenir une puissance de 0.80. Et le programme G*Power montre qu'il faudrait 59 personnes par échantillon pour obtenir une puissance de 0.99.

Une autre série de critiques porte sur la logique du test de l'hypothèse nulle. Pour plusieurs, la convention stricte consistant à rejeter l'hypothèse nulle lorsque p est inférieur à 0.05 et à la retenir lorsque p est supérieur à 0.05 n'a guère de sens. Cette critique ne porte pas sur la valeur spécifique de 0.05, mais sur l'idée qu'il devrait y avoir une ligne de démarcation rigide entre les résultats considérés comme significatifs, et ceux qui ne le sont pas. Imaginons deux études sur la même relation statistique avec des échantillons de taille similaire. L'une a une valeur p de 0.04 et l'autre, une valeur p de 0.06. Bien que les deux études produisent essentiellement le même résultat, la première sera probablement considérée comme intéressante et digne d'être publiée, tandis que la seconde sera jugée non significative. Cette convention est susceptible d'empêcher la publication de bonnes recherches et de contribuer au problème du tiroir à dossiers.

Une autre série de critiques porte sur l'idée que le test de l'hypothèse nulle — même lorsqu'il est compris et exécuté correctement — n'est tout simplement pas très informatif. Rappelons que l'hypothèse nulle stipule qu'il n'existe aucune relation entre les variables dans la population (par exemple le d de Cohen ou le r de Pearson est précisément égal à 0). Rejeter l'hypothèse nulle revient donc simplement à affirmer qu'il existe une *certaine* relation non nulle dans la population. Mais cette affirmation ne dit pas grand-chose. Imaginons que la chimie puisse seulement nous dire qu'il existe une *certaine* relation entre la température d'un gaz et son volume — au lieu de fournir une équation précise pour décrire cette relation. Certaines personnes critiques affirment même que la relation entre deux variables dans la population n'est jamais exactement

égale à 0, si elle est effectuée avec suffisamment de décimales. En d'autres mots, l'hypothèse nulle n'est jamais littéralement vraie. La rejeter ne nous apprend donc rien que nous ne sachions déjà !

Pour être juste, plusieurs scientifiques ont pris la défense des tests de l'hypothèse nulle. L'un d'entre eux, Robert Abelson, a fait valoir que lorsqu'il est correctement compris et réalisé, le test de l'hypothèse nulle remplit une fonction importante (Abelson, 1995). En particulier lorsqu'il s'agit de phénomènes nouveaux, il permet aux scientifiques de convaincre les autres, que leurs résultats ne doivent pas être considérés comme de simples hasards.

La fin des valeurs p ?

En 2015, le comité de rédaction de la revue *Basic and Applied Social Psychology* a annoncé l'interdiction d'utiliser le test de l'hypothèse nulle et des procédures statistiques connexes (Tramimow et Marks, 2015). Toutes les personnes voulant publier des articles sont invitées à les soumettre avec les valeurs p , mais celles-ci seront supprimées par la rédaction avant publication. Bien qu'aucun test statistique supérieur n'ait été proposé pour remplacer le test de l'hypothèse nulle, le comité de rédaction a souligné l'importance des statistiques descriptives, et des tailles d'effet. Ce point de vue, le rejet de « l'étalon-or » (*gold standard*, en anglais), bien que peu adopté, a contribué à alimenter le débat en psychologie sur ce que nous savons réellement, et sur la manière dont nous le savons.

Que faire?

Même les personnes qui défendent le test de l'hypothèse nulle reconnaissent ses limites. Mais que faut-il faire ? Quelques suggestions figurent désormais dans le manuel de publication de l'APA. L'une d'entre elles consiste à accompagner chaque test de l'hypothèse nulle par une mesure de la taille de l'effet, telle que le d de Cohen ou le r de Pearson. Ainsi les scientifiques fournissent une estimation de la force de la relation dans la population, et pas seulement de l'existence ou non d'une relation. (Rappelons que la valeur p ne peut se substituer à une mesure de la force de la relation, puisqu'elle dépend aussi de la taille de l'échantillon. Même un résultat très faible peut être statistiquement significatif si l'échantillon est suffisamment grand).

Une autre suggestion consiste à utiliser des **intervalles de confiance** plutôt que des tests d'hypothèse nulle. Un intervalle de confiance autour d'une statistique représente une étendue de valeurs calculée de telle sorte qu'un certain pourcentage du temps (généralement 95 %), le paramètre de la population se situe dans cette étendue. Par exemple un échantillon de 20 personnes étudiant à l'université peut avoir une estimation moyenne de 200 calories pour un biscuit aux pépites de chocolat, avec un intervalle de confiance de 95 % allant de 160 à 240. En d'autres mots, il y a une probabilité très élevée (95 %) que la moyenne de la population se situe dans cet intervalle. Les personnes encourageant l'utilisation des intervalles de confiance font valoir qu'ils sont beaucoup plus faciles à interpréter que les tests de l'hypothèse nulle. Un autre avantage des intervalles de confiance est qu'ils fournissent les informations nécessaires pour réaliser des tests d'hypothèse nulle si quelqu'un le souhaite. Dans cet exemple, la moyenne de l'échantillon de

200 est significativement différente, au seuil de 0.05, de toute moyenne hypothétique de la population qui se situe en dehors de l'intervalle de confiance. Ainsi l'intervalle de confiance de 160 à 240 nous indique que la moyenne d'échantillon est statistiquement significativement différente d'une moyenne de population hypothétique de 250 (parce que cet intervalle de confiance n'inclut pas la valeur de 250).

Enfin il existe des solutions plus radicales aux problèmes du test de l'hypothèse nulle qui impliquent d'utiliser des approches très différentes des statistiques inférentielles. Les **statistiques bayésiennes** par exemple consistent à spécifier la probabilité que l'hypothèse nulle et toutes les hypothèses alternatives importantes soient vraies avant de mener l'étude, puis mener l'étude pour ensuite mettre à jour les probabilités en fonction des données. Il est trop tôt pour dire si cette approche deviendra courante dans la recherche en psychologie. Pour l'instant, le test de l'hypothèse nulle — accompagnée de mesures de la taille de l'effet et d'intervalles de confiance — demeure l'approche dominante.

Description de l'image

Figure 13.3 description longue:

Tableau détaillant les deux types d'erreurs lors du test de l'hypothèse nulle. Le tableau ressemble à ceci:

Décision	L'état réel du monde	
	H_0 faux	H_0 vrai
Rejeter H_0	Décision correcte	Erreur de type I
Retenir H_0	Erreur de type II	Décision correcte

[\[Retourner à la Figure 13.3\]](#)

Figure 13.4 description longue

Représentation d'une erreur de type I (faux positif) et d'une erreur de type II (faux négatif). Dans l'erreur de type I, un médecin appuie un stéthoscope sur l'abdomen d'un patient et dit : « Vous êtes enceinte ». Dans l'erreur de type II, un médecin appuie un stéthoscope sur l'abdomen gonflé d'une patiente et dit, « Vous n'êtes pas enceinte. » [Retourner à la Figure 13.4\]](#)

72. De la « crise de répliquabilité » aux pratiques de science ouverte

Objectifs d'apprentissage

1. Décrire ce que signifie la « crise de répliquabilité » en psychologie.
2. Décrire quelques pratiques de recherche douteuses.
3. Identifier quelques moyens d'améliorer la rigueur scientifique.
4. Comprendre l'importance de la transparence dans la science psychologique.

Au début de ce livre, nous avons évoqué le « Projet de répliquabilité en plusieurs laboratoires » (en anglais "[Many Labs Replication Project](#)") qui n'a pas réussi à répliquer la découverte originale de Simone Schnall et de ses collègues selon laquelle le fait de se laver les mains incite les gens à considérer les transgressions morales comme moins répréhensibles (Schnall, Benton, et Harvey, 2008). Bien que ce projet illustre la nature collaborative et autocorrective de la science, il représente également une réponse spécifique à la récente « **crise de répliquabilité** » en psychologie, une expression qui renvoie à l'incapacité des scientifiques à reproduire les résultats de recherches antérieures. Considérons par exemple les résultats du [projet de répliquabilité](#), ayant mobilisé plus de 270 psychologues dans le monde entier coordonnant leurs efforts pour tester la fidélité de 100 expériences déjà publiées en psychologie (Aarts et al., 2015). Bien que 97 des 100 études originales aient trouvé des effets statistiquement significatifs, seulement 36 des répliquations en ont trouvé ! De plus la taille des effets observés dans les répliquations était en moyenne la moitié des effets rapportés dans les études originales. Bien entendu l'échec à répliquer un résultat en soi ne discrédite pas nécessairement l'étude originale, car des différences au niveau de la puissance statistique, des populations échantillonnées, des procédures utilisées ou encore la présence de variables modératrices pourraient expliquer les différents résultats (Yong, 2015).

Bien que plusieurs pensent que l'échec à répliquer les résultats de la recherche soit une caractéristique attendue du progrès scientifique cumulatif, d'autres y voient une preuve de problèmes systémiques dans la recherche en psychologie, notamment un biais de publication, qui favorise la découverte et la publication de résultats contre-intuitifs mais statistiquement significatifs, au lieu du processus plus ennuyeux (mais incroyablement vital) de répliquer les résultats antérieurs pour tester leur robustesse (Aschwanden, 2015 ; Frank, 2015 ; Pashler et Harris, 2012 ; Scherer, 2015). Pire encore on suggère que la faible répliquabilité de nombreuses études serait le signe de l'utilisation répandue de pratiques de recherche douteuses par les scientifiques en psychologie. Il peut s'agir de :

1. La suppression sélective des valeurs aberrantes afin d'influencer (souvent en gonflant artificiellement) les relations statistiques entre les variables mesurées.

2. La présentation sélective des résultats, en ne retenant que les résultats qui confirment les hypothèses.
3. L'exploration des données sans hypothèses *a priori*, puis la prétention qu'un résultat statistiquement significatif avait été prédit à l'origine, une pratique appelée «**HARKing**» (Hypothesizing After the Results are Known en anglais) ou émission d'hypothèses après que les résultats soient connus (Kerr, 1998).
4. Une pratique connue sous le nom de « [p-hacking](#) » (brièvement abordée dans la section précédente), dans laquelle des scientifiques pourraient effectuer des calculs statistiques inférentielles pour voir si un résultat est significatif avant de décider de recruter des personnes participantes supplémentaires et de collecter davantage de données (Head, Holman, Lanfear, Kahn et Jennions, 2015). Comme vous l'avez appris, la probabilité de trouver un résultat statistiquement significatif est influencée par le nombre de personnes participant à l'étude.
5. La fabrication pure et simple de données (comme dans le cas de Diederik Stapel décrit au début du chapitre 3), bien qu'il s'agisse d'un cas de fraude plutôt que d'une « pratique de recherche ».

Il est important de faire la lumière sur ces pratiques de recherche douteuses pour s'assurer que les scientifiques d'aujourd'hui et de demain (comme vous) comprennent les dommages qu'elles causent à l'intégrité et à la réputation de notre discipline (visionner la vidéo [Répliquer des études scientifiques pour tester leur fiabilité](#) de Radio Canada Info ou lire par exemple, « [l'index de répliquabilité](#) [en anglais] », un « test de dopage » statistique mis au point par Ulrich Schimmack en 2014 pour estimer la répliquabilité des études, des revues et même de scientifiques spécifiques). Cependant en plus de mettre en évidence ce qu'il ne faut pas faire, cette soi-disant « crise » a également souligné l'importance d'améliorer la rigueur scientifique :

1. La conception et la réalisation d'études ayant une puissance statistique suffisante, afin d'accroître la fidélité des résultats.
2. La publication des résultats nuls comme des résultats significatifs (ce qui permet de réduire le biais de publication et le problème du tiroir à dossiers).
3. La description détaillée des devis de recherche pour permettre à d'autres scientifiques de reproduire l'étude en suivant une procédure identique ou très similaire.
4. La réalisation et la publication de répliquations de haute qualité (Brandt et al., 2014).

Une réponse particulièrement prometteuse à la crise de la répliquabilité a été l'émergence de **pratiques de science ouverte** qui visent à accroître la transparence et l'ouverture du processus scientifique. Par exemple *Psychological Science* (la revue phare de l'[Association for Psychological Science](#)) et d'autres revues attribuent désormais des badges numériques aux scientifiques qui ont préenregistré leurs hypothèses et leurs plans d'analyse des données, partagé ouvertement leur matériel de recherche avec d'autres scientifiques (par exemple pour permettre des tentatives de répliquabilité) ou mis leurs données brutes à la disposition des autres scientifiques (voir la Figure 13.6).



Figure 13.6 Badges de reconnaissance de pratiques transparentes.

Ces initiatives menées par le « [Center for Open Science](#) » ont conduit à l'élaboration de « lignes directrices pour la promotion de la transparence et de l'ouverture » (voir une traduction au Tableau 13.7 du [TOP 2025](#) pour Transparency and Openness Promotion) qui ont depuis été formellement adoptées par plus de 3000 revues et 50 organisations, une liste qui s'allonge chaque semaine. Si l'on ajoute à cela les exigences récemment imposées par les organismes de financement fédéraux au Canada (les trois conseils) et aux États-Unis (National Science Foundation) concernant la publication des recherches financées par des fonds publics dans des revues en libre accès, il semble que l'avenir de la science et de la psychologie soit résolument tourné vers une plus grande « ouverture ».

Pour en lire davantage sur la science ouverte en psychologie, consultez Thériault (2024). L'importance de la science ouverte en recherche en psychologie. Accepté dans *Fragments : Revue de psychologie*. Preprint: <https://doi.org/10.31234/osf.io/758dx>

Tableau 13.7 Promotion de la transparence et l'ouverture : Lignes directrices

Pratiques de recherche			
Critère	Niveau 1: Divulger	Niveau 2: Partager et citer	Niveau 3: Certifier
Enregistrement de l'étude	Les autrices et auteurs indiquent si l'étude a été enregistrée et, le cas échéant, où et quand elle l'a été.	Les scientifiques enregistrent l'étude et citent cet enregistrement.	Une partie indépendante de scientifiques certifie que l'étude a été enregistrée au moment approprié et que l'enregistrement est complet, conformément aux bonnes pratiques pour le devis de recherche.
Protocole de l'étude	Les auteurs et autrices indiquent si le protocole de l'étude est disponible et, le cas échéant, où et quand il a été partagé.	Les scientifiques partagent publiquement le protocole de l'étude et le citent.	Une partie indépendante de scientifiques certifie que le protocole a été partagé au moment approprié et qu'il est complet, conformément aux bonnes pratiques pour le devis de recherche.
Plan d'analyse	Les autrices et auteurs indiquent si le plan d'analyse est disponible et, le cas échéant, où et quand il a été partagé.	Les scientifiques partagent publiquement le plan d'analyse et le citent.	Une partie indépendante de scientifiques certifie que le plan d'analyse a été partagé au moment approprié et qu'il est complet, conformément aux bonnes pratiques pour le devis de recherche.
Transparence du matériel	Les auteurs et autrices indiquent si le matériel est disponible et, le cas échéant, où.	Les scientifiques citent le matériel déposé dans un dépôt fiable par eux-mêmes ou par d'autres.	Une partie indépendante de scientifiques certifie que le matériel a été déposé et documenté conformément aux bonnes pratiques relatives au type de matériel.
Transparence des données	Les autrices et auteurs indiquent si les données sont disponibles et, le cas échéant, où.	Les scientifiques citent les données déposées dans un dépôt fiable par eux-mêmes ou par d'autres.	Une partie indépendante de scientifiques certifie que les données ont été déposées avec les métadonnées, conformément aux bonnes pratiques relatives au type de données.
Transparence du code analytique	Les autrices et auteurs indiquent si le code analytique est disponible et, le cas échéant, où.	Les scientifiques citent le code analytique déposé dans un dépôt fiable par eux-mêmes ou par d'autres.	Une partie indépendante de scientifiques certifie que le code analytique a été déposé et documenté, conformément aux bonnes pratiques pertinentes.
Transparence du rapport	Les auteurs et autrices indiquent si une ligne directrice de rédaction a été utilisée — et, le cas échéant, laquelle.	Les auteurs et autrices partagent publiquement une liste de vérification complétée de la ligne directrice de rédaction et citent cette ligne directrice.	Une partie indépendante de scientifiques certifie que les chercheuses et chercheurs ont respecté la ligne directrice appropriée pour le devis de recherche.
Transparences des résultats	Une partie indépendante de scientifiques vérifie que les résultats n'ont pas été rapportés de façon sélective en fonction de leur nature. Pour ce faire, la partie indépendante peut vérifier que l'enregistrement de l'étude, le protocole et le plan d'analyse correspondent au rapport final — et que le rapport final reconnaît toute déviation.		
Réplicabilité computationnelle	Une partie indépendante de scientifiques vérifie que les résultats rapportés peuvent être répliqués en utilisant les mêmes données et en suivant les mêmes procédures computationnelles. Pour ce faire, la partie indépendante peut vérifier qu'elle obtient les mêmes résultats en utilisant les données et le code déposés dans un dépôt fiable.		
Réplication	Une étude visant à fournir des preuves diagnostiques sur les affirmations d'une étude antérieure en répliquant les procédures originales avec un nouvel échantillon.		

Pratiques de recherche			
Critère	Niveau 1: Divulger	Niveau 2: Partager et citer	Niveau 3: Certifier
Rapport enregistré	Une étude enregistrée dans laquelle le protocole et le plan d'analyse sont évalués par les pairs, et l'étude est préacceptée par une revue scientifique avant même d'être menée.		
Multivers	Une étude dans laquelle une même équipe de recherche examine la question de recherche à travers différents choix raisonnables de traitement et d'analyse des mêmes données.		
Plusieurs analystes	Une étude dans laquelle des équipes d'analyse indépendantes effectuent des analyses alternatives plausibles d'une même question de recherche sur le même ensemble de données.		

Attributions aux médias

Thériault, R. (2025). Les badges numériques en français tirés de « Bonnes pratiques et science ouverte : Un document d'accompagnement au PRD », disponible au https://remi-theriault.com/files/science_ouverte.pdf. Reproduite avec la permission de l'auteur.

Grant, S., Corker, K. S., Mellor, D., Stewart, S. L. K., Cashin, A. G., Lagisz, M., Mayo-Wilson, E., Moher, D., Umpierre, D., Barbour, V., Buck, S., Collins, G. S., Hazlett, H. F., Hrynaskiewicz, I., Lee, C. J., Parker, T. H., Rethlefsen, M. L., Toomey, E. and Nosek, B. A. (2025). *TOP 2025: An Update to the Transparency and Openness Promotion Guidelines*. [Center for Open Science](https://osf.io/preprints/metaarxiv/nmfs6_v2). MetaArXiv Preprints https://osf.io/preprints/metaarxiv/nmfs6_v2

Vu Van, B. A. (journaliste), (2025, 16 mars). Les Jeux de la réplication: mettre des études à l'épreuve. Dans Lemyre, É. et Caron, S. (réalisation), Découverte, *Radio-Canada Info*. Repéré le 24 août 2025. https://www.youtube.com/watch?v=QcC2cFG_1so

73. Points clés et exercices

Points clés

- Le test de l'hypothèse nulle est une approche formelle permettant de décider si une relation statistique observée dans un échantillon reflète une relation réelle dans la population ou si elle est simplement due au hasard.
- La logique du test de l'hypothèse nulle consiste à supposer que l'hypothèse nulle est vraie, à déterminer la probabilité du résultat de l'échantillon si cette hypothèse est correcte, puis à prendre une décision. Si le résultat de l'échantillon était improbable si l'hypothèse nulle était vraie, alors celle-ci est rejetée en faveur de l'hypothèse alternative. Si le résultat n'est pas improbable, l'hypothèse nulle est retenue.
- La probabilité d'obtenir le résultat de l'échantillon si l'hypothèse nulle était vraie (la valeur p) repose sur deux facteurs : la force de la relation et la taille de l'échantillon. Il est souvent possible de juger raisonnablement si la relation trouvée avec un échantillon est statistiquement significative en considérant rapidement ces deux facteurs.
- La signification statistique n'est pas la même chose que la force ou l'importance de la relation. Même les relations faibles peuvent être statistiquement significatives si la taille de l'échantillon est suffisamment grande. Il est donc essentiel de considérer à la fois la force de la relation et la signification pratique d'un résultat, en plus de sa signification statistique.
- Pour comparer deux moyennes, le test de l'hypothèse nulle le plus courant est le test t . Le test t à un échantillon est utilisé pour comparer la moyenne d'un échantillon à une moyenne hypothétique de la population, le test t à échantillons dépendants est utilisé pour comparer deux moyennes dans un devis intrasujet, et le test t à échantillons indépendants est utilisé pour comparer deux moyennes dans un devis intersujets.
- Pour comparer plus de deux moyennes, le test d'hypothèse nulle le plus courant est l'analyse de la variance (ANOVA). L'ANOVA à un facteur est utilisée pour les devis intersujets avec une seule variable indépendante, l'ANOVA à mesures répétées est utilisée pour les devis intrasujets et l'ANOVA factorielle est utilisée pour les plans factoriels.
- Un test de l'hypothèse nulle du r de Pearson est utilisé pour comparer une valeur observée du r de Pearson dans l'échantillon avec une valeur hypothétique de la population égale à 0.
- La décision de rejeter ou de retenir l'hypothèse nulle n'est pas une garantie qu'elle soit correcte. Une erreur de type I se produit lorsque l'on rejette l'hypothèse nulle alors qu'elle est vraie. Une erreur de type II se produit lorsque l'on ne rejette pas l'hypothèse nulle alors qu'elle est fausse.
- La puissance statistique d'un devis de recherche est la probabilité de rejeter l'hypothèse nulle compte

tenu de la force attendue de la relation dans la population et de la taille de l'échantillon. Les scientifiques doivent s'assurer que leurs études disposent d'une puissance statistique suffisante avant de les mener.

- Le test de l'hypothèse nulle a fait l'objet de critiques, en raison de la mauvaise compréhension qu'en ont les scientifiques, de sa logique jugée discutable, et qu'il est peu informatif. D'autres soutiennent qu'il remplit une fonction importante, en particulier lorsqu'il est utilisé conjointement avec les mesures de taille de l'effet, les intervalles de confiance et d'autres techniques. Il demeure l'approche dominante des statistiques inférentielles en psychologie.
- Au cours des dernières années, la psychologie s'est trouvée confrontée à l'incapacité de répliquer des résultats des recherches. Plusieurs l'ont interprétée comme un aspect normal du progrès scientifique, mais d'autres ont suggéré que cela mettait en évidence des problèmes découlant de pratiques de recherche douteuses.
- Une des réponses à cette « crise de réplabilité » a été l'émergence de pratiques de science ouverte, qui favorisent la transparence et l'ouverture du processus de recherche. Ces pratiques incluent notamment l'attribution de badges numériques pour encourager le préenregistrement des hypothèses ainsi que le partage des données brutes et du matériel de recherche.

Exercices

- Discussion : Imaginez une étude montrant que les personnes qui mangent plus de brocoli ont tendance à être plus heureuses. Expliquez à quelqu'un qui ne connaît rien en statistiques pourquoi les scientifiques effectueraient un test de l'hypothèse nulle.
- Pratique : Utilisez le Tableau 13.1 pour décider si chacun des résultats suivants est statistiquement significatif.
 - La corrélation entre deux variables est $r = -.78$ sur la base d'un échantillon de 137 personnes.
 - Le score moyen d'une caractéristique psychologique est de 25 (écart-type = 5) chez les femmes et le score moyen est de 24 (écart-type = 5) chez les hommes. Il y a 12 femmes et 10 hommes dans cette étude.
 - Dans une expérience en mémoire, le nombre moyen d'éléments rappelés par les 40 personnes participantes de la Condition A est supérieur de 0,50 écart-type au nombre moyen rappelé par les 40 personnes de la Condition B.
 - Dans une autre expérience en mémoire, les scores moyens des personnes participantes dans les conditions A et B sont exactement les mêmes !
 - Une personne étudiante trouve une corrélation de $r = 0.04$ entre le nombre de cours suivis par la cohorte étudiante de sa classe de méthodes de recherche et leur niveau de stress.

- Pratique : Utilisez l'un des outils en ligne, Excel, JASP ou SPSS pour reproduire le test t à un échantillon, le test t à échantillons dépendants, le test t à échantillons indépendants et l'ANOVA à un facteur pour les quatre ensembles de données d'estimation des calories présentés dans cette section.
- Pratique : Un échantillon de 25 personnes étudiant à l'université a évalué leur amitié sur une échelle de 1 (*beaucoup plus faible que la moyenne*) à 7 (*beaucoup plus élevée que la moyenne*). Leur score moyen était de 5.30 avec un écart-type de 1.50. Effectuez un test t à un échantillon pour comparer leur score moyen à une moyenne hypothétique de 4 (*la moyenne*). La question est de savoir si les personnes étudiant à l'université ont tendance à se considérer comme plus amicales que la moyenne.
- Pratique : Déterminez si chacune des valeurs de r de Pearson suivantes est statistiquement significative pour un test unilatéral et un test bilatéral.
 - La corrélation entre la taille et le QI est de + 0.13 dans un échantillon de 35 personnes.
 - Pour un échantillon de 88 personnes étudiant à l'université, la corrélation entre le degré de dégoût ressenti et la sévérité de leurs jugements moraux est de + 0.23.
 - La corrélation entre le nombre de tracas quotidiens et l'humeur positive est de – 0.43 pour un échantillon de 30 adultes d'âge moyen.
- Discussion : Des scientifiques comparent l'efficacité de deux formes de psychothérapie pour la phobie sociale à l'aide d'un test t à échantillons indépendants.
 - Expliquez ce que cela signifierait pour les scientifiques de commettre une erreur de type I.
 - Expliquez ce que cela signifierait pour les scientifiques de commettre une erreur de type II.
- Discussion : Imaginez que vous réalisez un test t et que la valeur p est de 0.02. Comment pourriez-vous expliquer la signification de cette valeur p à une personne qui n'est pas familière avec le test de l'hypothèse nulle ? Assurez-vous d'éviter les interprétations erronées courantes de la valeur p .
- Pour de la pratique additionnelle avec les erreurs de type I et de type II, essayez [ces problèmes](#) de Carnegie Mellon [en anglais]
- Discussion : Selon vous, quels sont les principaux avantages de l'adoption des pratiques de science ouverte, telles que l'enregistrement préalable et le partage des données brutes et du matériel de recherche ? Pouvez-vous identifier certains inconvénients de ces pratiques ?
- Pratique : Lisez l'article en ligne « [Science isn't broken : It's just a hell of a lot more hard than we give it credit for](#) » [en anglais] et utilisez l'outil interactif intitulé « Hack your way to scientific glory » afin de mieux comprendre la mauvaise conduite scientifique du « p-hacking ».

74. Références

- Abelson, R. P. (1995). *Statistics as principled argument*. Mahwah, NJ: Erlbaum.
- Aarts, A. A., Anderson, C. J., Anderson, J., van Assen, M. A. L. M., Attridge, P. R., Attwood, A. S., ... Zuni, K. (2015, 21 septembre). *Reproducibility Project: Psychology*. Récupéré de osf.io/ezcuj
- Aschwanden, C. (2015, 19 août). *Science isn't broken: It's just a hell of a lot harder than we give it credit for*. Fivethirtyeight. Repéré dans <http://fivethirtyeight.com/features/science-isnt-broken/>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... can't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Cohen, J. (1994). The world is round: $p < .05$. *American Psychologist*, 49, 997-1003.
- Découverte (2025). *Répliquer des études scientifiques pour tester leur fiabilité*. Radio-Canada Info. https://www.youtube.com/watch?v=QcC2cFG_1so&t=585s
- Frank, M. (2015, 31 août). *The slower, harder ways to increase reproducibility*. Repéré dans <http://babieslearninglanguage.blogspot.ie/2015/08/the-slower-harder-ways-to-increase.html>
- Head M. L., Holman, L., Lanfear, R., Kahn, A. T., et Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol*, 13(3): e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16, 259-263.
- Kanner, A. D., Coyne, J. C., Schaefer, C., et Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine*, 4, 1-39.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196-217. https://doi.org/10.1207/s15327957pspr0203_4
- Lakens, D. (2017, 25 décembre). *About p-values: Understanding common misconceptions*. <https://correlaid.org/en/blog/understand-p-values/>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., et Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Chichester, UK: Wiley.

- Pashler, H., et Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments explained. *Perspectives on Psychological Science*, 7(6), 531-536. <https://doi.org/10.1177/1745691612463401>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 83, 638-641.
- Scherer, L. (2015, septembre). *Guest post by Laura Scherer*. Repéré dans <http://sometimesimwrong.typepad.com/wrong/2015/09/guest-post-by-laura-scherer.html>
- Schnall, S., Benton, J., et Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19(12), 1219-1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>
- Simonsohn U., Nelson L. D., et Simmons J. P. (2014). P-Curve: a key to the file drawer. *Journal of Experimental Psychology: General*, 143(2), 534-547. <https://doi.org/10.1037/a0033242>
- Thériault (2024). L'importance de la science ouverte en recherche en psychologie. Accepté dans *Fragments : Revue de psychologie*, 1-14. Preprint: <https://doi.org/10.31234/osf.io/758dx>
- Tramimow, D. et Marks, M. (2015). *Editorial. Basic and Applied Social Psychology*, 37, 1-2. <https://dx.doi.org/10.1080/01973533.2015.1012991>
- Wilkinson, L., et Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604
- Yong, E. (27 août, 2015). *How reliable are psychology studies?* Repéré dans <http://www.theatlantic.com/science/archive/2015/08/psychology-studies-reliability-reproducibility-nosek/402466/>

Glossaire

α (alpha)

Le critère qui indique à quel point une valeur p doit être basse avant que le résultat de l'échantillon soit considéré comme suffisamment improbable pour rejeter l'hypothèse nulle (habituellement fixé à 0,05).

α de Cronbach

Une statistique qui mesure la cohérence interne entre les éléments d'une mesure.

alternance

utilisation tantôt du féminin, tantôt du masculin

Analyse appliquée du comportement

Une application des principes de l'analyse expérimentale du comportement qui joue un rôle important dans la recherche contemporaine sur les troubles du développement, l'éducation, le comportement organisationnel et la santé, parmi de nombreux autres domaines d'application.

Analyse de contenu

Une famille d'approches systématiques de la mesure utilisant des méthodes qualitatives pour analyser des données d'archives complexes.

Analyse de variance (ANOVA)

Un test statistique utilisé lorsqu'il y a plus de deux groupes ou moyennes de conditions à comparer.

Analyse documentaire

Décrit les recherches antérieures pertinentes sur le sujet et peut aller de quelques paragraphes à plusieurs pages.

Analyse expérimentale du comportement

Un sous-domaine de la psychologie (behaviorisme) qui se concentre exclusivement sur les effets des récompenses, des punitions et d'autres facteurs externes sur le comportement.

Analyse exploratoire

Une analyse utilisée pour examiner la possibilité qu'il y ait des relations dans les données que vous n'avez pas supposées.

Analyse factorielle

Technique statistique complexe dans laquelle les personnes en recherche étudient les relations entre un grand nombre de variables conceptuellement similaires.

Analyse planifiée

Utilisée pour tester une relation que vous attendiez dans votre hypothèse.

Anonymat

Lorsque le nom d'une personne participante et d'autres informations personnelles identifiables ne sont pas collectés.

ANOVA à mesures répétées

Compare les moyennes des mêmes personnes participantes testées dans des conditions différentes ou à des moments différents dans laquelle la variable dépendante est mesurée plusieurs fois pour chaque personne.

ANOVA à un facteur

Utilisée pour les devis intersujets avec une seule variable indépendante.

ANOVA factorielle

Méthode statistique permettant de détecter les différences de moyennes entre les conditions lorsqu'il y a au moins deux variables indépendantes dans un plan factoriel. Elle permet de détecter les effets principaux et les effets d'interaction.

Article théorique

Un article de synthèse qui se consacre principalement à la présentation d'une nouvelle théorie.

Articles de synthèse

Articles qui résument des recherches déjà publiées sur un sujet et qui présentent généralement de nouvelles façons d'organiser ou d'expliquer les résultats.

Assignment aléatoire

Il s'agit d'utiliser un processus aléatoire pour décider quelles personnes participants sont testées dans quelles conditions.

Asymétrie

Lorsqu'un histogramme présente un sommet déplacé vers l'extrémité supérieure de son étendue avec une queue relativement longue du côté négatif (*asymétrie négative*), ou lorsque le sommet est déplacé

vers l'extrémité inférieure de son étendue avec une queue relativement longue du côté positif (*asymétrie positive*).

Autonomie

Le droit d'une personne de faire ses propres choix et d'entreprendre ses propres actions sans contrainte.

Avantages

Souligne l'importance de maximiser les avantages de la recherche tout en minimisant les risques pour les personnes participantes et la société.

Barres d'erreur

Barres représentant la variabilité dans chaque groupe ou condition.

Biais d'échantillonnage

Se produit lorsqu'un échantillon est sélectionné de telle manière qu'il n'est pas représentatif de l'ensemble de la population et qu'il produit donc des résultats inexacts.

Biais de confirmation

Tendance à se concentrer sur les cas qui confirment nos croyances intuitives et à négliger les cas qui ne confirment pas nos croyances.

Biais de non-réponse

Se produit lorsqu'il existe une différence systémique entre les non-répondants et les répondants à l'enquête.

Bloc de randomisation

Toutes les conditions apparaissent une fois dans la séquence avant que l'une d'elles ne soit répétée.

BRUSO

Acronyme qui signifie "bref", "pertinent", "sans ambiguïté", "spécifique" et "objectif", utilisé pour créer des questions de questionnaire efficaces, brèves et directes.

Cadre de l'échantillonnage

Liste de tous les membres de la population à partir de laquelle les personnes interrogées seront sélectionnées.

Caractéristiques intégrées de la demande

Indices subtils qui révèlent aux personnes participantes la manière dont la personne responsable de la recherche s'attend à ce qu'elles répondent dans l'expérience.

Carrés moyens à l'intérieur des groupes (CMintra)

Une estimation de la variance de la population, basée sur les différences entre les scores au sein de chaque groupe.

Carrés moyens entre les groupes (CMinter)

Une estimation de la variance de la population basée sur les différences entre les moyennes des échantillons.

Citation de référence

Une citation dans le texte renvoyant à l'ouvrage où cette idée est apparue à l'origine et une référence complète à cet ouvrage dans la liste des références.

Code de Nuremberg

Ensemble de 10 principes éthiques pour la recherche, rédigés en 1947 à l'occasion du procès de Nuremberg des médecins nazis accusés de crimes de guerre contre les prisonniers des camps de concentration.

Code éthique de l'APA

Il s'agit des principes éthiques des psychologues et du code de conduite de l'APA. Il a été publié pour la première fois en 1953 et comprend environ 150 normes éthiques spécifiques que les psychologues et leurs personnes étudiantes sont censés respecter.

Codification

Une partie de l'observation structurée qui consiste, pour les personnes qui observent, à utiliser un ensemble de directives clairement définies pour « coder » les comportements — c'est-à-dire classer les comportements spécifiques qu'ils observent dans une catégorie — et compter le nombre de fois ou la durée pendant laquelle le comportement se produit.

Coefficient de corrélation

Décrit la force et la direction de la relation entre deux variables (souvent mesurée par le r de Pearson).

Coefficient de corrélation de Pearson (ou r de Pearson)

Une statistique qui mesure la force d'une corrélation entre des variables quantitatives.

Cohérence interne

La constance des réponses des personnes aux différents items d'une mesure comportant plusieurs items.

Communication orale

La personne chargée de la présentation se tient devant un public d'autres personnes en recherche et leur parle de ses recherches.— généralement à l'aide d'un diaporama.

Communication par affiche

Une autre façon de présenter une recherche lors d'un congrès consiste à utiliser une affiche de grande taille qui illustre et résume l'étude réalisée par l'équipe de recherche.

Comparaisons post hoc

Un test non planifié (non formulé comme hypothèse) permettant de déterminer quelles paires de moyennes de groupe diffèrent les unes des autres.

Complice

Une personne aidant qui se fait passer pour une véritable personne participante à une étude.

Condition contrôle

La condition dans laquelle les personnes participantes ne reçoivent pas le traitement.

Condition contrôle placebo

Condition dans laquelle les personnes participantes reçoivent un placebo plutôt que le traitement.

Condition contrôle sans traitement

Condition dans laquelle les personnes participantes ne reçoivent aucun traitement.

Condition contrôle sur liste d'attente

Condition dans laquelle les personnes participantes sont informées qu'elles recevront le traitement mais elles doivent attendre que les personnes participantes de la condition traitement l'aient déjà reçu.

Condition traitement (ou condition de traitement)

La condition dans laquelle les personnes participantes reçoivent le traitement.

Conditions

Les différents niveaux de la variable indépendante attribués aux personnes participantes.

Confidentialité

Un engagement à ne pas divulguer les informations personnelles des personnes participantes sans leur consentement ou sans une autorisation légale appropriée.

Confounding variable

An extraneous variable that varies systematically with the independent variable, and thus confuses the effect of the independent variable with the effect of the extraneous one.

Congrès scientifique

Un congrès qui peut aller d'un petit à un grand événement, au cours duquel des personnes responsables de recherches en psychologie partagent leurs travaux entre elles au moyen de présentations.

Conseil d'examen institutionnel (CEI)

Un comité chargé d'examiner les protocoles de recherche pour déceler d'éventuels problèmes éthiques.

Consentement éclairé

Cela signifie que les personnes menant la recherche obtiennent et documentent l'accord des personnes pour participer à une étude après les avoir informées de tout ce qui pourrait raisonnablement affecter leur décision.

Construits (psychologiques)

Variables psychologiques représentant l'état mental ou l'expérience d'un individu, souvent non directement observables, telles que les traits de personnalité, les états émotionnels, les attitudes et les habiletés.

Contrebalancement

Variation de l'ordre des conditions dans lesquelles les personnes participantes sont testées pour résoudre le problème des effets d'ordre dans les expériences intrasujets.

Contrebalancement aléatoire

Méthode dans laquelle l'ordre des conditions est déterminé de manière aléatoire pour chaque personne participante.

Contrebalancement complet

Méthode dans laquelle un nombre égal de personnes participantes effectue chaque ordre possible des conditions.

Contrôle

Maintenir les variables externes constantes afin de séparer l'effet de la variable indépendante de l'effet des variables externes.

Contrôle statistique

Contrôler des variables tierces potentielles afin d'exclure d'autres interprétations plausibles.

Corrélation moitié-moitié

Un score obtenu en divisant les items d'une mesure en deux ensembles et en examinant la relation entre les deux ensembles de scores, afin d'évaluer la cohérence interne de la mesure.

Corrélation partielle

Méthode de contrôle des variables externes en les mesurant et en les incluant dans l'analyse statistique.

Corrélations faussées (ou corrélations illusoires)

Corrélations qui résultent non pas des deux variables mesurées, mais d'une troisième, variable non mesurée qui affecte les deux variables mesurées.

Crise de répliquabilité

Une expression qui fait référence à l'incapacité des personnes menant des recherches à répliquer les résultats de recherches antérieures.

Critère

Variable qui devrait théoriquement être corrélée avec le construit mesuré (pluriel : critères).

d de Cohen

La mesure de la taille de l'effet la plus largement utilisée pour les différences entre les moyennes de groupe ou de condition, qui est la différence entre les deux moyennes divisée par l'écart-type.

Débriefage

C'est le processus d'informer les personnes participantes le plus vite possible de l'objectif de l'étude, de révéler toute déception, et de régler toutes idées fausses qu'ils pourraient avoir à la suite de leur participation.

Déception

Désinformer les personnes participantes sur l'objectif d'une étude en ayant recours à des complices, en utilisant de l'équipement bidon comme le générateur de chocs de Milgram et la présentation de faux retour d'information à propos de leur performance (par exemple, leur dire qu'elles ont eu une mauvaise note à un test alors qu'en réalité, elles ont obtenu de bons résultats).

Déclaration d'Helsinki

Code d'éthique créé par le Conseil médical mondial en 1964.

Définition conceptuelle

Décrit les comportements et les processus internes qui constituent les construits psychologiques, ainsi que la façon dont elle est liée à d'autres variables.

Définition opérationnelle

Manière précise de définir une variable en indiquant comment elle mesurée.

Devis à groupes appariés

Un devis expérimental dans lequel les personnes participantes dans les différentes conditions sont appariées en fonction de la variable dépendante ou d'une ou de plusieurs variables externes, avant la manipulation de la variable indépendante.

Devis à groupes non équivalents

Un devis intersujets dans lequel les personnes participantes n'ont pas été assignées aléatoirement aux conditions.

Devis à un facteur avec deux niveaux

Un devis d'expérience impliquant une seule variable indépendante avec deux conditions.

Devis à un facteur avec plusieurs niveaux

Lorsqu'une expérience comporte une variable indépendante qui est manipulée pour produire plus de deux conditions.

Devis à un groupe avec post-test seulement

Un traitement est introduit (ou une variable indépendante est manipulée) et une variable dépendante est mesurée une fois le traitement introduit.

Devis à un groupe prétest post-test

Un devis d'expérience dans lequel la variable dépendante est mesurée une fois avant le début du traitement et une fois après sa mise en place.

Devis ABA (plan ABA)

Autre terme pour désigner le protocole d'inversion.

Devis de réplication alternée avec retrait du traitement

Dans ce devis, le traitement est retiré au premier groupe lorsqu'il est ajouté au second groupe.

Devis de traitements alternés

Dans ce devis, deux traitements ou plus sont alternés relativement rapidement selon un horaire régulier.

Devis post-test seulement avec groupes non équivalents

Les personnes participantes d'un groupe sont exposées à un traitement, un groupe non équivalent n'est pas exposé au traitement, puis les deux groupes sont comparés.

Devis prétest-post-test avec groupes non équivalents

Dans ce devis, il y a un groupe traitement qui reçoit un prétest, un traitement et un post-test. Parallèlement, il existe un groupe contrôle non équivalent qui reçoit un prétest, ne reçoit pas le traitement et reçoit un post-test.

Devis prétest-post-test avec réplication alternée

Dans ce devis, des groupes non équivalents sont soumis à un prétest de la variable dépendante, puis un groupe reçoit un traitement tandis qu'un groupe contrôle non équivalent ne reçoit pas de traitement, la variable dépendante est évaluée à nouveau, puis le traitement est ajouté au groupe contrôle, et enfin la variable dépendante est évaluée une dernière fois.

Diagramme à bandes

Présentation graphique de données sous forme de barres de taille variable, généralement utilisée pour présenter et comparer les scores moyens de deux ou plusieurs groupes ou conditions.

Diagramme de dispersion

Graphique qui présente les corrélations entre deux variables quantitatives, l'une sur l'axe des x et l'autre sur l'axe des y. Les scores sont représentés à l'intersection des valeurs sur chaque axe.

Distribution

La façon dont les scores se répartissent entre les différents niveaux d'une variable.

Doctorat de philosophie [Ph.D.]

Diplôme universitaire obtenu grâce à l'étude intensive d'une discipline particulière et à la réalisation d'un ensemble de travaux de recherche qui apportent de nouvelles connaissances à la littérature académique.

Données brutes

Données non analysées qui se présentent sous différentes formes: questionnaires papier-crayons remplis, fichiers informatiques remplis de chiffres ou de textes, vidéos ou notes écrites qu'il faut éventuellement organiser, coder ou combiner d'une manière ou d'une autre.

doublet

présentation des formes masculines et féminines d'un mot genré.

doublets abrégés

combinaison de la forme masculine et féminine d'un mot en utilisant un signe typographique comme le point médian

Écart-type

C'est la distance moyenne entre les scores et la moyenne dans une distribution.

Échantillon

Une petite partie de la population que la personne menant de la recherche souhaite étudier.

Échantillonnage aléatoire simple

Méthode d'échantillonnage probabiliste dans laquelle chaque individu de la population a une probabilité égale d'être sélectionnée pour faire partie de l'échantillon.

Échantillonnage aléatoire stratifié

Une alternative courante à l'échantillonnage aléatoire simple dans laquelle la population est divisée en différents sous-groupes ou "strates" (ce qui généralement basé sur des caractéristiques démographiques), puis un échantillon aléatoire est prélevé dans chaque "strate".

Échantillonnage aléatoire stratifié disproportionné

Est utilisé pour échantillonner des personnes répondantes supplémentaires dans des sous-groupes particulièrement petits — permettant de tirer des conclusions valables sur ces sous-groupes.

Échantillonnage aléatoire stratifié proportionnel

Est utilisé pour sélectionner un échantillon dans lequel la proportion de personnes répondantes dans chacun des différents sous-groupes correspond à la proportion dans la population.

Échantillonnage de commodité

Méthode courante d'échantillonnage non probabiliste dans laquelle l'échantillon est constitué de personnes facilement disponibles et intéressées à participer (par exemple, plusieurs de la population étudiante en psychologie).

Échantillonnage en boule de neige

Une forme d'échantillonnage non probabiliste dans laquelle les personnes participantes existantes à la recherche aident à recruter de nouvelles personnes participantes pour l'étude.

Échantillonnage en grappes

Un type d'échantillonnage probabiliste dans lequel de grandes grappes de personnes sont échantillonnées au hasard, puis les individus au sein de chaque grappe sont échantillonnés au hasard.

Échantillonnage non probabiliste

Se produit lorsque la personne menant la recherche ne peut pas spécifier la probabilité que chaque membre de la population soit sélectionné pour l'échantillon.

Échantillonnage par autosélection

Une forme d'échantillonnage non probabiliste dans laquelle les individus choisissent de participer à la recherche de leur propre gré, sans être contactés directement par la personne menant la recherche.

Échantillonnage par quotas

Forme d'échantillonnage non probabiliste dans laquelle les sous-groupes de l'échantillon sont recrutés de manière proportionnelle aux sous-groupes de la population.

Échantillonnage probabiliste

Se produit lorsque la personne menant la recherche peut spécifier la probabilité que chaque membre de la population soit sélectionné pour l'échantillon.

Échelle d'évaluation

Un ensemble ordonné de réponses parmi lesquelles les personnes participantes doivent choisir.

Effet d'attente de la personne menant l'expérimentation

Lorsque les attentes de la personne menant l'expérimentation quant à la manière dont les personnes participantes « devraient » se comporter influencent effectivement le comportement de ces personnes participantes dans l'expérience.

Effet de cohorte

Des différences entre les groupes peuvent refléter la génération à laquelle appartiennent les personnes plutôt qu'un effet direct de l'âge.

Effet de fatigue

Effet par lequel les personnes participantes exécutent moins bien une tâche dans des conditions ultérieures parce qu'elles sont fatiguées ou s'ennuient.

Effet de l'ordre (effet d'ordre)

Effet qui se produit lorsque les réponses des personnes participantes dans les différentes conditions sont affectées par l'ordre des conditions auxquelles elles ont été exposées.

Effet de pratique

Effet dans lequel les personnes participantes effectuent mieux une tâche dans des conditions ultérieures parce qu'elles ont eu l'occasion de la pratiquer.

Effet de report (ou effet carryover)

Un effet du fait d'avoir été testé-e dans une condition, qui influence ensuite le comportement des personnes participantes dans des conditions ultérieures.

Effet du contexte (ou effet de contraste)

Influences involontaires sur les réponses des personnes répondantes parce qu'elles ne sont pas liées au contenu de l'item mais au contexte dans lequel l'item apparaît.

Effet Hawthorne

Dans le cas d'une observation naturaliste non déguisée, il s'agit d'un type de réactivité : lorsque les gens savent qu'ils sont observés et étudiés, ils peuvent agir différemment de ce qu'ils feraient normalement.

Effet placebo

Effet dû au placebo plutôt qu'au traitement.

Effet principal

L'effet d'une variable indépendante sur la variable dépendante — en faisant la moyenne des niveaux de toutes autres variables indépendantes.

Effets simples

Sont un moyen de décomposer l'interaction pour comprendre précisément ce qui se passe.

Empirisme systématique

L'empirisme renvoie à l'apprentissage fondé sur l'observation, et les scientifiques acquièrent des connaissances sur le monde naturel de façon systématique, en planifiant, en réalisant, en enregistrant et en analysant soigneusement ces observations.

Entrevues

Méthode de recherche qualitative permettant de recueillir des informations longues et détaillées auprès des personnes participantes à l'aide de questions ouvertes structurées, semi-structurées ou non structurées.

Erreur d'échantillonnage

La variabilité aléatoire d'une statistique d'un échantillon à l'autre.

Erreur de type I

Un faux positif dans lequel l'équipe de recherche conclut que les résultats sont statistiquement significatifs alors qu'en réalité il n'existe aucun effet réel dans la population et que les résultats sont dus au hasard. En d'autres termes, il s'agit de rejeter l'hypothèse nulle alors qu'elle est en fait vraie.

Erreur de type II

Une occasion manquée dans laquelle l'équipe de recherche conclut que les résultats ne sont pas statistiquement significatifs alors qu'en réalité il existe un effet réel dans la population et qu'elle a

simplement échoué à le détecter. En d'autres termes, il s'agit de retenir l'hypothèse nulle alors qu'elle est en fait fausse.

Erreur standard

L'écart-type du groupe est divisé par la racine carrée de la taille de l'échantillon du groupe.

Essai clinique randomisé

Une expérience qui étudie l'efficacité des psychothérapies et des traitements médicaux.

Étendue

Une mesure de dispersion qui mesure la distance entre les scores les plus élevés et les plus bas dans une distribution.

Éthique

La branche de la philosophie qui s'intéresse à la moralité — ce que signifie un comportement moral et comment les gens peuvent atteindre cet objectif.

Étude de cas

Examen approfondi d'un individu.

Étude en double aveugle

Une méthode pour réduire le biais de la personne menant l'expérimentation, lorsque ni la personne participante ni la personne menant l'expérimentation ne connaissent la condition attribuée à la personne participante.

Étude en laboratoire

Une étude menée dans un environnement de laboratoire.

Étude pilote

Il s'agit d'une étude à petite échelle menée pour s'assurer qu'une nouvelle procédure fonctionne comme prévu.

Étude sur le terrain

Une étude menée dans un environnement « réel » en dehors du laboratoire.

Études longitudinales

Études dans lesquelles un même groupe de personnes est suivi au fil du temps, à mesure qu'il vieillit.

Études transversales

Études qui impliquent la comparaison de deux ou plusieurs groupes de personnes préexistants (par exemple, des enfants à différents stades de développement).

Études transversales séquentielles

Études dans lesquelles des personnes appartenant à différents groupes d'âge sont suivies par l'équipe de recherche sur une période de temps plus courte.

Examen par les pairs en double aveugle

Un processus dans lequel les personnes responsables de l'évaluation d'un article de recherche ne connaissent pas l'identité des personnes ayant mené l'étude, et inversement, ces dernières ne connaissent pas l'identité des personnes responsable de l'évaluation.

Expérience

Un type d'étude conçue spécifiquement pour répondre à la question de savoir s'il existe une relation causale entre deux variables.

Expérience intersujets

Expérience dans laquelle chaque personne participante est testée dans une seule condition.

Expérience intrasujet

Une expérience dans laquelle chaque personne participante est testée dans toutes les conditions.

Expérience sur le terrain

Un type d'étude sur le terrain où une variable indépendante est manipulée dans un environnement naturel et où les variables extérieures sont contrôlées autant que possible.

Faisabilité

Probabilité qu'une question de recherche puisse être répondue avec succès en fonction du temps, de l'argent, de l'équipement et du matériel disponibles, ainsi que des connaissances techniques, des compétences et de l'accès aux personnes participantes à la recherche.

Falsifiable

Une affirmation scientifique qui doit être exprimée de telle sorte qu'il existe des observations qui — si elles étaient faites — constitueraient des preuves contre l'affirmation.

Fichier de données

Données saisies dans une feuille de calcul et formatées pour être analysées.

Fidélité

Réfère à la cohérence d'une mesure.

Fidélité interjuges

La mesure dans laquelle différents responsables de l'observation sont cohérents dans leurs jugements.

Fidélité test-retest

Lorsque les personnes menant une recherche mesurent un concept qu'elles supposent cohérent dans le temps, les résultats qu'elles obtiennent devraient également être cohérents dans le temps.

Figures

Représentations graphiques de données, telles que des diagrammes à secteurs, des histogrammes ou des diagrammes de dispersion, utilisées pour rendre compte de manière claire et efficace d'un certain nombre de résultats.

Final manuscripts

Manuscripts that are prepared by the author in their final form and submitted for publication.

Formulaire de consentement

Le processus d'obtention du consentement éclairé en demandant aux personnes participantes de lire et de signer le formulaire.

formulation neutre

Mot ou formulation n'ayant pas de connotation genrée (sans alternance féminin/masculin) tout en ayant un genre grammatical qui LUI/LEUR? est propre.

Graphiques linéaires

Graphiques utilisés lorsque la variable indépendante est mesurée de manière plus continue (par exemple, le temps) ou pour présenter des corrélations entre des variables quantitatives lorsque la variable indépendante a ou est organisée en un nombre relativement restreint de niveaux distincts.

Groupes de discussion (focus groups)

Utilisé dans la recherche qualitative qui implique de petits groupes de personnes qui participent ensemble à des entretiens axés sur un sujet ou une question particulière.

HARKing

Émettre des hypothèses après que les résultats sont connus : Pratique consistant pour les chercheurs à analyser des données sans hypothèse a priori, en affirmant après coup qu'un résultat statistiquement significatif avait été prédit à l'origine. HARK vient de Hypothesizing After the Results are Known.

Heuristiques

Raccourcis mentaux pour former et maintenir nos croyances.

Histogramme

Représentation graphique d'une distribution de fréquences.

Histoire

Les événements extérieurs au modèle de recherche prétest-post test qui ont pu influencer plusieurs ou la totalité des personnes participantes entre le prétest et le post-test.

Hypothèse

Prédiction spécifique d'un nouveau phénomène qui devrait être observé si une théorie particulière est exacte.

Hypothèse alternative

Alternative à l'hypothèse nulle (souvent symbolisée par H_1), cette hypothèse propose qu'il existe une relation dans la population et que la relation dans l'échantillon reflète cette relation dans la population.

Hypothèse nulle

L'idée qu'il n'y a pas de relation dans la population et que la relation dans l'échantillon ne reflète que l'erreur d'échantillonnage (souvent symbolisée par H_0 et lue comme « H-zéro »).

Individus participants

Personnes participant à une enquête ou une étude.

Inspection visuelle

Cela signifie qu'il faut tracer les données de chaque personne participante, examiner attentivement ces tracés et juger si et dans quelle mesure la variable indépendante a eu un effet sur la variable dépendante.

Instrumentation

Une menace potentielle pour la validité interne lorsque les caractéristiques de base de l'instrument de mesure changent au cours de l'étude.

Interaction

Lorsque l'effet d'une variable indépendante dépend du niveau d'une autre variable.

Interaction croisée

Cela signifie que la variable indépendante a un effet aux deux niveaux, mais que les effets sont dans des directions opposées.

interaction diffuse

Signifie qu'il y a un effet d'une variable indépendante à un niveau de l'autre variable indépendante et que l'effet de cette variable indépendante est faible ou nul à l'autre niveau de l'autre variable indépendante (pluriel interactions diffuses)

Intérêt scientifique (ou interestingness)

Degré auquel une question est jugée intéressante, que ce soit pour le grand public ou pour la communauté scientifique. Trois aspects doivent être pris en considération : l'existence d'un doute quant à la réponse, la contribution à combler une lacune dans la littérature scientifique, et l'importance des retombées pratiques.

Intervalles de confiance

Étendue de valeurs qui est calculée afin qu'un certain pourcentage du temps (généralement 95%), le paramètre de la population se situe dans cette étendue.

introduction

Présentation de la question de recherche, de sa pertinence et de recherches antérieures pertinentes et introduit à la méthode de recherche qui sera utilisée.

Item-order effect

When the order in which the items are presented affects people's responses.

Items à réponse fermée

Items de questionnaire qui posent une question et proposent un ensemble limité d'options de réponse, parmi lesquelles les personnes participantes doivent sélectionner.

Items à réponse ouverte

Il suffit de poser une question et de permettre aux personnes participantes d'y répondre de la manière qu'elles choisissent.

Justice

L'importance de mener la recherche de manière à répartir équitablement les risques et les avantages entre les différents groupes au niveau de la société.

Line graphs

Graphs used when the independent variable is measured in a more continuous manner (e.g., time) or to present correlations between quantitative variables when the independent variable has, or is organized into, a relatively small number of distinct levels.

Littérature de recherche

Toutes les recherches publiées dans ce domaine.

Livres savants

Ouvrages rédigés par des personnes praticiennes et des personnes en recherches principalement à l'intention d'autres personnes praticiennes et des personnes en recherches.

Manipuler

Modification systématique du niveau ou de la condition de la variable indépendante de sorte que différents groupes d'individus participants soient exposés à différents niveaux de cette variable, ou que le même groupe d'individus participants soit exposé à différents niveaux à différents moments.

Matrice de corrélations

Indique le coefficient de corrélation entre les paires de variables de l'étude.

Maturation

Les personnes participantes peuvent avoir changé entre le prétest et le post-test de manière naturelle, simplement parce qu'elles grandissent et apprennent.

Médiane

Le point médian d'une distribution de notes, en ce sens que la moitié des notes de la distribution lui est inférieure et l'autre moitié lui est supérieure.

Mesure

Il s'agit de l'attribution de notes à des individus de manière à ce que ces notes représentent une caractéristique de ces individus.

Mesure d'auto-évaluation

Mesures dans lesquelles les personnes participantes font état de leurs propres pensées, sentiments et actions.

Mesures du comportement

Mesures dans lesquelles un autre aspect du comportement des personnes participantes est observé et enregistré.

Mesures physiologiques

Mesures impliquant l'enregistrement d'une grande variété de processus physiologiques, y compris la fréquence cardiaque et la pression artérielle, la réponse galvanique de la peau, les niveaux d'hormones, l'activité électrique et le flux sanguin dans le cerveau.

Méta-analyse

Un article de synthèse qui fournit un résumé statistique de tous les résultats précédents.

Méthode hypothético-déductive

Processus cyclique d'élaboration de théories, commençant par un phénomène observé, puis développant ou utilisant une théorie pour faire une prédiction spécifique de ce qui devrait se produire si cette théorie est correcte, testant cette prédiction, affinant la théorie à la lumière des résultats, et utilisant cette théorie affinée pour développer de nouvelles hypothèses, et ainsi de suite.

Méthode scientifique

La méthode scientifique est un processus de collecte et d'évaluation systématique de preuves pour tester des idées et répondre à des questions.

Mode

Le score le plus fréquent dans une distribution.

Monographie

Une présentation écrite cohérente d'un sujet qui ressemble à un article de synthèse écrit par un seul auteur (ou une seule autrice) ou un petit groupe d'auteurs ou d'autrices.

Moyenne

La moyenne d'une distribution de scores (symbolisée par M) où la somme des scores est divisée par le nombre de scores.

Niveau d'intervalle

Une mesure qui consiste à attribuer des scores à l'aide d'échelles numériques dans lesquelles les intervalles ont la même signification partout sur l'échelle.

Niveau de base

La phase initiale d'un plan ABA, qui sert de condition contrôle et dans laquelle est mesuré le niveau de réponses avant l'introduction de tout traitement.

Niveau de ratio

Une mesure qui consiste à attribuer des scores de manière à ce qu'il existe un vrai point zéro représentant l'absence totale de la quantité mesurée.

Niveau nominal

Mesure utilisée pour les variables catégorielles et consistant à attribuer des scores associés aux noms de catégories.

Niveau ordinal

Une mesure qui consiste à attribuer des scores de manière à ce qu'elles représentent l'ordre de classement des individus.

Niveaux de mesure

Quatre catégories, ou échelles, de mesure (c'est-à-dire nominale, ordinale, d'intervalle et de rapport) qui spécifie les types d'informations qu'un ensemble de scores peut contenir et les types de procédures statistiques qui peuvent être utilisées avec les scores.

Observation naturaliste

Méthode d'observation qui consiste à observer le comportement des personnes dans l'environnement où il se produit habituellement.

Observation naturaliste déguisée

Lorsque les personnes menant des recherches s'engagent dans une observation naturaliste en effectuant leurs observations aussi discrètement que possible afin que les personnes participantes ne soient pas conscientes qu'elles font l'objet d'une étude.

Observation naturaliste non déguisée

Lorsque les personnes participantes sont informées de la présence d'une personne menant une recherche et de la surveillance de leur comportement.

Observation participante

Les personnes menant la recherche deviennent des participantes actives au sein du groupe ou de la situation qu'elles étudient.

Observation participante déguisée

Les personnes menant la recherche se font passer pour des membres du groupe social qu'elles observent et dissimulent leur véritable identité de chercheuses ou chercheurs.

Observation participante non déguisée

Les personnes menant la recherche s'intègrent au groupe qu'elles étudient et révèlent leur véritable identité de personne responsable d'une recherche au groupe étudié.

Observation structurée

Lorsque la personne menant une recherche observe attentivement un ou plusieurs comportements spécifiques dans un cadre particulier qui est plus structuré que les cadres utilisés dans l'observation naturaliste ou l'observation participante.

Opérationnalisation

La spécification de la manière exacte dont la question de recherche sera étudiée dans le devis d'expérience.

Opérations convergentes

Lorsque les psychologues utilisent plusieurs définitions opérationnelles d'un même concept — que ce soit au sein d'une même étude ou d'une étude à l'autre.

ouverture

introduction à la question de recherche et en quoi elle est intéressante.

p-hacking

Lorsque les personnes menant une recherche prennent diverses décisions au cours du processus de recherche afin d'augmenter leurs chances d'obtenir un résultat statistiquement significatif (et une erreur de type I) en supprimant arbitrairement les valeurs aberrantes, en choisissant sélectivement de présenter les variables dépendantes, en ne présentant que les résultats significatifs, etc., jusqu'à ce que leurs résultats produisent une valeur *p* souhaitable.

page titre

Page au début d'un article qui contient le titre et le nom des auteurs ou autrices.

Paramètres

Valeurs correspondantes dans la population

paraphrase

phrase qui exprime dans votre propre mot l'idée originale présentée dans une autre source

Placébo

Traitement simulé dépourvu de tout ingrédient actif ou élément censé rendre le traitement efficace, mais qui est par ailleurs identique au traitement.

Plan factoriel inter-sujets

Toutes les variables indépendantes sont manipulées entre les sujets.

Plan factoriel mixte

Un plan qui manipule une variable indépendante entre les sujets et une autre à l'intérieur des sujets.

Plans factoriels

Expériences comprenant plus d'une variable indépendante dans lesquelles chaque niveau d'une variable indépendante est combiné avec chaque niveau des autres variables afin de produire toutes les combinaisons possibles.

Politique fédérale en matière de protection des êtres humains

Un ensemble de lois fondées sur le Rapport Belmont qui s'appliquent aux recherches menées, financées ou réglementées par le gouvernement fédéral.

Population

Un grand groupe de personnes au sujet duquel les personnes menant des recherches en psychologie cherchent généralement à tirer des conclusions, et à partir duquel l'échantillon est constitué.

Pourcentage de données qui ne se chevauchent pas

Il s'agit du pourcentage de réponses dans la condition de traitement qui sont plus extrêmes que la réponse la plus extrême dans une condition contrôle pertinente.

Pratique clinique en psychologie

Diagnostic et traitement des troubles psychologiques et des problèmes connexes.

Pratique de science ouverte

Une pratique dans laquelle les personnes menant des recherches partagent ouvertement leur matériel de recherche avec les autres du domaine dans l'espoir d'accroître la transparence et l'ouverture de l'entreprise scientifique.

Présélection

Une façon de minimiser les risques dans une étude et, d'identifier et d'éliminer les personnes participantes à haut risque.

Preuve convergente

Une idée qui nous invite à examiner le patron de failles présentes dans la littérature scientifique, puisque la nature de ce patron peut soit appuyer, soit fragiliser les conclusions que nous souhaitons tirer.

Problème de directionnalité

Problème dans lequel deux variables, X et Y, sont statistiquement liées soit parce que X cause Y, soit parce que Y cause X, et donc la direction causale de l'effet ne peut pas être connue.

Problème de directionnalité x

Problème dans lequel deux variables, X et Y, sont statistiquement liées soit parce que X cause Y, soit parce que Y cause X, et donc la direction causale de l'effet ne peut pas être connue.

Problème de la troisième variable

Deux variables, X et Y, peuvent être statistiquement liées non pas parce que X cause Y, ou parce que Y cause X, mais parce qu'une troisième variable, Z, cause à la fois X et Y.

Problème du tiroir à dossiers

Le problème des résultats de recherche qui ne sont pas publiés lorsqu'ils ne révèlent pas d'effet statistiquement significatif. En conséquence, la littérature publiée ne présente pas un portrait complet des résultats positifs et négatifs portant sur une question de recherche.

Protection de la vie privée

Le droit d'une personne de décider quelles informations la concernant sont partagées avec d'autres.

Protocole

Une description détaillée de la recherche — qui est examinée par un comité indépendant.

Protocole à niveaux de base multiples

Dans ce protocole, plusieurs niveaux de base sont établis pour une personne participante ou un niveau de base est établi pour plusieurs personnes participantes.

Protocole à séries temporelles interrompues

Série de mesures prises à intervalles réguliers, interrompue par l'introduction d'un traitement.

Protocole à séries temporelles interrompues avec groupes non équivalents

Implique de prendre un ensemble de mesures à intervalles réguliers sur une période de temps, à la fois avant et après une intervention d'intérêt, dans deux groupes ou plus qui ne sont pas équivalents.

Protocole d'inversion

Le plan de recherche à cas unique le plus simple, dans lequel l'équipe de recherche mesure la variable dépendante en trois phases : un niveau de base, avant l'introduction d'un traitement (A) ; après l'introduction du traitement (B) ; puis un retour au niveau de base après le retrait du traitement (A). On l'appelle souvent un plan ABA.

Protocole inversé à multiples traitements

Dans ce protocole, la phase du niveau de base est suivie de phases distinctes au cours desquelles différents traitements sont introduits.

Pseudoscience

Il s'agit d'activités et de croyances qui sont revendiquées comme scientifiques par les personnes partisans—et qui peuvent sembler scientifiques à première vue—mais qui ne le sont pas.

Psychologie folklorique

Croyances intuitives à propos du comportement, des pensées et des sentiments des gens.

Psychométrie

Un sous-domaine de la psychologie qui s'intéresse aux théories et aux techniques de la mesure psychologique.

PsycINFO

Une base de données électronique complète couvrant des milliers de revues professionnelles et d'ouvrages savants. Il s'agit de publications qui remontent à plus de 100 ans—dont le contenu est, dans la plupart des cas, synonyme de littérature de recherche en psychologie.

Puissance statistique

Dans un devis de recherche, la puissance statistique correspond à la probabilité de rejeter l'hypothèse nulle compte tenu de la taille de l'échantillon et de la force attendue de la relation.

Questions empiriques

Il s'agit de questions sur la façon dont le monde est réellement et, par conséquent, on peut y répondre en l'observant systématiquement.

Rang percentile

Pour un score donné, le pourcentage de scores dans la distribution qui sont inférieurs à ce score.

Rapport Belmont

Ensemble de lignes directrices fédérales rédigées en 1978 en réponse aux abus de l'étude Tuskegee, qui reconnaissent trois principes importants dans la recherche avec des êtres humains : la justice, le respect des personnes et les bienfaits, et qui constituent la base des réglementations fédérales appliquées à la recherche.

Rapport de recherche empirique

Un article qui présente les résultats d'une ou plusieurs nouvelles études.

Rapports de recherche empirique

Rapports de recherche décrivant une ou plusieurs nouvelles études empiriques menées par les auteurs ou autrices.

Réactivité

Il s'agit du moment où une mesure modifie le comportement des individus participants.

Réalisme du quotidien

Lorsque les individus participants et la situation étudiée sont similaires à ceux sur lesquels les personnes menant la recherche veulent faire des généralisations et que les individus participants rencontrent tous les jours.

Réalisme psychologique

Lorsque le même processus mental est utilisé à la fois en laboratoire et dans le monde réel.

Recherche à cas unique

Un type de recherche quantitative qui consiste à étudier en détail le comportement de chaque personne pour un petit nombre d'individus participants.

Recherche à méthodes mixtes

Recherche combinant des approches quantitatives et qualitatives.

Recherche à risque plus que minimal

Recherche qui présente un risque plus que minimal pour les individus participants et qui doit être examinée par l'ensemble des membres du CÉR.

Recherche accélérée

Recherche évaluée par un comité d'éthique (CÉR) qui n'est pas anonyme et/ou peut impliquer des informations potentiellement stigmatisantes, ou encore des procédures intrusives ou inconfortables, mais qui n'expose les personnes participantes qu'à des risques ne dépassant pas le risque minimal (c'est-à-dire les risques rencontrés par des personnes en bonne santé dans la vie quotidienne ou lors d'examens physiques ou psychologiques de routine).

Recherche appliquée

Recherche menée principalement pour répondre à un problème pratique.

Recherche corrélationnelle

Recherche non expérimentale parce qu'elle se concentre sur la relation statistique entre deux variables, mais n'inclut pas la manipulation d'une variable indépendante.

Recherche exemptée

Recherche portant sur l'efficacité d'activités éducatives normales, sur l'utilisation de mesures psychologiques standard et d'enquêtes de nature non sensible administrées de manière à préserver la confidentialité, ainsi que la recherche utilisant des données existantes provenant de sources publiques.

Recherche fondamentale

Recherche menée principalement dans le but d'acquérir une compréhension plus détaillée et plus précise du comportement humain, sans nécessairement chercher à répondre à un problème pratique particulier.

Recherche non expérimentale

Une recherche qui ne comporte pas de manipulation d'une variable indépendante.

Recherche par enquête

Il s'agit d'une méthode quantitative et qualitative qui présente deux caractéristiques importantes : les variables sont mesurées à l'aide d'auto-évaluations et une attention considérable est accordée à la question de l'échantillonnage.

Recherche par observation

Recherche non expérimentale, car elle se concentre sur l'enregistrement systématique des comportements dans un contexte naturel ou en laboratoire, sans manipulation d'aucune variable.

Recherche qualitative

Recherche qui commence par une question de recherche moins ciblée, qui recueille une grande quantité de données relativement « non filtrées » auprès d'un nombre relativement restreint d'individus, qui décrit ces données à l'aide de techniques non statistiques comme la théorie ancrée, l'analyse thématique, l'analyse critique du discours ou l'analyse phénoménologique interprétative, et qui vise à comprendre en détail l'expérience vécue par les personnes participantes à la recherche.

Recherche quantitative

Recherche qui commence généralement par une question de recherche ou une hypothèse ciblée, qui recueille une petite quantité de données numériques auprès d'un grand nombre d'individus, qui décrit les données obtenues à l'aide de techniques statistiques et qui tire des conclusions générales concernant une population plus vaste.

Recherche sur les groupes

Recherche qui consiste à étudier un grand nombre de personnes participantes et à examiner leur comportement principalement en matière de moyennes de groupe, d'écarts-types, et ainsi de suite.

Récit théorique

Une méthode de recherche qualitative qui implique une interprétation des données en fonction des thèmes identifiés par la personne menant la recherche.

reformulation

Modification d'un mot, d'un ensemble de mots ou de la structure de la phrase de manière à ce que la tournure soit non genrée; inclus les femmes, les hommes et les gens non binaires

Régression

Une technique statistique qui permet aux personnes en recherche de prédire la valeur d'une variable en fonction d'une autre.

Régression multiple

implique la mesure de plusieurs variables (X_1 , X_2 , X_3 ... X_i) et leur utilisation pour prédire un résultat, la variable (Y).

Régression simple

Une procédure statistique qui utilise la valeur d'une variable pour en prédire une autre. Parfois appelée « régression linéaire ».

Régression vers la moyenne

Fait référence au fait statistique qu'un individu qui obtient un score extrêmement élevé ou extrêmement bas pour une variable à une occasion donnée aura tendance à obtenir un score moins extrême à l'occasion suivante.

Rejeter l'hypothèse nulle

Une décision prise par les personnes en recherche dans le cadre d'un test d'hypothèse nulle, qui se produit lorsque la relation observée dans l'échantillon serait extrêmement improbable.

Relation négative

Une relation dans laquelle des scores plus élevés sur une variable tendent à être associés à des scores plus faibles sur l'autre.

Relation positive

Une relation dans laquelle des scores plus élevés sur une variable tendent à être associés à des scores plus élevés sur l'autre.

Relations linéaires

Relations entre deux variables dans lesquelles les points dans un diagramme de dispersion se rapprochent d'une seule ligne droite.

Relations non linéaires

Relations entre deux variables dans lesquelles les points dans un diagramme de dispersion ne sont pas proches d'une ligne droite unique, mais se situent souvent le long d'une ligne courbe.

Rémission spontanée

La tendance pour plusieurs problèmes médicaux et psychologiques à s'améliorer avec le temps sans aucune forme de traitement.

Réponse socialement désirable

Lorsque les personnes participantes répondent d'une manière qu'elles estiment socialement acceptable.

Réserve de sujets

Un groupe établi de personnes qui ont accepté d'être contactées pour participer à des études de recherche.

Respect des personnes

L'un des principes du rapport Belmont qui met l'accent sur la nécessité pour les personnes participantes d'exercer leur autonomie et sur la protection de celles dont l'autonomie est réduite, souvent au moyen du consentement éclairé.

Restriction de l'étendue

Se produit lorsque l'une ou les deux variables présentent une étendue limitée dans l'échantillon par rapport à la population, ce qui peut rendre la valeur du coefficient de corrélation trompeuse .

Résumé

Un bref résumé de la question de recherche, des méthodes, des résultats et des conclusions de l'étude.

Retenir l'hypothèse nulle ou conserver l'hypothèse nulle

Décision prise par les personnes chercheuses dans le cadre d'un test d'hypothèse nulle, qui se produit lorsque la relation observée dans l'échantillon ne serait pas extrêmement improbable.

Revue scientifique

Il s'agit de périodiques qui publient des articles de recherche originaux.

Scepticisme

Consiste à prendre le temps de considérer des alternatives et à rechercher des preuves — en particulier des données empiriques recueillies de manière systématique — lorsqu'il s'agit d'une question suffisamment importante pour le justifier.

Science

L'étude systématique de la structure et du comportement du monde physique et naturel par l'observation et l'expérimentation.

Score de différence

Une méthode pour réduire des paires de scores (par exemple, prétest et post-test) à un seul score en calculant la différence entre eux.

Score Z

Correspond à la différence entre le score d'une personne et la moyenne de la distribution, divisée par l'écart-type de cette distribution. Il représente le nombre d'écarts types qui sépare ce score de la moyenne.

Section des résultats

Partie où l'on présente les principaux résultats de l'étude, y compris ceux provenant des analyses statistiques.

Session de communications par affiches

Une session d'une à deux heures qui se déroule dans une grande salle sur le site du congrès où des dizaines de communications par affiches de recherche sont présentées.

Signification pratique

Renvoie à l'importance ou à l'utilité d'un résultat dans un contexte réel.

Statistique de test

Une statistique (p. ex., F , t , etc.) calculée afin d'être comparée à ce qui est attendu sous l'hypothèse nulle, et qui permet ainsi d'obtenir la valeur p .

Statistiquement significatif

Un effet qui est peu probable d'être attribuable au hasard et qui représente donc vraisemblablement un effet réel dans la population.

Statistiques

Données descriptives qui consistent à mesurer une ou plusieurs variables dans un échantillon et à

calculer des indices de synthèse descriptifs (p. ex., moyennes, coefficients de corrélation) pour ces variables.

statistiques bayésiennes

Approche dans laquelle la personne en recherche spécifie la probabilité que l'hypothèse nulle et toutes les hypothèses alternatives importantes soient vraies avant de mener l'étude, mène l'étude, puis met à jour les probabilités sur la base des données.

Statistiques descriptives

Renvoie à un ensemble de techniques permettant de résumer et d'afficher des données.

Statistiques inférentielles

Une méthode de recherche qui permet aux personnes en recherche de tirer des conclusions ou de faire des déductions sur une population à partir de données provenant d'un échantillon.

Stratégie de l'état stable

Lorsque la personne menant la recherche attend que le comportement de l'individu participant dans une condition devienne relativement cohérent d'une observation à l'autre avant de changer de condition.

Style APA

Un ensemble de lignes directrices pour la rédaction en psychologie et dans les domaines connexes.

Style de bas niveau

Est traité du chapitre 4 « Les mécanismes du style » jusqu'au chapitre 7 « Exemples de références » du *Publication Manual*, lequel comprend toutes les lignes directrices spécifiques concernant l'orthographe, la grammaire, les références et leurs citations, les nombres et les statistiques, les figures et les tableaux, et ainsi de suite.

Style de haut niveau

Lignes directrices du *Publication Manual* de l'APA concernant l'expression claire des idées, y compris une écriture formelle, simple et évitant un langage biaisé.

Symétrie

Lorsque les moitiés gauche et droite de l'histogramme sont des images miroirs l'une de l'autre.

Tableau de fréquence

Affichage de chaque valeur d'une variable et du nombre de personnes participantes ayant cette valeur.

Tableau des plans factoriels

Montre comment chaque niveau d'une variable indépendante est combiné avec chaque niveau des autres variables pour produire toutes les combinaisons possibles dans un plan factoriel.

Taille de l'effet

Décrit la force d'une relation statistique.

Tendance centrale

Est le point milieu d'une distribution — le point autour duquel les scores de la distribution tendent à se regrouper. (Un autre terme pour désigner la tendance centrale est la moyenne.)

Test bilatéral

Situation dans laquelle on rejette l'hypothèse nulle si la statistique de test de l'échantillon est extrême dans l'une ou l'autre direction (+/-).

Test de l'hypothèse nulle

Une approche formelle pour décider entre deux interprétations d'une relation statistique dans un échantillon.

test t

Un test qui consiste à examiner la différence entre deux moyennes.

Test t à échantillons dépendants

Utilisé pour comparer deux moyennes pour le même échantillon testé à deux moments différents ou dans deux conditions différentes (parfois appelé test *t* pour échantillons appariés).

Test t à échantillons indépendants

Utilisé pour comparer les moyennes de deux échantillons distincts (M_1 et M_2).

Test t à un échantillon

Utilisé pour comparer une moyenne d'échantillon (M) à une moyenne hypothétique de la population (μ_0) servant de référence intéressante pour la comparaison.

Test unilatéral

Situation dans laquelle on rejette l'hypothèse nulle seulement si la valeur *t* de l'échantillon est extrême dans une direction précise, spécifiée avant la collecte des données.

Testable et falsifiable

Capacité de tester une hypothèse à l'aide des méthodes scientifiques et possibilité de recueillir des données empiriques qui permettraient d'infirmer cette hypothèse si elle est effectivement fausse.

Testing

A threat to internal validity that occurs when the measurement of the dependent variable during the pretest affects participants' responses at posttest.

Théorie

Une explication ou une interprétation cohérente d'un ou de plusieurs phénomènes.

Théorie ancrée

Les personnes en recherche partent des données et développent une théorie ou une interprétation qui est « ancrée » dans ces données.

Tolérance à l'incertitude

Accepter qu'il y ait beaucoup de choses que nous ne savons tout simplement pas.

Traitement

Toute intervention visant à améliorer le comportement des personnes.

Traitements empiriquement validés

Traitements dont il a été démontré, par des observations systématiques, qu'ils mènent à de meilleurs résultats comparativement à l'absence de traitement ou à des groupes témoins recevant un placebo. On dit souvent « traitements fondés sur des données probantes ».

Triangulation

L'idée est d'utiliser simultanément des méthodes quantitatives et qualitatives pour étudier les mêmes questions générales et comparer les résultats.

Valeur aberrante

Un score extrême qui est beaucoup plus élevé ou plus bas que le reste des scores de la distribution.

Valeur critique

La valeur absolue qu'une statistique de test (par exemple, F , t , etc.) doit dépasser pour être considérée comme statistiquement significative.

valeur p

Probabilité d'obtenir le résultat de l'échantillon, ou un résultat plus extrême, si l'hypothèse nulle était vraie.

Validité

Degré auquel les scores issus d'une mesure représentent la variable qu'ils sont censés mesurer.

Validité apparente

Degré selon lequel une méthode de mesure semble, à première vue, mesurer le construit d'intérêt.

Validité concomitante

Une forme de validité de critère, où le critère est mesuré en même temps (de façon concomitante) que le construit.

Validité convergente

Une forme de validité de critère dans laquelle de nouvelles mesures sont corrélées avec des mesures déjà établies du même construit.

Validité de construit

L'une des « quatre grandes » validités, soit celle selon laquelle la question de recherche est clairement opérationnalisée par les méthodes utilisées dans l'étude.

Validité de contenu

Degré selon lequel une mesure reflète l'ensemble des aspects du construit d'intérêt.

Validité de critère

Degré selon lequel les scores des personnes à une mesure sont corrélés avec d'autres variables (appelées critères) avec lesquelles on s'attend à ce qu'ils soient corrélés.

Validité discriminante

Degré selon lequel les scores obtenus à une mesure d'un construit **ne** sont **pas** corrélés avec des mesures d'autres construits conceptuellement distincts, et permettent ainsi de les différencier.

Validité externe

Renvoie au degré selon lequel on peut généraliser les résultats d'une étude à d'autres circonstances ou contextes, comme des situations en milieu réel.

Validité interne

Renvoie au degré selon lequel on peut inférer avec confiance une relation causale entre des variables.

Validité prédictive

Une forme de validité dans laquelle le critère est mesuré à un moment donné dans le futur (après la mesure du construit), afin de vérifier que le construit « prédit » le critère.

Validité sociale

Renvoie à des traitements qui ont des effets substantiels sur des comportements importants et qui peuvent être appliqués de façon fiable dans les contextes réels où ils se manifestent.

Validité statistique

Concerne le traitement approprié des données par des méthodes statistiques et la validité des conclusions statistiques tirées par les personnes en recherche.

Variability

The extent to which the scores vary around their central tendency in a distribution.

Variable

Une quantité ou une qualité qui varie selon les personnes ou les situations.

Variable catégorielle

Variable qui représente une caractéristique d'une personne, telle que le programme d'études, et qui est généralement mesurée en attribuant la réponse de chaque individu à l'une de plusieurs catégories (par exemple, psychologie, anglais, sciences infirmières, ingénierie, etc.)

Variable de résultat ou variable de critère

Variable qui est prédite par une variable prédictive dans une équation de régression. En anglais, on la dénomme outcome variable.

Variable dépendante

Variable mesurée par la personne responsable de l'expérimentation (c'est l'effet présumé).

Variable indépendante

Variable manipulée par la personne responsable de l'expérimentation.

Variable indépendante non-manipulée

Une variable indépendante qui est mesurée, mais qui n'est pas manipulée.

Variable prédictive

Une variable dans une équation de régression que l'on suppose liée à la valeur d'une variable de résultat ou de critère (qu'elle « prédit »).

Variable quantitative

Une quantité, telle que la taille, qui est généralement mesurée en attribuant un nombre à chaque personne.

Variables confondantes

Un type particulier de variable externe qui varie systématiquement en même temps que les variables étudiées et qui fournit donc une explication alternative aux résultats.

Variables externes

Toute variable autre que la variable dépendante et la variable indépendante.

Variance

Mesure de la distance moyenne des scores par rapport à la moyenne.

Vérification de la manipulation

Consiste à s'assurer que la manipulation expérimentale a bien fonctionné en utilisant une mesure différente du construit que la personne en recherche tente de manipuler.

Volumes édités

Livres constitués de chapitres rédigés par différents/différentes auteurs ou autrices portant sur divers aspects d'un même sujet, et supervisés par un ou plusieurs personne-s responsable-s de l'édition.

Références

Chapitre 1

American Psychology Association (2020). *About APA*. Extrait de <https://www.apa.org/about>

Bushman, B. J. (2002). L'évacuation de la colère alimente-t-elle ou éteint-elle la flamme? Catharsis, rumination, distraction, colère et réponse agressive. *Personality and Social Psychology Bulletin*, 28, 724-731.

Gilovich, T. (1991). *How we know what isn't so : La faillibilité de la raison humaine dans la vie quotidienne*. New York, NY: Free Press.

Gladwell, M. E. (2005). *Blink: The power of thinking without thinking*. (9e éd.). New York: Little, Brown & Co.

Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports*, 83, 19-64.

Kassin, S. M. et Gudjonsson, G. H. (2004). The psychology of confession evidence: A review of the literature and Issues. *Psychological Science in the Public Interest*, 5, 33-67. <https://doi.org/10.1111/j.1529-1006.2004.00016.x>

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, S., Bernstein, M. J., . . . Nosek, B. A. (2013). Investigating variation in replicability : A "many labs" replication project. *Social Psychology*, 45(3), 142-152. <https://doi.org/10.1027/1864-9335/a000178>

Lilienfeld, S. O. (2011). Les 10 commandements pour aider les étudiants à distinguer la science de la pseudoscience en psychologie. *Revue Électronique de Psychologie Sociale*, 5, 23-29.

Lilienfeld, S. O., Lynn, S. J., Ruscio, J. et Beyerstein, B. L. (2010). *50 great myths of popular psychology* (50 grands mythes de la psychologie populaire). Malden, MA : Wiley-Blackwell.

Mann, T., Tomiyama, A. J., Westling, E., Lew, A., Samuels, B. et Chatman, J. (2007). Medicare's search for effective obesity treatments : Diets are not the answer. *American Psychologist*, 62, 220-233. <https://doi.org/10.1037/0003-066x.62.3.220>

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82. <https://psycnet.apa.org/doi/10.1126/science.1139940>

Norcross, J. C., Beutler, L. E. et Levant, R. F. (dir.). (2005). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. Washington, DC : American Psychological Association.

Popper, K. R. (2002). *Conjectures et réfutations : La croissance de la connaissance scientifique*. New York, NY : Routledge.

Schnall, S., Benton, J. et Harvey, S. (2008). With a clean conscience : La propreté réduit la sévérité des jugements moraux. *Psychological Science*, 19(12), 1219-1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>

Sexton, M., Cuttler, C., Finnell, J. et Mischley, L. (2016). Une enquête transversale sur les utilisateurs de cannabis médical : Patterns of use and perceived efficacy (Modes d'utilisation et efficacité perçue). *Cannabis and Cannabinoid Research*, 1, 131- 138. <https://doi.org/10.1089/can.2016.0007>

Chapitre 2

- Adair, J. G. et Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. (L'explosion des connaissances, des références et des citations : La réponse unique de la psychologie à une crise.). *American Psychologist*, 58, 15-23. <https://doi.org//10.1037/0003-066X.58.1.15>
- Collet, C., Guillot, A. et Petit, C. (2010). Phoning while driving I: A review of epidemiological, psychological, behavioral and physiological studies. (Téléphoner au volant I : examen des études épidémiologiques, psychologiques, comportementales et physiologiques). *Ergonomics*, 53, 589-601. <https://doi.org//10.1080/00140131003672023>
- Drews, F. A., Pasupathi, M. et Strayer, D. L. (2004). Passenger and cell-phone conversations in simulated driving. (Conversations entre passagers et téléphones portables lors d'une simulation de conduite). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48, 2210-2212. <https://doi.org//10.1037/a0013119>
- Milgram, S. (1963). Behavioral study of obedience. (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67, 371-378. <https://doi.org//10.1037/h0040525>
- Mueller, P. A., et Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: Advantages of longhand over laptop note taking. (La plume est plus puissante que le clavier : les avantages de la prise de notes manuscrite par rapport à la prise de notes sur ordinateur portable.) *Psychological Science*, 25(6), 1159-1168. <https://psycnet.apa.org/doi/10.1177/0956797614524581>
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H. et Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. (Facilité de récupération de l'information : Un autre regard sur l'heuristique de disponibilité.). *Journal of Personality and Social Psychology*, 61(2), 195-202. <https://doi.org//10.1037/0022-3514.61.2.195>
- Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius*. (Créativité : au-delà du mythe du génie). New York, NY : Freeman.
- Zajonc, R. B. (1965). Social facilitation: A solution is suggested for an old unresolved social psychological problem. (Facilitation sociale : Une solution est proposée pour un vieux problème socio-psychologique non résolu.). *Science*, 149, 269-274. <https://doi.org//10.1126/science.149.3681.269>
- Zajonc, R. B., Heingartner, A. et Herman, E. M. (1969). Social enhancement and impairment of performance in the cockroach. (Renforcement social et altération des performances chez le cafard). *Journal of Personality and Social Psychology*, 13(2), 83-92. <https://doi.org//10.1037/h0028063>
- Zajonc, R. B. et Sales, S. M. (1966). Social facilitation of dominant and subordinate responses. (Facilitation sociale des réponses dominantes et subordonnées). *Journal of Experimental Social Psychology*, 2(2), 160-168. [https://doi.org//10.1016/0022-1031\(66\)90077-1](https://doi.org//10.1016/0022-1031(66)90077-1)

Chapitre 3

- Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited (Recherche utilisant la duperie intentionnelle : Questions éthiques revisitées). *American Psychologist*, 40(2), 165-174. <https://doi.org/10.1037/0003-066X.40.2.165>
- Bowd, A. D. et Shapiro, K. J. (1993). The case against animal laboratory research in psychology (Les arguments contre la recherche animale en laboratoire dans le domaine de la psychologie). *Journal of Social Issues*, 49(1), 133-142. <https://doi.org/10.1111/j.1540-4560.1993.tb00913.x>
- Burger, J. M. (2009). Replicating Milgram: Would people still obey today? (Reproduction de Milgram : Les gens obéiraient-ils encore aujourd'hui ?). *American Psychologist*, 64(1), 1-11. <https://doi.org/10.1037/a0010932>
- Burns, J. F. (2010, 24 mai). British medical council bars doctor who linked vaccine to autism (Le conseil médical britannique exclut le médecin qui a établi un lien entre les vaccins et l'autisme). *The New York Times*. <http://www.nytimes.com/2010/05/25/health/policy/25autism.html>
- Carey, B. (2011, November 2). Fraud case seen as a red flag for psychology research. (Une affaire de fraude considérée comme un signal d'alarme pour la recherche en psychologie.) *The New York Times*. Repéré au <http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html>
- Conseil de recherches en sciences humaines du Canada, Conseil de recherches en sciences naturelles et en génie du Canada, Instituts de recherche en santé du Canada. (2022). *Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2 (2022)*. Gouvernement du Canada.
- Haidt, J., Koller, S. et Dias, M. (1993) Affect, culture, and morality, or is it wrong to eat your dog? (Affect, culture et moralité, ou est-ce mal de manger son chien ?) *Journal of Personality and Social Psychology*, 65(4), 613-628. <https://doi.org/10.1037/0022-3514.65.4.613>
- Jump, P. (2011, November 28). A star's collapse. (Une étoile s'écroule). *Inside Higher Ed*. Repéré au <https://www.insidehighered.com/news/2011/11/28/scholars-analyze-case-massive-research-fraud>
- Koocher, G. P. (1977). Bathroom behavior and human dignity (Comportement dans les toilettes et dignité humaine). *Journal of Personality and Social Psychology*, 35(2), 120-121. <https://doi.org/10.1037/0022-3514.35.2.120>
- Mann, T. (1994). Informed consent for psychological research: Do subjects comprehend consent forms and understand their legal rights? (Consentement éclairé pour la recherche en psychologie : Les sujets comprennent-ils les formulaires de consentement et leurs droits légaux ?) *Psychological Science*, 5(3), 140-143. <https://doi.org/10.1111/j.1467-9280.1994.tb00650.x>
- Milgram, S. (1963). Behavioral study of obedience (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67(4), 371-378. <https://doi.org/10.1037/h0040525>
- Middlemist, R. D., Knowles, E. S. et Matter, C. F. (1976). Personal space invasions in the lavatory : Suggestive evidence for arousal (Envahissement de l'espace personnel dans les toilettes : preuves suggestives de pulsion). *Journal of Personality and Social Psychology*, 33(5), 541-546. <https://doi.org/10.1037/0022-3514.33.5.541>

- Miller, N. E. (1985). The value of behavioral research on animals (La valeur de la recherche comportementale sur les animaux). *American Psychologist*, 40(4), 423-440. <https://doi.org/10.1037/0003-066X.40.4.423>
- Reverby, S. M. (2009). *Examining Tuskegee : The infamous syphilis study and its legacy* (Examiner Tuskegee : la tristement célèbre étude sur la syphilis et son héritage). University of North Carolina Press.
- Rosenthal, R. M. (1994). Science and ethics in conducting, analyzing, and reporting psychological research (Science et éthique dans la conduite, l'analyse et le rapport de la recherche en psychologie). *Psychological Science*, 5(3), 127-133. <https://doi.org/10.1111/j.1467-9280.1994.tb00646.x>
- Sieber, J. E., Iannuzzo, R. et Rodriguez, B. (1995). Deception Methods in Psychology: Have They Changed in 23 Years? (Les méthodes de duperie en psychologie : ont-elles changé en 23 ans ?) *Ethics Behavior*, 5(1), 67-85. https://doi.org/10.1207/s15327019eb0501_5.
- Singal, J. (2015, May 29). The Case of the Amazing Gay-Marriage Data: How a Graduate Student Reluctantly Uncovered a Huge Scientific Fraud. *New York Magazine*. Répéré au <http://nymag.com/scienceofus/2015/05/how-a-grad-student-uncovered-a-huge-fraud.html>

Chapitre 4

- Amir, N., Freshman, M. et Foa, E. (2002). Enhanced Stroop interference for threat in social phobia (Interférence Stroop accrue pour la menace dans la phobie sociale). *Journal of Anxiety Disorders*, 16(1), 1-9. [https://doi.org/10.1016/S08876185\(01\)00084-6](https://doi.org/10.1016/S08876185(01)00084-6)
- Bandura, A., Ross, D. et Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models (Transmission de l'agressivité par l'imitation de modèles agressifs). *Journal of Abnormal and Social Psychology*, 63(3), 575-582. <https://doi.org/10.1037/h0045925>
- Cacioppo, J. T. et Petty, R. E. (1982). The need for cognition (Le besoin de cognition). *Journal of Personality and Social Psychology*, 42(1), 116-131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cohen, S., Kamarck, T. et Mermelstein, R. (1983). A global measure of perceived stress (Une mesure globale de la perception du stress). *Journal of Health and Social Behavior*, 24(4), 386-396. <https://doi.org/10.2307/2136404>
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. (Une mesure globale de la perception du stress). *Journal of Health and Social Behavior*, 24(4), 386-396. <https://doi.org/10.2307/2136404>
- Costa, P. T., Jr., et McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory (L'évaluation de la personnalité normale dans la pratique clinique : L'inventaire de personnalité NEO). *Psychological Assessment*, 4(1), 5-13. <https://doi.org/10.1037/1040-3590.4.1.5>
- Delongis, A., Coyne, J. C., Dakof, G., Folkman, S. et Lazarus, R. S. (1982). Relationships of daily hassles, uplifts, and major life events to health status (Relation entre les tracasseries quotidiennes, les satisfactions et les événements majeurs de la vie et l'état de santé). *Health Psychology*, 1(2), 119-136. <https://doi.org/10.1037/0278-6133.1.2.119>

- Ginet, A. et Py, J. (2000). Le besoin de cognition : une échelle française pour enfants et ses conséquences au plan sociocognitif. *L'Année psychologique*, 100(4), 585-628. <https://doi.org/10.3406/psy.2000.28665>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains (Une mesure très succincte des cinq grands domaines de la personnalité). *Journal of Research in Personality*, 37(6), 504-528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Guillet, L. et Hermand, D. (2006). Critique de la mesure du stress. *L'Année psychologique*, 106, 129-164. <https://hal.science/hal-00755690>
- Hogan, T. P., Stephenson, R. et Parent, N. (2007). *Introduction à la psychométrie* (2e éd.). Chenelière éducation.
- Holmes, T. H. et Rahe, R. H. (1967). The Social Readjustment Rating Scale (L'échelle d'évaluation de la réadaptation sociale). *Journal of Psychosomatic Research*, 11(2), 213-218. [https://doi.org/10.1016/0022-3999\(67\)90010-4](https://doi.org/10.1016/0022-3999(67)90010-4)
- MacDonald, T. K., et Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? (L'estime de soi, l'humeur et l'intention d'utiliser des préservatifs : Quand une faible estime de soi conduit-elle à des comportements de santé à risque ?) *Journal of Experimental Social Psychology*, 38(3), 299-306. <https://doi.org/10.1006/jesp.2001.1505>
- Le Corff, Y. et Busque-Carrier, M. (2016). Structural validity of the NEO personality inventory 3 (NEO-PI-3) in a French-Canadian sample (Validité structurelle de l'inventaire de personnalité NEO 3 (NEO-PI-3) auprès d'un échantillon canadien-français). *International Journal of Arts & Sciences*, 9(3), 461-472.
- Le Corff, Y. (s.d.). NEO-PI3™ : *Inventaire de personnalité : Traduction et adaptation canadienne-française*. <https://www.irpcanada.com/neo-pi3>
- Paquette Houde, C., Abdel Baki, A., Lecomte, T., Lussier, Valade, M. et Ngô, T. L. (2018). *Guide de pratique pour le traitement cognitif-comportemental des troubles psychotiques* (1ère éd.). Jean Goulet – tccmontreal. <https://collections.banq.qc.ca/ark:/52327/3533838>
- Petty, R. E., Briñol, P., Loersch, C. et McCaslin, M. J. (2009). The need for cognition. Dans M. R. Leary et R. H. Hoyle (dir.) *Handbook of Individual Differences in Social Behavior* (p.318-329). Guilford Press
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press
- Rosenberg, M. (1989). *Society and the adolescent self-image* (éd. rév.). Wesleyan University Press.
- Segerstrom, S. E. et Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry (Le stress psychologique et le système immunitaire humain : Une étude méta-analytique de 30 ans de recherche). *Psychological Bulletin*, 130(4), 601-630. <https://doi.org/10.1037/0033-2909.130.4.601>
- Stevens, S. S. (1946). On the theory of scales of measurement (Sur la théorie des échelles de mesure). *Science*, 103(2684), 677-680.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions (Études de l'interférence dans les réactions verbales en série). *Journal of Experimental Psychology*, 18(6), 643-662. <https://doi.org/10.1037/h0054651>
- Vallière, E. F., et Vallerand, R. J. (1990). Traduction et validation canadienne-française de l'échelle de l'estime de soi de Rosenberg. *International Journal of Psychology*, 25, 305-316. <https://doi.org/10.1080/00207599008247865>

Chapitre 5

- Bauman, C.W., McGraw, A.P., Bartels, D.M. et Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology (Réexamen de la validité externe : Préoccupations concernant les problèmes de chariots et autres dilemmes sacrificiels en psychologie morale). *Social and Personality Psychology Compass*, 8(9), 536-554. <https://doi.org/10.1111/spc3.12131>
- Birnbaum, M.H. (1999). How to show that $g > 221$: Collect judgments in a between-subjects design (Comment montrer que $g > 221$: Recueillir des jugements dans un plan entre sujets). *Psychological Methods*, 4(3), 243-249. <https://doi.org/10.1037/1082-989X.4.3.243>
- Cialdini, R. (2005, April). Don't throw in the towel: Use social influence research (Ne jetez pas l'éponge : Utiliser la recherche sur l'influence sociale). *APS Observer*. <https://www.psychologicalscience.org/index.php/publications/observer/2005/april-05/dont-throw-in-the-towel-use-social-influence-research.html>
- Chekroun P. (2008). Pourquoi les individus aident-ils moins autrui lorsqu'ils sont nombreux ?. *Revue électronique de Psychologie Sociale*, 2, 9-16. <https://psychologiescientifique.org/wp-content/uploads/2018/02/Chekroun-2008-Pourquoi-les-individus-aident-ils-moins-autrui-lor.pdf>
- Darley, J. M. et Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility (L'intervention des passants dans les situations d'urgence : Diffusion de la responsabilité). *Journal of Personality and Social Psychology*, 8(4, Pt.1), 377-383. <https://doi.org/10.1037/h0025589>
- Fredrickson, B. L., Roberts, T.-A., Noll, S. M., Quinn, D. M. et Twenge, J. M. (1998). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance (Ce maillot de bain devient vous : Différences entre les sexes dans l'auto-objectivation, la restriction alimentaire et les performances en mathématiques). *Journal of Personality and Social Psychology*, 75(1), 269-284. <https://doi.org/10.1037/0022-3514.75.1.269>
- Goldstein, N. J., Cialdini, R. B. et Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels (Une chambre avec un point de vue : L'utilisation des normes sociales pour motiver la conservation de l'environnement dans les hôtels). *Journal of Consumer Research*, 35(3), 472-482. <https://doi.org/10.1086/586910>
- Guéguen, N. et de Gail, Marie-Agnès. (2003). The effect of smiling on helping behavior: Smiling and good Samaritan behavior (L'effet du sourire sur le comportement d'aide : Le sourire et le comportement du bon samaritain). *Communication Reports*, 16(2), 133-140. <https://doi.org/10.1080/08934210309384496>
- Ibolya, K., Brake, A., et Voss, U. (2004). The effect of experimenter characteristics on pain reports in women and men (Effet des caractéristiques de l'expérimentateur sur les rapports de douleur chez les femmes et les hommes). *Pain*, 112(1-2), 142-147. <https://doi.org/10.1016/j.pain.2004.08.008>

- Judd, C.M. et Kenny, D.A. (1981). *Estimating the effects of social interventions (Estimation des effets des interventions sociales)*. Cambridge University Press.
- Knecht, S., Draeger, B., Deppe, M., Bobe, L., Lohmann, H., Floel, A., Ringelstein, E.-B. et Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans (La main et la dominance hémisphérique du langage chez les humains en bonne santé). *Brain: A Journal of Neurology*, 123(12), 2512-2518. <https://doi.org/10.1093/brain/123.12.2512>
- Manning, R., Levine, M. et Collins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses (le meurtre de Kitty Genovese et la psychologie sociale de l'aide : La parabole des 38 témoins). *American Psychologist*, 62(6), 555-562. <https://doi.org/10.1037/0003-066X.62.6.555>
- Morling, B. (2014, April). Teach your students to be better consumers (Apprenez à vos étudiants à devenir de meilleurs consommateurs). *APS Observer*. <http://www.psychologicalscience.org/index.php/publications/observer/2014/april-14/teach-your-students-to-be-better-consumers.html>
- Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Hollingsworth, J. C., Carol M. Ashton, C. M. et Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee (Essai contrôlé de la chirurgie arthroscopique pour l'arthrose du genou). *The New England Journal of Medicine*, 347(2), 81-88. <https://doi.org/10.1056/NEJMoa013259>
- Price, D. D., Finniss, D. G. et Benedetti, F. (2008). A comprehensive review of the placebo effect: Recent advances and current thought (Une revue complète de l'effet placebo : Avancées récentes et réflexions actuelles). *Annual Review of Psychology*, 59, 565-590. <https://doi.org/10.1146/annurev.psych.59.113006.095941>
- Rosenthal, R. (1976). *Experimenter effects in behavioral research (Effets de l'expérimentateur dans la recherche comportementale)* (éd. augmentée). Wiley
- Rosenthal, R., et Fode, K. (1963). The effect of experimenter bias on performance of the albino rat (L'effet du biais de l'expérimentateur sur la performance du rat albinos). *Behavioral Science*, 8(3), 183-189. <https://doi.org/10.1002/bs.3830080302>
- Rosenthal, R. et Rosnow, R. L. (1976). *The volunteer subject (Le sujet volontaire)*. Wiley.
- Shapiro, A. K. et Shapiro, E. (1999). *The powerful placebo: From ancient priest to modern physician (Le puissant placebo : Du prêtre antique au médecin moderne)*. Johns Hopkins University Press.

Chapitre 6

- Abrams, L. S. et Curran, L. (2009). "And you're telling me not to stress?" A grounded theory study of postpartum depression symptoms among low-income mothers (« Et vous me dites de ne pas stresser ? » Une étude théorique sur les symptômes de la dépression post-partum chez les mères à faible revenu). *Psychology of Women Quarterly*, 33(3), 351-362. <https://doi.org/10.1111/pwqu.2008.33.issue-310.1111/j.1471-6402.2009.01506.x>
- Bryman, A. (2012). *Social Research Methods*, 4th ed. Oxford: OUP.

- Bell, E., Bryman, A. et Kleinknecht, S. (2022). *Social Research Methods (Les méthodes en recherche sociale)* (6e éd. canadienne). Oxford University Press.
- Bushman, B. J. et Huesmann, L. R. (2001). Effects of televised violence on aggression (Effets de la violence télévisée sur l'agressivité). Dans D. Singer et J. Singer (dir.), *Handbook of children and the media* (p. 223–254). Sage.
- Cacioppo, J. T. et Petty, R. E. (1982). The need for cognition (Le besoin de cognition). *Journal of Personality and Social Psychology*, 42(1), 116–131. <https://doi.org/10.1037/0022-3514.42.1.116>
- Cohen, D., Nisbett, R. E., Bowdle, B. F. et Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An "experimental ethnography" ("L'insulte, l'agression et la culture de l'honneur dans le Sud : Une « ethnographie expérimentale ».). *Journal of Personality and Social Psychology*, 70(5), 945–960. <https://doi.org/10.1037/0022-3514.70.5.945>
- Corbin, J. et Strauss, A. (2014). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (Les bases de la recherche qualitative : Techniques et procédures pour développer une théorie ancrée)* 4e éd. Sage Publications.
- Diener, E. (2000). Subjective well-being: The science of happiness, and a proposal for a national index (Le bien-être subjectif : La science du bonheur et une proposition d'indice national). *American Psychologist*, 55(1), 34–43. <https://doi.org/10.1037/0003-066X.55.1.34>
- Festinger, L., Riecken, H. et Schachter, S. (1956). *When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world (Quand la prophétie échoue : Une étude sociale et psychologique d'un groupe moderne qui a prédit la destruction du monde.)*. University of Minnesota Press.
- Freud, S. (1966). *Cinq leçons sur la psychanalyse*. Payot.
- Glaser, B. G. et Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research. (La découverte de la théorie ancrée : Stratégies de recherche qualitative)*. Aldine.
- Geertz, C. (1973). *The interpretation of cultures (L'interprétation des cultures)*. Basic Books.
- Geertz, C. (2017). *The interpretation of cultures (L'interprétation des cultures)* 3e éd. Basic Books.
- Jouriles, E. N., Garrido, E., Rosenfield, D. et McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress (Expériences d'agression psychologique et physique dans les relations amoureuses des adolescents : Liens avec la détresse psychologique). *Child Abuse & Neglect*, 33(7), 451–460. <https://doi.org/10.1016/j.chiabu.2008.11.005>
- Kraut, R. E. et Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach (Messages sociaux et émotionnels du sourire : Une approche éthologique). *Journal of Personality and Social Psychology*, 37(9), 1539–1553. <https://doi.org/10.1037/0022-3514.37.9.1539>
- Levine, R. V. et Norenzayan, A. (1999). The pace of life in 31 countries (Le rythme de vie dans 31 pays.). *Journal of Cross-Cultural Psychology*, 30(2), 178–205. <https://doi.org/10.1177/0022022199030002003>
- Lindqvist, P., Johansson, L. et Karlsson, U. (2008). In the aftermath of teenage suicide: A qualitative study of the psychosocial consequences for the surviving family members (Après le suicide d'un adolescent : Une étude

- qualitative des conséquences psychosociales pour les membres survivants de la famille). *BMC Psychiatry*, 8(26), 1-7. <https://doi.org/10.1186/1471-244X-8-26>
- Loftus, E. F. et Pickrell, J. E. (1995). The formation of false memories (La formation de faux souvenirs). *Psychiatric Annals*, 25(12), 720-725. <https://doi.org/10.3928/0048-5713-19951201-07>
- Messerli, F. H. (2012). Chocolate consumption, cognitive function, and Nobel laureates (Consommation de chocolat, fonctions cognitives et lauréats du prix Nobel). *New England Journal of Medicine*, 367(16), 1562-1564. <https://doi.org/10.1056/NEJMon1211064>
- Milgram, S. (1963). Behavioral study of obedience (Étude comportementale de l'obéissance). *Journal of Abnormal and Social Psychology*, 67(4), 371-378. <https://doi.org/10.1037/h0040525>
- Milgram, S. (1974). Obedience to authority: An experimental view (Soumission à l'autorité : Un point de vue expérimental). Harper & Row.
- Milgram, S. (2017). Soumission à l'autorité : un point de vue expérimental (E. Molinié, Trad.; Nouvelle éd. augmentée). Pluriel.
- Pelham, B. W., Carvallo, M. et Jones, J. T. (2005). Implicit egotism (Égoïsme implicite). *Current Directions in Psychological Science*, 14(2), 106-110. <https://doi.org/10.1111/j.0963-7214.2005.00344.x>
- Peterson, C., Seligman, M. E. P., et Vaillant, G. E. (1988). Pessimistic explanatory style is a risk factor for physical illness: A thirty-five year longitudinal study (Le style explicatif pessimiste est un facteur de risque de maladie physique : Une étude longitudinale sur trente-cinq ans). *Journal of Personality and Social Psychology*, 55(1), 23-27. <https://doi.org/10.1037/0022-3514.55.1.23>
- Plomin, R., DeFries, J. C., McClearn, G. E. et McGuffin, P. (2008). *Behavioral genetics* (5e éd.). Worth.
- Radcliffe, N. M. et Klein, W. M. P. (2002). Dispositional, unrealistic, and comparative optimism: Differential relations with knowledge and processing of risk information and beliefs about personal risk (Optimisme dispositionnel, irréaliste et comparatif : Relations différentielles avec la connaissance et le traitement des informations sur le risque et les croyances sur le risque personnel). *Personality and Social Psychology Bulletin*, 28(6), 836-846. <https://doi.org/10.1177/0146167202289012>
- Rentfrow, P. J. et Gosling, S. D. (2008). The do re mi's of everyday life: The structure and personality correlates of music preferences (The do re mi's of everyday life : The structure and personality correlates of music preferences). *Journal of Personality and Social Psychology*, 84(6), 1236-1256. <https://doi.org/10.1037/0022-3514.84.6.1236>
- Rosenhan, D. L. (1973). On being sane in insane places (Être sain d'esprit dans des lieux insensés). *Science*, 179(4070), 250-258. <https://doi.org/10.1126/science.179.4070.25>
- Todd, Z., Nerlich, B., McKeown, S. et Clarke, D. D. (dir.) (2004) Mixing methods in psychology: The integration of qualitative and quantitative methods in theory and practice. (Mélange de méthodes en psychologie : L'intégration des méthodes qualitatives et quantitatives dans la théorie et la pratique). Psychology Press.
- Trenor, J.M., Yu, S.L., Waight, C.L., Zerda, K.S et Sha T.-L. (2008). The relations of ethnicity to female engineering students' educational experiences and college and career plans in an ethnically diverse learning environment (Les relations entre l'ethnicité et les expériences éducatives des étudiantes en ingénierie et leurs plans d'études et de carrière

dans un environnement d'apprentissage ethniquement diversifié). *Journal of Engineering Education*, 97(4), 449-465. <https://doi.org/10.1002/j.2168-9830.2008.tb00992.x>

Watson, J. B. et Rayner, R. (1920). Conditioned emotional reactions (Réactions émotionnelles conditionnées). *Journal of Experimental Psychology*, 3(1), 1-14. <https://doi.org/10.1037/h0069608>

Wilkins, A. (2008). "Happier than Non-Christians": Collective emotions and symbolic boundaries among evangelical Christians (« Plus heureux que les non-chrétiens » : Émotions collectives et frontières symboliques chez les chrétiens évangéliques.). *Social Psychology Quarterly*, 71(3), 281-301. <https://doi.org/10.1177/019027250807100308>

Chapitre 7

Buhrmester, M., Kwang, T. et Gosling, S.D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high quality, data? (Amazon's Mechanical Turk : une nouvelle source de données peu coûteuses mais de grande qualité ?) *Perspectives on Psychological Science*, 6(1), 3-5. <https://doi.org/10.1177/1745691610393980>

Chang, L. et Krosnick, J.A. (2003). Measuring the frequency of regular behaviors: Comparing the 'typical week' to the 'past week' (Mesurer la fréquence des comportements réguliers : comparer "la semaine typique" et "la semaine dernière"). *Sociological Methodology*, 33(1), 55-80. <https://doi.org/10.1111/j.0081-1750.2003.t01-1-00127.x>

Converse, J. M. (1987). *Survey research in the United States : Roots and emergence, 1890-1960 (Recherche par sondage aux États-Unis : racines et émergence, 1890-1960)*. University of California Press.

Gosling, S. D., Vazire, S., Srivastava, S. et John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires (Faut-il faire confiance aux études basées sur le web ? Une analyse comparative de six idées reçues sur les questionnaires en ligne.). *American Psychologist*, 59(2), 93-104. <https://doi.org/10.1037/0003-066X.59.2.93>

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. et Tourangeau, R. (2004). *Survey methodology*. Wiley.

Krosnick, J.A. et Berent, M.K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format (Comparaisons de l'identification à un parti et des préférences politiques : L'impact du format des questions d'enquête). *American Journal of Political Science*, 27(3), 941-964. <https://doi.org/10.2307/2111580>

Lahaut, V. M. H. C. J., Jansen, H. A. M., van de Mheen, D. et Garretsen, H. F. L. (2002). Non-response bias in a sample survey on alcohol consumption (Biais de non-réponse dans une enquête par sondage sur la consommation d'alcool). *Alcohol and Alcoholism*, 37(3), 256-260. <https://doi.org/10.1093/alcalc/37.3.256>

Lerner, J. S., Gonzalez, R. M., Small, D. A. et Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment (Effets de la peur et de la colère sur la perception des risques de terrorisme : Une expérience nationale sur le terrain). *Psychological Science*, 14(2), 144-150.

Likert, R. (1932). A technique for the measurement of attitudes (Une technique de mesure des attitudes). *Archives of Psychology*, 140, 1-55.

Miller, J.M. et Krosnick, J.A. (1998). The impact of candidate name order on election outcomes (L'impact de l'ordre

des noms des candidats sur les résultats des élections). *Public Opinion Quarterly*, 62(3), 291-330.
<https://doi.org/10.1086/297848>

Natala@aws. (2011, Janvier 26). Re: MTurk CENSUS: About how many workers were on Mechanical Turk in 2010? Dans *Amazon Web Services Discussion Forums*. <https://forums.aws.amazon.com/thread.jspa?threadID=58891>

Peterson, R. A. (2000). *Constructing effective questionnaires*. Sage.

Schwarz, N. (1999). Self-reports: How the questions shape the answers (Auto-rapports : comment les questions façonnent les réponses). *American Psychologist*, 54(2), 93-105. <https://doi.org/10.1037/0003-066X.54.2.93>

Schwarz, N. et Strack, F. (1990). Context effects in attitude surveys: Applying cognitive theory to social research (Effets de contexte dans les sondages des attitudes : application de théorie cognitive à la recherche sociale). Dans W. Stroebe et M. Hewstone (dir.), *European review of social psychology* (Vol. 2, p. 31-50). Wiley.

Strack, F., Martin, L. L. et Schwarz, N. (1988). Priming and communication: The social determinants of information use in judgments of life satisfaction (L'amorçage et la communication : les déterminants sociaux de l'utilisation de l'information dans les jugements de satisfaction de la vie). *European Journal of Social Psychology*, 18(5), 429-442.
<https://doi.org/10.1002/ejsp.2420180505>

Sudman, S., Bradburn, N. M. et Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology* (Réfléchir aux réponses : l'application des processus cognitifs aux méthodes des sondages). Jossey-Bass.

Chapitre 8

Cook, T. D., et Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings* (Quasi-expérimentation : Questions relatives à la conception et à l'analyse sur le terrain). Houghton Mifflin.

Eysenck, H. J. (1952). The effects of psychotherapy: An evaluation (Les effets de la psychothérapie : une évaluation). *Journal of Consulting Psychology*, 16(5), 319-324. <https://doi.org/10.1037/h0063633>

Posternak, M. A. et Miller, I. (2001). Untreated short-term course of major depression : A meta-analysis of studies using outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, 66(2-3), 139-146.
[https://doi.org/10.1016/S0165-0327\(00\)00304-9](https://doi.org/10.1016/S0165-0327(00)00304-9)

Smith, M. L., Glass, G. V. et Miller, T. I. (1980). *The benefits of psychotherapy* (Les bienfaits de la psychothérapie). Johns Hopkins University Press.

Chapitre 9

Brown, H. D., Kosslyn, S. M., Delamater, B., Fama, A. et Barsky, A. J. (1999). Perceptual and memory biases for health-related information in hypochondriacal individuals (Biais de perception et de mémoire pour les informations relatives

à la santé chez les personnes hypocondriaques). *Journal of Psychosomatic Research*, 47(1), 67–78.
[https://doi.org/10.1016/S0022-3999\(99\)00011-2](https://doi.org/10.1016/S0022-3999(99)00011-2)

MacDonald, T. K. et Martineau, A. M. (2002). Self-esteem, mood, and intentions to use condoms: When does low self-esteem lead to risky health behaviors? (L'estime de soi, l'humeur et l'intention d'utiliser des préservatifs : Quand une faible estime de soi conduit-elle à des comportements de santé à risque ?) *Journal of Experimental Social Psychology*, 38(3), 299–306. <https://doi.org/10.1006/jesp.2001.1505>

Gilliland, K. (1980). The interactive effect of introversion-extraversion with caffeine induced arousal on verbal performance (L'effet interactif de l'introversion-extraversion et de l'excitation induite par la caféine sur la performance verbale). *Journal of Research in Personality*, 14(4), 482–492. [https://doi.org/10.1016/0092-6566\(80\)90006-9](https://doi.org/10.1016/0092-6566(80)90006-9)

Schnall, S., Benton, J. et Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments (Avec une conscience propre : La propreté réduit la sévérité des jugements moraux). *Psychological Science*, 19(12), 1219–1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>

Schnall, S., Haidt, J., Clore, G. L. et Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096–1109. <https://doi.org/10.1177/0146167208317771>

Chapitre 10

Baer, D. M., Wolf, M. M. et Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97.

Danov, S. E. et Symons, F. E. (2008). A survey evaluation of the reliability of visual inspection and functional analysis graphs. *Behavior Modification*, 32, 828–839.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (2e éd.). Oxford.

Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models. *Behavioral Processes*, 54, 137–154.

Hall, R. V., Lund, D. et Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis*, 1, 1–12.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Lanovaz, M. (2013). L'utilisation de devis expérimentaux à cas unique en psychoéducation. *Revue de psychoéducation*, 42(1), 161–183. <https://umontreal.scholaris.ca/items/4aa6cffa-1976-48a0-9dae-c45faeb63fa2>

Ross, S. W. et Horner, R. H. (2009). Bully prevention in positive behavior support. *Journal of Applied Behavior Analysis*, 42, 747–759.

Scruggs, T. E., et Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality*, 9, 227–244.

- Shadish, W. R., Cook, T. D. et Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Authors Cooperative.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.
- Wolf, M. (1976). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

Chapitre 11

- American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style (7th ed.)*. American Psychological Association.
- American Psychological Association. (s.d.) *Student Paper Setup Guide*.
<https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf>
- American Psychological Association. (s.d.). APA Style. <https://apastyle.apa.org>
- American Psychological Association. (s.d.). APA Style: *Paraphrasing and citation activities*. Repéré le 23 septembre 2025.
<https://apastyle.apa.org/instructional-aids/paraphrasing-citation-activities.pdf>
- Bem, D. J. (2003). Writing the empirical journal article. Dans J. M. Darley, M. P. Zanna, et H. R. Roediger III (dir.), *The complete academic: A practical guide for the beginning social scientist* (2e éd.). American Psychological Association.
- Bentley, M., Peerenboom, C. A., Hodge, F. W., Passano, E. B., Warren, H. C., et Washburn, M. F. (1929). Instructions in regard to preparation of manuscript. *Psychological Bulletin*, 26, 57-63.
- Couture, M. (2025, 25 mars). *Adaptation française de l'APA (7e éd.)*, sous licence CC BY-NC 4.0.
<http://profmcouture.ca/apa/index.htm>
- Darley, J. M., et Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 4, 377-383.
- Gouvernement du Canada. *Écriture inclusive — les lignes directrices et ressources*.
<https://www.noslangues-ourlanguages.gc.ca/fr/cles-de-la-redaction/ecriture-inclusive-lignes-directrices-ressources>
- Girardet, C, et Nuber, N. (s.d.) *Les normes APA (7e éd.)*. Haute école pédagogique Vaud, Suisse
<https://etudiant.hepl.ch/files/live/sites/files-site/files/filiere-ps/programme-formation/normes-apa-stage-maeps-fps-hep-vaud.pdf>
- Institut national de la recherche scientifique (s.d.) *Inclusivement vôtres! Guide de rédaction inclusive* Repéré le 22 avril 2026.
<https://inrs.ca/linrs/equite-diversite-et-inclusion/guide-de-redaction-inclusive/>

- Madigan, R., Johnson, S., et Linton, P. (1995). The language of psychology: APA style as epistemology. *American Psychologist*, 50, 428–436. <https://psycnet.apa.org/doi/10.1037/0003-066X.50.6.428>
- Office québécois de la langue française. (s.d.) *Féminisation et rédaction épicienne*, Dépannage linguistique. Repéré le 22 avril 2026. <https://vitrinelinguistique.oqlf.gouv.qc.ca/banque-de-depannage-linguistique/la-redaction-et-la-communication/feminisation-et-redaction-epicene>
- Université Laval. *Outils et formations pertinentes en matière d'équité, diversité et inclusion*. Repéré le 22 avril 2026. <https://nutriss.ulaval.ca/boite-a-outils/equite-diversite-et-inclusion-edi/outils-et-formations-pertinentes>
- Université de Montréal (s.d.). *Citer selon les normes APA, 7e édition*. Repéré le 23 septembre 2025. <https://bib.umontreal.ca/citer/styles-bibliographiques/apa>
- Université de Montréal (s.d.). *Inclusivement: Guide d'écriture pour toutes et tous*. Repéré le 22 avril 2026. <https://francais.umontreal.ca/soutiller-en-francais/ecriture-inclusive/>
- Université de Montréal (s.d.). *Quiz interactif sur l'écriture en paraphrases*. Repéré le 22 avril 2026. <https://integrite.umontreal.ca/quiz-generaux/paraphraser/>
- Université d'Ottawa (s.d.). *Citer et évaluer les sources*. Repéré le 23 septembre 2025. <https://www.uottawa.ca/bibliotheque/ecriture-citation>
- Université de Sherbrooke. *Le français langue commune à UdeS, le français en outil: La rédaction inclusive*. Repéré le 22 avril 2026. <https://www.usherbrooke.ca/langue/le-francais-en-outils/la-redaction-inclusive>
- Université du Québec (2023). *Guide de communication inclusive: pour des communications qui mobilisent, transforment et ont du style!* <https://reseau.uquebec.ca/system/files/documents/guide-communication-inclusive-universite-du-quebec-2023.pdf>

Chapitre 12

- American Psychological Association. (2020). Publication manual of the American Psychological Association (7e éd.). <https://doi.org/10.1037/0000165-000>
- American Psychological Association (s.d.). Student Paper Setup: APA style 7e edition. [Guidehttps://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf](https://apastyle.apa.org/instructional-aids/student-paper-setup-guide.pdf)
- Bem, D. J. (2003). Writing the empirical journal article. Dans J. M. Darley, M. P. Zanna et H. L. Roediger III (dir.), *The complete academic: A career guide* (2e éd., p. 185-219). American Psychological Association.
- Brown, N. R. et Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *The Journal of Sex Research*, 36(3), 292-297. <https://doi.org/10.1080/00224499909551999>

- Buss, D. M. et Schmitt, D. P. (1993). Sexual Strategies Theory: An Evolutionary Perspective on Human Mating. *Psychological Review*, 100(2), 204-232. <https://psycnet.apa.org/doi/10.1037/0033-295X.100.2.204>
- Carlson, K. A. et Conard, J. M. (2011). The last name effect: How last name influences acquisition timing. *Journal of Consumer Research*, 38(2), 300-307. <https://doi.org/10.1086/658470>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <https://psycnet.apa.org/doi/10.1037/0033-2909.112.1.155>
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16(5), 259-263. <https://doi.org/10.1111/j.1467-8721.2007.00516.x>
- MacDonald, T. K. et Martineau, A. M. (2002). Self-Esteem, Mood, and Intentions to Use Condoms: When does low self-esteem lead to risky health behaviors ? *Journal of Experimental Social Psychology*, 38(3), 299-306. <https://doi.org/10.1006/jesp.2001.1505>
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A. et Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222-243. <https://doi.org/10.1037/a0017619>
- Ollendick, T. H., Öst, L.-G., Reuterskiöld, L., Costa, N., Cederlund, R., Sirbu, C., Davis 3rd, T. E. et Jarrett, M. A. (2009). One-session treatment of specific phobias in youth: a randomized clinical trial in the United States and Sweden. *Journal of Consulting and Clinical Psychology*, 77(3), 504-516. <https://doi.org/10.1037/a0015158>
- Schmitt, D. P. et Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. *Journal of Personality and Social Psychology*, 89(4), 623-642. <https://doi.org/10.1037/0022-3514.89.4.623>

Chapitre 13

- Abelson, R. P. (1995). *Statistics as principled argument*. Mahwah, NJ: Erlbaum.
- Aarts, A. A., Anderson, C. J., Anderson, J., van Assen, M. A. L. M., Attridge, P. R., Attwood, A. S., ... Zuni, K. (2015, 21 septembre). *Reproducibility Project: Psychology*. Récupéré de osf.io/ezcuj
- Aschwanden, C. (2015, 19 août). *Science isn't broken: It's just a hell of a lot harder than we give it credit for*. Fivethirtyeight. Repéré dans <http://fivethirtyeight.com/features/science-isnt-broken/>
- Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., ... can't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, 50, 217-224. <https://doi.org/10.1016/j.jesp.2013.10.005>
- Cohen, J. (1994). The world is round: $p < .05$. *American Psychologist*, 49, 997-1003.
- Découverte (2025). *Répliquer des études scientifiques pour tester leur fiabilité*. Radio-Canada Info. https://www.youtube.com/watch?v=QcC2cFG_1so&t=585s

- Frank, M. (2015, 31 août). *The slower, harder ways to increase reproducibility*. Repéré dans <http://babieslearninglanguage.blogspot.ie/2015/08/the-slower-harder-ways-to-increase.html>
- Head M. L., Holman, L., Lanfear, R., Kahn, A. T., et Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol*, 13(3): e1002106. <https://doi.org/10.1371/journal.pbio.1002106>
- Hyde, J. S. (2007). New directions in the study of gender similarities and differences. *Current Directions in Psychological Science*, 16, 259–263.
- Kanner, A. D., Coyne, J. C., Schaefer, C., et Lazarus, R. S. (1981). Comparison of two modes of stress measurement: Daily hassles and uplifts versus major life events. *Journal of Behavioral Medicine*, 4, 1–39.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Lakens, D. (2017, 25 décembre). *About p-values: Understanding common misconceptions*. <https://correlaid.org/en/blog/understand-p-values/>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., et Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82.
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. Chichester, UK: Wiley.
- Pashler, H., et Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments explained. *Perspectives on Psychological Science*, 7(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 83, 638–641.
- Scherer, L. (2015, septembre). *Guest post by Laura Scherer*. Repéré dans <http://sometimesimwrong.typepad.com/wrong/2015/09/guest-post-by-laura-scherer.html>
- Schnall, S., Benton, J., et Harvey, S. (2008). With a clean conscience: Cleanliness reduces the severity of moral judgments. *Psychological Science*, 19(12), 1219–1222. <https://doi.org/10.1111/j.1467-9280.2008.02227.x>
- Simonsohn U., Nelson L. D., et Simmons J. P. (2014). P-Curve: a key to the file drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. <https://doi.org/10.1037/a0033242>
- Thériault (2024). L'importance de la science ouverte en recherche en psychologie. Accepté dans *Fragments : Revue de psychologie*, 1-14. Preprint: <https://doi.org/10.31234/osf.io/758dx>
- Tramimow, D. et Marks, M. (2015). *Editorial. Basic and Applied Social Psychology*, 37, 1–2. <https://dx.doi.org/10.1080/01973533.2015.1012991>
- Wilkinson, L., et Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604

Yong, E. (27 août, 2015). *How reliable are psychology studies?* Repéré dans <http://www.theatlantic.com/science/archive/2015/08/psychology-studies-reliability-reproducibility-nosek/402466/>

Ressources pour le corps enseignant

Lecture Slides

From the authors: "This ancillary set was created under a [Round Four ALG Textbook Transformation Grant](#).

Note: Most of the lecture slides for research methods are at this link: <http://grants.kennesaw.edu/psychscience/index.php>

Pearcey, Sharon; Kirsner, Beth; Randall, Christopher; Willard, Jen; Williamson, Adrienne; and Downtain, Tricia, "Psychology Research Methods and Statistics: Ancillary Set" (2017). *Psychology, Sociology, Anthropology, and Social Work Ancillary Materials*. 3. <https://oer.galileo.usg.edu/psychology-ancillary/3>

Terms, Definitions, External Resources, Questions, and Slide Presentations

From the author: "The current project helped create ancillary materials for each of the textbook's thirteen chapters. These include **slide presentations, test-banks, and online resources** (e.g., websites, videos). Many of the resources shared here were a culmination of OER-enabled pedagogy, as students were actively involved in the creation of exam questions and collection of resources."

Grissett, Judy Orton, "Research Methods in Psychology Ancillary Set" (2018). *Psychology, Sociology, Anthropology, and Social Work Ancillary Materials*. 9. <https://oer.galileo.usg.edu/psychology-ancillary/9>

Student Lab Guide

Dr. Derrick Wirtz, University of British Columbia, Okanagan
<https://blogs.ubc.ca/psy0270/>

Classroom Activities

TeachPsychScience.Org

Méthodes de recherche en psychologie

1re édition

Merci de faire partie du mouvement des REL (ressources éducatives libres) et d'avoir adopté notre récente 1re édition du livre *Méthodes de recherche en psychologie*. Votre adoption contribue à améliorer l'accès et l'abordabilité pour vos étudiants et libère des ressources financières dont ils ont besoin pour des dépenses essentielles comme la nourriture, le loyer, l'essence, la garde de personnes à charge, etc.

Nous recueillons des données d'adoption uniquement dans le but de quantifier l'impact de notre ouvrage. Les informations que vous fournissez nous aident à plaider de manière plus convaincante en faveur de l'adoption de manuels ouverts auprès de vos collègues et au sein de nos établissements.

AUCUNE DONNÉE DÉTAILLÉE NE SERA PARTAGÉE PAR NOUS. Jamais. Cela signifie que personne d'autre n'aura accès à vos informations d'adoption. Nous utilisons uniquement des données agrégées pour renforcer notre argumentaire quant à la valeur des REL et des livres ouverts. Très occasionnellement, il se peut que nous communiquions par courriel avec les personnes ayant adopté le manuel afin de vérifier les adoptions, d'annoncer des révisions ou de nouvelles éditions, etc.

Encore une fois, merci beaucoup, et n'hésitez pas à communiquer avec nous si vous avez des questions,

Rajiv Jhangiani, Université polytechnique Kwantlen; Carrie Cuttler, Washington State University; et Dana Leighton, Texas A&M University—Texarkana; Molly A. Metz, University of Toronto; Josée Turcotte, Université Laurentienne; et Laura Auchinleck, Université Laurentienne Auteurs et autrices de la 1re édition