

Corrigé du chapitre 3

Comprendre les données sociales

Mark Paradis, Ph. D.

2 septembre 2026

Voici le corrigé détaillé des exercices de *fin de chapitre* du chapitre 3 de *Comprendre les données sociales*. Le corrigé des exercices situés à l'intérieur de chaque section se trouve dans le livre lui-même, immédiatement après les exercices auxquels il répond.

Faites l'exercice avant de lire le corrigé. Un corrigé lu en premier n'est qu'un texte de plus ; un corrigé lu en second est une vérification de votre raisonnement, et seul le second enseigne quelque chose.

Le corrigé complet, pour tous les chapitres, se trouve à <https://www.markpara-dispolitics.com/books/comprendre-les-donnees-sociales/online/solutions/>.

Chapitre 3 Introduction au devis de recherche

Corrigé de l'exercice 3.E1. (a) VI : déjeuner — telle que le titre la suppose, une variable binaire (déjeune ou ne déjeune pas), même si la fréquence vaudrait mieux. VD : les notes, vraisemblablement une moyenne générale ou un score à une épreuve.

(b) H_1 : Les élèves qui déjeunent ont de meilleures notes que les élèves qui ne déjeunent pas. H_0 : Il n'y a aucune différence de notes entre les élèves qui déjeunent et ceux qui ne déjeunent pas.

(c) Deux variables confondantes, chacune avec son mécanisme.

Le revenu du ménage. Les familles qui ont plus de ressources ont plus de chances d'avoir de quoi déjeuner et une routine qui laisse le temps de le faire, et leurs enfants ont plus de livres, un espace d'étude plus calme, du tutorat et des écoles mieux pourvues. Le revenu produit à la fois le déjeuner et les notes.

L'organisation du ménage, ou l'engagement parental. Un ménage qui produit un déjeuner de façon fiable est vraisemblablement un ménage qui fait respecter les heures de coucher, vérifie les devoirs et se présente aux rencontres de parents. Le déjeuner est un marqueur du ménage, non un traitement.

Le sommeil se défend aussi : les élèves qui se lèvent à temps pour manger sont des élèves qui se sont couchés plus tôt, et le sommeil agit directement sur le rendement scolaire.

(d) Un essai contrôlé randomisé. Recruter des élèves et les assigner au hasard à recevoir ou non un programme de déjeuner gratuit. La randomisation rend les groupes comparables quant au revenu, à l'organisation du ménage, au sommeil et à tout le reste, mesuré ou non. Comparer les notes à la fin de la session.

Une difficulté pratique mérite d'être signalée : vous ne pouvez pas empêcher le groupe témoin de déjeuner à la maison, de sorte que la comparaison porte en réalité sur l'offre et l'absence d'offre plutôt que sur le fait de manger ou non. C'est un devis bien réel — c'est ainsi que les programmes de déjeuner scolaire sont effectivement évalués — mais la question à laquelle il répond diffère légèrement de celle que posait le titre.

Corrigé de l'exercice 3.E2. (a) *Fallacieuse.* Z = le revenu national, ou le développement. Les pays plus riches ont une alphabétisation plus élevée et une lecture des journaux plus répandue, et aussi de meilleurs soins maternels, un meilleur assainissement et une meilleure nutrition, qui réduisent la mortalité infantile. Rien dans le fait de lire un journal n'agit sur la survie d'un nourrisson.

(b) *Véritablement causale, probablement — mais avec une causalité inverse à l'œuvre elle aussi.* L'exercice améliore vraisemblablement l'humeur par des voies physiologiques bien documentées. Mais les personnes plus heureuses ont aussi plus d'énergie et de motivation pour s'entraîner, et la dépression réduit l'activité

physique. Les deux flèches sont réelles, ce qui fait de ce cas une causalité réciproque. Notez qu'une corrélation transversale ne peut pas répartir les deux.

(c) *Causalité inverse*. Les villes embauchent plus de policiers *parce que* la criminalité est élevée, et non l'inverse. Lire cela comme si la police causait la criminalité renverse la séquence. C'est la configuration à surveiller chaque fois que la variable indépendante est une réponse des pouvoirs publics — les réponses sont allouées aux problèmes.

(d) *Confondue, et partiellement véritable*. La prise de notes aide probablement, par la répétition et la récupération. Mais les élèves qui prennent plus de notes ont aussi plus de chances d'assister régulièrement aux cours, d'y être attentifs et d'être motivés — autant de facteurs qui agissent sur le résultat aux examens de façon indépendante. La réponse honnête est qu'une partie de l'association est causale et qu'une autre ne l'est pas, et que la corrélation brute surestime l'effet de la prise de notes.

Ce que les quatre cas ont en commun : « fallacieux » et « causal » ne forment pas toujours une opposition nette. Les relations réelles sont fréquemment les deux à la fois, et c'est ce qui les rend difficiles.

Corrigé de l'exercice 3.E3. (a) *Question de recherche* : Fréquenter l'université réduit-il les préjugés envers les groupes extérieurs ? H_1 : Les adultes ayant fréquenté l'université déclarent un niveau de préjugés envers les groupes extérieurs plus faible que les adultes ne l'ayant pas fréquentée. H_0 : Il n'y a aucune différence de préjugés entre les adultes qui ont fréquenté l'université et ceux qui ne l'ont pas fréquentée.

(b) VI : la fréquentation universitaire — binaire, ou mieux, le nombre d'années d'études postsecondaires complétées. VD : les préjugés, qui demandent du soin. Une échelle validée, comme une mesure de distance sociale, ou des thermomètres de sentiment envers des groupes nommés. L'autodéclaration sur ce sujet est vulnérable à la désirabilité sociale (chapitre 2), ce qu'il vaut la peine de signaler au moment de proposer la mesure.

(c) Le devis idéal est une expérimentation, et elle ne peut pas être menée — on ne peut pas assigner au hasard des personnes à fréquenter l'université. Le meilleur devis réalisable est un panel longitudinal mesurant les préjugés avant l'entrée et de nouveau après l'obtention du diplôme, avec un groupe de comparaison de personnes semblables qui n'ont pas fréquenté l'université.

(d) Deux variables confondantes. *Le milieu familial* : les familles scolarisées et plus aisées ont à la fois plus de chances d'envoyer leurs enfants à l'université et plus de chances d'entretenir et de transmettre des attitudes moins empreintes de préjugés. *Les attitudes antérieures* : les jeunes déjà moins portés aux préjugés peuvent avoir plus de chances de choisir l'université, ou d'y choisir certains programmes — une sélection plutôt qu'un effet.

(e) Un devis longitudinal mesure les préjugés *avant* l'université, ce qui répond directement à la seconde variable confondante : si vous connaissez le niveau de départ de chaque personne, vous pouvez demander s'il a changé plutôt que s'il diffère. Il établit aussi l'antériorité temporelle.

Ce qu'il ne règle pas : le milieu familial diffère toujours entre celles et ceux qui fréquentent l'université et les autres, et tout ce qui agit à la fois sur la fréquentation et sur le changement d'attitude au cours de la même période reste confondu. Les données longitudinales suppriment les différences stables entre les personnes ; elles ne suppriment pas les différences entre leurs trajectoires.

Corrigé de l'exercice 3.E4. (a) Transversal — 2 400 adultes mesurés une seule fois.

(b) *Covariation* : satisfaite. $r = ,31$ avec $p < ,001$ établit une association. Notez que $r = ,31$ est modéré : la cohésion rend compte d'une part modeste de la variation de la santé.

Antériorité temporelle : non satisfaite, et elle ne peut pas l'être avec ce devis. Les deux variables ont été mesurées au même moment. L'inverse est plausible — les personnes en meilleure santé sortent davantage, connaissent leurs voisins et perçoivent leur quartier comme plus cohésif.

Élimination des explications concurrentes : non satisfaite. Aucune variable confondante n'a été traitée.

(c) Non. La conclusion est une affirmation causale — *renforcer* la cohésion *améliorerait* la santé — et le devis ne soutient qu'une association. C'est l'implication pour les politiques publiques qui rend le dépassement lourd de conséquences : si la relation est confondue, un programme qui augmente la cohésion n'améliorera pas la santé, et l'argent aura été dépensé pour rien.

(d) Deux variables confondantes. *Le revenu ou la défavorisation du quartier* : les secteurs aisés ont une population plus stable, plus d'équipements collectifs et une meilleure santé par l'alimentation, l'accès aux soins et les conditions de travail. *L'état de santé individuel* : les personnes malades participent moins à la vie collective et déclarent une moins bonne santé, ce qui produit l'association par la VD elle-même.

La stabilité résidentielle, l'âge et le fait d'être propriétaire sont autant de solutions de rechange défendables.

(e) Un panel longitudinal suivant les mêmes adultes sur plusieurs années établirait la séquence et permettrait à chaque personne de servir de point de comparaison à elle-même. Plus solide encore serait une expérimentation naturelle — un changement externe de cohésion sans rapport avec la santé, comme le réaménagement d'un quartier —, mais ces cas sont rares et rarement nets.

Corrigé de l'exercice 3.E5. (a) *Données nécessaires*. Les dépenses en éducation, idéalement en part du PIB plutôt qu'en dollars absolus, pour que des pays de tailles

différentes soient comparables (§2.6.5). Le PIB par habitant. Les deux pour le même ensemble de pays et — c'est déterminant — mesurés à plus d'un moment si l'affirmation doit être éprouvée plutôt que décrite. Sources : la Banque mondiale ou l'OCDE.

(b) *Devis*. Un devis comparatif international et, si la mesure se fait à un seul moment, transversal. Cela soutient la description et la corrélation, non la causalité.

(c) *Vraisemblablement causale ?* Partiellement, au mieux, et les explications concurrentes sont fortes.

La causalité inverse est la plus sérieuse. Les pays riches peuvent se permettre de dépenser davantage en éducation. Comme les dépenses se mesurent habituellement en part du PIB, un pays plus riche qui dépense la même part dépense bien plus par élève — et la direction va de la richesse vers la dépense aussi facilement que l'inverse.

La confusion. La capacité de l'État, la stabilité politique et l'état de droit font vraisemblablement monter à la fois la dépense publique et la production économique. Une population déjà qualifiée aussi.

Le décalage. Les dépenses en éducation agissent sur l'économie par la productivité des personnes qu'elles ont formées, ce qui représente un décalage de quinze à vingt-cinq ans. Une corrélation contemporaine mesure un intervalle entièrement inadéquat.

(d) *Passer à la causalité*. Établir la séquence avec des données de panel sur des décennies, de sorte qu'on puisse montrer que les changements de dépense précèdent les changements de croissance. Contrôler les variables confondantes mesurables. Chercher une expérimentation naturelle — une réforme de l'éducation motivée par quelque chose d'étranger à l'économie. Et préciser le mécanisme, puisque « la dépense » recouvre les salaires des enseignants, les bâtiments et les subventions universitaires, qui n'ont aucune raison d'agir de la même façon.

C'est l'une des questions les plus étudiées de l'économie du développement et elle reste disputée, ce qui est en soi instructif sur la difficulté de l'inférence causale à partir de données comparatives.

Corrigé de l'exercice 3.E6. Votre question sera la vôtre : ce corrigé expose donc ce que chacun des six éléments doit accomplir, puis donne un spécimen travaillé auquel confronter le vôtre.

1. *L'unité d'analyse, énoncée*. C'est l'omission la plus courante. Les individus, les ménages, les quartiers et les pays font des études différentes, même avec les deux mêmes variables.

2. *Une opérationnalisation précise*. « Mesurer le bien-être » n'est pas une opérationnalisation. Nommer une échelle, un ensemble d'items ou un indicateur administratif en est une.

3. Un H_1 directionnel, et un H_0 qui est l'absence de relation plutôt que son renversement. H_0 énonce qu'il n'y a aucune différence; il n'énonce pas que la différence va dans l'autre sens.

4. Une justification qui nomme un mécanisme. C'est l'élément le plus souvent laissé mince. « Des recherches antérieures suggèrent un lien » n'est qu'une justification empirique. Un exposé de la manière dont la cause produirait l'effet est ce qui rend une hypothèse digne d'être éprouvée, et il engendre habituellement d'autres prédictions vérifiables.

5. Un devis accordé à la question, avec une menace reconnue. Chaque devis de la §3.2 laisse quelque chose d'ouvert. Si vous ne savez pas nommer une faiblesse du vôtre, c'est que vous ne l'avez pas encore trouvée, non que vous l'avez évitée.

6. Une variable confondante plausible et justifiée. Elle doit agir à la fois sur la variable indépendante et sur la variable dépendante, et vous devez dire pourquoi.

Gardez l'ensemble réalisable. Un projet exigeant un panel de vingt ans sur quarante pays n'est pas un devis, c'est un souhait.

Un spécimen travaillé.

(1) Chez les cégépiennes et les cégépiens du Québec, le fait de travailler contre rémunération pendant la session abaisse-t-il le rendement scolaire ? L'unité d'analyse est l'étudiante ou l'étudiant pris individuellement.

(2) VI : le nombre d'heures rémunérées travaillées par semaine pendant la session, autodéclaré sous forme de dénombrement. VD : la moyenne des notes de la session, tirée des dossiers de l'établissement plutôt que de l'autodéclaration, puisque les étudiants déclarent leurs notes dans un sens qui les flatte.

(3) H_1 : les étudiants qui travaillent plus d'heures rémunérées par semaine ont une moyenne de session plus faible. H_0 : il n'y a aucune relation entre les heures rémunérées travaillées par semaine et la moyenne de session.

(4) Mécanisme : le temps est fini. Les heures travaillées sont des heures qui ne vont pas aux travaux scolaires, au sommeil ou à la présence en classe, qui agissent chacun sur le rendement. Le mécanisme prédit aussi quelque chose de plus — l'effet devrait être plus grand aux nombres d'heures élevés qu'aux nombres faibles, puisque les premières heures se prennent sur les loisirs et les dernières sur l'étude.

(5) Devis : une enquête transversale appariée aux dossiers de notes, ce qui est réalisable pour un travail étudiant. Menace à la validité interne : la sélection. Les étudiants qui choisissent de travailler de longues heures peuvent différer des autres d'une manière qui agit sur les notes par elle-même, de sorte que le devis ne peut pas séparer le travail du type d'étudiant qui travaille.

(6) Variable confondante : le revenu du ménage. Les étudiants de ménages à plus faible revenu travaillent davantage et affrontent d'autres désavantages — trajets plus longs, espace d'étude moins calme, moins de soutiens payants — qui abaissent

les notes indépendamment de l'emploi.

Corrigé de l'exercice 3.E7. (a) Un devis *comparatif* — des pays comme unités, comparés les uns aux autres. Comme il compte trois vagues, c'est aussi un panel, quoique de pays plutôt que d'individus.

(b) *Covariation* : oui, le jeu de données peut l'établir. Une forte association positive entre la scolarité et l'espérance de vie est mesurable.

Antériorité temporelle : en partie. Trois vagues permettent de demander si les pays dont la scolarité a monté plus tôt ont vu leur espérance de vie monter plus tard. C'est plus que ce qu'offre une coupe transversale, et moins qu'un long panel.

Élimination des explications concurrentes : non. Le jeu de données contient une poignée de variables, et n'importe quelle caractéristique nationale non mesurée pourrait produire l'association.

L'affirmation honnête est donc que l'éducation et l'espérance de vie sont associées d'un pays à l'autre, et que l'association persiste d'une vague à l'autre — non que l'éducation fait monter l'espérance de vie.

(c) Le revenu national est la variable confondante évidente : les pays plus riches ont à la fois plus de scolarité et des vies plus longues. En vérifiant à l'annexe B, *worldsim* contient *bel et bien* *gdppc*, de sorte que celle-là est mesurable. Bien d'autres — la qualité du système de santé, les conflits, l'assainissement — ne le sont pas, et c'est là tout le propos de l'exercice : les variables confondantes que vous pouvez traiter sont celles que quelqu'un a pensé à mesurer.

(d) Trois vagues ajoutent la possibilité d'observer le changement dans le temps et d'établir une séquence approximative. Elles ne fournissent pas de randomisation, ne saisissent pas les variables confondantes non mesurées, et ne permettent pas d'écarter qu'une troisième caractéristique nationale ait changé en même temps que les deux autres. Trois observations forment aussi une série courte pour détecter des effets décalés.

(e) Les chiffres ne concorderont pas, et ils ne sont pas censés concorder. *worldsim* donne au Canada une espérance de vie supérieure à quatre-vingt-quatre ans à la première vague et un peu plus de quatre-vingts à la deuxième — une chute de plusieurs années en une seule vague, qu'aucun pays réel n'a connue sur la période que couvre la Banque mondiale. Les chiffres de scolarité se situent dans une fourchette plausible, mais ils ne suivent pas non plus les séries publiées.

La leçon est celle qu'il faut emporter dans tout le reste du livre. *worldsim* a été construit de sorte que l'éducation et l'espérance de vie y soient fortement associées ; l'association est dans les données parce qu'on l'y a mise. Une conclusion tirée de ce jeu de données est une conclusion sur la simulation, et elle vous dit si vous avez appliqué la méthode correctement — ce à quoi le jeu de données sert. Elle n'est pas une donnée sur le Canada, et aucun de ses nombres ne devrait jamais être cité

comme s'il l'était.

Corrigé de l'exercice 3.E8. L'article est le vôtre : ce qui suit expose donc ce que chaque partie demande, et la configuration que vous avez le plus de chances de trouver.

(a) L'écart à chercher se tient entre le verbe du titre et celui de l'étude. « Lié à » dans le corps du texte devenant « cause » dans le titre est la configuration décrite à la §3.5.4, et elle est très courante. Reformuler l'affirmation précisément suppose de nommer la population, les deux variables telles qu'elles ont été réellement mesurées, et la force de la relation si elle est rapportée.

(b) Bien des articles ne rapportent pas du tout le devis. Le dire est une bonne réponse et non un échec à en trouver un — et l'absence est elle-même instructive, puisqu'une affirmation dont le devis n'est pas divulgué ne peut pas être évaluée par un lecteur.

(c) Appliquez les quatre questions à cet article-ci plutôt qu'en général. « Il pourrait y avoir une variable confondante » ne fait pas avancer l'analyse ; en nommer une qui soit plausible, avec le mécanisme par lequel elle agirait sur les deux variables, oui.

(d) La phrase à viser a cette forme : *cette étude montre que X et Y sont associés chez [population], et n'établit pas que X cause Y*. Si vous savez écrire cette phrase au sujet de votre article, vous avez la compétence à laquelle ce chapitre servait.

Une chose que vous remarquerez souvent : le texte de l'article est plus prudent que son titre, et les chercheurs qu'on y cite le sont plus encore. Cette chaîne d'escalade — du chercheur, au communiqué, au titre — est le point d'éthique de la §3.5.