

Corrigé du chapitre 1

Comprendre les données sociales

Mark Paradis, Ph. D.

2 septembre 2026

Voici le corrigé détaillé des exercices de *fin de chapitre* du chapitre 1 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 1 Introduction aux méthodes quantitatives

Corrigé de l'exercice 1.E1. (a) Le mot qui fait des dégâts, c'est *prouve*. Une étude montrant que les adolescents qui utilisent davantage les médias sociaux déclarent plus de symptômes dépressifs a établi une association. Trois autres explications restent ouvertes : la flèche causale peut aller dans l'autre sens, des adolescents déprimés se repliant sur leur téléphone ; un troisième facteur — perturbation du sommeil, isolement social, situation familiale — peut produire les deux ; ou l'association peut être un artefact de la composition de l'échantillon. Le chapitre 3 traite cela comme il se doit.

(b) Appuyer une affirmation causale exige un devis, et non davantage de données. Une étude longitudinale mesurant les deux variables chez les mêmes personnes sur plusieurs années pourrait au moins établir ce qui est venu en premier. Une expérimentation assignant au hasard une partie des personnes participantes à un usage réduit serait plus forte encore, puisque la randomisation rend les groupes comparables sur tout, mesuré ou non. Les deux exigent des instruments validés plutôt qu'un seul énoncé autodéclaré, et les deux exigent un échantillon qui représente une population définie.

(c) Plusieurs enjeux éthiques se présentent. La recherche auprès de personnes mineures exige le consentement parental en plus de l'assentiment de la jeune personne elle-même. Interroger sur des symptômes dépressifs peut causer de la détresse et crée l'obligation de fournir de l'information sur les ressources d'aide. Une expérimentation restreignant l'accès aux médias sociaux intervient dans la vie sociale des personnes participantes, ce qui demande une justification proportionnée à la connaissance acquise.

Remarquez que (b) et (c) tirent en sens contraire : le devis qui appuierait le mieux l'affirmation causale est aussi celui qui soulève les questions éthiques les plus sérieuses. Cette tension est réelle et reviendra tout au long du livre.

Corrigé de l'exercice 1.E2. *Devis quantitatif.* Question : la fréquence de consommation de contenu politique sur les médias sociaux prédit-elle le vote et d'autres formes de participation chez les 18 à 29 ans ? Données : une enquête auprès d'un échantillon représentatif de ce groupe d'âge, mesurant l'usage des plateformes, l'exposition au contenu politique et plusieurs formes de participation, avec des variables de contrôle démographiques. Force : les résultats se généralisent à la population échantillonnée, et des explications concurrentes comme la scolarité peuvent être prises en compte. Limite : l'enquête saisit ce que les gens déclarent plutôt que ce qu'ils font, et même une association forte n'établira pas la direction.

Devis qualitatif. Question : comment les jeunes adultes comprennent-ils le rapport entre leur vie en ligne et leur action politique ? Données : vingt à trente entrevues en profondeur, échantillonnées de façon raisonnée pour varier les niveaux

de participation. Force : ce devis peut mettre au jour des mécanismes et des significations qu'une enquête ne peut pas anticiper — y compris des raisons que personne n'aurait pensé à inscrire dans un questionnaire. Limite : il ne peut pas établir la fréquence d'un motif.

Ce ne sont pas deux réponses concurrentes à la même question. L'enquête établit qu'un motif existe et à quel point il est répandu ; les entrevues expliquent ce qui s'y passe. Un devis mixte menant les entrevues en premier améliorerait l'enquête, parce qu'il lui fournirait le vocabulaire des personnes concernées pour formuler les questions.

Corrigé de l'exercice 1.E3. Réponse ouverte. Quatre critères pour vérifier votre réponse.

Les trois questions viennent-elles vraiment de disciplines différentes ? Trois variations sur le comportement électoral, c'est une discipline trois fois. Visez des questions qu'une sociologue, un économiste et une psychologue reconnaîtraient chacun comme relevant de leur domaine.

Les questions trouvent-elles réponse dans ce jeu de données ? Une question sur l'évolution dans le temps exige un jeu de données à plusieurs vagues. Une question d'effet causal ne peut pas être tranchée par des données d'enquête transversales ; dites-le plutôt que d'en réclamer trop.

Les variables sont-elles nommées précisément ? C'est ici que parcourir le dictionnaire des variables rapporte : nommez de vraies variables, ou citez la formulation exacte des questions.

Le plan d'analyse est-il esquissé raisonnablement ? Il n'est pas nécessaire de connaître le nom de la technique. « Comparer la confiance moyenne entre les personnes ayant fait des études postsecondaires et celles qui n'en ont pas fait » est une réponse complète et juste à ce stade-ci.

Corrigé de l'exercice 1.E4. (a) Trois enjeux, au minimum.

Le consentement. Des publications écrites pour ses abonnés n'ont pas été offertes à la recherche. La visibilité publique décrit qui *peut* voir quelque chose, et non à qui cela était destiné ; et des conditions d'utilisation ne sont pas un consentement éclairé.

La sensibilité et les torts. Une information sur la santé mentale porte une charge de stigmatisation et peut avoir des conséquences sur l'emploi, l'assurance et les relations. Les personnes qui publient pendant une catastrophe sont, par définition, en détresse et moins en mesure de considérer jusqu'où leurs mots peuvent voyager.

L'identifiabilité. Une citation textuelle peut généralement être retracée au moyen d'un moteur de recherche, et une poignée d'attributs — tranche d'âge, ville, profession — réduit souvent une population à une seule personne. « Aucun nom n'a été employé » n'est pas une anonymisation.

(b) Mesures d'atténuation : ne rapporter que des tendances agrégées ; paraphraser plutôt que citer ; demander l'examen d'un comité d'éthique de la recherche même si les données sont publiques ; exclure les comptes appartenant à des personnes mineures ; et se demander si la question de recherche pourrait trouver réponse dans des données dont les personnes concernées ont consenti à l'usage.

(c) Les deux réponses se défendent ; c'est le raisonnement qui compte. Un *oui* défendable met en balance le bénéfice public d'une meilleure compréhension de la santé mentale en situation de catastrophe et le faible risque individuel lorsque seuls des agrégats sont diffusés, et il accepte les mesures d'atténuation ci-dessus comme des conditions. Un *non* défendable soutient qu'on ne peut pas se passer du consentement au seul motif qu'il est malcommode à obtenir, et que la vulnérabilité pendant une catastrophe renforce l'obligation plutôt qu'elle ne l'affaiblit. Trancher sans avoir pesé l'autre côté laisse l'exercice à moitié fait.

Corrigé de l'exercice 1.E5. L'argument contient une description juste et une conclusion fausse, et la réponse doit les séparer.

Ce qui est juste. Le postpositivisme accepte bel et bien que la mesure soit imparfaite et que les valeurs de la personne qui fait la recherche façonnent ce qu'elle étudie. C'est une description exacte, et non une concession arrachée sous la pression.

Où l'argument échoue. La conclusion suppose que, sans certitude, il n'existe aucune base pour comparer des études. Cela ne suit pas. Des études peuvent se comparer sur des bases qui n'ont rien à voir avec la certitude : l'échantillon représente-t-il la population qu'il prétend représenter ? la mesure a-t-elle été validée ? l'analyse a-t-elle été spécifiée avant que les données ne soient vues ? le résultat se reproduit-il ? l'incertitude déclarée correspond-elle à ce que montrent les données ? Une étude dotée d'un échantillon représentatif et d'un instrument validé vaut mieux qu'une étude qui n'a ni l'un ni l'autre, et aucun recours à la certitude n'est nécessaire pour le dire.

Le point de fond. Le postpositivisme remplace la certitude individuelle par une procédure collective. L'objectivité n'est pas un état qu'une personne atteint en étant sans préjugés ; c'est ce que produisent ensemble la méthode, la transparence, la réplication et la critique par des personnes aux présupposés différents. L'argument traite l'objectivité comme une propriété des individus, ce qui est précisément le présupposé que le postpositivisme abandonne.

Être d'accord avec l'énoncé de départ reste défendable, à condition que la réponse affronte ce contre-argument plutôt que de répéter l'énoncé.

Corrigé de l'exercice 1.E6. Plusieurs réponses se défendent ; c'est la colonne des limites qui vaut la réflexion.

Tranche de revenu (7 catégories). Économie, sociologie. Limite : les tranches font perdre l'information à l'intérieur de chaque intervalle, et la tranche supérieure

est ouverte, de sorte que la queue supérieure — là où loge l'inégalité — devient invisible.

Confiance envers le gouvernement (0–10). Science politique. Limite : « gouvernement » est ambigu — le gouvernement en poste, la fonction publique, ou le régime lui-même — et les personnes interrogées peuvent répondre à propos d'objets différents.

Pays de naissance. Sociologie, géographie, démographie. Limite : la variable enregistre où une personne est née et rien sur le moment de son arrivée, son statut ou son attachement ; elle sert souvent d'indicateur indirect pour des choses qu'elle ne mesure pas.

Nombre d'enfants. Démographie, économie, sociologie. Limite : la variable ne dit pas s'il s'agit des enfants nés, des enfants vivants ou des enfants présents dans le ménage — trois nombres différents.

Santé autoévaluée (5 points). Santé publique, psychologie, sociologie. Limite : la variable mesure la santé perçue, et une même condition objective est déclarée différemment selon les cultures, les âges et les attentes.

Identification partisane. Science politique. Limite : les catégories sont propres à chaque pays et instables dans le temps, ce qui rend difficiles la comparaison internationale et la comparaison de longue durée.

Le motif à dégager : chacune de ces limites porte sur ce que la mesure laisse de côté ou brouille, et non sur l'arithmétique appliquée ensuite.

Corrigé de l'exercice 1.E7. Votre réponse dépend des pays que vous avez choisis. Le raisonnement ci-dessous vaut pour n'importe lesquels des trois.

(a) Les écarts seront importants — plusieurs années dans la plupart des cas — et cela tient au devis et non à une erreur.

(b) L'ordre devrait tenir. Un pays dont l'espérance de vie réelle est plus élevée devrait avoir une espérance de vie simulée plus élevée.

(c) Oui, et c'est le but de l'exercice. Les données de ce livre reproduisent la *structure* du monde — quels pays se classent au-dessus de quels autres, et grosso modo comment les relations s'infléchissent — tout en faussant délibérément les *niveaux*.

(d) Trois raisons ; deux suffisent. D'abord, les données simulées ont une vérité connue : comme elles ont été construites plutôt qu'observées, les vraies valeurs de la population sont disponibles, et une estimation peut donc être comparée à la bonne réponse, ce qu'aucun jeu de données réel ne permet. Ensuite, elles sont stables — les chiffres réels sont révisés, ce qui ferait cesser de correspondre les exercices et leur corrigé. Enfin, les licences : reproduire de vrais jeux de données dans un livre vendu entraîne des coûts et des restrictions que les données synthétiques évitent.

Il y en a une quatrième qui mérite d'être ajoutée. Ayant vérifié la distorsion

vous-même, vous avez maintenant une raison concrète de vous méfier de tout nombre de ce livre pris comme un fait sur le monde — ce qui est exactement l'habitude que la préface demande.

Corrigé de l'exercice 1.E8. Réponse ouverte. Le but est de lire de près une section méthodologique, une habitude qui vaut la peine d'être prise tôt.

(a) *Qu'a-t-on mesuré, et comment.* Repérez l'instrument lui-même — une échelle précise, une donnée administrative, une série d'énoncés d'enquête — plutôt que le concept. « Ils ont mesuré le bien-être » ne fait que répéter le nom de la variable ; « ils ont employé une échelle de cinq énoncés portant sur les deux dernières semaines » décrit la mesure.

(b) *Qui a été étudié, et comment sélectionné.* Distinguez la population de l'échantillon, et notez la méthode de recrutement. Les études sont souvent muettes là-dessus ; remarquer ce silence fait partie de la réponse.

(c) *Une chose que l'étude ne peut pas vous apprendre.* Les limites les plus utiles sont celles qui découlent du devis plutôt que du sujet : qu'une étude transversale ne peut pas établir la séquence ; qu'un échantillon tiré d'une seule ville ne peut pas parler pour un pays ; qu'une mesure autodéclarée ne peut pas distinguer un vrai changement d'un changement dans la disposition à le déclarer.

Une section méthodologique est souvent plus ardue à lire que les résultats, et contient souvent moins de détails qu'on ne s'y attend. Les deux constats méritent d'être faits.

Corrigé de l'exercice 1.E9. 1. *La confiance envers les gouvernements nationaux a-t-elle décliné dans les démocraties riches depuis trois décennies ?* (a) Quantitative. La question porte sur une ampleur et sur une évolution dans le temps, pour de nombreux cas — soit exactement ce à quoi servent les enquêtes à vagues répétées. (b) L'unité d'analyse est le couple pays-année, même si les données sous-jacentes sont recueillies auprès d'individus. (c) L'obstacle est la comparabilité : il faut que la question ait voulu dire la même chose dans chaque pays et à chaque décennie, et la traduction, l'évolution du vocabulaire politique et le changement des modes d'enquête menacent tous cette exigence.

2. *Pourquoi certaines personnes qui se méfient de leur gouvernement continuent-elles néanmoins de voter ?* (a) Qualitative, ou mixte à noyau qualitatif. La question porte sur un raisonnement et sur un sens, et l'éventail des réponses ne peut pas être anticipé assez bien pour rédiger des questions fermées. (b) L'unité est l'individu. (c) L'obstacle est que les gens ne sont pas toujours des narrateurs fiables de leurs propres motifs, et qu'une justification formulée après coup se distingue mal d'un raisonnement réel.

3. *Une campagne d'information publique augmente-t-elle la confiance envers une autorité sanitaire locale ?* (a) Quantitative, et expérimentale si possible — « X

augmente-t-il Y » est une question causale, et la randomisation en est la meilleure réponse disponible. (b) L'unité est l'individu, ou peut-être le quartier si la campagne est déployée géographiquement. (c) L'obstacle est la contamination : les personnes du groupe témoin peuvent voir la campagne malgré tout, et la confiance peut bouger pour des raisons sans rapport avec elle.

Le motif à dégager, c'est que la forme de la question, et non le sujet, détermine l'approche. Les trois portent sur la confiance envers le gouvernement ; elles appellent trois devis différents.