



Évaluer les données à l'ère de la désinformation : guide étape par étape

Cet aide-mémoire vous aide à évaluer de manière critique les données à une époque où la désinformation (renseignements dont on peut vérifier la fausseté ou l'inexactitude) est de plus en plus courante. Bien qu'aucun ensemble de données ne soit parfait, comprendre ses forces et ses limites vous permet d'éviter d'être induit en erreur et de prendre des décisions plus éclairées, fondées sur des preuves.

Source

Pouvez-vous identifier la source spécifique des données?

Si l'estimation provient d'un rapport, pouvez-vous en identifier la source originale?

La source est-elle fiable et digne de confiance?

Les données utilisées pour produire l'estimation sont-elles assorties de limitations ou de mises en garde connues en ce qui a trait à leur utilisation?

Méthodes

Comment les données ont-elles été obtenues - la méthodologie est-elle transparente?

Les données proviennent-elles d'une enquête par sondage auprès d'une population d'intérêt bien définie, d'un sondage de type marketing, d'une base de données administratives, ou reflètent-elles les conclusions d'une étude qualitative telle qu'un groupe de discussion?

Actualité

Ces données sont-elles d'actualité?

Quand cet ensemble de données a-t-il été acquis?

Savez-vous quand les données ont été mises à jour pour la dernière fois?

La période de collecte a-t-elle une incidence sur la validité des données?

L'échantillon

Quelle est la taille de l'échantillon?

Quelle est l'unité d'échantillonnage (par exemple, individus, ménages, entreprises)?

Quel est le taux de réponse (sur l'ensemble des unités échantillonnées, quelle est la proportion de celles qui ont répondu)?

Comment les répondants ont-ils été obtenus (sont-ils issus d'un échantillonnage aléatoire ou se sont-ils portés volontaires au départ)?

L'échantillon représente-t-il la population concernée (par exemple, tous les Canadiens) ou certaines unités sont-elles systématiquement exclues de l'échantillon, ce qui se traduit par une couverture inadéquate de cette population?

Définitions et mesures

Comment le concept d'intérêt est-il défini et comment est-il mesuré?

La définition repose-t-elle sur une norme reconnue¹?

Si vous comparez des estimations provenant de sources différentes, les définitions et les mesures du concept sous-jacent sont-elles les mêmes, ou du moins comparables?

Exactitude

Comment l'estimation a-t-elle été calculée?

Des indicateurs de qualité² ont-ils été mis à disposition des utilisateurs pour évaluer l'adéquation des données à l'utilisation prévue?



Le contexte

Quel est le contexte général des données?

Comment ces estimations se comparent-elles...

... aux estimations provenant d'autres sources de données?

... aux estimations similaires des années précédentes, si elles sont disponibles?

... entre les groupes sur la base des connaissances actuelles sur les différences entre ceux-ci?

Un contexte suffisant a-t-il été fourni pour reconnaître l'expérience vécue ou les réalités historiques de la (des) population(s) à laquelle (auxquelles) les données se réfèrent?

Désagrégation

Les données sont-elles disponibles au bon niveau de désagrégation³ pour votre analyse?

Sources de données supplémentaires

Existe-t-il d'autres sources de données susceptibles d'apporter une compréhension plus nuancée du concept mesuré?

Le couplage⁴ de plusieurs ensembles de données pourrait-il révéler de nouvelles informations?

Considérations éthiques⁵

Quel est le bénéfice public attendu de l'utilisation de ces données?

Quelles sont les conséquences négatives potentielles de l'utilisation de ces données (p. ex. atteinte à la vie privée, manque de transparence, préjudices pour des individus ou des groupes, érosion de la confiance, enjeux de durabilité, préoccupations en matière de sécurité, etc.)?

Les bénéfices potentiels sont-ils crédibles compte tenu des conséquences possibles?

¹ Pour plus d'informations sur les normes utilisées par Statistique Canada, consultez Normes, sources de données et méthodes (<https://www.statcan.gc.ca/fr/concepts/index>).

² Les indicateurs de qualité devraient décrire la précision (par exemple, les intervalles de confiance) ainsi que les biais potentiels pouvant découler de faibles taux de réponse (si les répondants diffèrent des non-répondants) ou d'une représentativité / couverture inadéquate de la population cible dans les données échantillonées. Pour une introduction aux notions fondamentales de la qualité des données, regardez la vidéo *Qualité des données en six dimensions* (<https://www.statcan.gc.ca/fr/afc/litteratie-donnees/catalogue/892000062020001>). Pour plus d'informations sur les intervalles de confiance, consultez *Comprendre les intervalles de confiance dans le cadre du Recensement de 2021* (<https://www12.statcan.gc.ca/census-recensement/2021/ref/98-20-0001/982000012021003-fra.cfm>).

³ Les données désagrégées sont des données qui sont ventilées au niveau de détail le plus bas possible. La désagrégation permet de mettre en évidence des différences que la vue d'ensemble peut masquer. Statistique Canada dispose d'un ensemble de normes en matière de données désagrégées, que l'on peut les Normes relatives aux données désagrégées (<https://www.statcan.gc.ca/fr/concepts/ndd>).

⁴ Statistique Canada dispose de l'infrastructure nécessaire pour coupler les données de plusieurs domaines, tels que la santé, la justice, l'éducation et le revenu. Pour plus d'informations, consultez:

- Environnement de couplage de données sociales (ECDS) (<https://www.statcan.gc.ca/fr/ecds/index>)
- Environnement de fichiers couplables – Entreprises (<https://www.statcan.gc.ca/fr/apercu/statcan/efc>)

⁵ Pour plus d'information sur l'éthique des données, regardez la vidéo *L'éthique des données* (<https://www.statcan.gc.ca/fr/afc/litteratie-donnees/catalogue/892000062022001>). Vous pouvez aussi consulter le rapport *Les fondements des examens éthiques à Statistique Canada* (<https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2024002-fra.htm>).