

Mesures

Mettre votre nom

21-10-2021

Un élément clé de votre design de recherche

Concepts et mesures

Que mesurer

Comment mesurer

L'erreur de mesure et ses conséquences

Limiter l'erreur de mesure

Un élément clé de votre design de recherche

La mesure

- ▶ La mesure est une partie essentielle de votre design de recherche.
- ▶ La mesure découle de votre théorie, de la façon dont vous pensez que le monde fonctionne et dont vous pensez que votre traitement manipule ce monde. Autrement dit, la mesure est un exercice théorique autant que technique et logistique.
- ▶ Lorsque nous enregistrons un nombre, un mot ou une lettre dans un ensemble de données pour **représenter** quelque chose de plus ou moins abstrait dans le monde (comme “la stabilité financière” , “la faim” ou “la capacité en mathématiques”), nous **mesurons**.

Une mauvaise mesure est problématique

- ▶ Des problèmes de mesure peuvent vous amener à tirer des inférences (causales) incorrectes de votre étude (erreur systématique).
- ▶ Un bruit de mesure réduit la puissance statistique (erreur aléatoire).

Collecte de données

- ▶ La collecte de données prend souvent une très grande partie du temps et des ressources financières disponibles dans le budget du projet.
- ▶ De nouvelles données peuvent être un résultat de recherche utile en soi et une base importante pour la recherche future. Les données sont un bien public !

Concepts et mesures

Du concept aux données/scores I

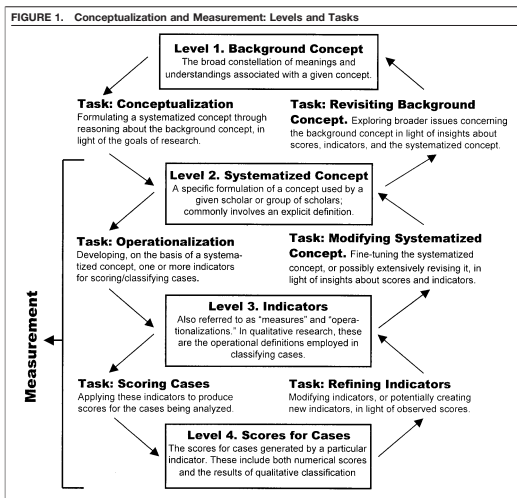


Figure 1: La mesure, depuis le concept jusqu'aux scores (ou données enregistrées). De **Adcock and Collier (2001)**.

Du concept aux données/scores II

- ▶ Quel exemple de concept, comme un résultat, aimeriez-vous changer avec un traitement expérimental ?
- ▶ Comment sauriez-vous qu'une unité (comme une personne ou un village) affiche une "capacité" ou un "rang" élevé ou faible pour ce concept ?
- ▶ Que pourriez-vous observer qui vous ferait sentir plus ou moins confiant que cette personne était plus affamée ou moins favorable à la violence que cette autre personne ?
- ▶ À quelles critiques pourriez-vous faire face si vous disiez : "Je pense que cette unité diffère de cette autre unité sur mon résultat conceptuel" ?

Validité de la mesure

- ▶ On parle ici de “**validité** de la mesure.”
- ▶ Une mesure valide est une mesure pour laquelle nous pouvons argumenter de manière convaincante qu'elle représente ce que nous disons qu'elle représente, et qu'elle ne représente pas d'autres aspects de l'unité. (voir **Shadish et al. (2002)** pour en savoir plus sur la validité et la fiabilité de la mesure).

Question I

- ▶ *“Un bateau qui peut faire quarante milles à l’heure en eau calme fait un voyage de cent milles le long d’un certain cours d’eau. Si ce voyage dure deux heures, combien de temps le retour prendra-t-il ?” (à partir de la SAT de 1926 aux USA)*
- ▶ Est-ce que cela mesure la “capacité en mathématiques” ? la “capacité verbale” ? ou la connaissance en navigation ?

Question II

Choisissez la paire de mots qui ressemble le plus à la paire du haut :

Coureur : Marathon

- (a) envoyé : ambassade
- (b) martyr : massacre
- (c) rameur : régates
- (d) cheval : écurie

La bonne réponse est : (c) rameur : régates. (à partir des années 1980 SAT aux USA)

- Est-ce que cela mesure la connaissance des sports universitaires ? Ou la "capacité verbale" ?

Que mesurer

Vous devez tout mesurer

- ▶ Résultat pour toutes les unités de votre étude (pour chaque série avec les données manquantes).
- ▶ Traitement (assignation et conformité).
- ▶ Des indications que votre traitement est reçu et interprété comme vous l'attendez (contrôle de manipulation).
- ▶ Indications des dommages possibles causés par vos recherches.
- ▶ Covariables, y compris le contexte qui pourrait (a) affecter le fonctionnement de votre intervention ou (b) influencer la variabilité de votre résultat.

Les indicateurs doivent être connectés aux théories et aux mécanismes

- ▶ Nous avons souvent plusieurs théories sur *comment* une intervention peut affecter un résultat (mécanismes différents).
- ▶ Mesurer des indicateurs qui sont *uniques* à chaque mécanisme et qui peuvent aider à les *différencier*.
- ▶ Ces indicateurs peuvent inclure des résultats intermédiaires qui sont réalisés *avant* le résultat final.
- ▶ Ils peuvent inclure des résultats secondaires tels que
 - ▶ les résultats pour lesquels nous attendons des effets uniquement selon certaines théories.
 - ▶ les résultats placebo pour lesquels nous n'attendons aucun effet.

Indicateurs

- ▶ Le cas idéal est la mesure directe du concept ou du phénomène d'intérêt sans erreur (rarement possible).
 - ▶ Une assignation de traitement sous votre contrôle est l'exception. Il est crucial de l'enregistrer.
- ▶ Nous ne pouvons souvent mesurer que des **indicateurs** liés mais pas entièrement déterminants pour le concept sous-jacent ou le phénomène d'intérêt.
 - ▶ Des réponses correctes à des problèmes spécifiques (indicateurs) pour l'aptitude mathématique sous-jacente (le phénomène réel).
 - ▶ Nombre de jours sans nourriture (indicateurs) pour la faim (le phénomène réel).

Sélectionnez des indicateurs valides

- ▶ Des personnes raisonnables peuvent être en désaccord sur la conceptualisation. Qu'entendons-nous vraiment par aptitude mathématique ou faim ? Qu'est-ce que nos scores pourraient dire d'autre sur les gens, villages, etc. en plus de ce que nous espérons représenter ?
- ▶ Sélectionnez des indicateurs étroitement liés à ce que nous **entendons** lorsque nous parlons du phénomène d'intérêt.
- ▶ Sélectionnez des indicateurs **valides**, i.e. la connaissance mathématique ne devrait pas également mesurer la connaissance en navigation.

Sélectionnez des indicateurs fiables

- ▶ Sélectionnez des indicateurs **fiables**.
- ▶ Exemple : Un mètre doit toujours mesurer un mètre quelle que soit la température. Un mètre en caoutchouc peut produire des mesures **peu fiables** du concept abstrait de longueur, surtout s'il est utilisé par un enfant de 5 ans. Un laser peut produire des mesures plus fiables qu'un mètre en bois entre les mains d'un utilisateur expérimenté pour mètre en bois.

Mesures multiples

- ▶ Si vous avez *plusieurs indicateurs pour le même phénomène*, vous devrez déterminer comment agréger ces indicateurs.
 - ▶ La moyenne n'aura pas toujours de sens. Avez-vous des exemples ?
- ▶ Avoir plusieurs indicateurs peut souvent améliorer la **fiabilité** de la mesure et peut-être renforcer les arguments en faveur de sa **validité**.
- ▶ La combinaison d'un bâton en bois et d'un laser pourrait être idéale pour nous convaincre que nous avons mesuré la longueur d'une manière que nous comprenons comme **longueur** (plutôt que, disons, une température) et que le score particulier ne dépend pas de manière cruciale du contexte de la mesure.

Comment mesurer

Outils de mesure et sources de données

- ▶ Après avoir déterminé quel concept mesurer et comment nous pourrions le reconnaître lorsque nous le voyons, nous devons déterminer **comment** le mesurer.
- ▶ Nous disposons de divers outils et sources :
 - ▶ Mesures d'enquête
 - ▶ Mesures comportementales
 - ▶ Données administratives (fiches fiscales, résultats des élections, etc.)
 - ▶ Images, télédétection
 - ▶ Texte (transcriptions de discours, journaux, etc.)
 - ▶ Capteurs électroniques, téléphones, autres appareils (appareils ménagers)
 - ▶ Autres

Considérations dans le choix des outils et des sources

Différents outils impliquent différents compromis :

- ▶ Validité : le score capture-t-il le concept et uniquement le concept ?
- ▶ Fiabilité : une procédure de notation utilisée deux fois donnerait-elle la même réponse ?
- ▶ Biais (erreur systématique)
- ▶ Précision (erreur aléatoire)
- ▶ Échantillon pour lequel vous pouvez faire des mesures
- ▶ Timing des mesures
- ▶ Coût

Maintenir la symétrie entre les bras de traitement

- ▶ Vous devez maintenir la symétrie entre les groupes de traitement et de contrôle pendant que vous effectuez la mesure.
 - ▶ Le nombre de fois et pendant combien de temps vous interagissez avec les participants doit être le même dans tous les bras de traitement.
 - ▶ Les questions doivent être les mêmes.
 - ▶ Soyez particulièrement vigilant sur les points précédents pour les indicateurs utilisés lors des contrôles de manipulation.
- ▶ **La mesure ne doit pas être l'intervention expérimentale.** Nous voulons que la seule différence entre les bras de traitement soit l'intervention, pas la mesure.

L'erreur de mesure et ses conséquences

Erreur de mesure I

- ▶ Nous voulons éviter/minimiser deux types d'erreur de mesure :
 - ▶ Erreur systématique (biais)
 - ▶ Erreur aléatoire (manque de précision)

Erreur de mesure II

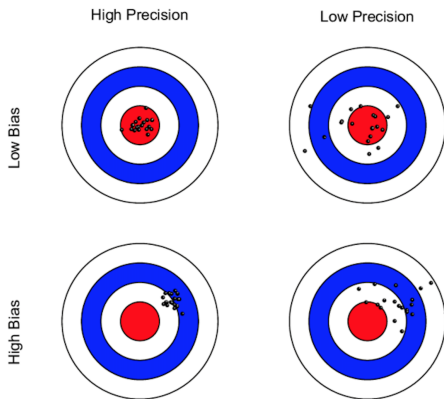


Figure 2: Biais et précision

(Source : Précision et biais de mesure)

Exemples d'erreurs de mesure

- ▶ Erreur systématique
 - ▶ Une balance mal calibrée, donc tout le monde semble peser 2 kg de moins qu'ils ne le sont en réalité
 - ▶ Un journal alimentaire qui sous-évalue systématiquement les collations
 - ▶ Effet de demande
 - ▶ Effet Hawthorne
- ▶ Erreur aléatoire
 - ▶ Une main tremblante pour mesurer une distance
- ▶ Notez que ce sont des exemples de mesure **peu fiable**. Elles peuvent également être **invalides**, mais ce n'est pas obligatoire.

Conséquences d'une mauvaise mesure du traitement

- ▶ En plus de générer une description incorrecte du niveau de certains phénomènes, l'erreur de mesure peut affecter nos inférences causales.
- ▶ Si la variable de traitement est binaire (une unité peut être en traitement ou en contrôle seulement), alors l'erreur de mesure est négativement corrélée avec la vraie variable. (Un 1 est mal codé comme 0, donc l'erreur est -1 ; un 0 est mal codé comme 1, donc l'erreur est 1.)
- ▶ Si vous utilisez la méthode des moindres carrés pour calculer les estimateurs d'effet moyen du traitement, ce type d'erreur conduit à des estimations plus petites de l'effet causal (coefficient sur la variable de traitement).

Conséquences d'une erreur de mesure aléatoire

- ▶ Avec la méthode des moindres carrés, une erreur aléatoire plus importante dans le résultat conduit à des estimations moins précises de l'effet causal (coefficient sur la variable de traitement).
- ▶ La réduction de l'erreur de mesure aléatoire dans le résultat peut augmenter la puissance statistique (car le résultat a moins de bruit non lié au traitement).

Conséquences de l'erreur de mesure systématique I

- ▶ Si toutes les mesures sont décalées du même montant, comme -2kg pour tout le monde :

- ▶ cela ne fait aucune différence lorsque les effets du traitement sont définis comme des différences de résultats potentiels,

$$\tau_i = Y_i(1) - Y_i(0), \tilde{\tau}_i = (Y_i(1) - 2) - (Y_i(0) - 2), \text{ donc}$$

$$\tau_i = \tilde{\tau}_i$$

- ▶ mais cela est problématique si l'effet du traitement est défini comme le rapport des résultats potentiels ($Y_i(1) > 0$ et $Y_i(0) > 0$).

$$\tau_i = Y_i(1)/Y_i(0), \tilde{\tau}_i = (Y_i(1) - 2)/(Y_i(0) - 2), \text{ donc } \tau_i \neq \tilde{\tau}_i \text{ sauf quand } Y_i(1) = Y_i(0).$$

- ▶ La distance entre $\tilde{\tau}_i$ et τ_i dépend de la taille de 2 (l'erreur) par rapport aux valeurs réelles $Y_i(0)$, $Y_i(1)$.
- ▶ Remarque : les coefficients de régression logistique sont des ratios de résultats potentiels.

Conséquences de l'erreur de mesure systématique II

- ▶ L'erreur de mesure peut être corrélée à la vraie valeur de Y .
- ▶ Par exemple, les personnes qui se livrent à un comportement désapprouvé, embarrassant ou illégal peuvent sous-déclarer ce comportement, tandis que celles qui ne le font pas le rapportent correctement. (Ceci est connu sous le nom de biais de désirabilité sociale.)
 - ▶ Cette sous-déclaration peut aussi se produire avec les victimes de ce type de comportement, qui se blâment ou craignent des sanctions de ceux qui préfèrent ne pas savoir (par exemple, les victimes de violence entre partenaires intimes).
- ▶ Cela rend plus difficile la détection d'un effet d'une intervention visant à réduire ce comportement.

Conséquences de l'erreur systématique de mesure III

- ▶ Une autre forme de biais de désirabilité sociale peut également conduire à corrélér l'erreur de mesure avec le traitement.
- ▶ Par exemple, votre intervention pourrait viser à réduire les attitudes hostiles envers les membres d'autres groupes sociaux. Si les participants peuvent comprendre les objectifs de votre étude, ils peuvent (inconsciemment) essayer de plaire au chercheur en lui disant ce qu'il veut entendre. Les personnes du groupe de traitement peuvent sous-estimer leur hostilité envers les autres groupes par rapport au groupe témoin.
- ▶ Cela rend difficile de savoir si la différence dans les résultats observés entre les groupes de traitement et de contrôle est due à l'intervention réduisant réellement l'hostilité ou à la connaissance du traitement modifiant le signalement de l'hostilité.

Limiter l'erreur de mesure

Quelques options pour limiter les erreurs de mesure

- ▶ L'auto-déclaration par un sujet (sur une enquête) est plus problématique que l'observation discrète du sujet ("dans la nature") par quelqu'un d'autre.
- ▶ Les mesures comportementales sont moins sujettes aux biais de désirabilité sociale.
- ▶ Les dossiers administratifs pour lesquels les fausses déclarations entraînent des sanctions légales pourraient être plus précis.
- ▶ Offrir plus de confidentialité afin que la notation puisse se faire sans observation du chercheur ou autre tiers.
- ▶ Garder certaines hypothèses et objectifs de l'étude cachés aux participants à l'étude.
- ▶ Si vous ne pouvez pas contrôler l'erreur de mesure, étudiez-la — déterminez sa propension et s'il s'agit d'un problème. Envisager des études pilotes axées sur la mesure.

Exemple - dossiers administratifs

- ▶ Utiliser des feuilles de présence à une réunion, au lieu de demander si quelqu'un y a assisté.
 - ▶ Il se peut qu'il n'y ait des registres de présence que pour les réunions qui ne présentent pas d'intérêt pour votre population cible d'origine.
 - ▶ Vous devrez peut-être planifier à l'avance la collecte de données lors d'une prochaine réunion.

Exemple - mesures comportementales I

- ▶ Demandez aux sujets de vrais efforts comme la signature d'une pétition, un don ou une autre tâche qui a un petit coût personnel, au lieu de demander aux sujets s'ils soutiennent un problème particulier.
 - ▶ Cela peut ne capturer que les sujets qui se soucient fortement de cette question.
 - ▶ Exemple : le travail de Pedro Vicente sur l'achat de voix à Sao Tomé-et-Principe utilise des enquêtes, des dossiers administratifs et indique si les personnes interrogées ont envoyé une lettre prépayée pour mesurer les résultats. Voir la [Note d'orientation des politiques publiques EGAP 20 : L'achat de votes est-il efficace ?](#)

Exemple - mesures comportementales II

- ▶ Organiser des “jeux de laboratoire” pour mesurer la coopération ou la générosité envers les groupes externes, au lieu de demander aux sujets s'ils coopéreraient avec les autres.
 - ▶ L'étude de Scacco et Warren sur les préjugés et la discrimination utilise des variantes du jeu du dictateur. Voir [Le contact social peut-il réduire les préjugés et la discrimination ? Une expérience de terrain au Nigeria](#)

Exemple - couverture ou confidentialité I

- ▶ Fournir une “couverture” pour que les répondants croient que leurs réponses ne peuvent pas leur être attribuées.
 - ▶ Réponse aléatoire : le hasard détermine si le répondant doit répondre honnêtement à une question ou répondre “oui” indépendamment de la vérité. L'enquêteur ne sait pas dans quelle condition le hasard a mis le répondant.

Exemple - couverture ou confidentialité II

- ▶ Liste d'expériences : donnez aux répondants une liste d'items ou d'énoncés et demandez combien sont vrais pour eux. Les répondants sont randomisés pour recevoir une liste différente : une d'entre elles contient un élément sensible supplémentaire (par exemple, "mon mari me bat"). Cela permet au chercheur d'estimer la prévalence d'un élément particulier. Notez que cette approche réduit la puissance statistique pour une taille d'échantillon donnée.
 - ▶ Confidentialité simple : pour des questions telles que le choix du vote, demandez au répondant de remplir un bulletin de vote fictif et de le placer dans une boîte verrouillée au lieu de répondre directement à l'enquêteur.

Exemple - Cacher les hypothèses aux répondants

- Voir l'étude de Scacco et Warren sur la théorie du contact social à Kaduna, au Nigeria, où les participants étaient recrutés pour un programme de compétences en informatique, aucun programme n'était présenté comme un programme visant à réduire les préjugés et la discrimination.

Quelle quantité d'erreur est trop ?

- ▶ La réduction des erreurs de mesure est importante, mais peut être assez coûteuse. Alors qu'avez-vous besoin de faire ?
- ▶ Cela dépend de l'échelle de l'expérience et de vos objectifs.
 - ▶ Comparer la taille des erreurs à la taille de l'effet du traitement. Changer les mentalités est difficile. Le biais de désirabilité sociale peut être important par rapport aux faibles effets du traitement sur les attitudes.
 - ▶ Comparez la taille des erreurs à l'étendue possible de ce résultat. Oublier un centime sur votre solde bancaire n'affecte pas sensiblement la mesure de votre patrimoine global.

Conseils généraux pour la mesure I

- ▶ Commencez par une utilisation standard des indicateurs que la communauté scientifique a acceptés pour représenter le concept d'intérêt.
 - ▶ Ceux-ci auront été testés sur le terrain pour vous et la comparabilité des mesures entre différentes études et sites d'étude est un plus.
- ▶ Mais attention à bien vérifier que les indicateurs standard ont du sens.
 - ▶ Par exemple, la façon dont on mesure le revenu peut différer entre les zones riches et les zones pauvres.
 - ▶ De même, la façon dont on mesure les attitudes politiques peut différer dans des régimes plus ou moins démocratiques.

Conseils généraux pour la mesure II

- ▶ Rappelons qu'une mesure combinée des capacités mathématiques et verbales liées à la navigation de plaisance était une pratique courante de 1920 à 1980 aux États-Unis. N'oubliez pas que vous mesurez des constructions sociales et que vous devez donc faire attention, autant que possible, aux croyances antérieures, partiellement ou non examinées, que vous et la communauté des chercheurs apportez à l'étude (pour des exemples plus frappants de ce problème, voir Gould ([1981] 1996)) .
- ▶ Contactez d'autres chercheurs dans votre domaine.
- ▶ Mesurez plus finement pour la plage de la variable de résultat où vous attendez un changement.

Ressources pour les enquêtes dans la recherche

- ▶ Les principales archives de données sont consultables par sujet. L'ICPSR est une archive majeure :

<https://www.icpsr.umich.edu/web/pages/ICPSR/>

- ▶ Pew Research Center sur le design de questionnaires

<https://www.pewresearch.org/methods/u-s-survey-research/questionnaire-design/>

- ▶ Tools4Dev pour la rédaction de questions

<https://www.tools4dev.org/resources/how-to-write-awesome-survey-questions-part-1/>

- ▶ Les bibliothèques universitaires ont souvent des guides pour trouver des données :

<https://guides.libraries.emory.edu/c.php?g=944707&p=6810109>

Références

10 choses à savoir sur la mesure dans les expériences

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95 (3): 529–46.

Gould, Stephen Jay. [1981] 1996. *The Mismeasure of Man*. WW Norton & Company.

Shadish, William R., Thomas D. Cook, Donald Thomas Campbell, and others. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*/William r. Shadish, Thomas d. Cook, Donald t. Campbell. Boston: Houghton Mifflin,.