

Le cours de mesurage a pour objectif principal de permettre aux étudiants d'apprendre à utiliser les mots « mesurer », « mesure » etc. avec une signification précise. Le cours distingue deux significations dont les implications sont distinctes tant théoriquement que méthodologiquement. Pour qualifier ces deux significations, on dira que l'une est *classique* (cas du mesurage de la quantité de masse d'un tas de farine avec une balance à aiguille par exemple) tandis que l'autre est non-classique ou encore *évaluative* (cas de l'évaluation de l'intelligence avec un test d'intelligence par exemple).

Prérequis (cours de L1 à réviser)

1. Mesurer, mesurant, mesuré

Dans l'expression « la balance mesure le poids », qui a la forme « [sujet] mesure [complément d'objet direct] », le sujet est le *mesurant* ; le COD (complément d'objet direct) du verbe est le *mesuré*.¹ De même, dans l'expression « les matrices de Raven mesurent l'intelligence », « les matrices de Raven » est le mesurant et « l'intelligence » le mesuré. La grammaire du verbe « mesurer » veut que le sujet du verbe « mesurer » assure l'action de mesurer et que le COD du verbe « mesurer » soit l'objet de l'action de mesurer. Mais on ne peut pas s'en tenir à la grammaire.

2. Dépendance causale

Comment est-il possible que la balance mesure le poids ? La balance mesure le poids parce que la position indiquée par son aiguille varie *en fonction de* la quantité de masse des objets qu'on pose sur son plateau.² La position de l'aiguille (mesurant) *dépend de* la quantité de masse (mesuré). Dans les limites de fonctionnement de la balance, plus la quantité de masse augmente, plus l'aiguille penche à droite ; plus elle diminue, plus l'aiguille penche à gauche ; si la quantité de masse ne varie pas, alors l'aiguille ne bouge pas. La variation de la quantité de masse *cause* le mouvement de l'aiguille.³

Il serait incorrect d'inverser la dépendance : la quantité de masse des objets pesés *ne dépend pas* de la position de l'aiguille car le mouvement de l'aiguille de la balance n'a aucun *effet* sur la masse des objets pesés.

3. Formalisation fonctionnelle

La fonction $F: D \rightarrow A$ [lire « F de (grand) D dans (grand) A »] est croissante au sens large si $d < d' \Rightarrow F(d) \leq F(d')$ [lire « (petit) d est strictement inférieur à (petit) d prime implique F de d est inférieur ou égal à F de d prime »], où d et d' sont des éléments quelconques de D, D l'ensemble (ou domaine) de

¹ Quelque chose, le mesurant, mesure autre chose, le mesuré.

² Mathématiquement, la masse est un segment, la quantité de masse un point sur ce segment.

³ Mathématiquement, on dit que la position de l'aiguille dépend de manière *croissante* de la quantité de masse.

définition (ou de départ) de F et A l'ensemble dans lequel F prend ses valeurs (son ensemble d'arrivée).⁴

La balance mesure le poids parce qu'il existe la fonction de mesurage $F: Q \rightarrow Y$ croissante au sens large, où Q est l'ensemble des quantités de masse des objets qu'on peut poser sur la balance sans la dérégler et Y l'ensemble des positions que l'aiguille de la balance peut occuper sur son cadran lorsque la balance est bien réglée. Soient deux objets de masse q et q' , avec $q < q'$ et $F(q) = y$ et $F(q') = y'$ les positions respectives de l'aiguille. Alors $y \leq y'$ – l'aiguille obéit à l'effet du poids (force) que la masse exerce sur la balance.⁵

Si les mouvements de l'aiguille ne dépendaient pas de la masse des objets pesés ou s'ils dépendaient aussi d'autres facteurs, on ne pourrait pas, en se basant sur l'observation de l'aiguille de la balance *déduire*, au sens logique du terme, quoique ce soit concernant la masse de l'objet pesé. Admettre l'existence d'une fonction de mesurage constitue une *condition nécessaire* à l'interprétation classique du verbe « mesurer ». D'où, sur le plan de la méthodologie scientifique, la nécessité de tester l'hypothèse qu'il existe une fonction de mesurage dans telle ou telle condition expérimentale, si on veut employer le verbe « mesurer » au sens classique, notamment lorsqu'on met au point un instrument de mesure.

4. Application à la proposition « un test psychologique donné mesure l'intelligence »

Que signifie qu'un test psychologique donné mesure l'intelligence ? Si on veut assumer l'acception classique du verbe « mesurer », il faut affirmer qu'une fonction de mesurage existe. Quel serait son ensemble de définition ? Son ensemble d'arrivée ? L'intelligence est-elle définie comme une grandeur par les théories de l'intelligence ? Si non, l'intelligence n'est pas mesurable au sens classique du terme puisque le COD du verbe « mesurer » est par définition une grandeur. Admettons que l'intelligence soit définie comme une grandeur, ce qui permet de considérer que l'ensemble des quantités d'intelligence soit le domaine de définition d'une fonction de mesurage. Quel est l'ensemble d'arrivée de cette fonction ? Autrement dit, quels sont les observables ('mesurants') possibles étant donné l'emploi de tel test psychologique ?

5. Réponse au test psychologique

La réponse à un test psychologique a deux formes. La première, dans l'ordre chronologique, est la suite des réponses à tous les items du test.⁶ La seconde est un *score*, dont le calcul nécessite la connaissance de la première et l'utilisation d'une règle de codage numérique des réponses aux items du test. Dans le cours de L1, on se limitera à considérer le score. Ainsi, l'ensemble d'arrivée d'une fonction de mesurage psychologique peut être défini comme l'ensemble des scores possibles. Si le score à tel test d'intelligence mesure l'intelligence, c'est qu'il existe une fonction de mesurage $F: Q \rightarrow Y$ où Q est l'ensemble des quantités d'intelligence possibles et Y l'ensemble des scores possibles, telle que $q > q' \Rightarrow F(q) \geq F(q')$ où $F(q) = y$ est un score. Ainsi, le score dépend par hypothèse de manière croissante d'une quantité d'intelligence.⁷

⁴ Pour une présentation plus détaillée, voir les points 2.1 à 2.6 du cours vidéo accessible ici : https://www.youtube.com/watch?v=wd0UIOpBz6l&feature=em-upload_owner. Pour l'usage des notations voir si nécessaire <https://epistemo.hypotheses.org/3470> et <https://epistemo.hypotheses.org/3951>.

⁵ La vérité d'une telle affirmation d'existence est, comme toute vérité scientifique, hypothétique. Ce qui importe est que l'hypothèse soit testable.

⁶ Voir <https://epistemo.hypotheses.org/3951> pour une approche détaillée qui sera abordée en L2.

⁷ Pour la différence entre croissance au sens strict et croissance au sens large d'une fonction, voir note 4.

6. Test expérimental, falsification, corroboration de l'hypothèse de mesurage

Du point de vue logique, tester l'existence d'une fonction croissante dans un certain contexte expérimental implique de définir à l'avance un ensemble (on dit aussi une classe) d'observations théoriquement absurdes étant donnée l'hypothèse de l'existence d'une fonction de mesurage.⁸ Par exemple, on considère la quantité de masse q d'un récipient de farine posé sur la balance et la position de l'aiguille correspondante $F(q) = y$; on ajoute de la farine dans le récipient, ce qui implique que la quantité de masse résultante, q' , est telle que $q' > q$. Théoriquement la position de l'aiguille, $F(q') = y'$, est telle que $y' \geq y$ (F est croissante). Les observations de la forme $y' < y$ (l'aiguille se déplace vers la gauche quand on ajoute de la farine) sont théoriquement absurdes (ou encore impossibles). Elles *falsifient* logiquement l'hypothèse. Si les observations ne falsifient pas l'hypothèse, on dit qu'elles la *corroborent*, c'est-à-dire qu'on ne sait pas déduire de ces observations que l'hypothèse est fautive, et qu'on ne sait pas déduire non plus que l'hypothèse est vraie, ce qui permet néanmoins de conserver l'hypothèse en tant qu'hypothèse *plausible* et de dire qu'on l'a testée avec succès, le succès étant d'autant plus impressionnant que le test a été sévère (ou difficile).

Le raisonnement qui précède a une forme logique qu'il s'agit de connaître et d'utiliser pour pouvoir s'expliquer analytiquement. On considère la *prémisse* A qui implique logiquement la *conclusion* B et on note l'implication $A \Rightarrow B$. La *table de vérité* de « $A \Rightarrow B$ » est reproduite ci-dessous, où 0 et 1 signifient respectivement « faux » et « vrai ».⁹

A	B	$A \Rightarrow B$	Diagramme
0	0	1	z
0	1	1	y
1	0	0	w
1	1	1	x

Lorsque « $A \Rightarrow B$ » est vraie (lignes 1, 2 et 4 de la table), vérifier la conclusion B (i.e., attester que B est vraie) n'apporte aucune information concernant la vérité de la prémisse A : A peut être fautive quand B est vraie (ligne 2), ou vraie quand B est vraie (ligne 4). Falsifier B lorsque « $A \Rightarrow B$ » est vraie permet de déduire que A est fautive.¹⁰ Le raisonnement hypothético-déductif qui permet de conclure à l'inexistence d'une fonction de mesurage dans des conditions expérimentales données repose sur une implication de la forme $A \Rightarrow B$ et plus précisément sur ce qu'on appelle sa *contraposée*, $\text{non-}B \Rightarrow \text{non-}A$.¹¹

7. Hypothèses auxiliaires

Dans un test expérimental, la prémisse A est la conjonction logique d'au moins deux propositions.¹² Revenons à notre exemple : on déduit que $y' \geq y$ (B) en acceptant la conjonction logique de trois propositions :

⁸ Ces observations possibles sont les *falsificateurs* de l'hypothèse.

⁹ Pour une interprétation graphique ou plus de détails, voir <https://epistemo.hypotheses.org/3470>.

¹⁰ C'est ce simple fait logique qui a conduit Karl Popper à développer une philosophie critique rationnelle de la connaissance scientifique – voir par exemple cette citation : <https://epistemo.hypotheses.org/132>.

¹¹ Pour montrer que les deux expressions sont équivalentes voir <https://epistemo.hypotheses.org/3470>.

¹² La conjonction logique est la concaténation de plusieurs propositions par « et », voir <https://epistemo.hypotheses.org/3470>.

A1. q et q' existent (i.e., la grandeur « masse » des objets posés sur la balance existe et les objets possèdent une quantité de masse donnée),

A2. la position de l'aiguille dépend de manière croissante de la quantité de masse de l'objet posé sur la balance,

A3. $q < q'$ (puisqu'on ajoute de la farine).

Ainsi, on a l'implication (A1 et A2 et A3) \Rightarrow B – où B signifie « $F(q) \leq F(q')$ » ou encore « $y \leq y'$ ». Supposons qu'on observe que $y > y'$ (non-B). D'où, comme non-B \Rightarrow non-(A1 et A2 et A3), non-(A1 et A2 et A3). Il faut que les hypothèses auxiliaires A1 et A3 soient considérées comme vraies pour qu'on puisse déduire non-A2 de non-(A1 et A2 et A3), A2 étant l'hypothèse générale qui fait l'objet du test expérimental.¹³

Récapitulons. On raisonne avec $A = (A1 \text{ et } A2 \text{ et } A3)$, $A \Rightarrow B$. Si on observe non-B on déduit non-A et on sait qu'alors non-A \Rightarrow non-A2 puisqu'on admet A1 et A3. En vérifiant non-B, on a falsifié l'hypothèse de mesurage (A2). Si on avait vérifié B, on aurait *corroboré* l'hypothèse de mesurage.¹⁴

Revenons au cas où on se demande si un score de QI mesure l'intelligence. Comment s'assurer de la validité des hypothèses auxiliaires qui sont nécessaires pour tester que le score de QI dépend de manière croissante de la quantité d'intelligence, sachant qu'on ne sait pas encore mesurer les quantités d'intelligence ? On verra dans le cours de L2 comment, en considérant non pas le score, mais la réponse à tous les items du test, on peut tester l'hypothèse de mesurage sans spécifier comment q et q' , dont on accepte l'existence par hypothèse, sont ordonnées.

8. Interprétation évaluative d'un score de QI

Il n'est pas nécessaire d'admettre qu'un score de QI mesure l'intelligence pour s'en servir de moyen d'évaluer l'intelligence. Le verbe « évaluer » signifie littéralement « conférer une valeur à ». Ainsi, la valeur économique d'un lingot d'or, qu'on mesure (au sens non-classique du terme) en monnaie, est une propriété extrinsèque à l'objet « lingot » : c'est parce que l'objet est plongé dans une culture qu'il reçoit une valeur monétaire. Il en va de même pour la valeur intellectuelle qui est conférée à Paul qui a un QI de 80 points et à Julie qui a un QI de 110 points par exemple. La valeur intellectuelle des gens, telle qu'elle est mesurée (au sens non-classique du terme tant qu'on n'a pas attesté une fonction de mesurage) par leurs scores de QI, a un sens parce que la culture permet qu'on associe les mots « intelligent » aux mots « plus », « moins », « très » par exemple dans des phrases dont l'intelligibilité suppose qu'on accepte que l'intelligence est quelque chose qui varie en plus ou moins (même si un examen conduit à dire que cela n'est pas si simple). La valeur intellectuelle n'est pas la cause du score, c'est son effet culturel.¹⁵ La relation de dépendance causale entre le mesuré (valeur intellectuelle) et le mesurant (score) est inversée par rapport à l'acception classique du terme « mesurer ».

Bien entendu, on peut soutenir que le score résulte d'un certain nombre de phénomènes psychologiques mais il faut reconnaître qu'en l'état actuel des connaissances, on ne sait pas décrire

¹³ Une hypothèse générale est dite *générale* parce qu'elle s'applique sans exception à tous les éléments de son *domaine de validité* ; ici, le domaine de validité de l'hypothèse d'une fonction de mesurage est l'ensemble de toutes les quantités de masse de tout objet pouvant être posé sur la balance sans la dérégler. *Quels que* soient q et q' de Q , $q < q' \Rightarrow F(q) \leq F(q')$, où Q est le domaine de validité de l'hypothèse.

¹⁴ Rappel : corroborer une hypothèse signifie qu'on ne peut pas déduire qu'elle est vraie ni qu'elle est fausse. Déduire A de B et de $A \Rightarrow B$ serait raisonner de manière logiquement *invalid*.

¹⁵ Le principe de l'évaluation psychologique quantitative est celui de la comparaison à une norme sociale – norme précisée par ce qu'on appelle les étalonnages en psychologie.

correctement ces phénomènes ni, a fortiori, définir une fonction dont les scores soient l'ensemble d'arrivée puisqu'on ignore ce que serait son ensemble de départ.

9. Erreur de mesure et propriétés psychométriques

La psychométrie utilise le verbe « mesurer » d'une manière ambiguë. La modélisation statistique des réponses aux tests psychologiques utilise des grandeurs dites latentes (non-observées) qui sont reliées aux données (observables) avec la spécification explicite d'une « erreur de mesure ». Par exemple, on dit qu'un score de QI mesure l'intelligence avec une certaine erreur de mesure, ce qui empêche de déduire que Paul, qui a un QI de 80 points soit moins intelligent que Julie, qui a un QI de 110 points, parce que ces scores sont des mesures entachées d'une erreur qu'on ne sait pas encadrer utilement.¹⁶ Ainsi, la psychométrie laisse aux utilisateurs des scores psychologiques la responsabilité de leur interprétation théorique. D'où la nécessité que le psychologue soit au clair avec les termes qu'il utilise.

Il faut aussi remarquer que les propriétés dites psychométriques ne décrivent pas les scores mais les objets statistiques qui font l'objet de la modélisation statistique (par exemple, des matrices de variance-covariance¹⁷). Ainsi, la notion psychométrique de fidélité (ou fiabilité, *reliability* en anglais) ne décrit pas des scores particuliers mais des variables statistiques, de telle manière que la précision de la mesure de ρ par tel score demeure une inconnue pour la psychométrie.

10. Ressources numériques

Support vidéo. Pour développer votre capacité à imaginer des tests expérimentaux fondés sur le mesurage dans la vie quotidienne, étudiez les vidéos 0, 1 et 2 disponibles ici :

<https://epistemo.hypotheses.org/cours-vidéo>.

ENT. Exercez-vous avec les exercices (appelés « tests ») et travaillez le cours sur le forum de mesurage mis à disposition sur l'ENT.

Cours de L2

11. Réponse au test psychologique : approche multivariée¹⁸

Un test psychologique permet d'observer le comportement d'une personne¹⁹ dans un langage de description standardisé. L'observation résultante est appelée la réponse au test. Cette réponse est une *suite* de réponses dès que le test est constitué d'au moins deux items puisque chaque item génère une réponse à l'item. On notera m le nombre d'items d'un test et on dira que la réponse est m -variée ou encore multivariée ; c'est un m -uplet.

Par exemple, considérons un test d'intelligence composé de $m = 5$ items, dont les réponses sont cotées (ou codées) 0 pour « réponse incorrecte ou manquante » et 1 pour « réponse correcte ». La réponse au test est un 5-uplet. L'expression « (0, 1, 1, 0, 1) » signifie que la réponse observée au test

¹⁶ L'intervalle $[3 ; 5]$ encadre le nombre 4 et aussi le nombre 4,45, mais pas le nombre 2. Les nombres 3 et 5 sont respectivement les bornes inférieure et supérieure de l'intervalle $[3, 5]$.

¹⁷ Voir par exemple <http://www.jybaudot.fr/Stats/covariance.html>.

¹⁸ Multivarié signifie qui varie selon plusieurs points de vue.

¹⁹ Personne qu'on appellera le répondant parce que son comportement est une réponse à une stimulation.

est la suite « réponse incorrecte au premier item, réponse correcte au second item, etc. ».²⁰
 L'expression « (0, 1, 2, 0, 0) » est incorrecte du point de vue descriptif car la valeur « 2 » n'est pas définie dans le langage de description du test. L'expression « (0, 1, 1, 1, 0, 1) » est incorrecte car c'est un 6-uplet alors que le test ne comprend que cinq items.

12. Produit cartésien, cardinal

Le langage descriptif ou, en un mot, le *référentiel* du test, est un ensemble de réponses possibles qui a la structure d'un produit cartésien. Dans l'exemple précédent, on considère l'ensemble {0, 1} des réponses à un item quelconque. Le référentiel du test, noté R, s'obtient en combinant cinq fois l'ensemble {0, 1}. On note

$$R = \{0, 1\} \times \{0, 1\} \times \{0, 1\} \times \{0, 1\} \times \{0, 1\} = \{0, 1\}^5$$

et on lit « \times » « croix » et non pas « multiplié par » car la multiplication n'est pas définie sur les ensembles d'une part et l'opération « \times » n'est pas commutative d'autre part.²¹

Cardinal de R. Le cardinal d'un ensemble est son nombre d'éléments. On note $|E|$ le cardinal de l'ensemble E.

Ainsi, $|\{0, 1\}| = 2$ et $|R| = |\{0, 1\}^5| = 2^5$. On remarquera que le signe « \times » peut être interprété comme une multiplication lorsqu'on l'applique au cardinal des ensembles – et non aux ensembles.

13. Hypothèse de mesurage appliquée au référentiel du test

Supposons que le test d'intelligence utilisé dans l'exemple précédent permette de mesurer l'intelligence. Alors il existe une fonction croissante $F: Q \rightarrow R$, où Q est l'ensemble des quantités d'intelligence possibles et R le référentiel du test.

Quels sont les falsificateurs de l'hypothèse ? Si F existe, les mesures de deux quantités q et q' par F, $F(q)$ et $F(q')$ sont nécessairement ordonnées au sens large, puisque q et q' sont ordonnées au sens large (c'est-à-dire $q \leq q'$ ou $q \geq q'$). D'où l'impossibilité théorique que deux réponses distinctes au test ne soient pas ordonnées au sens large puisque F est croissante.

Considérons les réponses $x = (1, 0, 0, 0, 0)$ et $x' = (1, 1, 0, 0, 0)$. Peut-on les ordonner ? Peut-on ordonner $(1, 1, 1, 1, 0)$ et $(0, 1, 1, 1, 1)$?

14. Ordre produit direct, ordre partiel

On peut définir une relation d'ordre applicable à un produit cartésien d'ensembles en utilisant les relations d'ordre définies sur ces ensembles. Prenons un exemple : on considère les attributs « coût » et « consommation ». Deux voitures ont des coûts comparables (qu'on peut ordonner au sens large) d'une part, des consommations comparables d'autre part. Si la première voiture est décrite par (3000, 9) et la seconde par (2500, 7), l'expression « $(3000, 9) \geq (2500, 7)$ » est correcte car $3000 \geq 2500$ d'une part et $9 \geq 7$ d'autre part.

Ordre produit direct. D'une manière générale, soit le produit cartésien $E = E_1 \times E_2 \times \dots \times E_m$ et $\geq_1, \geq_2, \dots, \geq_m$ les relations d'ordre définies respectivement sur E_1, E_2, \dots, E_m . Soient $x = (x_1, x_2, \dots, x_m)$ et $x' =$

²⁰ Voir <https://epistemo.hypotheses.org/1243>.

²¹ En effet, soient les ensembles A et B, $A \times B \neq B \times A$.

$(x'_1, x'_2, \dots, x'_m)$ deux éléments de E . La relation dite d'ordre produit direct \geq est définie sur E de la manière suivante : $x \geq x'$ si et seulement si $x_1 \geq_1 x'_1, x_2 \geq_2 x'_2, \dots, x_m \geq_m x'_m$.²²

Revenons à l'exemple du test d'intelligence composé de cinq items cotés dans $\{0, 1\}$. La réponse 0 vaut moins que la réponse 1 mais cela ne nous aide pas parce que nous voulons interpréter la réponse non pas comme une évaluation mais comme une mesure.

Fonction de mesurage multivariée. Notons $F_1: Q \rightarrow \{0, 1\}$ la fonction croissante qui par hypothèse mesure q quelconque de Q dans $\{0, 1\}$, l'ensemble des réponses possibles au premier item, et ainsi de suite pour F_2, \dots, F_5 . Supposons, pour un item donné, $F(q) = 0$ et $F(q') = 1$. Alors $q < q'$.

Démonstration.

- i. $F(q) = 0$ et $F(q') = 1$ (hypothèse).
- ii. $q > q' \Rightarrow F(q) \geq F(q')$ (F croissante par hypothèse).
- iii. $F(q) < F(q') \Rightarrow q \leq q'$ (contraposée de ii).
- iv. $i \Rightarrow F(q) < F(q')$ (car $0 < 1$).
- v. D'où $i \Rightarrow q \leq q'$ (car iii).
- vi. Si $q = q'$ alors $F(q) = F(q')$: contradiction avec i. D'où $q \neq q'$.
- vii. D'où $q < q'$ (car v).

Ainsi les valeurs 0 et 1 s'interprètent comme des mesures (ordinales et non pas quantitatives ou métriques) d'une quantité d'intelligence et non pas comme des évaluations (ou encore des *préférences*), sous réserve bien entendu que la théorie soit plausible (i.e., corroborée).

La fonction $F = F_1 F_2 F_3 F_4 F_5: Q \rightarrow \{0, 1\}^5$, qui par hypothèse associe tout q de Q à un 5-uplet dans l'ensemble des 5-uplets possibles doté de la relation d'ordre produit direct, est définie croissante : $q < q' \Rightarrow F(q) \leq F(q')$.

Falsificateur de F , incomparabilité de deux m -uplets. On peut maintenant répondre aux questions posées en 13 :

$$(1, 0, 0, 0, 0) \leq (1, 1, 0, 0, 0),$$

$$\text{ni } (1, 1, 1, 1, 0) \leq (0, 1, 1, 1, 1)$$

$$\text{ni } (1, 1, 1, 1, 0) \geq (0, 1, 1, 1, 1).$$

On notera l'incomparabilité de deux m -uplets par le signe « $\langle \rangle$ » : $(1, 1, 1, 1, 0) \langle \rangle (0, 1, 1, 1, 1)$.

15. Reformulation du problème de mesurage en psychologie

Si on veut interpréter les réponses aux tests psychologiques comme des mesures au sens classique du terme, il faut préalablement élaborer l'hypothèse que la variation des réponses (le mesurant) dépend de manière croissante de la variation de quantités dont on admet l'existence (le mesuré) – sans quoi il n'y a pas d'hypothèse à tester. Pour cela, on peut définir, sans qu'il soit nécessaire de la déterminer ou encore de l'identifier, une fonction de mesurage multivariée dont la croissance implique qu'il est théoriquement impossible d'observer deux mesures multivariées (m -uplets) incomparables. On dispose ainsi d'une hypothèse testable : l'observation de deux réponses incomparables falsifie l'existence d'une telle fonction de mesurage.

²² Voir aussi <https://epistemo.hypotheses.org/2650>.

Support ontologique de l'hypothèse de mesurage. Encore faut-il que l'hypothèse de mesurage (multivarié) porte sur une réalité : supposer que les réponses de Paul à tel test sont des mesures n'implique pas que les réponses de Julie au même test soient des mesures, tant qu'on n'a pas de raison de supposer que les fonctions de mesurage respectives sont identiques.²³ De même, admettre que les réponses de Paul ne sont pas des mesures n'implique pas que les réponses de Julie ne soient pas des mesures.

L'observation de réponses incomparables provenant de plusieurs répondants n'implique pas que ces réponses ne sont pas des mesures : cela implique qu'il n'existe pas de fonction de mesurage commune aux répondants. Si les réponses aux tests sont des mesures, c'est que les répondants fonctionnent comme des instruments de mesure lorsqu'ils produisent leurs réponses et, du point de vue méthodologique, une telle hypothèse doit être testée à l'échelle de chaque répondant.

16. Ordinalité des mesures

Une échelle de mesure dite ordinale est, par définition, un ensemble d'éléments ordonnés tels que deux éléments quelconques de cet ensemble (de cette « échelle ») sont non seulement comparables mais encore « conservent » la comparabilité des éléments qu'ils *représentent*.²⁴

Ordinalité d'une réponse à l'item sous l'hypothèse de mesurage. La réponse à l'item étant définie comme l'image par une fonction de mesurage d'une quantité de grandeur, le rapport de deux réponses n'indique pas quelque chose comme une « distance » entre les deux quantités mesurées, il indique seulement que les quantités sont distinctes et ordonnées lorsque les réponses sont différentes – si les deux réponses sont égales, on ne sait pas déduire si les quantités qu'elles mesurent sont égales ou distinctes, d'où si on veut être précis, l'ordinalité au sens large de la réponse. Les nombres qui codent les réponses aux items sont par hypothèse ordinaux, ce sont des *rangs* : la liste de ces nombres est « premier, second, troisième, etc. ».

Non-ordinalité des scores au test. Par conséquent, l'utilisation arithmétique de ces nombres²⁵ pour calculer un score n'a pas de sens par rapport aux quantités qu'ils mesurent. C'est une des raisons qu'on peut invoquer pour justifier l'interprétation évaluative des scores au test, puisque l'interprétation classique empêche de les *calculer*.²⁶ Par hypothèse en effet dans le cadre du mesurage qualitatif multivarié développé ci-dessus, les réponses d'un répondant à tous les items, bien qu'elles puissent être différentes en raison des différences qui existent entre les items, mesurent la même quantité *q*.

Un autre argument débouchant sur la non-ordinalité des scores au test consiste à remarquer que la phrase « première réponse à l'item 1 plus troisième réponse à l'item 2 » par exemple n'a pas de sens dans le langage descriptif (référentiel) du test.

²³ La psychologie différentielle est un secteur de la psychologie qui, contrairement à la psychologie générale, permet l'hypothèse que les lois psychologiques, si elles existent, ne soient pas nécessairement communes à toutes les entités auxquelles s'intéresse la psychologie.

²⁴ Soient *a* et *b* mesurés par *x* et *y* respectivement ; la conservation de l'ordre signifie que $a < b \Leftrightarrow x < y$, ce qui signifie qu'il existe une fonction croissante au sens strict. Une telle restriction n'est pas nécessaire quand on adopte l'approche classique du mesurage. Dans l'approche dite représentationnelle de la mesure, une échelle ordinale (e.g., Stevens, 1946, voir une critique de son article ici : <https://epistemo.hypotheses.org/3696>) est une suite de nombres dont les rapports d'ordre conservent les rapports d'ordre des éléments qu'ils « représentent ».

²⁵ Leur emploi avec l'addition et les opérations dérivées, soustraction, multiplication, division.

²⁶ Ce qui n'interdit pas que, dans le cas d'un mesurage ordinal multivarié, les *m*-uplets puissent être remplacés par des ordinaux puisqu'ils forment une échelle.

17. Ressources numériques

Support vidéo. Etudier la vidéo 3 disponible ici : <https://epistemo.hypotheses.org/cours-video>.

ENT. Exercez-vous avec les exercices (appelés « tests ») et travaillez le cours sur le forum de mesurage mis à disposition sur l'ENT.

Vocabulaire

Ci-dessous une liste de termes utilisés de manière technique pour les besoins du cours.

cardinal : taille (d'un ensemble)

comparable (incomparable) : propriété de deux m-uplets, voir <https://epistemo.hypotheses.org/2650>

conclusion : la conclusion d'un argument de la forme « prémisses implique conclusion »

condition nécessaire (vs. suffisante) : dans $A \Rightarrow B$, A est la condition suffisante de B, B la condition nécessaire de A.

conditions (expérimentales) : circonstances dans lesquelles se trouve un objet sur lequel on fait une expérience

contraposée : $\text{non-B} \Rightarrow \text{non-A}$ est la contraposée de $A \Rightarrow B$

corroborer, corroboration : ne pas falsifier (une hypothèse, proposition), le fait de ne pas falsifier

déduire, déduction : raisonnement logiquement valide (calcul de la valeur vérité)

décrire, description : caractériser un objet dans un langage descriptif donné

ensemble (de définition, de départ, d'arrivée, de valeurs) : collection d'objets possédant au moins une propriété commune

ensemble de définition ou de départ : étant donnée une relation qui relie un x à un y, ensemble des x

ensemble d'arrivée ou de valeurs : étant donnée une relation qui relie un x à un y, ensemble des y

erreur (de mesure, de description) : phrase fautive

évaluer, évaluation : donner une valeur à

expérience, expérimentation : observation pour tester une hypothèse

falsifier, falsificateur : démontrer faux, observation qui falsifie une hypothèse (proposition) falsifiable

fonction : relation d'un ensemble de départ D dans un ensemble d'arrivée A telle que tout élément de D est relié à un et un seul élément de A

fonction croissante : fonction F telle que $x < x' \Rightarrow F(x) \leq F(x')$

hypothèse (auxiliaire, générale) : proposition (ou conjonction de propositions) supposée vraie ou fautive

implication (logique) : voir dans le cours la table de vérité qui définit l'implication

item (de test) : tâche élémentaire et format de réponse et format de cotation de la réponse

langage (de description) : ensemble de termes caractérisant ou de modalités descriptives

mesurer, mesure, mesurage : utiliser une fonction croissante reliant le mesuré au mesurant

m-uplet : série de valeurs, voir <https://epistemo.hypotheses.org/1608>

ontologie : qui traite des possibilités de l'existence

ordre, ordonné, ordinal : ensemble doté d'une relation d'ordre, qui peut être ordonné

ordre produit direct : relation d'ordre définie sur un produit cartésien d'ensembles ordonnés

plausible : qui peut être vrai

préférence : cas particulier d'ordre

prémisse : proposition dont on déduit une conclusion

produit cartésien : combinaison des éléments de plusieurs ensembles

raisonnement (hypothético-déductif) : raisonnement fondé sur l'implication

rang : numéro dans une échelle

réponse, répondant : comportement émis en réponse à une stimulation

référentiel (d'un test) : ensemble des réponses possibles

score : nombre calculé à partir d'une règle d'agrégation de plusieurs réponses qualitatives

stimulation : événement supposé déclencher une réponse

valide : vrai