Document generated on 07/19/2025 4:28 a.m.

Mesure et évaluation en éducation

Une théorie des seuils psychométriques à double contrôle d'erreur – Partie III : les applications illustrées

Louis Laurencelle

Volume 40, Number 3, 2017

URI: https://id.erudit.org/iderudit/1048912ar
DOI: https://doi.org/10.7202/1048912ar

See table of contents

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)
2368-2000 (digital)

Explore this journal

Cite this article

Laurencelle, L. (2017). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie III : les applications illustrées. *Mesure et évaluation en éducation*, *40*(3), 105–133. https://doi.org/10.7202/1048912ar

Article abstract

Both test interpretation and test-based ruling depend on comparing the respondent's score to a standard (or a scale of standards), a comparison by which the respondent is to be ranked somehow among the reference population. This operation goes beyond the field of psychological testing, and is of relevance to school testing, medical assessing, quality control, etc. However, in many cases, the individual's score is altered by a "measurement error", and also the standard used has its own uncertainty, ascribable to its sample data basis. In parts I and II (Laurencelle 2015, 2016a) of this series of articles, the problem was explicitly identified and solved, and practical numerical procedures were proposed. This third part, focused on the user, summarizes the theory and presents a series of worked out examples that may serve as solution models (both exact and approximate) to the user, while illustrating various instances of the theory.

Tous droits réservés © ADMEE-Canada - Université Laval, 2017

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Une théorie des seuils psychométriques à double contrôle d'erreur
– Partie III : les applications illustrées

Louis Laurencelle
Université du Québec à Trois-Rivières

Mots clés : décision psychométrique, erreur de mesure, incertitude de la norme, norme sûre, illustrations

L'interprétation et la décision psychométriques reposent toutes deux sur la confrontation entre le score obtenu par le répondant et la norme (ou l'échelle normative) par laquelle on veut le situer dans la population de référence. Cette confrontation déborde le seul domaine de la psychométrie pour intéresser la docimologie, le testing médical, le contrôle de qualité. Or, dans bien des cas, la mesure individuelle est marquée d'une « erreur de mesure », de même que la norme à appliquer est incertaine parce que basée sur un simple échantillon de la population ciblée. Les parties I et II (Laurencelle 2015, 2016a) de cette série d'articles ont permis d'identifier explicitement ce problème et d'en proposer des procédures de solution exactes et approximatives. Cette troisième partie, axée davantage sur la pratique, récapitule la théorie et présente surtout une série d'exemples travaillés qui pourront servir de modèles de solution aux intéressés, tout en illustrant diverses applications de la théorie.

Key words: test-based ruling, measurement error, norm's sampling error, safe norm, worked problems

Both test interpretation and test-based ruling depend on comparing the respondent's score to a standard (or a scale of standards), a comparison by which the respondent is to be ranked somehow among the reference population. This operation goes beyond the field of psychological testing, and is of relevance to school testing, medical assessing, quality control, etc. However, in many cases, the individual's score is altered by a "measurement error", and also the standard used has its own uncertainty, ascribable to its sample data basis. In parts I and II (Laurencelle 2015, 2016a) of this series of articles, the problem was explicitly identified and solved, and practical numerical procedures were proposed. This third part, focused on the user, summarizes the theory and presents a series of worked out examples that may serve as solution models (both exact and approximate) to the user, while illustrating various instances of the theory.

Palavras-chave: decisão psicométrica, erro de medição, incerteza da norma, norma segura, ilustrações

A interpretação e a decisão psicométricas apoiam-se no confronto entre o resultado obtido pelo respondente e a norma (ou a escala normativa), pela qual se pretende situá-lo na população de referência. Este confronto ultrapassa o campo da psicometria, sendo relevante para a docimologia, os testes médicos, o controlo de qualidade, etc. Ora, em muitos casos, não só a medida individual é alterada por um "erro de medição", como a norma usada apresenta a sua própria incerteza porque é baseada numa amostra simples da população alvo. As Partes I e II (Laurencelle 2015, 2016a) desta série de artigos permitiram identificar explicitamente este problema e propor procedimentos de solução exatos e aproximados. Esta terceira parte, que se concentra mais na prática, recapitula a teoria e apresenta, acima de tudo, uma série de exemplos trabalhados que poderão servir como modelos de solução para os interessados, ao mesmo tempo que ilustram várias aplicações da teoria.

Note de l'auteur : La correspondance liée à cet article peut être adressée à [louis.laurencelle@gmail.com].

Introduction

C'est un truisme que de dire l'importance de la mesure et de l'évaluation dans la conduite des affaires humaines, que ce soit pour le commerce, la médecine, les travaux scientifiques, l'évaluation scolaire, le testing psychométrique. À chaque circonstance donnée, mesurer et évaluer nous permet de constater, de décider et d'agir « en connaissance de cause », l'action prise étant alors responsable, imputable et vérifiable, comme il se doit.

Dans le processus de mesure-évaluation, la mesure fournit une grandeur, une valeur numérique[1], censée décrire une caractéristique de la personne évaluée : c'est le matériau de base. Pour ce matériau, trois questions doivent être satisfaites : 1) Sur quoi ou sur quelle caractéristique la mesure informe-t-elle ? 2) À quoi doit-elle être comparée ? Sur quel critère baser son jugement ? 3) Quel est le degré de sûreté de ce jugement et sur quelles données probantes cette sûreté repose-t-elle ? La réponse à la première question concerne la validité de la mesure (Cronbach, 1971 ; André, Loye & Laurencelle, 2015). Pour répondre à la deuxième question, le fabricant de l'instrument de mesure fournit habituellement un critère, une procédure d'évaluation comparative et d'interprétation de la donnée obtenue. Quant au degré de sûreté du jugement à faire, lequel dépend à la fois de la qualité de mesure et de la légitimité du critère, la littérature à ce jour ne l'a abordé que de façon qualitative, par des mises en garde ou par des règles de prudence, notamment en ce qui touche l'incertitude échantillonnale du critère (numérique) de jugement. C'est à ce flou artistique que notre théorie du double contrôle d'erreur veut répondre.

Pourquoi un double contrôle d'erreur, ou une double erreur ? C'est que, dans l'interprétation du résultat d'un test ou d'une mesure, l'évaluateur a à sa disposition deux ingrédients : une donnée et un critère, ou une grille de critères ou normes, à quoi confronter la donnée. Or, sauf dans des cas improbables, chaque donnée, chaque mesure contient un degré plus ou moins grand d'imprécision. Cette imprécision est souvent moins grande dans les mesures physiques (la force du bras, la glycémie, le tour de taille), tandis qu'elle est souvent plus grande dans la plupart des mesures psychométriques (échelle d'anxiété situationnelle, d'introversion, de névrose obsessionnelle), d'autres mesures occupant une position intermédiaire

(quotient intellectuel, échelles d'aptitude, mesures de rendement scolaire). Cette imprécision, dont la source est diverse, a pour effet, chez le répondant, qu'il obtienne une mesure flottant autour de sa vraie valeur, de manière aléatoire. En outre, le critère auquel cette valeur sera confrontée a lui-même souvent été produit sur une base empirique, en utilisant quelques dizaines ou centaines de mesures de référence, le critère, appelé aussi « norme », présentant lui aussi sa dose d'incertitude.

La connaissance et le traitement de l'incertitude, ou imprécision, qui découle de cette double source d'erreur donnent à l'évaluateur le moyen de nuancer son jugement, en quantifiant en quelque sorte la marge d'approximation sous laquelle la mesure du répondant se trouve placée. L'évaluateur peut aussi, grâce à une norme qui incorpore cette marge, arriver à un classement ou à un jugement plus sûr, plus sûr parce que débordant la marge d'incertitude que l'erreur ou les erreurs présentes ont pu y ajouter.

Dans les parties I (Laurencelle, 2015) et II (Laurencelle, 2016a) de cette série d'articles, nous avons posé les bases mathématiques de la théorie du double contrôle d'erreur pour la décision psychométrique. Notre ouvrage, *L'étalonnage et la décision psychométrique : exemples et tables* (Laurencelle, 2016b), traite aussi *in extenso* cette question aux sections F, G et H (p. 117-196). Dans la présente partie III, nous avons tenté de donner au lecteur un mode d'emploi, un guide pratique pour la compréhension et l'application des règles et calculs développés dans nos ouvrages préalables. Nonobstant la nature fortement numérique et mathématique des sujets traités, nous nous sommes efforcé d'en rendre l'exposé le plus sobre et simple possible : le lecteur pourra juger lui-même de notre résultat. Nous aborderons en premier le cas le plus simple, celui d'une situation dans laquelle la norme appliquée, le seuil, est édictée ou stable et la mesure, marquée d'erreur (correspondant à $n \to \infty$, $\sigma_e > 0$). Nous examinerons ensuite le cas d'un seuil estimé par échantillonnage et utilisant une mesure pure ($n < \infty$, $\sigma_e = 0$). Nous nous pencherons enfin sur le cas, typique, d'un seuil échantillonné associé à une mesure fluctuante ($n < \infty$, $\sigma_e > 0$). Le tout sera récapitulé à la faveur d'exemples supplémentaires.

Cas $n \to \infty$, $\sigma_e > 0$

Soit un examen de certification, au seuil de réussite C prescrit tel que, en principe, l'individu obtenant un score $x \geq C$ est considéré comme ayant réussi. Cependant, chaque mesure individuelle x comporte une valeur d'incertitude e telle que $x = V + e$, V étant sa valeur vraie[2]. Le seuil de réussite ou de classement peut être établi en mesurant toute la population (équivalant à un échantillon de taille $n \to \infty$) ou, plus généralement, être décrété par une instance décisionnelle, par exemple un organisme normatif : le ministère de l'Éducation, une régie de normes du travail ou un comité d'experts en santé publique. Pour ce cas de figure, afin de mettre en place le vocabulaire de la présentation et parce que ce cas est le plus simple, nous exposerons le raisonnement mathématique de façon explicite. Les cas suivants seront traités de manière plus laconique.

En admettant un modèle normal simple (pour e) : la norme sûre linéaire

La mesure considérée ici est sujette à erreur, c.-à-d. que, pour chaque évaluation de la même personne dans les mêmes conditions, le score obtenu fluctuera : c'est ce que reflète le coefficient de fidélité R (ou r_{XX}) lorsque $R < 1$. Le composant e responsable de cette fluctuation s'ajoute à l'opération de mesure et varie généralement selon une distribution normale[3] autour de 0. Son écart-type, l'erreur-type de mesure[4] σ_e, est estimé par un échantillon de taille suffisante, cet estimateur étant dénoté s_e.

Probabilité de réussite

Considérons un individu donné à valeur (ou mesure) vraie V_o. Son score à composant fluctuant e sera $x_o = V_o + e$, et la probabilité qu'il « réussisse » (atteigne ou excède le critère C) est simplement :

$$
\begin{aligned}
(1)\quad \Pr\{ V_o + e \geq C \} &= \Pr\{ e \geq C - V_o \} \\
&= \Pr\{ e / s_e \geq (C - V_o)/s_e \} \\
&= \Pr\{ z \geq (C - V_o)/s_e \} = 1 - \Pr\{ z < (C - V_o)/s_e \} \\
&= 1 - \Phi\{ (C - V_o)/s_e \},
\end{aligned}
$$

la notation $\Phi(x)$ désignant l'intégrale normale standard à x. Rappelons que, pour effectuer ce calcul, il faut connaître V_o, la mesure sans erreur, une condition généralement utopique.

Norme sûre linéaire

Pour un répondant qui serait tout juste non qualifié, soit $V < C$ et $V \to C$, la probabilité que son score $V + e$ déborde C tendra vers ½ puisque e fluctue autour de 0, une éventualité ambiguë. De là émane, dans certaines circonstances (sélection, qualifications compétitives, préjudice d'une qualification erronée), le besoin d'un seuil protégé, une «norme sûre» affectée d'un coefficient de sûreté $1 - \alpha$, où par exemple $\alpha = 0{,}05$ ou $0{,}01$. Le cas contraire est intéressant aussi, soit celui d'un répondant vraiment qualifié (selon $V > C$), mais dont le score $x = V + e$ risque d'être décalé par en bas, en deçà du critère C. En fait, tout comme dans le premier cas, s'il est tout juste qualifié selon $V \to C$, sa probabilité d'être retenu, ou rejeté, tend vers ½!

Pour accommoder cette situation, il est possible de décaler le critère C en lui ajoutant ou retranchant une quantité Λ, l'énoncé de probabilité devenant:

$$\begin{aligned}(2)\quad \Pr\{ V_o + e \geq C \pm \Lambda \} &= \Pr\{ e \geq C \pm \Lambda - V_o \} \\ &= \Pr\{ e / s_e \geq (C \pm \Lambda - V_o)/s_e \} \\ &= 1 - \Phi\{ (C \pm \Lambda - V_o)/s_e \}.\end{aligned}$$

Prenons une situation dans laquelle on veut produire une norme allongée, $C' = C + \Lambda$, telle qu'elle protège contre la rétention d'un candidat *non qualifié* (selon $V < C$) selon un risque ne débordant pas α, autrement dit, avec un coefficient de sûreté d'au moins $1 - \alpha$. Cette norme allongée sera désignée norme exigeante. On cherche alors la valeur de Λ telle que la probabilité que $V + e$ déborde $C + \Lambda$ n'excède pas le seuil α, *sous condition que* $V < C$, soit:

$$(3)\quad \Pr\{ V + e \geq C + \Lambda \} = \Pr\{ e \geq C + \Lambda - V \mid V < C \} \leq \alpha.$$

Le risque maximal de rétention erronée advient lorsque $V \to C$ ou $V \approx C$; comme il est maximal, c'est ce risque que nous égalons à α, d'où:

$$\begin{aligned}(4)\quad \Pr\{ C + e \geq C + \Lambda \} &= \Pr\{ e \geq \Lambda \} = \alpha \\ &= \Pr\{ e / s_e \geq \Lambda/s_e \} = \alpha \\ &= 1 - \Phi(\lambda_{exig}) = \alpha,\end{aligned}$$

en utilisant la forme standard $\lambda_{exig} = \Lambda / \sigma_e$, la solution étant:

$$(5a)\quad \Phi(\lambda_{exig}) = 1 - \alpha$$

ou

$$(5b)\quad \lambda_{exig} = \Phi^{-1}(1 - \alpha) = z_{1-\alpha}$$

et enfin

(6) $C'_{exig} = C + \lambda_{exig} \cdot s_e$.

La norme C′ exigeante à appliquer garantit que, pour un candidat vraiment non qualifié, selon V < C, le risque qu'il soit retenu n'excédera pas α. Quant à la norme dite « permissive », étant donné la symétrie de l'erreur normale *e*, sa forme standard est simplement :

(7) $\lambda_{perm} = -\lambda_{exig} = z_\alpha$,

associée à $C'_{perm} = C + \lambda_{perm} \cdot s_e$, et elle garantit qu'un candidat vraiment qualifié n'ait pas plus que 100α % de chances d'être refusé par le test.

En dehors du modèle normal : la norme linéaire non normale et la norme binomiale

Le cas d'une norme absolue décrétée discuté ici admet deux extensions raisonnables et simples, en dehors du modèle de la loi normale : la première concerne une erreur linéaire (continue) de distribution non normale, l'autre un modèle (discontinu ou discret) binomial.

Modèle à distribution linéaire non normale de e

Selon la structure du test et la nature de la capacité évaluée, la fluctuation du score *x* autour d'une valeur vraie hypothétique V peut prendre diverses formes. Par exemple, si *x* est lognormale (loi à variable X continue et positive et à asymétrie positive), l'erreur attachée est vraisemblablement elle aussi lognormale ; ou bien, si X est (effectivement) bornée entre une valeur minimale et une maximale, elle obéit peut-être à une loi Bêta, tout comme l'erreur échantillonnale qui lui est attachée ; et cætera. Le même raisonnement et les mêmes mathématiques que ceux appliqués pour le modèle linéaire normal peuvent être transférés ici, cela en les accommodant mathématiquement au modèle de distribution approprié. Une autre voie, souvent plus simple, consiste à transformer numériquement la variable X en une autre X′ qui présente des indices de normalité (p. ex., γ_1 et γ_2)[5] acceptables, puis à fixer la norme sur X par le biais de la variable normalisée X′. Il reste que, pour une variable continue (ou quasi continue[6]), le modèle normal emporte la préférence et, sauf dans des circonstances requérant une grande rigueur, on peut l'appliquer sans souci.

Modèle binomial

Certaines mesures, par exemple plusieurs formats de tests ou les épreuves psychométriques ou docimologiques, ont une structure de type binomial : le score est basé sur plusieurs composants ou items à scoring dichotomique et il peut se définir par :

$$x = y_1 + y_2 + \ldots + y_k,$$

où y_i vaut 0 ou 1 selon la valeur de la réponse à l'item i, et où x, la somme brute des y_i, varie de 0 à n. La probabilité de bonne réponse ($y = 1$), dénotée π, caractérise le niveau d'habileté du répondant[7] ; pour un ensemble binomial homogène, π est unique et constante pour tous les items. Il est aussi possible que π varie d'un item à l'autre : on parle alors d'une mixture de binomiales, ou « binomiale de Poisson ». Comparativement au modèle linéaire $x = \mathrm{V} + e$, dans lequel le composant e représente l'incertitude, celle-ci émane directement de la probabilité π caractérisant le modèle binomial. L'appendice présenté en fin d'article donne des indications supplémentaires pour le calcul de ces modèles.

Exemple 1 : Norme sûre linéaire (critère stable, mesure avec erreur)

Les examens scolaires au Québec sont notés sur 100, le seuil de réussite imposé étant de 60, soit C = 60 : il s'agit ici d'une norme décrétée. Disons que le ministère de l'Éducation a préparé une épreuve d'habileté en mathématiques pour les élèves de 4e secondaire, laquelle comporte 40 questions à réponse courte. Une préexpérimentation a permis de déterminer que la distribution du score à l'examen suit approximativement une loi normale (voir Appendice), l'écart-type estimé étant $s_X = 10$ et le coefficient α de Cronbach[8], qui servira d'estimation de fidélité R, étant égal à 0,84.

Écart-type et fidélité nous permettent d'estimer l'erreur-type de mesure σ_e par[9] :

(8) $\sigma_e \approx s_e = s_x \sqrt{1-r_{xx}}$;

ici, $s_e \approx 10 \times \sqrt{1-0{,}84} = 4{,}00$.

Robert obtient $x = \mathrm{V}_o + e = 57$. Si l'on suppose que le niveau réel de Robert approche C, c.-à-d. $\mathrm{V}_o \approx \mathrm{C}$, une mesure sans erreur lui aurait donné $x \approx 60$, avec une chance sur deux de réussir. Dans ce cas, le composant e de son score, égal à -3 (= $x - \mathrm{V}_0 = 57 - 60$), l'aurait desservi. Supposons

plutôt que l'erreur de mesure dans son score est à peu près nulle ($e \approx 0$), alors $V_o \approx 57$, et la formule (1) donnera $1 - \Phi\{ (57 - 60)/4{,}00 \} = 1 - \Phi\{ 0{,}75 \} \approx 0{,}227$. C'est la probabilité que, tenant compte de l'erreur de mesure supposée (et inconnue), Robert ait le niveau de capacité correspondant au seuil de réussite demandé.

Comme le montre l'exemple ci-dessus, l'appréciation des probabilités de succès ou d'échec à une qualification psychométrique dépend directement de la valeur vraie V du répondant, valeur qu'il est généralement impossible de connaître pour un individu donné[10]. L'agence de sélection ou de qualification peut néanmoins gérer le risque d'une décision incorrecte par le biais de la norme sûre, ou protégée : il s'agit de limiter à la probabilité α le risque que toute personne soit classée dans la catégorie incorrecte, c.-à-d. retenue alors qu'elle ne le mérite pas (selon $V < C$) ou bien rejetée alors qu'elle le mérite (selon $V > C$). Dans le cas présent, et adoptant un seuil d'erreur α de 5 % (équivalant à un taux de protection de 0,95), la table d'intégrale normale appliquée à (5b) nous fournit $\lambda_{exig} = z_{1\alpha} = z_{0,95} = 1{,}645$. Au seuil de 1 %, on aura $z_{0,99} = 2{,}326$.

La norme sûre permissive, $C'_{perm} = C - z_{1-\alpha} \times s_e$, serait ici $C'_{perm} = 60 - 1{,}645 \times 4{,}00 \approx 53{,}4$ ($\alpha = 5\%$), ou 50,7 ($\alpha = 1\%$) : avec son $x = 57$, Robert passerait. Quant à la norme exigeante, $C'_{exig} = C + z_{1-\alpha} \times s_e$, on trouve = 66,6 (5 %) et 69,3 (1 %), et Robert ne passerait pas.

Exemple 2 : Norme sûre binomiale (critère stable, mesure avec erreur)

Un exemple très simple illustrera l'utilité, réelle mais réduite, du modèle binomial pour le contexte étudié. Soit un test de connaissances présenté sous forme de $n = 20$ items à choix binaire, chaque item offrant deux réponses valant 1 ou 0 point (c.-à-d. *bonne* vs *mauvaise*). Le test envisagé a pour but de filtrer les candidats qui ont acquis quelque chose du champ de connaissances sondé, les parfaits « ignorants » devant répondre au hasard, avec probabilité $\pi = ½$ de choisir la bonne réponse. Pour ces candidats ignorants, le score moyen attendu est $\mu = n \cdot \pi = 20 \times 0{,}5 = 10$, la variance, $\sigma^2 = n \cdot \pi(1 - \pi) = 20 \times 0{,}5 \times 0{,}5 = 5$ et l'écart-type, $\sigma = \sqrt{5} \approx 2{,}236$. Ainsi, un candidat informé espère obtenir un résultat plus haut que 10 ! Enfin, la distribution théorique de x est quasi normale (avec $\gamma_1 = 0$ et $\gamma_2 = -0{,}1$; voir Appendice).

Supposons qu'on veuille s'assurer que le candidat répond ou ne répond pas vraiment mieux qu'au hasard, et ce, avec une assurance de 0,95, c.-à-d. un risque maximal de se tromper de $\alpha = 0{,}05$. Alors, en utilisant la formule A1 (voir Appendice), on cherche le score minimal x^* tel que $\Pr\{ x \geq x^* \mid \pi = ½, n = 20 \} \leq \alpha$, d'où $\Pr\{ z_x \leq (x^* - \mu - ½)/\sigma\} \geq 1 - \alpha$ et, finalement, $x^* \geq \mu + ½ + \sigma \cdot z_{1-\alpha}$. Avec $\alpha = 0{,}05$ et $z_{1-0{,}05} \approx 1{,}645$, on obtient $x^* \geq 10 + ½ + 2{,}236 \times 1{,}645 \geq 14{,}18 = 15$, la réponse. Pour $x^* = 15$, le calcul binomial exact donne $\Pr\{ x \geq 15 \} = 0{,}0207$, alors que cette probabilité serait de 0,0577 pour $x^* = 14$. Soit dit en aparté, on peut formuler l'hypothèse que la règle de passage du 60% appliquée dans les organisations scolaires ait émané, intuitivement ou expressément, d'un seuil implicite et naïf de 50% («l'élève connaît-il la réponse, oui ou non?»), le décret du 60% visant à déborder prudemment par le haut la possible contribution du hasard.

Un calcul pareil à celui illustré plus haut peut évidemment être fait pour des questionnaires à choix comportant $k > 2$ options (et $\pi = 1/k$), tout comme pour d'autres dans lesquels la probabilité de bonne réponse (π_i) varie d'un item i à l'autre. Dans chaque cas, évidemment, la structure du questionnaire et la préparation des items doivent respecter les postulats du modèle binomial envisagé.

Cas $n < \infty$, $\sigma_e = 0$

Ici, le critère de sélection ou de qualification est empirique, est basé sur la mesure d'un échantillon de n personnes (l'échantillon normatif) et réfère explicitement à un taux de sélection[11] f, complément du seuil quantile $P = 1 - f$, auquel on veut faire correspondre un score liminal ou norme, que nous désignons C_P. La règle de sélection est encore $x \geq C_P$. Le seuil quantile, par exemple $P = 0{,}80$, délimite la fraction supérieure $f (= 1 - P)$ de population qu'on prévoit retenir, par exemple 0,20 ou 20%. Dans le contexte présent, l'erreur de mesure e est réputée nulle, avec $\sigma_e = 0$, de sorte que le score x du répondant reflète exactement sa compétence (ou $x = V$), tandis que, pour la norme C_P, elle doit être estimée à partir de données empiriques et, par conséquent, elle est imprécise.

La référence populationnelle à laquelle la norme doit correspondre n'est pas connue. Si elle l'était, son quantile P fournirait la valeur précise X_P qui servirait de norme, et l'obtention d'un score $x > X_P$ garantirait la

compétence du répondant, l'erreur de mesure étant nulle. Or, pour fixer la norme, l'utilisateur dispose seulement d'un échantillon ou groupe normatif[12], à partir duquel il doit conjecturer, c.-à-d. estimer C_P, la valeur produite héritant de l'incertitude de l'échantillon dont elle provient.

L'estimation du seuil se fait à partir de la série normative de n données et procède généralement de l'une de deux façons, soit en utilisant un composé linéaire à référence normale $C_P = \overline{X} + s_X \times z_{1-f}$, soit par le quantile ordinal $C_P = X_{[P\cdot(n+1)]}$,[13] lequel flotte autour du vrai quantile inconnu X_P selon son erreur échantillonnale.

Le score X lui-même étant pur, la probabilité que le score x de Robert déborde C_P dépend uniquement et directement de la distribution de C_P autour de X_P, selon l'une des deux formes d'estimation utilisées.

Exemple 3 : Norme sûre linéaire (critère échantillonnal, mesure sans erreur)

En stipulant le modèle normal linéaire

Supposons que notre base normative repose sur un échantillon de n = 150 éléments[14], leur mesure ayant pour moyenne $\overline{X} = 28{,}36$ et pour écart-type $s_X = 5{,}12$. La fraction de population retenue (f) est 0,05 (ou 5 %), donnant un seuil naïf C_P de $\overline{X} + z_{1-f} \cdot s_X \approx 28{,}36 + 1{,}645 \times 5{,}12 \approx 36{,}78$, $z_{1-f} = z_{0,95} \approx 1{,}645$ étant le quantile normal au rang quantile $P = 1 - f = 0{,}95$. Abstraction faite du léger biais négatif de l'écart-type s,[15] l'application de cette norme ferait en sorte qu'un candidat produisant exactement un score $x = 36{,}78$ aurait une chance sur deux d'être retenu ou rejeté, cela en raison de l'imprécision de C_P, laquelle découle de l'imprécision de $\overline{X}$ et de s_X. Dans le but de réduire à un niveau de probabilité α l'erreur attachée à une décision désavantageuse, on peut, en se basant sur cette incertitude, imposer une norme exigeante (réduire le risque d'un faux positif, consistant à retenir un répondant qui serait non qualifié) ou une norme permissive (réduire le risque d'un faux négatif, consistant à rejeter un répondant qualifié), cela en décalant la norme.

Norme sûre exacte

Selon les conditions décrites ci-dessus, la mesure, ou score x, ne comporte aucune part d'erreur. Dans le langage de la théorie de la double erreur, cela se traduit par le fait que la fidélité de mesure est parfaite, soit R = 1. Consultant les tables appropriées de Laurencelle (2016b), soit les tables F4b et F9b, pour $f = 0{,}05$ au taux d'incertitude $\alpha = 0{,}05$, R = 1,

n = 150, nous trouvons respectivement $\lambda = 1{,}870$ pour le seuil exigeant et $\lambda = 1{,}454$ pour le seuil permissif. En appliquant alors la formule générale :

(9) $C_P = \overline{X} + s_X \times \lambda$,

nous établissons les seuils $C_{Pexig} = 28{,}36 + 5{,}12 \times 1{,}870 \approx 37{,}93$ et $C_{Pperm} \approx 28{,}36 + 5{,}12 \times 1{,}454 \approx 35{,}80$ pour les critères exigeant et permissif respectivement. On peut aussi approximer les seuils λ par une formule[16]. En appliquant la règle : « Retenir Robert si son score x atteint ou déborde 37,93 », on est sûr à 95 % (= $1 - \alpha$) ou mieux que Robert fait bien partie de la portion des 5 % (= f) supérieurs de la population.

Norme approchée

Il est possible d'approximer ces valeurs, d'abord en estimant l'erreur-type de l'estimateur utilisé, puis en appliquant au seuil naïf 36,78 une marge d'erreur relative au taux d'erreur α consenti. La variance du seuil (9) (voir Laurencelle, 2016b, formule A-15) étant :

(10) $\mathrm{var}(C_P) = \mathrm{var}(\overline{X} + z_{1-f} \cdot s_X) \approx \sigma_X^2 / n \times [1 + z_P^2/2]$,

c.-à-d. $5{,}12^2 / 150 \times (1 + 1{,}645^2/2) \approx 0{,}4112$, nous obtenons l'erreur-type $\sigma_e(C_P) = \sqrt{0{,}4112} \approx 0{,}641$ associée à l'estimateur naïf C_P. Si l'on applique la marge de protection $z_{1-\alpha} = z_{0,95} = 1{,}645$, la norme permissive serait alors $C_{Pperm}(\alpha) = C_P - z_{1-\alpha}\, \sigma_e(C_P) \approx 36{,}78 - 1{,}645 \times 0{,}641 \approx 35{,}73$, alors que la norme exigeante serait $C_{Pexig}(\alpha) \approx 36{,}78 + 1{,}645 \times 0{,}641 \approx 37{,}83$, deux valeurs assez proches des valeurs exactes tirées des tables F.

Selon le modèle normal appliqué aux statistiques d'ordre

Le modèle normal peut être invoqué aussi pour les statistiques d'ordre, notamment dans les cas, fréquemment rencontrés en psychométrie, où les scores bruts des tests ont été transférés vers la forme normale, ce en projetant chaque quantile (ou chaque statistique d'ordre) de la distribution d'origine vers un quantile correspondant du modèle normal[17]. En outre, en fonction d'une perspective pragmatique souvent vraisemblable, le modèle normal peut convenir à décrire directement une seule *aile* de la distribution plutôt que toute la distribution, aile dans laquelle la sélection doit se faire. Nos normes « ordinales normales » (Laurencelle, 2016b) conviennent à ces deux situations.

La norme est basée sur une statistique d'ordre, notée $X_{[r]}$, plus complètement $X_{[r:n]}$: c'est, une fois la série statistique placée en ordre de valeurs croissantes, soit $X_{[1]} \leq X_{[2]} \leq \ldots \leq X_{[n]}$, la valeur qui occupe le rang r. Les statistiques d'ordre (SO) d'une série statistique émanant d'une loi de probabilité ont leur distribution façonnée par cette loi. Par exemple, les SO issues d'une loi normale ont une variance qui augmente à mesure que le rang r s'écarte du rang médian ½n, alors que le contraire se produit pour une loi uniforme. De plus, l'asymétrie (mesurée par l'indice γ_1) augmente quand r s'écarte de ½n, mais elle le fait en sens contraire pour les deux lois, par exemple s'écartant positivement de 0 pour la normale et négativement pour l'uniforme à mesure que r approche n. [Pour les calculs, voir Laurencelle, 2016b, ou David & Johnson, 1954]. La statistique d'ordre qui approche le P^e^ quantile de la population est donnée par $P = r / (n + 1)$, soit $X_{[r]}$, où $r = P \cdot (n + 1)$.

Exemple 4 : Norme sûre ordinale (critère échantillonnal, mesure sans erreur)

Reprenant le contexte de l'exemple 3, nous cherchons, dans notre série de $n = 150$ données, la SO située approximativement au 95e rang centile de la population, soit $r / (n + 1) \approx 0{,}95$ ou $r = 0{,}95 \times (150 + 1) = 143{,}45$, la valeur cherchée étant située dans le voisinage des données $X_{[143]}$ et $X_{[144]}$. Cette valeur supposerait que la série normative observée est stable, alors qu'on sait qu'elle est sujette aux fluctuations échantillonnales, comme l'étaient plus tôt la moyenne et l'écart-type de cette série. En fait, la variance d'erreur du quantile échantillonnal $Z_{[P:n]}$ dans une série standardisée de $n = 150$ données peut être estimée[18] par :

(11) $\mathrm{var}(Z_{[P:n]}) \approx \dfrac{P(1-P)}{n[\varphi(z_P)]^2}$,

où $\varphi(z) = \exp(-z^2/2) / \sqrt{2\pi}$ est la densité normale standard à la valeur z. Ici, avec $P = 0{,}95$, $z_P \approx 1{,}645$, $\varphi(1{,}645) \approx 0{,}1031$ et $n = 150$, l'estimation fournit $\mathrm{var}(Z_{[0,95]}) \approx 0{,}0298$. Un calcul approximatif plus précis (voir Laurencelle, 2015, p. 99 ou, plus extensivement, 2016b, p. 22, éq. 14) donne 0,0306 ; c'est ce calcul que nous recommandons.

Norme sûre exacte

Notre norme sûre ordinale basée sur le modèle normal (Laurencelle, 2016b) tient compte de cette fluctuation échantillonnale du quantile, les tableaux F11 à F20 présentant des valeurs de référence. Pour un taux de sélection $f = 0{,}05$, une erreur échantillonnale selon $\alpha = 0{,}05$ et une erreur

de mesure nulle (avec R = 1), la table F14b pour une norme exigeante indique $r_b = 99$ pour $n_b = 100$ et $r_h = 244$ pour $n_h = 250$. (Les indices b et h dénotent les valeurs basse et haute trouvées dans la table.) L'interpolation pour $n = 150$, par la formule :

(12) $$\hat{r} = \left[\frac{r_b}{n_b+1} + \left(\frac{r_h}{n_h+1} - \frac{r_b}{n_b+1}\right) \cdot \left(\frac{1/n - 1/n_b}{1/n_h - 1/n_b}\right)\right] \times (n+1),$$

donne $\hat{r} \approx 147{,}33$. Un calcul direct montre que $147 < r \leq 148$. La norme interpolée fait donc l'affaire, soit:

$C_{Pexig} \approx X_{[147,33]} = X_{[147]} + (X_{[148]} - X_{[147]}) \times 0{,}33$;

le facteur 0,33, à droite, est simplement la partie décimale de $\hat{r}$. Notons que, grâce à l'interpolation, la valeur obtenue est (légèrement) moins sévère, ou exigeante, que celle correspondant au 148 indiqué.

Quant à la norme permissive correspondante, on la trouve à la table F19b, soit $r_b = 91$ pour $n_b = 100$ et $r_h = 232$ pour $n_h = 250$, d'où $\hat{r} \approx 138{,}01$, le calcul direct donnant 138. Nous suggérons alors :

$C_{Pperm} = X_{[138]}$.

Norme approchée

Le rang approximatif $r = P \times (n + 1) = 143{,}45$ nous fournissant une base, il est légitime d'y adjoindre une marge d'erreur échantillonnale basée sur la variance d'erreur du quantile, ici $\mathrm{var}(Z_{[0,95:150]}) = \mathrm{var}(C_P) \approx 0{,}0306$, telle qu'indiquée plus haut, d'où l'erreur-type $\sigma(C_P) = 0{,}175$. Ignorant ici le léger degré d'asymétrie positive (estimé à 0,181) de la distribution de $X_{[143,45:150]}$ et appliquant le modèle normal à la variation du quantile, nous proposons donc :

(13) $C_P = X_{[r]} \pm z_{1-\alpha} \times \sigma(C_P) \cdot s_X$.

Ici, $C_P \approx X_{[143,45]} \pm 1{,}645 \times 0{,}175 \times 5{,}12 = X_{[143]} + (X_{[144]} - X_{[143]}) \times 0{,}45 \pm 1{,}47$, l'addition ou la soustraction du dernier terme fournissant la norme exigeante ou permissive.

Sans stipuler le modèle normal[19]

Que ce soit par une démonstration convaincante de son inadéquation ou bien par prudence, l'utilisateur peut renoncer tout à fait au modèle normal et se replier plutôt sur les seules données enregistrées, hors modèle, soit les statistiques d'ordre obtenues à partir de la série normative, dont les conditions d'adéquation sont plus libérales, et ce, aux dépens d'une variance un peu plus grande dans les valeurs produites. L'incertitude

dépend ici seulement du rang de $X_{[r]}$, quels que soient la valeur X correspondante ou son intervalle numérique par rapport aux autres données de l'échantillon[20].

Exemple 5 : Norme-paramétrique (critère échantillonnal, mesure sans erreur)

Norme sûre exacte

Reprenant le même exemple qu'en 3 et 4 et à partir de notre série normative de taille $n = 150$, nous devons, sans référence à un modèle de distribution, trouver le rang s d'une norme permettant de discriminer les meilleurs 5 % (= f) de la population, en gardant sous un taux α d'au plus 5 % l'incertitude attachée à cette norme. Les tables G1 et G2 (Laurencelle, 2016b) fournissent l'information nécessaire : pour chaque couple f:α donné est inscrite la taille n requise telle que le ratio s:n respecte les contraintes f et α prescrites.

La table G1 (*op. cit.*, p. 169), pour la norme non-paramétrique exigeante, présente la taille *minimale* requise pour que la statistique d'ordre $X_{[s]}$ indiquée[21] discrimine la fraction f voulue selon le taux d'erreur α maximal prescrit. Soit $f = 0{,}05$ et $\alpha = 0{,}05$, et la taille $n = 150$. À la colonne $\gamma = 0{,}95 = 1 - f$, on lit 124 (une taille admissible) pour la norme $s = n - 2$ ou $X_{[n-2]}$, et 153 (une taille que n'atteint pas notre $n = 150$) pour $X_{[n-3]}$. Ici, donc, la valeur correspondant à la statistique d'ordre $X_{[150-2]} = X_{[148]}$ satisferait nos conditions, avec la règle « Retenir le candidat si son score $x \geq X_{[148]}$ ».

Le lecteur minutieux devinera que, puisque notre $n = 150$ est campée entre les bornes $s = n - 2$ pour $n = 124$ et $s = n - 3$ pour 153, voire qu'elle est plus près de celle appropriée à 153 qu'à 124, une norme s plus précise pour $n = 150$ se situerait entre les rangs $n - 2$ et $n - 3$. En effet, et un calcul binomial direct le confirme, on peut raffiner la norme non-paramétrique soit en procédant par calcul direct et interpolation complexe sur α, soit par interpolation simple à partir des bornes offertes à la table G1. Soit n_b et s_b, la taille (inférieure à n) et le rang s correspondant, et n_h et s_h, la taille supérieure et son rang, le rang intermédiaire $\hat{s}$ estimé s'obtient par :

$$(14)\quad \hat{s} = \left[\frac{s_b}{n_b+1} + \left(\frac{s_h}{n_h+1} - \frac{s_b}{n_b+1}\right) \cdot \left(\frac{1/n - 1/n_b}{1/n_h - 1/n_b}\right)\right] \times (n+1),$$

Selon nos données ci-dessus (n_b = 124, s_b = 122, n_h = 153, s_h = 150, n = 150), nous obtenons $\hat{s}$ = 147,10, le calcul probabiliste indiquant aussi 147,10. La norme résultante, plus précise, serait alors interpolée entre les statistiques d'ordre $X_{[147]}$ et $X_{[148]}$, simplement par :
$\hat{s} = X_{[147,10]} \approx X_{[147]} + (X_{[148]} - X_{[147]}) \times 0{,}10$.

Quant à la norme permissive, la table G2 (*op. cit.*, p. 170) fournit la taille *n maximale* assurant que le candidat appartenant à la fraction supérieure *f* et confronté à la norme de rang *s* n'a qu'un risque de grandeur α d'être refoulé. Pour notre même exemple, avec un échantillon normatif de taille 150, avec un taux de sélection de 5% et avec un risque d'erreur de 5%, la table donne n_h = 155 et $s_h = n - 12$, pour une norme égale à la valeur de $X_{[138]}$. Si l'on passe à l'interpolation afin d'obtenir une valeur plus précise et en exploitant dans la table n_b = 140 et $s_b = n - 11$, on aboutit à $\hat{s} \approx$ 138,39, le calcul probabiliste indiquant la même valeur.

Norme approchée

L'incertitude de la statistique d'ordre non-paramétrique, $X_{[r]}$ ou $X_{[s]}$, ne fait référence à aucun modèle, sinon le modèle uniforme (aussi appelé *rectangulaire*), pour lequel on peut déterminer des moments. Soit une population de variable U, dont les mesures sont uniformément réparties entre les bornes *a* et *b*. Les moments de U sont alors $\mu = ½(a + b)$, $\sigma^2 = (a - b)^2/12$, $\gamma_1 = 0$ et $\gamma_2 = -1{,}2$. Modèle de rencontre plutôt rare en psychométrie, voire en biométrie, son invocation justifiée permet d'approcher la norme calculée ci-dessus, en en exploitant les propriétés (Johnson et al., 1994 ; Laurencelle, 1993).

Utilisant la variance de la statistique d'ordre uniforme $U_{[r]}$, soit $\sigma^2(U_{[r]}) = r\cdot(n + 1 - r) / [(n + 1)^2\cdot(n + 2)]$ et projetant sa valeur sur le domaine du rang *r* (de 1 à *n*), le calcul

(15) $\hat{r} = r \pm z_{1-\alpha}\cdot n\cdot\sigma(U_{[r]})$

fournit une estimation possible du rang de la norme sûre non-paramétrique, en mode exigeant (+) ou permissif (–). Tout de suite avec *r* = 143,45 et α = 0,05, nous obtenons $\sigma^2(U_{[r]}) \approx 0{,}0003125$ et σ = 0,0177, d'où $\hat{r}$ = 143,45 ± 1,645 ×150 × 0,0177 ≈ 147,81 pour le critère exigeant (+) ou = 139,08 pour le critère permissif. Nous aurions donc, en approximation, $C_{Pexig} = X_{[147,81]}$ et $C_{Pperm} = X_{[139,08]}$.

Les calculs approchés ci-dessus semblent tenir la route, voire ils concordent assez bien avec ceux effectués en invoquant le modèle normal (voir Exemple 4), ce malgré l'insertion cavalière de l'intervalle normal ($z_{1-\alpha}$) dans l'axe des rangs qui relève en fait du modèle uniforme[22].

Cas $n < \infty$, $\sigma_e > 0$

Le contexte de mesure sans doute le plus courant est celui dans lequel l'information normative provient de données mesurées sur un échantillon, plutôt qu'elle ne soit imposée par décret d'une autorité administrative ou d'un comité d'experts, et où la mesure utilisée, notée ici X, est sujette à fluctuation. C'est ce que symbolisent les indications $n < \infty$ et $\sigma_e > 0$, cette seconde notation correspondant à une fidélité imparfaite ($R < 1$).

Si la norme C_P attendue, quelle que soit sa forme, doit être fixée par calcul sur la base d'un échantillon normatif de taille n, elle retiendra une part de l'incertitude dégagée par l'échantillon. Puisque, de plus, le score x d'un candidat évalué doit être comparé à cette norme C_P, l'erreur e contenue dans le score, selon le modèle $x = V + e$, influencera aussi cette comparaison et elle déteindra éventuellement sur le classement du candidat, sur la décision de le retenir ou non. C'est à l'analyse de ce contexte et à la protection de la décision normative qui s'ensuit que s'attache la théorie de la double erreur, ou du double contrôle d'erreur, théorie exposée sommairement dans Laurencelle (2016a, 2016b).

Il s'agit donc de repérer et de sélectionner un candidat si sa mesure le classe parmi la fraction supérieure f de la population. Or, ce titre d'appartenance à une classe de la population se base et doit se baser sur les valeurs vraies des éléments qui la composent, et non sur leurs mesures fluctuantes. Si l'on retient la notation $P = 1 - f$ et occulte l'erreur e, il s'agit donc de repérer les candidats pour qui $V \geq V_P$, ceux dont le mérite réel les place dans la « vraie » fraction f visée.

Empruntant la même démarche que plus haut (4) et prenant le cas d'une norme exigeante, il s'agit de déterminer la valeur seuil C_P telle que la probabilité qu'un candidat non qualifié selon $V < V_P$ soit retenu ait pour maximum α, :

(16) $\max \Pr\{ x \geq C_P \} = \Pr\{ V + e \geq C_P \mid V < V_P \} = \alpha$;

d'où, puisque ce maximum se produit lorsque $V \rightarrow V_P$:

(17) $\Pr\{ V_P + e \geq C_P \} = \alpha$.

La solution de (17) diffère selon qu'il s'agit d'une norme linéaire, $C_P = \overline{X} + \lambda \cdot s_X$, ou ordinale, $C_P = X_{[r]}$. Les exemples suivants en illustrent l'usage.

En admettant un modèle normal complet (pour e *et X) : la norme sûre linéaire*

Le contexte statistique traité ici concerne, rappelons-le, une mesure à fidélité imparfaite ($R < 1$), confrontée éventuellement à une norme basée sur un échantillon normatif de taille limitée n. Ce contexte suppose naturellement que le composant d'erreur e suive une loi normale[23], et c'est aussi une condition pour X (et V, bien entendu), dont les propriétés normales sont exploitées pour la norme linéaire (distribution échantillonnale de $\overline{X}$ et s_X) et pour la norme ordinale (distribution de $X_{[r]}$). Il s'agit donc d'assurer, avec un coefficient de confiance de $1 - \alpha$, que la personne non qualifiée pour la fraction f supérieure de la population n'est pas retenue par le critère C_P ou, inversement, que la personne qualifiée n'est pas rejetée.

Exemple 6 : Norme sûre linéaire (critère échantillonnal, mesure avec erreur)

Nous reprenons notre exemple 3, la base normative comptant $n = 150$ mesures X, avec $\overline{X} = 28{,}36$, $s_X = 5{,}12$ et l'assurance d'une distribution (quasi) normale pour X.[24] La fraction de sélection est $f = 0{,}05$ (pour $P = 0{,}95$) et le niveau d'incertitude toléré est $\alpha = 0{,}05$. Dans le cas présent, la mesure offrirait une fidélité de $R = 0{,}80$, ce qui correspond à une erreur-type de mesure de $\sigma_e \approx s_e = s_X \cdot \sqrt{1-R} \approx 5{,}12 \times \sqrt{1\text{-}0{,}80} \approx 2{,}290$. Quelle norme proposer ?

Norme sûre exacte

La section F dans Laurencelle (2016b) présente les normes λ_{exig} et λ_{perm} dans un assortiment de tables, selon f (ou P), n, R et α, de même que des procédures d'estimation par interpolation pour les conditions f, n, R et α non cataloguées. Pour le critère exigeant, la table F4b (*op. cit.*, p. 139) donne $\lambda_{exig} = 2{,}253$, d'où, en appliquant (9), nous calculons $C_P = 28{,}36 + 2{,}253 \times 5{,}12 \approx 39{,}90$. Dans les conditions données, on est sûr à 95 % qu'un candidat obtenant $x \geq 39{,}90$ relève de la fraction des 5 % supérieurs de la population. Quant au critère permissif, qui assure à 95 % que la norme n'échappe pas un candidat possiblement méritant, la table F9b (*op. cit.*, p. 144) donne $\lambda_{perm} = 0{,}722$ pour $C_P \approx 32{,}06$.

Norme approchée

Une méthode un peu grossière pour approcher la norme correcte ci-dessus consiste à prolonger la norme naïve, $\overline{X} + z_{1-f} \cdot s_X$, par l'intervalle d'erreur approprié, celui-ci étant une combinaison simple des erreurs-types de chacun des composants[25]. Nous avons déjà calculé $s_e = 2{,}290$ et obtenu plus haut (10), où $s_{Cp} \approx 0{,}641$. L'erreur-type conjointe est donc :

(18) $s_{\text{conjointe}} = \sqrt{s_e^2 + s_{CP}^2}$,

soit $s_{\text{conjointe}} = \sqrt{2{,}290^2 + 0{,}604^2} \approx 2{,}378$, d'où $C_P \approx \overline{X} + z_{1-f} \cdot s_X + z_{1-\alpha} \cdot s_{\text{conjointe}} \approx 28{,}36 + 1{,}645 \times 5{,}12 + 1{,}645 \times 2{,}378 \approx 40{,}69$ (vs la norme exacte, 39,90). Si l'on soustrait plutôt qu'additionne la part des erreurs, la norme permissive approchée devient 32,87 (vs 32,06).

Exemple 7 : Norme sûre ordinale (critère échantillonnal, mesure avec erreur)

Norme sûre exacte

Toujours le même exemple, où l'on a une banque normative de $n = 150$ mesures normalement distribuées et de statistiques $\overline{X} = 28{,}36$ et $s_X = 5{,}12$, d'où on veut tirer une statistique d'ordre $X_{[r]}$ qui serve de norme de sélection, le taux de sélection étant $f = 0{,}05$ (avec $P = 1 - f = 0{,}95$), la fidélité étant estimée à 0,80 et l'erreur-type de mesure étant estimée à $s_e = 2{,}290$.

Les tables F (Laurencelle, 2016b) fournissent aussi des indications permettant de déterminer r, le rang de la statistique d'ordre requise. Pour le critère exigeant, à la table F14b, nous trouvons $r = 100$ pour $n = 100$ et $r = 248$ pour $n = 250$. Par interpolation suivant (12), nous obtenons $\hat{r} \approx 149{,}33$ pour $n = 150$, le calcul direct montrant $149 < r \leq 150$. La norme correspondrait ici à $C_{Pexig} = X_{[149,33]}$, avec interpolation. Quant à la norme permissive, la table F19b propose $r = 76$ pour $n = 100$ et $r = 191$ pour $n = 250$, d'où le rang interpolé devient 114,33 et $C_{Pperm} = X_{[114,33]}$.

Norme sûre approchée

Le rang de la norme naïve, estimé par $P \cdot (n + 1)$, serait $\hat{r} = 143{,}45$, tel que déjà vu. La protection contre les deux erreurs (l'erreur de mesure e et l'erreur échantillonnale de la statistique $X_{[P]}$) doit être ajoutée. L'erreur-type de mesure, déjà obtenue, est de $\sigma_e = 2{,}290$. Quant à l'erreur-type de $X_{[0,95]}$, l'exemple 4 nous a permis de trouver $s_{CP} = \sqrt{0{,}0306} \approx 0{,}175$. L'erreur

combinée devient alors $s_{\text{conjointe}} = \sqrt{s_{\text{e}}^2 + s_{\text{CP}}^2} = \sqrt{2{,}290^2 + 0{,}175^2} \approx 2{,}297$. Comme en (18), la norme exigeante proposée serait alors $X_{[143,45]} + z_{1-\alpha} \cdot s_{\text{conjointe}} \approx X_{[143,45]} + 1{,}645 \times 2{,}297 \approx X_{[143,45]} + 3{,}78$.

La norme non-paramétrique avec erreur de mesure n'est pas calculable, l'obstacle consistant dans l'impossibilité d'abouter une marge (linéaire) d'incertitude normale sur la mesure à une statistique d'ordre (SO) définie uniquement par son rang. Pour obvier à ce problème, il convient par exemple de transférer les SO originales vers le modèle normal (par le biais d'une règle de correspondance) et d'y appliquer ensuite la méthode de la norme sûre ordinale.

Un résumé des erreurs-types et d'autres exemples de leur usage

Dans les sections précédentes, différents contextes ont été présentés dans lesquels l'erreur de mesure, l'erreur du seuil comparatif ou les deux erreurs à la fois ont été prises en compte *dans la perspective d'un jugement comparatif*. Ces deux types d'erreur ont aussi une incidence dans la simple interprétation d'un score psychométrique ou d'une mesure quelconque. Le fait de connaître la marge d'erreur applicable à une situation donnée permettra à l'évaluateur de mieux visualiser la place probable du répondant dans la population et de prononcer éventuellement une interprétation mieux fondée et nuancée. Nous récapitulons ici, très sommairement, les grandes lignes de la théorie du double contrôle de l'erreur psychométrique, à la faveur de quelques nouveaux exemples, lesquels pourront aussi servir d'exercices au lecteur intéressé.

Cas $n \to \infty$, $\sigma_e > 0$ ($R < 1$)

Robert a passé une évaluation de son QI et a obtenu un score $x = 118$. L'échelle conventionnelle est de forme normale, par construction, avec comme paramètres $\mu = 100$ et $\sigma = 15$. La fidélité test-retest pour le test appliqué est de $r_{XX} = R = 0{,}90$. L'erreur-type (8) est donc ici $\sigma_e = 15 \times \sqrt{1 - 0{,}90} \approx 4{,}74$.

Les données présentées suggèrent que la valeur vraie de Robert occupe l'intervalle[26] $\{ V \pm \sigma_e \} = \{ 118 \pm 4{,}74 \} \approx \{ 113 ; 123 \}$. Un guide d'interprétation considère que les personnes occupant l'intervalle $\{ 85 ; 115 \}$ se situent *dans la moyenne*, alors que, au-dessus de 115, elles seraient *dans*

la haute moyenne. Avec son $x = 118$, peut-on affirmer que Robert se range dans la haute moyenne ? Le lecteur pourra vérifier que la probabilité qu'il s'y situe est d'environ 0,737 (en se rappelant qu'elle serait de 0,500 si son score avait été de 115 !) et qu'il aurait fallu qu'il soit de 126 ou plus pour qu'il soit classé *haute moyenne* selon un risque d'erreur d'au plus 5 %.

Exemple de test à format binomial[27]

Un chargé de cours de l'École polytechnique de Montréal veut repérer les étudiants qui ne connaissent rien aux logarithmes. Pour ce faire, il conçoit un questionnaire composé de 12 items à choix multiples simples, à raison de 4 items pour $k = 2$, 3 et 4 choix respectivement. Chaque bonne réponse vaut 1 point[28], de sorte que les scores possibles vont de 0 à 12. Pour s'assurer à 99 % que l'étudiant n'a pas répondu complètement au hasard, quel score-seuil le chargé de cours doit-il fixer ?

La probabilité binomiale ici est mixte (soit les trois valeurs π également réparties, $1/k = ½$, ⅓ et ¼), avec une moyenne $\overline{\pi} \approx 0{,}3611$ et une variance $\sigma^2_{\overline{\pi}} \approx 0{,}0108$. La formule A2 (voir Appendice) permet d'obtenir $\mu(X) = 4{,}333$ et $\sigma(X) \approx 1{,}624$. Pour réduire à 1 % le risque de non-repérage, il faut que le score-seuil x^* satisfasse l'inéquation : $(x^* - \mu(X) - ½) / \sigma(X) \geq z_{1-0{,}01} = 2{,}326$, d'où $x^* \geq \mu(X) + ½ + z_{0{,}99} \times \sigma(X)$. Le lecteur vérifiera que le score-seuil devra être $x^* = 9$ (ou $x^* = 8$ pour un risque à 5 %). Pour cet examen, la probabilité précise d'obtenir au hasard un score $x \geq 9$ est de 0,006 (et de 0,027 pour $x \geq 8$).[29]

Cas $n < \infty$, $\sigma_e = 0$ ($R = 1$)

Le poids d'un enfant amérindien mâle de 12 mois est de 14,5 kg : est-il en surpoids ? Pour en juger, le pédiatre se réfère à des normes de croissance, normes qui, idéalement, sont appropriées pour le sexe et l'ethnie concernés. Le manuel de normes indique, pour un garçon de cet âge, un poids moyen de 9,3 kg et un écart-type de 3,1 (données fictives), la norme étant basée sur un petit échantillon de $n = 100$ enfants.

Le poids, dans une population statistiquement homogène, se distribue normalement, de sorte qu'il est possible de vérifier l'exceptionnalité d'une valeur donnée par un test simple, soit $z = (x - \mu) / \sigma$, test qu'il faut réaliser ici par un test t en raison de la substitution obligée de l'écart-type échantillonnal ($s = 3{,}1$) pour l'écart-type populationnel σ inconnu. Le test $t = (x - \overline{X}) / s \approx 1{,}677$, avec $n - 1 = 99$ degrés de liberté, présente une pro-

babilité extrême de 0,048, ce qui indique que l'enfant dans notre exemple s'écarte significativement de la moyenne de sa population, selon un risque d'erreur de 5%. Mais est-ce là du surpoids ?

Supposons que, par surpoids, on entende avoir un poids logé parmi les 5% plus élevés de la population. Ce niveau de 5% constitue une sorte de taux de sélection, d'où $f = 0{,}05$, et alors les tables F (Laurencelle, 2016b ; voir aussi exemple 3) indiquent $\lambda_{exig} = 1{,}927$ pour un taux d'erreur $\alpha = 0{,}05$ et $n = 100$, avec $R = 1$. Le seuil de surpoids serait alors $C_p = 9{,}3 + 1{,}927 \times 3{,}1 \approx 15{,}3$, d'où, selon les critères appliqués, l'enfant ne serait pas jugé en surpoids. En norme approchée, le seuil naïf serait $\overline{X} + z_{1-f} \cdot s_X = 9{,}3 + 1{,}645 \times 3{,}1 \approx 14{,}000$, à quoi il faut, pour le critère exigeant, ajouter la part de l'erreur échantillonnale. Selon (10), $\mathrm{var}(C_P) \approx 3{,}1^2/100 \times (1 + 1{,}645^2/2) \approx 0{,}226$ et $\sigma(C_P) \approx 0{,}475$, la marge de protection étant $z_{1-\alpha} \cdot \sigma(C_P) = 1{,}645 \times 0{,}476 \approx 0{,}781$, fournissant la norme sûre approchée $C_P = 14{,}000 + 0{,}781 \approx 14{,}78$. La solution par la normale ordinale (inutile ici) serait $C_P \approx X_{[95,95]} + 1{,}645 \times 0{,}413 \times 3{,}1$ selon la formule (13), l'erreur-type $\sigma(X_{[P:n]}) \approx 0{,}413$ étant donnée par (11).

Cas $n < \infty$, $\sigma_e > 0$ ($R < 1$)

Une commission scolaire du Québec, préoccupée par les difficultés des élèves du secondaire en mathématiques, veut leur offrir un programme d'enseignement modifié, plus convivial et ludique. Dans ce but, on a mis sur pied un comité formé d'enseignants, de deux psychologues scolaires et d'un psychométricien afin qu'il élabore un test d'anxiété et de sentiment d'inaptitude vis-à-vis des mathématiques. Le test a été soumis à tous les élèves admissibles de 1re secondaire ($n = 450$). La distribution des scores, à peu près normale, va de $X = 12$ à $X = 87$ et a pour moyenne $\overline{X} = 53{,}86$ et pour écart-type $s_X = 19{,}75$. La fidélité de mesure a été estimée à $R = 0{,}80$. Considérant qu'environ 20% des élèves sont à risque, le taux d'identification (c.-à-d. de sélection) des élèves à risque est fixé à $f = 0{,}20$. Quelle norme devrait-on appliquer pour identifier les élèves à risque et pour leur proposer le programme modifié ?

Afin de ne pas laisser échapper des élèves potentiellement à risque, le comité opte pour un critère permissif, au seuil d'erreur de 10% (ou $\alpha = 0{,}10$). Les tables F (Laurencelle, 2016b) proposent des seuils λ pour $f = 0{,}50$, 0,25, 0,10, 0,05 et 0,01, de même que pour quelques valeurs pivots des paramètres α, R et n. La valeur λ_{perm} pour $f = 0{,}20$ doit donc être interpolée. L'interpolation proposée sur f est (*op. cit.*, équation F19) :

$\hat{\lambda} = \lambda_1 + (\lambda_2 - \lambda_1)\cdot[(u - u_1)/(u_2 - u_1)]$,
où $u_f = \sqrt{-\ln(-\ln(1-f))}$. La table F7a (*op. cit.*, p. 142), pour $n = 500$, R = 0,80 et $\alpha = 0{,}10$ et $f_1 = 0{,}25$, fournit $\lambda_{perm} = 0{,}027$; la valeur correspondante de la table F8a pour $f_2 = 0{,}10$ est 0,570. Calculant u_1 (pour f_1) ≈ 1,116, u_2 ≈ 1,500 et u (pour f) ≈ 1,225, l'interpolation ci-dessus donne $\hat{\lambda} \approx 0{,}181$. La norme permissive à appliquer serait donc $C_P \approx 53{,}86 + 0{,}181 \times 19{,}75 \approx 57{,}43 = 57$ ou 58, le risque d'« échapper » un cas problème étant d'environ 10 %. En approximation, la formule (13) s'applique encore, cette fois avec l'erreur-type conjointe (18), qui peut s'écrire aussi $s_{conjointe} = s_X \cdot \sqrt{1-R + (1-z^2_{1-f}/2)/n}$, et qui vaut ici 8,898. Appliquant alors (13), nous obtenons $C_P \approx 53{,}86 + 0{,}842 \times 19{,}75 - 1{,}282 \times 8{,}898 \approx 59{,}08 \approx 59$, une valeur permissive (notons le signe – dans la formule) légèrement trop haute[30].

Épilogue

Deux nuances importantes doivent conclure les procédures et illustrations élaborées dans cet article et dans les deux précédents. D'abord, comme c'est souvent le cas dans les applications statistiques, on devrait parler ici de vraisemblance plutôt que de probabilité, le degré de vraisemblance référant à un procédé inférentiel inductif (de l'observation au modèle), tandis que la probabilité est déductive (du modèle à l'observation) et suppose que le modèle ait été démontré vrai. Ainsi, les calculs de « probabilité » que nous faisons sous-tendent d'abord un modèle à partir duquel certains calculs permettent d'apprécier des conséquences *sans que le modèle ait été préalablement avéré*. La seconde nuance se rapporte aussi au modèle ou, plus précisément, au modèle de mesure. Dans plusieurs cas, comme ici, nous supposons que les mesures prises sont continues, alors que les valeurs rapportées ont, la plupart du temps, une définition numérique finie (p. ex., QI = 127, score d'anxiété = 32, poids = 65,4 kg) et que, dans la « population », elles répondent à un modèle de distribution déterminé, ici le modèle normal ou quasi normal. Cette inversion, de déduction à induction, et ce caractère au mieux approximatif du modèle de mesure et de distribution n'ont pas empêché une production scientifique et pratique abondante, concluante et utile, et il est d'usage de s'en contenter. Il reste tout de même prudent de garder ces nuances à l'esprit.

La « théorie des seuils psychométriques » présentée ici n'est évidemment qu'un essai de systématisation et de structuration mathématique des enjeux qui entourent la décision basée sur le testing, quel qu'il soit. Au Canada, l'affaire Meiorin (Colombie-Britannique c. BCGSEU, 1999 ; voir aussi Bosset, 2007) a constitué, par son importance et par ses retentissements déontologiques et légaux, un point de non-retour : toute décision à base de jugement, décision qui affecte la carrière ou la vie d'une personne, doit reposer sur des fondements explicites, explicitement validés et rigoureusement démontrés, le seuil de décision psychométrique en faisant intégralement partie. Notre effort en ce sens reste modeste et, bien sûr, très incomplet. Espérons que, par la créativité ou par la controverse qu'il pourra susciter, il aura contribué à actualiser le problème et incité d'autres chercheurs à en faire avancer la solution.

Réception : 7 juillet 2017

Version finale : 4 décembre 2017

Acceptation : 4 décembre 2017

NOTES

1. Certaines évaluations se basent sur le relevé de données non numériques, notamment en recherche dite qualitative. Voir par exemple Poisson (1992) et Lessard-Hébert, Goyette et Boutin (1996).
2. Rappelons que l'échelle de mesure des valeurs vraies V est liée à celle des mesures X, selon $\mu_V = \mu_X$, $\sigma_v^2 = R \cdot \sigma_x^2$, où R (habituellement noté r_{XX}) est le coefficient de fidélité. D'où la relation entre centiles de rang P est : $V_P = \mu_V + (X_P - \mu_X) \times \sqrt{R}$
3. Avec des exceptions. Voir en particulier le modèle binomial et l'exemple 2, plus bas.
4. Pour une définition plus complète de l'erreur-type de mesure et ses méthodes d'estimation, voir Laurencelle (1998), Bertrand et Blais (2004) et Allaire et Laurencelle (1998).
5. Il s'agit des indices d'asymétrie (γ_1) et d'aplatissement ou voussure (γ_2), dont les estimateurs échantillonnaux sont g_1 et g_2 ; pour une distribution normale, $\gamma_1 = 0$ et $\gamma_2 = 0$.
6. Les scores de test sont typiquement discontinus, tels les niveaux de QI ou des échelles d'anxiété ou de motivation, mais ils expriment une quantité sous-jacente continue et peuvent généralement être traités comme tels.
7. Le modèle simpliste et commode présenté ici est à comparer avec celui de la théorie des réponses aux items (TRI) attribué à F. M. Lord, dans lequel la probabilité de réponse dépend de l'interaction entre le niveau d'habileté du répondant et la difficulté de l'item présenté. Voir Bertrand et Blais (2004).
8. À ne pas confondre avec le seuil de probabilité α, utilisé ailleurs. D'autres indices momentanés (obtenus en une seule séance) sont aussi disponibles, tels ω (oméga) et GLB. Voir Trizano-Hermosilla et Alvarado, 2016.
9. La formule donnée n'est généralement pas le meilleur estimateur. Voir Allaire et Laurencelle (1998) ou Laurencelle (1998).
10. Généralement, mais pas absolument impossible à connaître, la valeur vraie peut soit être approximée par l'accumulation de mesures successives, soit être connue avant d'être intégrée à un système de mesure générateur d'erreur.
11. Le terme « taux de sélection » est utilisé à toutes fins utiles et sans exclusion, et il peut être généralisé au concept de borne de classement (comme celles appliquées pour les niveaux cliniques de compétence intellectuelle par le biais du QI), une mesure donnée pouvant présenter à la fois plusieurs bornes distinctes.
12. Par groupe normatif, nous entendons un échantillon, représentatif et de taille suffisante, tiré de la population de référence parmi laquelle et pour laquelle la mesure utilisée doit permettre d'effectuer un jugement comparatif juste. Voir à ce sujet Laurencelle (1998, p. 197 et suiv.).
13. Il s'agit d'une statistique d'ordre $X_{[r]}$, $r = 1$ à n, $X_{[3]}$ désignant par exemple la troisième valeur plus petite de la série ordonnée de n données. Notons que des formules autres que $r = P \cdot (n + 1)$ sont possibles, selon la forme $r = (P + a) \cdot (n + 1 - 2a)$, où $a \geq 0$, celle utilisée ici ($a = 0$) étant coutumière et correspondant exactement à l'espérance d'une statistique d'ordre de la loi uniforme.

14. Tout en étant d'un niveau acceptable, la taille 150 est plutôt modeste et ne conviendrait idéalement qu'à des mesures d'une grande stabilité et de dispersion modérée dans la population. Les normes psychométriques typiques ont avantage à reposer sur des échantillons normatifs de plusieurs centaines d'individus. Nous avons ici fixé la taille $n = 150$ afin que les propriétés de variance de la norme ressortent mieux.
15. En fait, $E\{ s \} < \sigma$ et $E\{ s \} = q_n \cdot \sigma$, où $q_n = \sqrt{2/(n-1)}\ \Gamma\,[n/2]/\Gamma[(n-1)/2] \approx 1 - 1/[4n - 4)]$. Utilisant $s = 5{,}12$ et $n = 150$, l'estimateur débiaisé serait $\sigma = E\{ s / q_n \} = 5{,}12 / 0{,}99832 \approx 5{,}129$, une différence sans doute négligeable. Les normes sûres dans Laurencelle (2016b) ne sont pas touchées par ce biais.
16. Laurencelle (2016b) présente en page 126 une formule d'approximation (F21), qui est $\hat{\lambda} = \delta b + z_\alpha \sqrt{(b^2+(1-b^2)(\delta^2-z_\alpha^2)}/\,[\sqrt{n}(b^2-z_\alpha^2\,(1-b^2)]$, dans laquelle $\delta = \sqrt{n} \cdot \Phi^{-1}(1-f)$, $b = 1 - 1/(4n - 4)$ et $z_\alpha = \Phi^{-1}(1-\alpha)$ pour une norme exigeante ou $\Phi^{-1}(\alpha)$ pour une norme permissive. Cette formule (adaptée de Johnson, Kotz & Balakrishnan, 1994) se rapporte à la loi t non centrale, laquelle a une parenté structurelle avec la norme protégée. Pour l'exemple présenté, les normes λ dérivées de la formule sont respectivement 1,871 et 1,454, d'où les quantités 38,47 et 35,43.
17. Les situations dans lesquelles les données sont mathématiquement transformées (grâce à une formule de transformation non linéaire) vers la forme normale (comme pour le QI, pour certaines échelles du MMPI-2 et dans d'autres cas) permettent d'exploiter toutes les propriétés linéaires du modèle normal, tel que dans l'exemple 3, le recours aux statistiques d'ordre devenant vain. Voir Laurencelle (1998, chap. 6) pour un exemple développé de traitement d'un modèle non normal.
18. Cette approximation fléchit dans le cas des centiles extrêmes, c.-à-d. $P > 0{,}90$ ou $P < 0{,}10$.
19. Les normes non-paramétriques discutées dans cette section font abstraction de tout modèle courant, notamment le modèle normal : le modèle de calcul, et seul modèle de référence, serait la loi uniforme (voir plus loin dans le texte). Toutefois, ces normes s'appliquent à une variable continue, si possible à une mesure, pour laquelle idéalement chaque élément de la population de référence a une valeur distincte (p. ex., poids, taille, force physique, glycémie, etc.). D'autres cas approchants sont aussi plausibles, tels les scores d'échelles psychométriques (non normaux) à dispersion suffisante. Les échelles ordinales de type Likert (p. ex., scores de 1, 2, ..., 7) ou les variables présentant un répertoire sévèrement limité de valeurs ne conviennent pas à ces normes.
20. Dans ce contexte non-paramétrique, le modèle de distribution de X est supposé inconnu, tout comme sa densité $f(X)$ et sa fonction de répartition ou intégrale $F(X)$. Tout se passe comme si les valeurs de X étaient traduites en leurs quantiles équivalents, selon $U \leftarrow F^{-1}(X)$, où $0 \leq U \leq 1$, la population des quantiles ayant par définition une distribution uniforme aux propriétés bien établies. Sous cette forme, hormis l'obligation de la continuité des valeurs, toute référence à l'axe de mesure X original est perdue.
21. Les tables G1 et G2 présentent les rangs r (bas) et s (haut), leur relation étant $s = n + 1 - r$. La norme (supérieure) applicable ici correspond à un rang « haut ».
22. À cet effet, il est intéressant et intrigant de juxtaposer les statistiques d'ordre uniforme $U[r{:}n]$ et normale $N[r{:}n]$ pour $r = 143{,}45$ et $n = 150$, l'asymétrie de la première, avec $\gamma_1 \approx -0{,}65$, s'opposant à celle de la seconde, $\gamma_1 \approx 0{,}18$.

23. Les moments de la distribution de e sont $\mu_e = 0$, $\sigma_e^2 = (1 - R)\cdot\sigma_x^2$, $\gamma_1 = \gamma_2 = 0$.
24. La littérature spécialisée propose différents indices et procédures attestant de la normalité d'une distribution empirique (mesures des moments empiriques g_1 et g_2, test de Shapiro-Wilk, etc.). En cas de doute, mieux vaut vérifier !
25. Les deux variables, e et X, étant ici strictement indépendantes (ne relevant pas du même échantillonnage), leur variance conjointe est la somme de leurs variances respectives.
26. Cet intervalle, $V \pm \sigma_e$, contient environ 68 % des valeurs probables (selon le modèle normal), pour ce cas particulier. Les publications en recherche scientifique utilisent généralement un intervalle à 95 %, soit $\mu \pm 1{,}960\cdot\sigma_e$.
27. Même s'il est traité sans référence à un axe de mesure proprement dit, cet exemple tombe dans la catégorie $\sigma_e > 0$, la raison étant qu'il y a incertitude sur la réponse (qui suppose un choix au hasard) et qu'il est posé comme tel.
28. Une approche plus « équitable » eût été de pondérer chaque item en lui accordant un pointage égal à la réciproque de sa probabilité, c.-à-d. égal au nombre d'options qu'il présente.
29. Probabilités évaluées par la méthode d'échantillonnage de Monte-Carlo.
30. La taille d'échantillon de l'exemple étant $n = 450$, nous avons utilisé les normes λ calculées pour $n = 500$. Un calcul plus précis (à plusieurs décimales) fournit le seuil 57,424 (au lieu de 57,43). Le recours aux tables F7a et F8a (*op. cit.*) pour $n = 250$ ($\lambda = 0{,}025$ et 0,567) nous permet d'appliquer la formule d'interpolation sur n (F17), fournissant pour $n = 450$ $\hat{\lambda} = 0{,}1802$ (au lieu de 0,1805) et un seuil 'exact' de 57,419. La réponse « CP = 57,42 » est très satisfaisante!

RÉFÉRENCES

Allaire, D. & Laurencelle, L. (1998). Comparaison Monte Carlo de la précision de six estimateurs de la variance d'erreur d'un instrument de mesure. *Lettres statistiques*, *10*, 27-50.

André, N., Loye, N. & Laurencelle, L. (2015). La validité psychométrique : un regard global sur le concept centenaire, sa genèse, ses avatars. *Mesure et évaluation en éducation*, *37*, 125-148. doi: 10.7202/1036330ar

Bertrand, R. & Blais, J.-G. (2004). *Modèles de mesure : l'apport de la théorie des réponses aux items.* Québec : Presses de l'Université du Québec.

Bosset, P. (2007). *Les fondements juridiques et l'évolution de l'obligation d'accommodement raisonnable*. Montréal : Commission des droits de la personne et des droits de la jeunesse. Repéré à www.cdpdj.qc.ca/Publications/accommodements_fondements_juridiques.pdf

Colombie-Britannique (Public Service Employee Relations Commission) c. British Columbia Government and Service Employees' Union (BCGSEU), [1999] 3 R.C.S. 3. Repéré à https://scc-csc.lexum.com/scc-csc/scc-csc/fr/item/1724/index.do

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed). Washington, DC: American Council on Education.

David, F. N. & Johnson, N. L. (1954). Statistical treatment of censored data: Part I – Fundamental formulae. *Biometrika*, *41*, 228-240. doi: 10.1093/biomet/41.1-2.228

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous univariate distributions* (2 vol.) (2nd ed.). New York: Wiley.

Johnson, N. L., Kotz, S. & Kemp, A. W. (1992). *Univariate discrete distributions*. New York: Wiley.

Laurencelle, L. (1993). La loi uniforme : propriétés et applications. *Lettres statistiques*, *9*, 1-23.

Laurencelle, L. (1998). *Théorie et techniques de la mesure instrumentale*. Québec : Presses de l'Université du Québec.

Laurencelle, L. (2015). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie I : L'imprécision échantillonnale des centiles. *Mesure et évaluation en éducation*, *38*, 87-110. doi: 10.7202/1036764ar

Laurencelle, L. (2016a). Une théorie des seuils psychométriques à double contrôle d'erreur – Partie II : L'erreur de mesure et le concept de norme sûre. *Mesure et évaluation en éducation*, *39*, 45-66. doi: 10.7202/1036705ar

Laurencelle, L. (2016b). *L'étalonnage et la décision psychométrique : exemples et tables* (2e éd.). Québec : Presses de l'Université du Québec.

Lessard-Hébert, M., Goyette, G. & Boutin, G. (1996). *La recherche qualitative : fondements et pratiques*. Montréal : Éditions Nouvelles.

Poisson, Y. (1992). *La recherche qualitative en éducation*. Québec : Presses de l'Université du Québec.

Trizano-Hermosilla, I. & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions : Congeneric and asymmetrical measurements. *Frontiers in Psychology, 7,* 769. doi: 10.3389/fpsyg.2016.00769

Appendice : Les calculs du modèle binomial

La distribution de probabilité $f(x)$ d'une variable binomiale $B(x \mid \pi, n)$ est $f(x) = {}_nC_x\pi^x(1 - \pi)^{n-x}$ pour $x = 0, 1, 2, \ldots, n$. Son espérance, ou moyenne, est $\mu = n{\cdot}p$, sa variance est $\sigma^2 = n{\cdot}\pi(1 - \pi)$, son asymétrie est $\gamma_1 = (1 - 2\pi)/\sigma$ et son aplatissement ou voussure est $\gamma_2 = [1 - 6\pi(1 - \pi)]/\sigma^2$. La variable transformée

(A1) $z = (x - n{\cdot}\pi \pm ½)/\sigma$

approche la loi normale standard selon que les indices de forme γ_1 et γ_2 avoisinent 0, soit $|\gamma_1| < 0{,}1$ et $|\gamma_2| < 0{,}1$. Pour cautionner cette approximation, on recommande généralement que $n \times \min\{\pi, (1 - \pi)\} \geq 10$. Par exemple, pour $n = 20$ et $\pi = 0{,}5$, on observe $\gamma_1 = 0$, $\gamma_2 = -0{,}100$; pour $n = 60$ et $\pi = 0{,}7$, $\gamma_1 \approx -0{,}113$ et $\gamma_2 \approx -0{,}021$. Dans la formule de calcul de z ci-dessus, la correction $\pm ½$ a pour fonction de réduire la valeur du numérateur vers 0.

De probabilité π constante et à n valeurs y_i indépendantes, la loi binomiale classique se voit être un cas particulier d'une loi plus générale, à π_i variables et à valeurs y_i corrélées, cette loi se développant soit en mixture de binomiales (binomiale de Poisson), soit en chaîne de Markov. Le modèle à π_i variables représente mieux que celui à π fixe les tests et examens scolaires typiques, mais on l'y a peu appliquée. Pour cette famille de lois, soit $\overline{\pi}$, la moyenne des n valeurs de π_i ; l'espérance et la variance de x sont:

(A2) $E(X) = n{\cdot}\overline{\pi}$; $\mathrm{var}(X) = n{\cdot}\overline{\pi}[(1 - \overline{\pi}) - \sigma^2(\pi_i)]$,

la variance étant donc plus faible que pour la loi à p fixe, comme c'est aussi le cas pour l'asymétrie. Le calcul exact de probabilité reste laborieux (par énumération complète de la combinatoire). Néanmoins, et si la probabilité moyenne $\overline{\pi}$ ne s'écarte pas trop de ½, l'application de l'approximation normale (et utilisant les valeurs paramétriques indiquées en A2) y supplée aisément. Voir aussi Johnson, Kotz et Kemp (1992).