Revue de psychoéducation

Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions
Why statistics is difficult to teach: A few considerations

Denis Cousineau and Bradley Harding

Volume 46, Number 2, 2017

URI: https://id.erudit.org/iderudit/1042257ar
DOI: https://doi.org/10.7202/1042257ar

See table of contents

Publisher(s)

Revue de Psychoéducation

ISSN

1713-1782 (print)
2371-6053 (digital)

Explore this journal

Cite this article

Cousineau, D. & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions. *Revue de psychoéducation*, *46*(2), 397–419. https://doi.org/10.7202/1042257ar

Article abstract

Statistics is a content matter notoriously difficult to teach to students of the social sciences. Statistical anxiety, a well-documented form of anxiety, is present from the beginning of the course and explains in part the learning difficulties experienced by these students. However, we believe that statistics anxiety is a consequence and not the source of these difficulties. To initiate a discussion on how statistical teaching could be improved in the social sciences, we consider herein three causes that might explain that anxiety. To initiate a discussion which could lead to improvements in the way statistics is taught, we examine herein three possible causes of this anxiety. First, human beings have very limited intuitions of randomness (as can be seen from erroneous conceptions from compulsive players) and consequently, distributions and sampling remains opaque concepts for many. Second, some statistical concepts are truly "meta-statistical" concepts in which one must conceive of statistics onto statistics. Finally, procedures for decision taking in contexts where information is limited and inaccurate are often poorly understood. In this text, we detail these three conceptual difficulties and propose recommendations to weaken their consequences. The propositions presented here still need to be validated with formal studies; this text hope to be a catalyst of discussions aiming at improving the teaching of quantitative methods and statistics.

Revue de psychoéducation
Volume 46, numéro 2, 2017, 397-419

Mesure et évaluation

Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? Quelques réflexions

Why statistics is difficult to teach: A few considerations

D. Cousineau[1]
B. Harding[1]

[1] Université d'Ottawa

Nous tenons à remercier vivement les nombreux collègues qui nous ont donnés leurs commentaires sur ce texte, en particulier André Achim, Dominic Beaulieu-Prévost, Sébastien Béland, Michael Cantinotti, Éric Frenette, Léon Harvey, et Daniel Lalande. Les discussions tenues aux colloques *Méthodes Quantitatives et Sciences Humaines* au fil des années ont aussi été très utiles; je souligne en particulier les apports de Rachad Antonius, Marcos Balbinotti, Nathalie Loye, Walter Marcontoni et Sophie Parent. Je tiens aussi à remercier certains de mes étudiants, dont Alexandra Turgeon et Alexandre Lafrenière. Cette recherche est appuyée en partie par des bourses de l'*Ontario Graduate Scholarship* et du *Conseil pour la Recherche en Sciences Naturelles et en Génie* accordée au second auteur et par une subvention du *Conseil pour la Recherche en Sciences Naturelles et en Génie* accordée au premier auteur.

Correspondance :

Denis Cousineau
École de psychologie
Université d'Ottawa
136 Jean-Jacques Lussier
Ottawa, Canada, K1N 6N5

Tél : (613) 562-5800 poste 7910

denis.cousineau@uottawa.ca

Résumé

Les statistiques sont une matière notoirement difficile à enseigner aux étudiants des sciences humaines. L'anxiété statistique, une forme d'anxiété bien documentée chez ces étudiants, est présente dès le début du cours et explique donc une partie des difficultés rencontrées par ces étudiants. Cependant, nous croyons que l'anxiété statistique est la conséquence plutôt que la source de ces difficultés. Pour enclencher une discussion qui, à terme, pourrait bonifier la façon d'enseigner les statistiques dans les sciences humaines, nous examinons ici trois causes qui pourraient expliquer cet anxiété. D'une part, les humains ont très peu d'intuition du hasard (comme le montrent les études sur les jeux d'argent dans lesquels des conceptions erronées sont dévoilées), et donc, les notions de distribution et d'échantillonnage restent des concepts opaques pour plusieurs. De plus, certains concepts statistiques reposent sur des raisonnements « méta-statistiques » dans lesquels il faut concevoir des statistiques sur des statistiques. Finalement, la notion de prise de décision dans un contexte où l'information est partielle et incertaine est souvent mal comprise. Dans ce texte, nous précisons ces trois difficultés et suggérons des recommandations pour en amoindrir les conséquences. Cependant, les pistes présentées ici nécessitent d'être validées par des études formelles; ce texte se veut avant tout un catalyseur de discussions visant l'amélioration de l'état de nos classes de méthodes quantitatives.

Mots-clés : **Éducation aux statistiques; fondements conceptuels; hasard; échantillonnage; méta-statistique.**

NDLR : Voilà un texte qui ne sera probablement pas lu par les cliniciens, mais dont le contenu devrait en rassurer plusieurs sur les difficultés qu'ils ont eu à jongler avec les statistiques. Par contre, si vous connaissez des professeurs de statistiques, offrez leur cet article en cadeau.

Abstract

Statistics is a content matter notoriously difficult to teach to students of the social sciences. Statistical anxiety, a well-documented form of anxiety, is present from the beginning of the course and explains in part the learning difficulties experienced by these students. However, we believe that statistics anxiety is a consequence and not the source of these difficulties. To initiate a discussion on how statistical teaching could be improved in the social sciences, we consider herein three causes that might explain that anxiety. To initiate a discussion which could lead to improvements in the way statistics is taught, we examine herein three possible causes of this anxiety. First, human beings have very limited intuitions of randomness (as can be seen from erroneous conceptions from compulsive players) and consequently, distributions and sampling remains opaque concepts for many. Second, some statistical concepts are truly "meta-statistical" concepts in which one must conceive of statistics onto statistics. Finally, procedures for decision taking in contexts where information is limited and inaccurate are often poorly understood. In this text, we detail these three conceptual difficulties and propose recommendations to weaken their consequences. The propositions presented here still need to be validated with formal studies; this text hope to be a catalyst of discussions aiming at improving the teaching of quantitative methods and statistics.

Keywords: Statistics education; concepts grounding; randomness; sampling, meta-statistic.

Introduction

Les statistiques et les méthodes quantitatives sont une matière obligatoire dans le cursus universitaire de plusieurs programmes en sciences humaines (tels que la sociologie, la criminologie, la psychologie, la psychoéducation et l'éducation). Bien que tous les cours universitaires soient exigeants et demandent un investissement de la part des étudiants, les cours de statistiques sont souvent vues comme une « bête noire » dans leurs cheminements universitaires. En fait, le cours de statistiques génère plus d'anxiété chez les étudiants en sciences humaines que tous leurs autres cours (Baloglu, 2003), un phénomène bien documenté qui a même reçu un nom spécifique, l'*anxiété statistique* (Benson, 1989) pour lequel il existe quelques instruments de mesures (Vigil-Colet, Lorenzo-Seva et Condon, 2008). Tout comme l'anxiété d'évaluation et l'anxiété généralisée, l'anxiété statistique s'exprime parfois par des réactions émotionnelles qui nuisent à l'apprentissage (Onwuegbuzie et Daley, 1999) et qui génèrent des sentiments négatifs face à la matière et à son enseignant (Onwuegbuzie et Seaman, 1995). D'ailleurs, dans une étude sur les étudiants québécois, l'anxiété statistique corrèle négativement avec la note obtenue (Cantinotti, Lalande, Ferlatte et Cousineau, 2017). Finalement, cette forme d'anxiété se distingue des autres types d'anxiété, ne se manifestant que dans les cours de statistiques (Finney et Schraw, 2003; Furnham et Chamorro-Premuzic, 2004; Hair et Hampson, 2006).

Évidemment, il est souhaitable de réduire cette anxiété statistique. Par exemple, (i) en donnant rapidement des rétroactions sur les apprentissages; (ii) en donnant des évaluations facultatives tel qu'un tout premier devoir qui n'est pas contributoire à la note finale; (iii) en donnant des activités en classe (Marson, 2007). Malgré tout, à notre avis, ces solutions (et d'autres, voir le GAISE College Report de l'American Statistical Association et al., 2005) ne vont pas à la source du problème. Nous croyons que l'anxiété statistique est une conséquence de la matière à apprendre, et la réduire par les techniques de gestion de classes ci-dessus ne touche pas le fond du problème. Nous croyons que pour attaquer le problème à sa source, il faut identifier les concepts « qui coincent », des concepts obstacles nécessaires pour comprendre les statistiques, sans être forcément acquis par les étudiants.

Dans ce qui suit, nous avons cherché à identifier certains concepts obstacles dans le premier cours de statistiques de niveau universitaire. Notre cible est le programme de psychologie même si, selon toute vraisemblance, les mêmes obstacles se retrouvent dans la plus part des programmes de sciences humaines et sociales. Notre réflexion n'est pas le fruit d'une démarche empirique, mais plutôt d'une mise en forme de nos expériences d'enseignement au cours des dernières années (15 ans pour l'auteur sénior). Il faut donc voir ces idées comme des hypothèses et non des conclusions.

Premièrement, nous examinons le concept de hasard. Ce concept est à la base des notions d'échantillonnage et de variabilité mais nous argumentons qu'il est peu compris par les étudiants. De plus, rares sont les cours de statistiques qui donnent l'occasion aux étudiants de voir le *hasard à l'œuvre*. Deuxièmement, nous examinons la notion de méta-statistiques (des statistiques sur des statistiques), dont la plus connue est l'erreur type de la moyenne. Ces méta-statistiques sont souvent mal comprises car la notion de ré-échantillonnage sur laquelle sont construites ces méta-statistiques est trop abstraite pour s'en faire une représentation. Finalement, nous examinons brièvement la notion de prise de décision et les fausses croyances qui l'accompagnent. À la fin de chacune de ces sections, nous proposons des recommandations qui pourraient améliorer l'enseignement de ces concepts. Pour conclure, nous questionnons la place des mathématiques (l'autre bête noire de plusieurs étudiants) dans l'enseignement des statistiques.

Fondamentalement, notre but n'est pas de réduire l'anxiété statistique directement, mais plutôt de changer la façon dont la matière est enseignée. En nous assurant que les étudiants ont les connaissances prérequises et en abordant ces connaissances plus tôt dans l'enseignement (avec des manipulations concrètes et appliquées), nous croyons que l'anxiété statistique devrait diminuer d'elle-même.

Cadre de la réflexion

Dans l'enseignement des mathématiques et de la physique, l'approche pédagogique à privilégier consiste à identifier les concepts nécessaires à l'acquisition d'une nouvelle compréhension (Allen, 2006; Halloun et Hestenes, 1985). Les théories du *constructivisme*, qui ont leurs racines notamment chez le philosophe Kant, postulent qu'il est impossible d'apprendre *ex nihilo* – de commencer totalement à zéro. Selon cette approche, une nouvelle habileté est toujours le raffinement ou la spécialisation d'habiletés antérieures. On devrait donc pouvoir, pour l'apprentissage de chaque nouveau concept, être capable d'identifier les concepts préalables.

Par exemple, dans le domaine des mathématiques, pour résoudre un problème tel que :

Que vaut 3 + 5 × 4?

il faut, au préalable, avoir trois compétences (George et Robitzsch, 2015), soit

i) être capable de multiplier des petits nombres (comme 5 et 4)
ii) être capable d'additionner des petits nombres (comme 3 et 20)
iii) savoir qu'il faut donner priorités aux multiplications sur les additions.

Le problème en statistiques revient donc à identifier les concepts et les habiletés nécessaires à la compréhension d'énoncés plus abstraits telles que :

La moyenne de la population se trouve dans l'intervalle [95.0, 105.0] avec une probabilité de 95 %.

Cet énoncé n'évoque pas de façon évidente une liste de compétences requises, contrairement à l'énoncé tiré des mathématiques ci-dessus. Pourtant, l'intervalle de confiance est une connaissance à acquérir dans tous les cours d'introduction aux statistiques de niveau universitaire, que ce soit en sciences humaines ou en sciences naturelles. Le fait qu'on ne puisse pas énumérer les concepts préalables à cet énoncé est la source d'un incroyable défi pour celui ou celle qui souhaite enseigner cette matière.

L'intuition du hasard

Le hasard est superbement difficile à définir (voir Nickerson, 2002). Il est donc hors de question de donner une définition formelle du hasard aux apprenants, surtout à ceux sans expérience en méthodes quantitatives. Cependant, qu'ils acquièrent une meilleure intuition de celui-ci est possible. Or, bien que le hasard soit omniprésent, beaucoup d'étudiants ont très peu de conceptions sur celui-ci et surtout, ils sont très mal outillés pour en parler. Les expressions du langage populaire reflètent bien cette méconception du hasard. Par exemple, l'adage « Jamais deux sans trois » parle de certitude alors que « p't'être bien qu'oui, p't'être bien que non » laisse sous-entendre que tout est binaire et même, que les deux alternatives ont autant de chance de se produire (équiprobabilité).

Les cours de statistiques, tels qu'ils sont souvent enseignés, n'aident probablement pas à raffiner le concept de hasard. Quelques exemples : (i) Lorsqu'interrogés au sujet d'une population dont la moyenne est de 100, le professeur demande à une classe quelle est la moyenne d'un échantillon tiré de cette population : nombreux sont ceux qui répondent 100. En fait, la seule réponse légitime ici est « à peu près 100 »; la seule façon d'être plus précis est de donner un intervalle de valeur avec une probabilité d'être dans cet intervalle, comme par exemple, « Entre 95 et 105 avec 95 % de probabilité ». En lisant ceci, on pourrait penser que les étudiants étaient novices en la matière. Or, on a obtenu cette même réponse chez des étudiants des cycles supérieurs à leur quatrième cours de statistiques, ce qui suggère que le concept de hasard échantillonnal n'a guère évolué au fil des cours reçus. (ii) De la même façon, nombreux sont les étudiants qui pensent qu'une population aura toujours forcément une répartition qui suit la distribution normale (une courbe en cloche). Cette distribution n'est pourtant pas universelle à toutes les populations. (iii) En donnant trop de poids aux formules, plusieurs étudiants n'acquièrent pas l'intuition du concept enseigné. Par exemple, plusieurs étudiants concluent de façon erronée qu'avec un plus grand échantillon, la variance sera plus petite puisqu'on divise par $n - 1$; ils oublient que l'on additionne aussi un plus grand nombre de différences aux carrées.

Pour aller plus loin, Konold (1995) a suggéré d'examiner le hasard tel qu'il se manifeste chez les étudiants de sciences humaines. Voici une adaptation libre de son « Weather problem » ou « Problème de la Météo » :

> **Problème de la météo** : Le centre météorologique du Canada cherche à déterminer l'exactitude de ses prédictions météorologiques. Il a fouillé dans ses archives des années 2004 à 2014 pour trouver toutes les journées où la prédiction pour le lendemain indiquait 70 % de chances de pluie. Puis il a comparé ces prédictions avec le temps réellement observé et s'il a plu ou pas lors de ces journées.
>
> La prédiction de 70 % de pluie peut être considérée comme précise s'il a plu
>
> a) lors de 95 % à 100 % de ces journées;
> b) lors de 85 % à 95 % de ces journées;
> c) lors de 75 % à 85 % de ces journées;
> d) lors de 65 % à 75 % de ces journées;
> e) lors de 55 % à 65 % de ces journées.

Dans cette étude, les réponses obtenues auprès de près de 200 répondants, étudiants universitaires débutant une formation en statistiques, indique qu'un tiers des répondants donne la réponse adéquate (d). Un second tiers répond (a). Le dernier tiers des répondants sont répartis uniformément parmi les trois autres choix de réponse (b, c et e).

Konold (1995) interprète ces résultats en supposant que l'étudiant cherche à visualiser une situation : or, demain, il pleuvra ou il ne pleuvra pas en un lieu donné. D'où le fait que le novice pense le résultat en termes binaires et est porté à forcer une dichotomie : une probabilité inférieure à 50 % signifie *certainement pas de pluie*, et une probabilité supérieure à 50 % signifie *sûrement de la pluie*. Plusieurs rajouteraient une zone grise plus ou moins étroite autour de 50 %, où la probabilité signifie *je n'en sais strictement rien*.

On pourrait faire l'hypothèse qu'il s'agit d'un problème avec la notion de probabilité et que les étudiants n'ont pas une formation adéquate avec ce terme technique. Cependant, telle n'est pas notre impression. Pour valider notre impression, une expérience similaire a été montée et testée en 2015 auprès de 45 étudiants en psychologie dans une classe d'introduction aux statistiques de l'Université d'Ottawa :

Problème inverse de la météo : Le centre météorologique du Canada observe un ensemble de paramètres (humidité, vent, etc.) afin de prédire la météo du lendemain. Or, pour demain, les ordinateurs n’arrivent pas à prédire s’il pleuvra ou non. En consultant les archives des années 2004 à 2014, un météorologiste trouve que 7 fois sur 10, avec des conditions identiques à celles présentement relevées, il a plu le lendemain.

Selon vous, que devrait annoncer le centre météorologique du Canada pour demain :

a) un risque de 95 % à 100 % de pluie?
b) Un risque de 85 % à 95 % de pluie?
c) un risque de 75 % à 85 % de pluie?
d) Un risque de 65 % à 75 % de pluie?
e) un risque de 55 % à 65 % de pluie?

La majorité de la classe a donné la bonne réponse (d). Ce problème est l’opposé du problème précédent. Ces deux items, pris conjointement, révèlent donc un problème dans le concept même de hasard, mais pas dans celui de probabilité.

Recommandations

1) Pour favoriser l’intégration de la notion de hasard, il faut pouvoir jouer avec celui-ci. Une première approche consiste à générer des évènements aléatoires (par exemple, lancer des pièces de monnaie, lancer des dés, ouvrir des sacs de Smarties), puis de noter les occurrences d’évènements comme l’ont suggéré plusieurs (Dunn, 2005; Küchenhoff, 2008;). Un autre outil pertinent serait la « Machine de Galton », un appareil qui laisse tomber des billes de verre sur un plan incliné où de nombreuses rangées de clous sont disposées. Ce dispositif permet d’observer un phénomène qui répond à la distribution binomiale. Lorsque la bille tombe, il n’existe que deux alternatives pour chaque rangée de clous, un bond à gauche ou un bond à droite.

En donnant l’occasion de jouer avec de tels appareils, les étudiants devraient être en mesure de voir le hasard à l’œuvre. Cependant, malgré l’utilité pédagogique d’outils tactiles et visuels, notre expérience suggère qu’avec des dispositifs physiques nous sommes limités dans le nombre de reproductions qui peuvent être réalisées avant que les étudiants ne commencent à s’ennuyer

2) Par la suite, il faut que l’étudiant puisse générer rapidement un grand nombre d’évènements aléatoires, les effacer et recommencer aussi souvent qu’il le veut. Pour cela, des générateurs d’échantillons aléatoires sur ordinateur se révèlent utiles. La plupart des logiciels d’analyses statistiques en sont dotés, que ce soit *SPSS* (*GRD*, Harding et Cousineau, 2014, 2015), *Excel* (les compléments *ToolPak*) ou encore *R* (les fonctions *r** où * est le nom d’une distribution théorique), etc. En quelques lignes de codes (ou clics de la souris), l’étudiant peut générer un échantillon complètement aléatoire aussi grand que nécessaire, aussi variable que souhaité et tiré une distribution théorique de leurs choix. Nos essais passés, où l’étudiant devait réaliser des exercices avec un tel outil, indiquent qu’ils restent concentrés et ne montrent pas de signes d'ennuis, contrairement aux activités avec des lancers de pièces de monnaie ou de dés. Idéalement, les exercices en question doivent simuler des évènements concrets. Cette approche n'est pas nouvelle puisque certains auteurs enseignent les statistiques uniquement avec des simulations de ce type (Lane, n. d.). Cette approche donne l'opportunité à l'étudiant de générer des échantillons aléatoires tirés de la population de leurs choix afin de voir le hasard à l'œuvre. Ils deviennent acteurs actifs ce qui peut contrecarrer le sentiment d'impuissance perçu par certains dans un cours traditionnel.

À titre d'exemple, un exercice visait à examiner les temps de livraison de pizzas venant de chez Gino. Dans cet exercice, l'outil GRD devenait un répertoire des « Gino's Recent Delivery times ». Les paramètres d'une distribution normale était mis à μ = 45 minutes, σ = 15 minutes. Ainsi, la majorité (75 %) des pizzas sont livrées en moins de 55 minutes. Le but de l'exercice est de trouver après combien de temps il faut rappeler la pizzéria si la pizza n'a pas encore été livrée. La réponse nécessite de définir un temps critique (un seuil) au-delà duquel le temps est exceptionnellement long à l'aide d'échantillons de temps de livraisons (par exemple, 95 % des temps de livraisons proposés par GRD pourraient être inférieurs à 70 minutes).

3) représentation visuelle Pour concevoir les résultats du hasard, l'étudiant doit l'expérimenter et le représenter. À cet effet, nous croyons que le graphique des fréquences (dans l'univarié) et le graphique de dispersion (dans le bivarié, voir le trivarié) sont des outils essentiels. Avec le graphique des fréquences sous les yeux, l'étudiant doit être invité à visualiser approximativement la tendance centrale (la moyenne, la médiane ou encore le mode), la dispersion (l'étendue des données, l'étendue interquartile, ou encore l'écart type), et finalement, le degré d'asymétrie (on peut sans doute omettre l'aplatissement, qui a très peu d'applications pratiques). À chaque fois que l'étudiant reçoit un échantillon, il devrait l'illustrer et essayer d'estimer visuellement la valeur de ces trois caractéristiques, avant de vérifier ces estimés avec des méthodes formelles. Ainsi, il pourra développer une intuition de ces grandeurs. L'exemple de la Figure 1 illustre un échantillon et ce qu'on peut déduire visuellement de celui-ci.

D'autres auteurs (tel que Lane, n. d.) vont privilégier le graphique à moustache (*Box plot*). Certaines informations sont plus difficiles à lire sur cette représentation (notamment l'asymétrie), mais il est par contre plus aisé de mettre plusieurs conditions sur le même graphique. Une alternative qui regroupe le meilleur des deux mondes est peut-être le graphique en violon (Marmolejo-Ramos et Matsunaga, 2009).

En employant une approche par graphiques, l'étudiant doit être en mesure de réaliser que chaque personne mesurée contribue un point au graphique. Utilisez les histogrammes plutôt qu'une courbe continue et indiquez que chaque personne observée est une brique Lego 2 × 2 qui est empilé sur les autres (recommandation 2 du GAISE College Report, American Statistical Association et al., 2005).

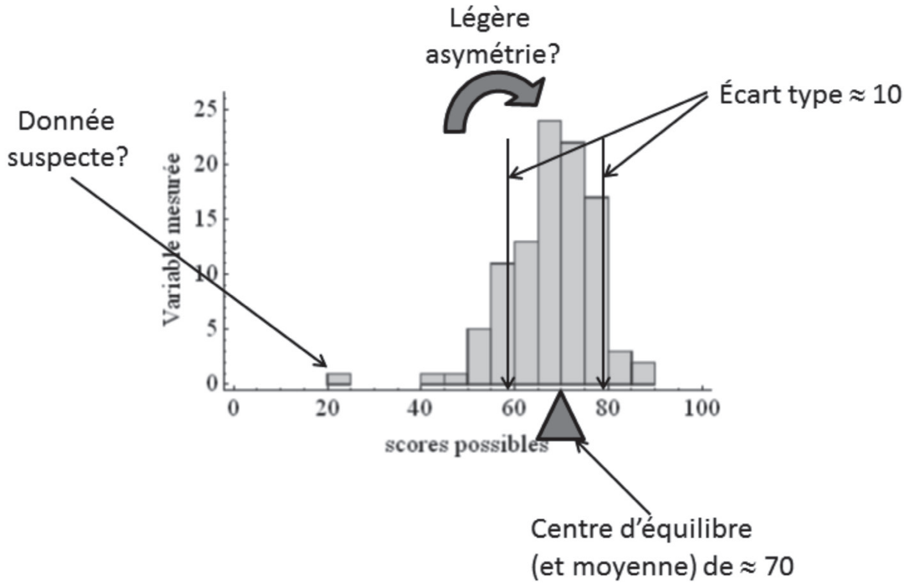

Figure 1. Exemple de graphique de distribution ainsi que les observations approximatives qui peuvent en être tirées. La moyenne est le centre d'équilibre si les histogrammes étaient disposés sur un plateau; l'écart type est approximativement la moitié de l'étendue couvrant la majorité des données. On voit aussi sur ce graphique qu'il y a une donnée suspecte.

En combinant cette recommandation avec les deux précédentes, l'étudiant peut non seulement voir le hasard en action mais aussi réaliser que hasard n'équivaut pas à « complétement aléatoire » surtout s'il simule de grands échantillons.

4) La variabilité est sans doute le concept central de l'échantillonnage. En effet, chaque nouvelle mesure est différente de la précédente, un concept portant plusieurs noms : *hasard échantillonnal, erreur de mesure, différences individuelles*, etc. On peut aussi parler de *bruit dans les données*, de *fluctuation naturelle*, de *dispersion*, d'*hétérogénéité*, etc. Les multiples synonymes ici sont essentiels car il faut accrocher la notion de variabilité à des concepts préalables. Par contre, il faut faire attention de ne pas submerger les étudiants de synonymes et adopter rapidement une définition précise (l'écart type ou la variance).

L'étudiant devrait être en mesure de développer l'intuition que les données sont généralement peu éloignées de la tendance centrale (si on ignore les populations multimodales). L'écart type doit devenir pour l'étudiant l'écart typique, et réaliser que la majorité (les deux tiers environ) des scores est typiquement à un écart type de la tendance centrale. On peut aussi aller plus loin et inviter l'étudiant à concevoir que presque tous les scores sont à moins de deux écarts types de la tendance centrale. Il s'agit d'heuristiques, mais qui ont l'avantage d'introduire la notion de fourchette de valeurs et d'intervalle probable (un écart type) ou très probable (deux écarts types) pour un futur score.

Dans la Figure 1, l'écart type semble être à peu près de 10 et presque toutes les données sont à plus ou moins 20 de la moyenne. La variance (qui vaut à peu près 100 dans la Figure 1) est dans cette représentation peu intuitive.

Les méta-statistiques

L'écart type peut être conceptualisé comme une prédiction sur une future observation : « si on prend une personne au hasard, elle a de grandes chances de se trouver à plus ou moins un écart type de la moyenne de son groupe ». Dans ce contexte, il est aisé de concevoir ce qu'est une « future observation » (on peut aussi dire une *reproduction* de la mesure). Cependant, pour saisir le concept plus avancé d'erreur type, une méta-statistique très connue, il faut passer à un stade plus abstrait. L'erreur type est une statistique descriptive sur une collection de statistiques descriptives (voir Harding, Tremblay et Cousineau, 2014, 2015, pour une revue d'un grand nombre d'erreurs types). Nous nous concentrons ici sur l'erreur type de la moyenne. L'erreur type de la moyenne donne la dispersion des moyennes, si on procédait à la mesure de plusieurs moyennes sur plusieurs échantillons. Cependant, il est difficile de maintenir ce genre de concept de second ordre car il faut concevoir des milliers d'échantillons contenant chacun de nombreuses mesures. Formellement, l'erreur type est l'écart type de moyenne, ce qui permet de dire que « l'erreur type est aux moyennes ce que l'écart type est aux individus ».

L'erreur type de la moyenne est un concept central. L'erreur type permet 1) de définir une fourchette de valeurs où devrait se trouver une moyenne d'échantillon selon l'hypothèse envisagée (utilisée dans les tests d'hypothèse nulle); 2) de définir une fourchette de valeurs où devrait se trouver la moyenne de la population selon les données disponibles (utilisée dans les intervalles de confiance), et 3) d'évaluer la précision d'un échantillon (une information complémentaire à la taille d'effet). En fait, la grandeur d'effet (brute ou standardisée), l'erreur type et la taille de l'échantillon sont les trois seules informations nécessaires pour la très vaste majorité des procédures statistiques et sont les informations recherchées lors d'une méta-analyse.

Notons au passage que l'erreur type ne se limite pas seulement à l'erreur type de la moyenne. On peut aussi souhaiter connaître l'erreur type de la médiane, l'erreur type du mode (une quantité qu'il n'est pas possible de chiffrer avec une formule), mais aussi, et là, c'est plus abstrait, l'erreur type de l'écart type et de l'asymétrie (voir Harding, Tremblay et Cousineau, 2014 pour les formules).

Une façon d'enseigner le concept d'erreur type est d'inviter les apprenants à concevoir la notion de *reproduction*, i.e., une situation dans laquelle de nouveaux échantillons sont colligés. Dans cette situation, les différentes moyennes obtenues peuvent être visualisées avec un graphique de dispersion.

Recommandations

1) Le ré-échantillonnage est le terme technique pour parler de reproduction. Pour se familiariser avec le ré-échantillonnage, il faut que l'étudiant puisse obtenir un échantillon, puis en obtenir un nouveau dans les mêmes conditions, puis encore un autre, etc. Encore ici, un outil informatique est avantageux, puisqu'il peut produire les échantillons rapidement.

Après avoir observé un grand nombre d'échantillons indépendants, l'étudiant devra être en mesure d'observer que la statistique d'intérêt (par exemple, la moyenne de l'échantillon) fluctue, mais que le degré de fluctuation est moindre (a) lorsque les échantillons sont grands plutôt que petits; (b) lorsque les individus composant la population simulée sont très homogènes plutôt qu'hétérogènes. Une formule compacte résume ces deux facteurs :

> L'erreur type est proportionnelle à l'écart type des individus et inversement proportionnelle à la taille de l'échantillon, inversement proportionnelle à la racine carrée de la taille de l'échantillon pour être exact.

2) Le sous-échantillonnage est aussi un outil qui peut être utilisé. Plusieurs sources mentionnent l'importance de faire travailler les étudiants sur des données réelles reliées à leur domaine d'étude plutôt que sur des données simulées (voir recommandation 2 du GAISE College Report, American Statistical Association et al., 2005). Utiliser des données réelles permet d'explorer une problématique concrète et d'inviter les étudiants à poser eux-mêmes les questions pour lesquelles ils souhaitent une réponse, sans connaître la réponse d'avance.

Une façon d'exploiter des données réelles est d'effectuer du sous-échantillonnage. De cette façon, chaque étudiant est exposé à une facette légèrement différente de la réalité contenue dans l'échantillon complet. Les étudiants peuvent ainsi voir que les statistiques extraites de leur portion des données diffèrent de celles de leurs voisins. Cependant, comme les étudiants ont été familiarisés avec le hasard (voir recommandations de la section précédente), ils devraient aussi comprendre que les statistiques extraites ne peuvent guère différer entre elles puisqu'elles proviennent toutes du même bassin de données.

Afin que les étudiants puissent pratiquer le sous-échantillonnage, il est possible d'installer ou coder un module de sous-échantillonnage (le module payant *Bootstrap* pour *SPSS*; une alternative gratuite à ce dernier, *GSD* pour *SPSS*; Harding et Cousineau, 2016; le module B*oot* dans *R*, ou encore le module *ToolPak* dans *Excel*) qui permet de créer un ou plusieurs sous-échantillons de n'importe quelle taille à partir de données existantes. Ces outils offrent l'occasion de pratiquer les méta-statistiques et de susciter des intuitions par rapport à l'échantillonnage et la validité d'un échantillon par rapport à un autre.

3) (représentation visuelle) Pour représenter l'erreur type, la meilleure option reste peut-être le graphique de moyennes avec des barres d'erreur. Si la longueur des barres d'erreur est donnée par l'erreur type, la moyenne d'une future reproduction se situe probablement à l'intérieur de cette fourchette; si la longueur des barres d'erreur est de deux erreurs types, la moyenne d'une future reproduction se situe *très* probablement à l'intérieur de cette fourchette.

La prise de décision en présence de données contenant du hasard

Plusieurs cours de méthodes quantitatives au premier cycle universitaire culminent avec les tests d'hypothèses comme le test *t*. Un tel test sur un seul groupe s'exprime par cette formule très compacte et trop absconse :

$$\text{Rejet de } H_0 \text{ si } \frac{\left|\overline{\mathbf{X}} - \mu_0\right|}{ET_{\overline{\mathbf{X}}}} > t_c \qquad (1)$$

où $H_0 : \mu = \mu_0$ est une hypothèse, sur la moyenne de la population étudiée, μ_0 est la moyenne supposée, $\overline{\mathbf{X}}$ est la moyenne de l'échantillon, $ET_{\overline{\mathbf{X}}}$ est l'erreur type de la moyenne, et t_c est une valeur critique lue à partir d'une table de Student.

Cette formule, que plusieurs étudiants apprennent par cœur, est cependant sujette à un très grand nombre d'interprétations erronées qui se retrouvent jusque dans la documentation spécialisée. Par exemple, un grand nombre de commentateurs parlent d'une hypothèse alternative, souvent notée $H_1 : \mu \neq \mu_0$ (erreur que l'on retrouve par exemple dans Loftus, 1996; Wagenmakers, 2007). Or, quand on décortique le test *t* par exemple, on voit qu'il est basé sur une comparaison entre un modèle où on postule que la moyenne vaut μ_0 et un modèle où on postule que la moyenne est la moyenne observée (Cousineau et Allan, 2015, Appendice 1, Listing 3). Certes, le second modèle est toujours le meilleur modèle; le test *t* essaye de chiffrer de combien l'adéquation du premier modèle est inférieure à celle du second modèle, et si cette réduction est significative. Si on voulait être exact, il faudrait donc écrire l'hypothèse alternative ainsi : $H_1 : \mu = \text{moyenne observée}$. Le rejet de l'hypothèse nulle indique qu'il est nécessaire d'utiliser la moyenne observée pour décrire les données.

Autre interprétation erronée : on entend souvent parler d'*acceptation de l'hypothèse nulle.* Cohen a à de nombreuses occasions (1969, 1990) tenté d'alerté l'attention de la communauté sur cette erreur conceptuelle, une situation qui n'a guère changée (Lecoutre, Poitevineau, et Lecoutre, 2003). Finalement, en faisant un test d'hypothèse nulle, on obtient aussi une valeur *p*. Or, cette quantité est souvent interprétée à tort comme la probabilité de l'hypothèse nulle étant donné les données observées, alors qu'il s'agit de l'inverse : la probabilité, sous cette hypothèse nulle, des données observées (Béland, Cousineau et Loye, 2016; Cohen, 1994; Rozeboom, 1960) .

Nous suggérons ici une interprétation qui devrait faciliter la compréhension de H_0 : avant de débuter une collecte de données, le chercheur se positionne dans un état d'ignorance. Il exprime son état d'ignorance en énonçant un état possible de la population qu'il désire étudier (l'hypothèse nulle). Par exemple, s'il examine la performance scolaire à la suite d'une intervention dont l'efficacité est inconnue, il suppose que la population est inchangée par cette intervention. Par la suite, il part en quête de données pour tester cette hypothèse. Un résultat probant indique qu'une moyenne telle que celle obtenue a peu de chance d'être observée si l'hypothèse nulle était fondée. Dans ce cas, le chercheur peut choisir de quitter l'état d'ignorance et de se commettre en disant que la moyenne est trop différente de ce qui est attendu si l'intervention est sans effet. *A contrario*, si les données ne sont pas probantes, le chercheur reste dans l'état d'ignorance qui le caractérisait avant qu'il n'entreprenne l'étude.

Les expressions « Rejet » et « Non rejet » doivent donc être comprises comme « Les données sont concluantes à l'encontre de H_0 » et « Les données sont non-concluantes » ou encore « Les données sont défavorables à l'hypothèse nulle » et « Les données sont muettes par rapport à l'hypothèse nulle ». Quant à l'hypothèse nulle, elle représente, pour reprendre Hoijtink (2012), la situation où les résultats n'auraient rien de spécial, le traitement n'apporterait rien de nouveau.

Une autre difficulté dans l'enseignement des tests statistiques tient à une confusion qui remonte à Aristote (Blanché, 1967). Le premier, il a proposé *l'axiome du tiers exclu* selon lequel une chose ou son contraire est forcément vraie. En vertu de cet axiome, d'où découle la grande majorité de la logique moderne formalisée par George Boole (1854), toutes les affirmations que l'on peut faire sur un système sont booléennes, c'est-à-dire binaire avec uniquement la valeur VRAI ou FAUX. Cette logique booléenne cependant exclut un troisième état, l'état d'ignorance, c'est-à-dire NI VRAI NI FAUX, qui est un état de non-décision, transitoire mais bien réel dans l'état de nos connaissances scientifiques.

Recommandations

1) Il peut être utile de décrire la prise de décision comme débutant dans un état d'ignorance. Cet état d'ignorance peut par la suite être exprimé comme une hypothèse nulle. Il faut à tout prix éviter l'expression « Hypothèse alternative » ou « contre-hypothèse » qui perpétue le principe du tiers exclu. Fisher (1925) et Cousineau(2009) sont d'avis que ces expressions relèvent d'une incompréhension des travaux de Neyman et Pearson (1933). Utiliser les expressions « Rejet probant » **ou** « Maintien du *statu quo* » pour décrire les résultats possibles d'un test d'hypothèse nulle pour souligner que l'on quitte l'état d'ignorance ou que l'on y reste.

2) Une façon de mieux saisir la logique derrière un test d'hypothèse nulle est d'écrire au long, dès le début du module d'enseignement, l'argument sous-tendant la décision de rejeter l'hypothèse. Par exemple, dans le cas d'un rejet probant, l'argument utilisé est :

> Selon H_0, la très grande majorité des moyennes sont dans un intervalle de longueur $ET_M \times t_c$ autour de μ_0
> Or, cette moyenne observée n'est pas dans l'intervalle
> Donc, cette moyenne observée ne fait pas partie de la très grande majorité des moyennes.

Autrement dit, une moyenne observée très éloignée de μ_0 est inhabituelle, bizarre, atypique, surprenante sous l'hypothèse nulle (cet argument utilise un intervalle $\mu_0 \pm ET_M \times t_c$ centré sur l'hypothèse nulle; cet intervalle n'est pas un intervalle de confiance car il n'est pas centré sur la moyenne observée). Suivant la procédure des tests d'hypothèse nulle, une moyenne inhabituelle (selon l'hypothèse nulle) entraîne cette conclusion :

Si H_0 est vraie, cette moyenne observée est inhabituelle
Cette moyenne observée ne peut pas être inhabituelle
Conséquemment, H_0 ne peut pas être vraie.

Ce deuxième syllogisme est erroné (car le second terme est faux) et reflète en réalité un acte de foi - la primauté de l'empirique sur le théorique (une foi héritée de la Renaissance). Il a l'avantage de souligner le fait qu'une moyenne observée est réellement inhabituelle que si on reproduit l'expérience. Ce second syllogisme est donc en réalité une invitation à reproduire pour vérification les expériences qui donnent des résultats inattendus (Cousineau, 2014; Pashler et Wagenmakers, 2012).

3) représentation visuelle Lorsque l'étudiant élabore une hypothèse, il devrait dessiner le graphique de moyennes le plus vraisemblable selon l'hypothèse nulle et anticiper la grandeur de la fourchette de valeurs dans laquelle le résultat devrait se trouver. Un tel graphique est illustré à la Figure 2a.

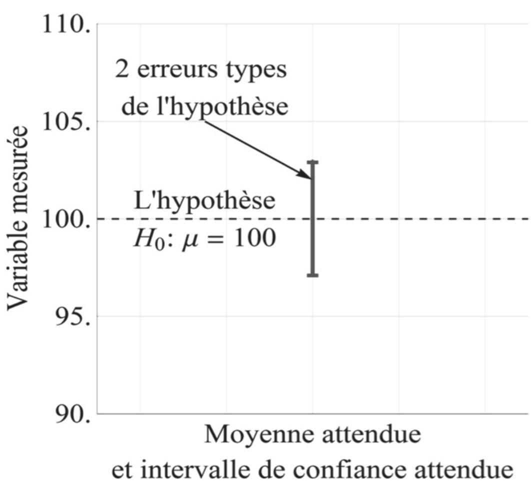

Figure 2a : Graphique anticipant le résultat possible d'un échantillon si l'hypothèse nulle H_0 : μ =100 est vraie. Ce graphique montre aussi un intervalle de ± 2 erreurs types autour de l'hypothèse nulle.

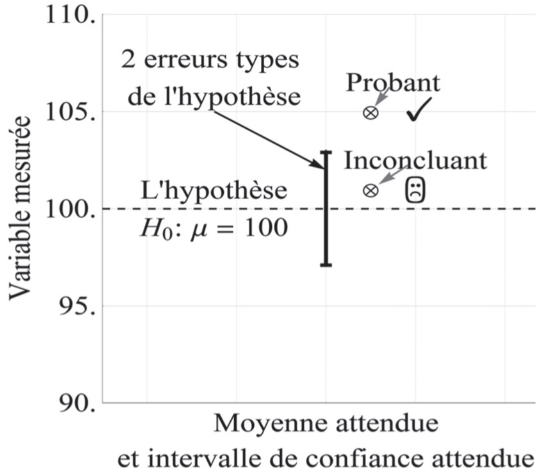

Figure 2b : Même graphique mais montrant en plus deux résultats possibles, un résultat de 106 (étiquetté « probant »), l'autre de 101 (étiquetté « Inconcluant »). Le second est très probable suivant l'hypothèse nulle, et ne mets pas en cause cette hypothèse; le premier est très inattendu selon l'hypothèse nulle (avec un *p* de .01) et remet en question la pertinence de l'hypothèse nulle.

En dessinant la dispersion de moyennes attendues, l'étudiant devrait être en mesure d'acquérir l'intuition que même si un échantillon semble inhabituel, une reproduction de l'étude devrait ramener la mesure à l'intérieur de la fourchette anticipée.

4) Ne pas hésiter à récrire les règles de décision en scores bruts plutôt qu'en scores *standardisés*. Par exemple, l'équation 1 devient

$$\begin{aligned}\text{Rejet de } H_0 \text{ si } \overline{\mathbf{X}} &> \mu_0 + t_c \times SE_{\mathbf{M}} \\ \text{ou si } \overline{\mathbf{X}} &< \mu_0 - t_c \times SE_{\mathbf{M}}\end{aligned} \tag{2a}$$

Par exemple, si l'hypothèse est que la moyenne des QI d'un groupe est de 100, que l'erreur type est de 1.5 points de QI, et que la taille d'échantillon est de 60 (d'où un t_c de 2.0 obtenue dans une table de Student), alors on obtient :

$$\begin{aligned}\text{Rejet de } H_0 \text{ si } \overline{\mathbf{X}} &> 100 + 2 \times 1.5 \text{ i.e.,} \quad \text{si } \overline{\mathbf{X}} > 103 \\ \text{ou si } \overline{\mathbf{X}} &< 100 - 2 \times 1.5 \text{ i.e., ou si } \overline{\mathbf{X}} < 97\end{aligned} \tag{2b}$$

Cette formulation a un premier avantage de s'exprimer avec l'échelle de mesure originale, et non en scores standardisés qui peuvent être situés directement sur le graphique anticipé ci-dessus. La Figure 2b indique deux scores possibles, un qui se trouve à l'intérieur de la fourchette, l'autre à l'extérieur. Cette règle est beaucoup plus concrète que la version de l'Équation 1 et risque moins d'être ritualisée que la version basée sur la valeur p :

$$\text{Rejet de } H_0 \text{ si } p < .05 \tag{3}$$

Cette formulation a aussi comme avantage que la taille des effets est directement visible. Si l'hypothèse de la Figure 2 avait une fourchette de valeurs allant de 99.9 à 100.1 et que le résultat observé avait été 100.2, la trivialité de la découverte aurait été flagrante (même si $p < .001$). Ce genre de « découverte » négligeable est possible lorsque l'échantillon est trop grand. Par exemple, pour des QI où l'écart type est de 15, un échantillon contenant 90 000 personnes serait suffisant pour obtenir un effet insignifiant (moins de 0.1 point de QI) mais significatif. De nos jours, on trouve fréquemment des études examinant des échantillons dépassant cet ordre de grandeur. Par exemple, Belmont and Marella (1973) ont examinés le QI de 386 336 jeunes hommes. Un examen informel de 100 articles contenant l'expression « large sample » dans leur résumé, publié en 2013 tel que référencés dans la base de données PsycINFO, montre que 6 articles ont examinés entre 10 000 et 100 000 sujets, et 4 avaient plus de 100 000 sujets. Le plus grand échantillon contenait des données d'un sondage pan-européen avait 1 465 630 adultes (Becchetti, Castriota, Corrado et Giachin Ricca, 2013).

Encore une fois, nous croyons qu'il est primordial de garder un lien avec les données brutes. En utilisant ce type de formulation où la prise de décision est un intervalle, on évite de multiplier les entités abstraites difficiles à saisir, et finalement, on évite l'utilisation de la valeur *p* qui donne à une valeur arbitraire beaucoup trop d'importance (Béland et al., 2016).

5) Décrire correctement la valeur *p* comme étant la probabilité des données observées si l'hypothèse était vraie (Cohen, 1994). La valeur *p* est une mesure qui indique à quel point la moyenne observée est inhabituelle si H_0 est vraie : plus le *p* est proche de zéro, plus la moyenne observée est inattendue et s'écarte de façon importante de la moyenne supposée dans l'énoncé de l'hypothèse nulle.

Il faut aider les étudiants à comprendre qu'une moyenne observée avec un faible *p* est possible et qu'ultimement, la décision de rejeter l'hypothèse nulle dépend, avant tout, des enjeux et des risques de cette décision. Éviter que le seuil de décision de .05 ne devienne un moyen de se commettre sans s'impliquer personnellement – de prendre une décision seulement car un livre de statistique nous dit qu'il faut qu'on le fasse.

En somme, nous suggérons de réduire considérablement l'importance du seuil de décision de .05 (Wasserstein et Lazar, 2016) et de façon générale, la place des tests statistiques dans un premier cours de statistique (Cumming, 2014). Il faut aussi prendre en compte l'importance des tailles d'effets et regarder non seulement la signification statistique mais aussi la signification clinique. Un savon qui accroît votre QI d'un millionième de point reste un savon.

L'estimé d'un effet et sa précision (un intervalle de confiance) apportent beaucoup d'informations et un graphique des moyennes avec les barres d'erreurs appropriées permettent d'avoir beaucoup plus de nuances dans les inférences. La Figure 3, inspirée de Wiens et Nilsson (sous presse) montre une gamme possible de résultats avec une zone bien visible où les résultats sont jugés sans signification clinique.

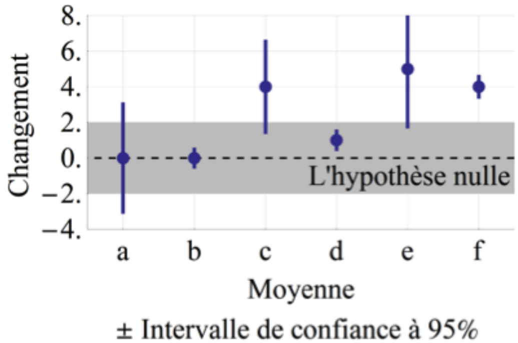

Figure 3. Six exemples de résultats possibles illustrés à l'aide de barres d'erreurs représentant un intervalle de confiance à 95 %, traduit de Wiens et Nilsson (sous presse). La zone grise indique une zone où le changement serait jugé négligeable. Le résultat (a) n'a pas de direction claire et a peu de précision; le résultat (b) n'a pas de direction claire, mais est très précis; le résultat (c) indique une amélioration mais est peu probant; le résultat (d) indique une amélioration concluante mais négligeable; le résultat (e) indique une amélioration concluante et importante, tout comme (f) bien que (f) soit plus précis.

Et les maths dans tout cela?

Certains auteurs admettent que, pour des raisons historiques, les cours de statistiques sont plus compliqués que nécessaire (Yilmaz, 1996). Stuart (1995) en particulier fait remarquer que les statistiques sont traditionnellement enseignées comme un cours de mathématiques. Il ajoute : "not only is this unnecessary, it is counterproductive for most students" (p. 45). Par exemple, en physique, les mathématiques sont très utilisées mais personne n'assimile la physique aux mathématiques du fait principalement que les grandeurs manipulées sont concrètes (en mécanique : la masse, la vitesse, etc.). Par ailleurs, nombreux sont ceux qui pensent que les statistiques sont une branche des mathématiques du fait cette fois que les objets statistiques (le hasard, l'échantillonnage) sont rendus abstraits, ne sont pas expliqués comme des phénomènes concrets mais par des formules algébriques. Coulter (1995) a diminué la place des mathématiques dans le cours de statistiques et affirme que cela a été positif pour les étudiants. Devrait-on complètement éliminer les mathématiques dans les cours de statistiques?

Au cours de l'année académique 2008-2009, le premier auteur a essayé cette expérience dans une classe de première année du baccalauréat en psychologie. Étonnamment, les étudiants ont été nombreux à demander des formules mathématiques. Loin de réduire l'anxiété, elle l'a augmentée de façon palpable pour une majorité d'étudiants! Deux interprétations sont possibles. D'une part, sans formule, les étudiants sont à la merci d'une boîte noire, le logiciel d'analyse statistique. N'ayant pas les moyens de vérifier le code informatique, ils n'ont pas de prise sur le déroulement de l'analyse. D'autre part, comme l'anxiété tend à diminuer les ressources cognitives nécessaires pour apprendre de nouvelles habiletés (Chanquoy, Tricot et Sweller, 2007), les personnes surchargées cognitivement recherchent des stratégies stéréotypées pour résoudre des problèmes qui devraient demander de la réflexion. Une stratégie stéréotypée d'apprentissage est le « par cœur ». Aussi, nous croyons que les étudiants anxieux cherchent à apprendre par cœur le maximum de choses. Or, une formule se prête bien à ce type d'exercice. La formule « la somme des x-i divisée par n » devient facilement un mantra qui rassure (faussement) l'étudiant puisqu'il peut réciter sans erreur la formule (sans retirer quelque intuition que ce soit de cet exercice futile). Ceci expliquerait aussi en partie la ritualisation du fameux Rejet-de-H_0-si-p-<-.05.

Recommandations

1) Pour éviter que les statistiques ne soient confondues avec la mathématique ou l'algèbre, il s'agit de mettre les phénomènes concrets en premier, de jouer avec eux, de les illustrer. Par la suite, une formule mathématique pourra être donnée pour structurer quantitativement le concept (par exemple, la moyenne comme le centre de gravité du graphique de fréquence; l'écart type comme l'écart typique à la moyenne, etc.).

De cette façon les étudiants réalisent que la formule n'est pas un problème mathématique mais une sonde qui quantifie les caractéristiques d'un échantillon.

2) Présenter les calculs comme une procédure à suivre pour obtenir un chiffre plutôt que comme une équation algébrique intimidante. Donner cette procédure sous trois formes différentes :

(en français) **Pour calculer** *la moyenne*, **faire** la somme des données puis diviser par le nombre de données;

(en *mathais*) $\overline{\mathbf{X}} \leftarrow \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$ (ou « $\leftarrow$ » se prononce : « reçoit la valeur »; il ne faut pas mettre le signe « = » car il s'agit d'une *définition* de $\overline{\mathbf{X}}$, pas d'une comparaison).

(en *colonnais*) Voir Figure 4

	A	B	C	D
1	**Pour calculer une moyenne:**			
2	no de la donnée	x_i		
3	1	67		
4	2	84		
5	3	77		
6	4	72		
7	5	69		
8	6	91		
9	la somme	460		
10	divisée par 6	76.7		
11	donne la moyenne			
12				
13	1- Mettre les données dans une colonne			
14	2- Additionner la colonne			
15	3- Diviser le total par le nombre de données.			
16	Ceci est la moyenne			
17				
18				

	A	B	C	D	E	F	G	H	I
1	**Pour calculer l'écart type**								
2	no de la donnée	x_i	→	la colonne précédente moins la moyenne	→	la colonne précédente au carré			
3	1	67		-9.7		93.44			
4	2	84		7.3		53.78			
5	3	77		0.3		0.11			
6	4	72		-4.7		21.78			
7	5	69		-7.7		58.78			
8	6	91		14.3		205.44			
9	la somme	460		la somme		433.33		la racine carrée	
10	divisée par 6	76.7		divisée par 5		86.67		de la variance	9.31
11	donne la moyenne			donne la variance				donne l'écart type	
12									
13	1- Mettre les données dans une colonne et obtenir la moyenne								
14	2- Dans une seconde colone, mettre la données moins la moyenne								
15	3- Dans une troisième colonne, mettre la colonne deux au carré								
16	4- Additionner cette dernière colonne								
17	5- Diviser par un de moins que le nombre de données. Ceci est la variance								
18	6-Prendre la racine carrée de la variance. Ceci est l'écart type								

Figure 4. **En haut, la procédure à suivre pour calculer une moyenne en *colonnais*, en bas, la procédure à suivre pour calculer l'écart type.**

Le *colonnais* est la langue qui consiste à mettre les nombres dans des colonnes et manipuler ces colonnes. Un chiffrier est donc l'outil parfait pour parler le *colonnais*, et des exemples avec *Excel* ou *Calc* peuvent être facilement donnés aux étudiants.En multipliant le nombre d'approches, l'étudiant a plus de chances de trouver une approche qui connecte avec ses acquis antérieurs.

3) représentation visuelle À chaque fois qu'une nouvelle statistique est présentée, montrez un graphique de fréquence et indiquer comment visualiser cette statistique. Donnez un second graphique et invitez l'étudiant à se commettre en donnant une valeur approximative à la statistique. Les graphiques peuvent provenir d'une panoplie d'échantillons similaires ou différents afin de montrer un large éventail de graphiques et apprendre à visualiser les statistiques descriptives sans nécessairement apprendre une formule par cœur.

Des heuristiques simples existent :

Pour la moyenne, trouver le point d'équilibre du graphique s'il était déposé sur un plateau;

Pour l'écart type, trouver les points d'inflexion, et diviser l'étendue entre ces points par deux.

4) Pour éviter que les étudiants perdent leur temps à apprendre par cœur les formules, donnez-leurs la possibilité d'apporter une feuille lors des évaluations sur laquelle ils peuvent noter ce qu'ils souhaitent, incluant les formules. De cette façon, leur étude peut se centrer davantage sur les concepts à comprendre.

En somme…

Un test d'hypothèse nulle comme celui en (1) peut se récrire comme :

$$\text{Rejet de } H_0 \text{ si } p < \alpha \tag{4}$$

où α est un seuil de décision (généralement 5 %) et la valeur *p* est la probabilité associée à la statistique obtenue à partir des données observées étant donné l'hypothèse nulle. Cette règle de décision revient à dire :

Si la probabilité de la statistique observée étant donné une hypothèse est faible (et inférieure à un seuil de décision α), (5)
Alors rejeter l'hypothèse.

Peut-on énumérer les concepts requis pour comprendre un tel énoncé? Un début d'arbre conceptuel est donné en Figure 5. Cet arbre conceptuel est certainement incomplet. Il est aussi assez peu opérationnel en ce sens qu'il donne peu d'indices sur la façon dont l'étudiant peut assimiler ces concepts et comment l'enseignant doit les aborder. Les recommandations qui précèdent énumèrent des exercices que l'étudiant peut réaliser et des attitudes que l'enseignant peut adopter.

Pour comprendre l'énoncé (4), il faut les concepts:

- Probabilité
  - Hasard
  - Échantillonnage
    - Ré-échantillonnage
    - Sous-échantillonnage
- Erreur type
  - Variabilité
  - Méta-statistique
    - Reproduction
- Décision
  - État d'ignorance
  - Données probantes

Figure 5. Début d'arbre conceptuel indiquant certaines connaissances requises afin de comprendre l'énoncé (4) en particulier et les tests d'hypothèse nulle en général

Conclusion

L'approche proposée ici aux enseignants, qui consiste à identifier les concepts prérequis pour comprendre les statistiques, devrait permettre de mieux enseigner les statistiques, de réduire l'anxiété et d'accroître la compétence des étudiants. Cependant, il reste encore beaucoup de chemin à parcourir.

1- L'arbre conceptuel de la Figure 5 est certainement incomplet. Quels autres concepts sont nécessaires? Sont-ils hiérarchisés comme la Figure 5 le suggère?

2- Quelles activités (et dans quel ordre) sont nécessaires pour que l'étudiant assimile un concept donné? Les concepts peuvent-ils être appris séparément, ou certains concepts sont-ils indissociables d'autres concepts?

3- Comment peut-on évaluer si le concept est acquis? Quelles questions pourraient vérifier l'acquisition de chaque concept isolément?

4- Est-ce que, ayant acquis les concepts de base, les étudiants manifesteront autant d'anxiété statistique? Si l'anxiété traduit un manque de contrôle, des activités où l'étudiant a le contrôle réduisent-elles l'anxiété?

L'enseignement des statistiques doit être évalué, revu et amélioré. Il est temps de tirer des enseignements de nos enseignements statistiques. Il n'y a pas que les étudiants qui peuvent avoir des concepts erronés. Une des erreurs importantes commises par les enseignants est qu'ils passent probablement beaucoup de temps à expliquer les notions de variables, de variables indépendantes, de mesures, d'échelles de mesure et qu'ils jugent inutiles d'aborder le concept de hasard.

Nous avons au bout des doigts une panoplie de ressources et d'outil afin d'améliorer la pédagogie statistique, tels que des générateur de nombres aléatoires qui donnent l'occasion de visualiser le hasard à l'œuvre (*GRD*, Harding et Cousineau, 2014, 2015; *RANDOM.org*; Haar, 1998-2015; *Rice Virtual Lab*, Lane et al., n.d.; *Rossman/Chance Applet Collection*, Rossman et Chance, n.d.); des modules de sous-échantillonnage pour mieux travailler les notions de reproductibilité (tel que *GSD*, Harding et Cousineau, 2016); des modules pour télévotage qui s'intègre à *PowerPoint* et à *SPSS* (tel que SurveyMonkey, Adobe PDFForms, ou *ivote*, T. Groulx et Cousineau, 2016) pour récolter des échantillons directement dans la classe sans délai.

Ce texte n'offre pas de solutions définitives mais des pistes pour formaliser l'enseignement des statistiques tout en réduisant le formalisme. Nous espérons que ce texte suscitera une discussion au niveau de l'éducation des méthodes quantitatives dans le but d'éliminer le préjugé que l'enseignement et la compréhension des statistiques est difficile.

Références

Allen, K. (2006). *The statistics concept inventory: Developpement and analysis of a cognitive assesment instrument in statistics* (Technical report series No. unpublished dissertation). Norman, OK: University of Oklahoma.

American Statistical Association with Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T… Witmer, J. (2005). *Guidelines for Assessment and Instruction in Statistics Education [GAISE] College Report.* Alexandria, VA: American Statistical Association.

Baloglu, M. (2003). Individual differences in statistics anxiety among college students. *Personality and Individual Differences, 34*, 855-865. doi: 10.1016/S0191-8869(02)00076-4

Becchetti, L., Castriota, S., Corrado, L. et Giachin Ricca, E. (2013). Beyond the Joneses: Inter-country income comparisons and happiness. *The Journal of Socio-Economics, 45*, 187-195.

Béland, S., Cousineau, D. et Loye, N. (2016). Les dix commandements du nouvel homo statisticus. *McGill journal of education/Revue des sciences de l'éducation de McGill. 51*, 947-960.

Belmont, L. et Marolla, F. A. (1973) Birth order, family size, and intelligence. *Science, 182*, 1096-1101.

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education, 57*, 247-261. doi: 10.1080/00220973.1989.10806509

Blanché, R. (1967) *La science actuelle et le rationalisme*. Paris, France : Presse Universitaire de France

Boole, G. (1854). *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities.* Macmillan. Reprinted with corrections, Dover Publications, New York, NY, 1958 and reissued by Cambridge University Press, 2009

Cantinotti , M., Lalande , D., Ferlatte, M.-A. et Cousineau, D. (2017) Validation de la version francophone du Questionnaire d'anxiété statistique (SAS-F-24). *Revue canadienne des sciences du comportement, 49*, 133–142.

Chanquoy, L., Tricot, A. et Sweller, J. (2007). *La charge cognitive : théorie et applications.* Paris, France : Armand Colin.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003. doi: 10.1037/0003-066X.49.12.997

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.

Cohen, J. (1990). Things I have learned (So far). *American psychologist, 45*, 1304-1312. doi: 10.1037/0003-066X.45.12.1304

Coulter, X. (1995). A new way of teaching statistics. *Journal of Educational Statistics, 24*, 181-187. doi: 10.2190/K906-0M2X-CGTE-7XGV

Cousineau, D. (2009). *Panorama des statistiques pour les sciences humaines*. Bruxelles, Belgique : De Boeck, 445 p.

Cousineau, D. (2014). Restoring confidence in psychological science findings: A call for direct replication studies. *The Quantitative Methods for Psychology, 10*, 77-79. doi: 10.20982/tqmp.10.2.p077

Cousineau, D. et Allan, T. A. (2015). Likelihood and its use in Parameter Estimation and Model Comparison. *Mesures et Évaluation en Éducation, 37*, 63-98. doi: 10.7202/1036328ar

Dunn, P. K. (2005). We can still learn about probability by rolling dice and tossing coins. *Teaching Statistics, 27*, 37-41. doi: 10.1111/j.1467-9639.2005.00205.x

Finney, S.J. et Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161-186. doi: 10.1016/S0361-476X(02)00015-2

Fisher, R. A. (1925) *Statistical methods for research workers*. Edinburgh, Écosse: Oliver and Boyd.

Furnham, A. et Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences, 37*, 943-955. doi: 10.1016/j.paid.2003.10.016

George, A. C. et Robitzsch, A. (2015). Cognitive diagnosis models in R: A didactic. *The Quantitative Methods for Psychology, 11*, 189-205. doi: 10.20982/tqmp.11.3.p189

Haar, M. (1998-2015). RANDOM.ORG: True Random Number Service. Retrieved October 7th, 2016, from https://www.random.org/

Hair, P. et Hampson, S. E. (2006). The role of impulsivity in predicting maladaptative behaviour among female students. *Personality and Individual Differences, 40*, 943-952. doi: 10.1016/j.paid.2005.10.002

Halloun, H. A. et Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53*, 1056-1065. doi: 10.1119/1.14031

Harding, B., Cousineau, D. (2014). GRD: An SPSS extension command for generating random data. *The Quantitative Methods for Psychology, 10*, 80-94. doi: 10.20982/tqmp.10.2.p080

Harding, B. et Cousineau, D. (2015). GRD 2.0: An extended SPSS extension command for generating random data. *The Quantitative Methods for Psychology, 11*, 127-138. doi: 10.20982/tqmp.11.3.p127

Harding, B. et Cousineau, D. (2016). GSD: An SPSS extension command for sub-sampling and bootstrapping datasets. *The Quantitative Methods for Psychology, 12*, 145-153. doi: 10.20982/tqmp.12.2.p138

Harding, B., Tremblay, C. et Cousineau, D. (2014). Standard errors: A review and evaluation of standard error estimators using Monte Carlo simulations. *The Quantitative Methods for Psychology, 10*, 107-123. doi: 10.20982/tqmp.10.2.p107

Harding, B., Tremblay, C., Cousineau, D. (2015) The Standard Error of the Pearson Skew. *The Quantitative Methods for Psychology. 11*, 32-36. doi: 10.20982/tqmp.11.1.p032

Hoijtink, H. (2012) *Informative hypotheses: Theory and applications for behavioral and social scientists*. London, UK: Chapman et Hall/CRC.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education, 3*, 1-12.

Küchenhoff, H. (2008). *Coin Tossing and Spinning – Useful Classroom Experiments for Teaching Statistics*, dans Shalabh, Heumann, C. (Dir.). *Recent Advances in Linear Models and Related Areas: Essays in Honour of Helge Toutenburg* (pp. 417-426). Heidelberg: Psysica-Verlag HD.

Lane, D. (n.d.) Online statistics education: A multimedia course of study. http://Onlinestatbook.com (last retrieved 22/06/2016).

Lecoutre, M.-P., Poitevineau, J. et Lecoutre, B. (2003) Even statisticians are not immune to miinterpretations of null hypothesis significance tests. *International Journal of Psychology, 38*, 37-45.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyse data. Current Directions in *Psychological Science, 5*, 161-171. doi: 10.1111/1467-8721.ep11512376

Marmolejo-Ramos, F. et Matsunaga, M. (2009) Getting the most from your curves: Exploring and reporting data using informative graphical techniques. *Tutorials in Quantitative Methods for Psychology, 5*, 40-50. doi: 10.20982/tqmp.05.2.p040

Marson, S. M. (2007). Three empirical strategies for teaching statistics. *Journal of Teaching in Social Work, 27*, 199-213. doi: 10.1300/J067v27n03_13

Neyman, J. et Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 231*, 289–337. doi:10.1098/rsta.1933.0009.

Nickerson, R. S. (2002) The production and perception of randomness. *Psychological Review, 109*, 330-357. doi: 10.1037/0033-295X.109.2.330

Onwuegbuzie, A. .J. et Daley, C. E. (1999). Perfectionism and statistics anxiety. *Personality and Individual Differences, 26,* 1089-1102. doi: 0.1016/S0191-8869(98)00214-1

Onwuegbuzie, A. J. et Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *Journal of Experimental Education, 63,* 115-124. doi: 10.1080/00220973.1995.9943816

Pashler, H. et Wagenmakers, E.-J. (2012) Editor's introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528-530. doi: 10.1177/1745691612465253

Rossman, A.J. et Chance, B.L. (n.d.) Rossman/Chance Applet Collection. Page consultée le 7 Octobre 2016, depuis http://www.rossmanchance.com/applets/index.html.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin, 57*, 416-428. doi: 10.1037/h0042040

Stuart, M. (1995). Changing the teaching of statistics. *The Statistician, 44,* 45-54. doi: 10.2307/2348615

T. Groulx, J. et Cousineau, D. (2016). Ivote: A system for televoting. *The Quantitative Methods for Psychology, 13*, 57-64.

Vigil-Colet, A., Lorenzo-Seva, U. et Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema, 20*, 174-180.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin and Review, 14*, 779-804. doi: 10.3758/BF03194105

Wasserstein, R. L. et Lazar, N. A. (2016). ASA's statement on statistical significance and p-values. *The American Statistician, version en libre accès*, 1-6. doi: 10.1080/00031305.2016.1154108

Wiens, S. et Nilsson, M. E. (sous presse) A tutorial on contrast analyses in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement*.

Yilmaz, M. R. (1996). The challenge of teaching statistics to non-specialists. *Journal of Statistics Education, 4*, 1-9.