



Université Libre de Bruxelles

Faculté des Sciences Psychologiques et de l'Éducation

**TITRE,
TITRE SUITE**

Par

MARIE DELACRE

En vue de l'obtention du grade de docteur

Septembre 2021

Abstract

Abstract (voir nombre de mots requis)

Contents

1	Introduction	iv
1.1	Truc qui n'ont rien à voir mais dont je pourrais avoir besoin à la défense	iv
1.2	Début de la vraie intro	iv
1.2.1	Limite 1: conditions d'application	v
1.3	Limite 2: hypothèse nulle	ix
1.4	Pourquoi jusque là la sauce n'a pas pris?	x
2	Conclusion	2
2.0.1	Usage des articles méthodo	2
2.0.2	Importance des simulations et des logiciels modernes pour enseigner les statistiques fréquentistes	2
2.0.3	Comment écrire/transmettre l'info aux psys	3
2.0.4	Recommandations générales	3
3	Bibliographie	4
4	Annexe(s)	6
4.1	Annexe A: additional tables	6
4.2	Annexe B: additional figures	7
4.3	Annexe C: code	8

Acknowledgements

Thank you for following this tutorial!

I hope you'll find it useful to write a very professional dissertation.

1 Introduction

1.1 Truc qui n'ont rien à voir mais dont je pourrais avoir besoin à la défense

“A review by van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, and Depaoli (2017) revealed that 31% of articles in the psychological literature that used Bayesian analyses did not even specify the prior that was used, at least in part because the defaults by the software package were used. Mindless statistic are not limited to pvalues” (dans l'article de Daniel... j'adore cet argument!)

1.2 Début de la vraie intro

On attend des chercheurs en psychologie, et des psychologues en général, qu'ils soient capables de produire des connaissances, fondées sur des preuves scientifiques (et non sur des croyances et opinions), et également de comprendre et évaluer les recherches menées par d'autres (Haslam and McGarty 2014). Or, dans un domaine dominé par les analyses quantitatives¹ (Counsell and Harlow 2017), les connaissances statistiques s'avèrent fondamentales pour comprendre, planifier et analyser une recherche (Howitt and Cramer 2017; Everitt 2001). Les statistiques font dès lors partie intégrante du cursus de formation des psychologues et jouent un rôle très important dans leur parcours.

Traditionnellement, depuis plus de 50 ans, les tests-*t* et les *ANOVA* se trouvent au coeur de la grande majorité des programmes dans les domaines des Sciences Psychologiques et de l'Education (Aiken, West, and Millsap 2008; Golinski and Cribbie 2009; Curtis and Harwell 1998) et des livres d'introduction aux statistiques pour psychologues (EXEMPLES?). Cela pourrait vraisemblablement expliquer pourquoi ils sont si persistants dans la recherche en psychologie (Counsell and Harlow 2017). Ces tests sont, depuis plus de 60 ans (Nunnally 1960; Byrne 1996), les tests les plus fréquemment cités dans la littérature scientifique (Golinski and Cribbie 2009). Dans une revue de 486 articles publiés en 2000 dans des journaux populaires en psychologie², Golinski and Cribbie (2009) avaient relevé 140 articles ($\approx 29\%$) au sein desquels les auteurs avaient mené au moins une ANOVA à un ou plusieurs facteurs. Plus récemment, Counsell and Harlow (2017) mentionnaient que parmi un ensemble de 151 études soumises dans 4 revues canadiennes en 2013, environ 40% incluaient une comparaison de moyennes. Peut-être est-ce en raison de leur grande fréquence d'usage, ajoutée à leur apparente simplicité, qu'on tend à croire que la plupart des chercheurs, si pas tous, ont une bonne maîtrise des tests de comparaisons de moyennes (Aiken, West, and Millsap 2008; Hoekstra, Kiers, and Johnson

¹parmi 68 articles analysés en 2013 par Counsell et ses collaborateurs (2017) dans 4 revues canadiennes, 92.7% incluaient au moins une analyse quantitative (contre 7.3% incluant une analyse qualitative)

²Les revues analysées étaient les suivantes: "Child Development", "Journal of Abnormal Psychology", "Journal of Consulting and Clinical Psychology", "Journal of Experimental Psychology: General", "Journal of Personality" et "Social Psychology"

2012). Pourtant, certains indices semblent contredire cette conviction. A travers cette thèse, nous allons adresser deux aspects de l’usage des tests de comparaison de moyennes qui peuvent être améliorés. *D’abord, ces sont souvent utilisés dans des contextes inadéquats, ce qui a pour conséquences d’augmenter le nombre de conclusions invalides et non répliquables (Hoekstra, Kiers, and Johnson 2012). Ensuite, ils sont presque invariablement utilisés en définissant comme hypothèse nulle une absence de différence entre les moyennes des groupes (Meehl 1990; Steyn 2000).* A AMELIORER.

1.2.1 Limite 1: conditions d’application

Bien qu’il existe plusieurs types de tests t , les chercheurs en psychologie tendent à privilégier par défaut le test t de Student et l’*ANOVA* de Fisher. La statistique t de Student se calcule comme suit (Student 1908):

$$t_{Student} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{1}{N-2} \sum_{j=1}^2 (n_j - 1) S_j^2\right) \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (1)$$

où N = le nombre total de sujets, et n_j et \bar{X}_j sont respectivement la taille et la moyenne du $j^{\text{ème}}$ échantillon ($j = 1, 2$). Sous l’hypothèse de normalité, la statistique t de Student suit une distribution t avec $n_1 + n_2 - 2$ degrés de liberté. La statistique F de Fisher est une généralisation de la statistique t de Student, et se calcule comme suit:

$$F = \frac{\frac{1}{k-1} \sum_{j=1}^k \left[n_j (\bar{x}_j - \bar{x}_{..})^2 \right]}{\frac{1}{N-k} \sum_{j=1}^k \left[(n_j - 1) S_j^2 \right]} \quad (2)$$

où k est le nombre d’échantillons indépendants, et S_j^2 est la variance du $j^{\text{ème}}$ échantillon ($1 \leq j \leq k$). Sous l’hypothèse de normalité, la statistique F suit loi de Fisher caractérisée par 2 paramètres:

$$df_1 = k - 1$$

$$df_2 = \sum_{j=1}^k n_j - k$$

Comme le révèlent les équations (1) et (2), la variance poolée intervient au dénominateur des statistiques t de Student et F de Fisher. Or, utiliser ce terme n’a de sens que si la condition d’homogénéité des variances est respectée. Dans certains cas, la violation de ces conditions peut entraîner une augmentation du risque alpha mais également amener à une perte de puissance (Hoekstra, Kiers, and Johnson 2012; Osborne and Waters 2002). Pour cette raison, les chercheurs devraient toujours s’assurer de systématiquement faire mention de ces conditions dans un article, afin de rassurer le lecteur sur la confiance qu’il peut accorder à la fiabilité des résultats (Osborne and

Waters 2002). Pourtant, malgré la demande de plus de transparence dans la transmission des analyses de données (Counsell and Harlow 2017), on constate que dans les articles publiés, il n'est que rarement fait mention de ces conditions d'application. Osborne and Christianson (2001), par exemple, avaient trouvé que seulement 8% des auteurs reportaient des informations sur les conditions d'application des tests, soit à peine 1% de plus qu'en 1969. Plus récemment, Hoekstra, Kiers, and Johnson (2012) ont montré que sur 50 articles publiés en 2011 dans *Psychological Science* utilisant au moins une ANOVA, test-*t* ou régression, seulement trois discutaient des questions de normalité et d'homogénéité des variances. Par ailleurs, les informations reportées sont souvent non exhaustives (Counsell and Harlow 2017), et la condition de normalité est plus fréquemment discutée que celle d'homogénéité des variances. Parmi les 61 articles analysés par Keselman et al. (1998), seulement 5% des articles mentionnaient simultanément les conditions de normalité et d'homogénéité des variances (alors que 6% mentionnaient exclusivement la condition de normalité, et 3% mentionnaient exclusivement celle d'homogénéité des variances). Golinski and Cribbie (2009) ont fait un constat similaire: parmi les 140 articles qu'ils ont analysé, seulement 11 mentionnaient explicitement la condition de normalité, contre 3 qui mentionnaient celle d'homogénéité des variances.

Il est possible que dans certains cas, les conditions d'application aient été vérifiées et respectées, mais que les auteurs n'aient pas reporté l'information dans leur article (Counsell and Harlow 2017).

Golinski and Cribbie (2009): *En ce qui concerne l'homogénéité des variances, seulement 3 mentionnaient l'hypothèse d'homogénéité des variances, alors que parmi les 65 articles donnant l'info sur les variances de groupe, 27 avaient un SDR de plus de 2 (dans une étude, le SDR était même de 104), et que la plupart d'entre elles avaient des designs non balancés.*

Cependant, cela ne constitue probablement pas la majorité des cas. Par exemple, parmi les 282 études parues dans SPPS entre avril 2015 et avril 2016, au sein des 100 études qui utilisaient le test *t* pour échantillons indépendants, une seule mentionnait une transformation des données de sorte à rendre la distribution normale (log-linéaire). Seulement 2 études ont privilégié l'alternative non-paramétrique : le *U* de Mann-Whitney. Autrement dit, sur 102 études, seulement 3 ont pris une décision inhérente au rejet de la normalité. Il s'agit pourtant d'une condition d'application dont le respect est souvent peu réaliste (Micceri, 1989). Par ailleurs, si le test *t* s'avère robuste à de légers à moyens écarts à la normalité, il ne l'est plus en cas de forte asymétrie (Delacre, Lakens & Leys, 2017), or des distributions fortement asymétriques représentent une part non négligeable des données. Micceri (1989) par exemple, mettait en évidence, à travers l'étude de 440 distributions, un taux de près de 20% de distributions fortement asymétriques.

Hoekstra et ses collaborateurs (2012) ont également constaté une absence de vérification des conditions d'application des tests ou, le cas échéant, une vérification inappropriée, par exemple à travers l'usage de tests statistiques (Hoekstra et al., 2012). Afin de le justifier, ils ont envisagé et testé quatre hypothèses explicatives par questionnaire. Premièrement, les auteurs pourraient considérer que vérifier les conditions soit inutile, parce qu'ils surestiment la robustesse des tests (Hoekstra et al., 2012). Deuxièmement, il est probable que les auteurs manquent de connaissance à propos de ces hypothèses (Hoekstra et al., 2012), ou en aient une compréhension erronée. Le théorème central limite, par exemple, fait l'objet d'incompréhensions, si bien que même des chercheurs expérimentés ont l'intuition erronée que la loi des grands nombres s'applique également à de petits nombres (Braver et al., 2014). Troisièmement, il est possible que les auteurs aient conscience de la nécessité de vérifier les conditions d'application, mais ne savent pas comment vérifier si une condition a été violée (Hoekstra et al., 2012) ou encore, quatrièmement, qu'ils n'aient simplement pas connaissance d'alternatives existantes dans les cas où une condition semble violée (Hoekstra et al., 2012). Ces quatre hypothèses ont pour point commun d'envisager un manque de connaissance ou une compréhension erronée des conditions des tests.

“It is interesting that, of the 11 articles that mentioned the normality assumption, 10 found distributions that were nonnormal. Although it is possible that the remaining articles that did not mention the normality assumption all found no evidence of nonnormality, it seems highly unlikely given that Micceri (1989) who examined 440 variables from published articles in education and psychology, found that 84% showed moderate to extreme skew”.

- voir dans mes 2 articles sur le Welch, je donne des exemples aussi de ça
- Counsell and Harlow (2017): sur 151 études analysées, 44 reportaient des informations relatives aux conditions d'application des tests utilisés. Parmi celles-ci, seulement 2 étaient exhaustives: toutes les autres soit reportaient seulement une partie des conditions, soit le faisaient de manière inappropriée (ex.: vérifier la normalité de la VD au lieu de celle des résidus, dans le cas d'une régression).

Il arrive fréquemment que les chercheurs annoncent avoir réalisé un test- t ou une $ANOVA$ sans spécifier de quel type de test- t ou d' $ANOVA$ il s'agit (Counsell and Harlow 2017). Il existe pourtant plusieurs variations de ces tests

N'EST-CE PAS PARCE QU'ILS UTILISENT L'OPTION DISPO PAR DEFAUT DANS LES LOGICIELS CLIC BOUTON?*

A cause du manque de transparence des chercheurs quant au respect ou non respect des conditions

d'application, il est très compliqué de vérifier si le test qu'ils ont choisi est approprié ou non [counsell_reporting_2017]. CA JE PEUX COMPLETER AVEC MON PROJET DE RECHERCHE: *le non report peut vouloir dire que la condition était ok mais ça semble très peu crédible qu'elle le soit dans autant d'études.*

→ POURQUOI ELLES NE SONT PAS VERIFIEES? → 1) Leur non report peut être dû à un désir de limiter le nombre de pages (on se contente de reporter ce que les reviewers/éditeurs nous demandent de reporter; Counsell and Harlow (2017)) 2) Par manque de connaissances, les chercheurs se contentent souvent des informations fournies dans les logiciels clic/bouton. *for example, if software does not report a CI on Cohen's d, it is unlikely that a researcher will calculate one his or herself* (Counsell and Harlow (2017)).

Ce constat est loin d'être récent. Par exemple, Keselman et al. (1998) mettaient en évidence le fait que les chercheurs tendaient à utiliser des tests non robustes aux violations des conditions d'application, sans vérifier au préalable si ces conditions étaient respectées.

Même lorsqu'un chercheur souhaite vérifier les conditions d'application, il reste confronté à plusieurs problèmes.

1. Les conditions reposent sur les paramètres de *population* et non sur les paramètres d'échantillon. Or, ces paramètres de population ne sont pas connus (s'ils l'étaient, on n'aurait pas besoin des statistiques) (Hoekstra, Kiers, and Johnson 2012).
2. Les conditions sont souvent très irréalistes.

Il existe des tests dit "tests robustes" qui ne sont théoriquement pas affectés par une violation des conditions d'application (Erceg-Hurn and Mirosevich 2008), mais ces derniers ne sont que peu ou pas utilisés (Sharpe 2013). Malgré de nombreuses tentatives (Keselman et al. 1998), leur succès reste très mitigé.

Comment améliorer les pratiques de recherche facilement, d'une manière qui assure que les chercheurs appliqueront les conseils? En proposant des switchs faciles. L'usage des tests de Welch est un bel exemple de switch facile. Ceci dit, ce n'est pas parce que le switch est facile qu'il est forcément fait: Keselman et al. (1998) écrit ceci: "Despite these repeated cautionary notes, behavioral science researchers have clearly not taken this message to heart. It is strongly recommended that test procedures that have been designed specifically for use in the presence of variance heterogeneity and/or nonnormality be adopted on a routine basis" (p.358). Rem.: ils parlent d'un article de Lix et al. (1996) qui mentionne des packages qui permettent de le faire mais l'article est introuvable sur google scholar.

L'open access est une des clés pour moi. w

ARTICLE1 ARTICLE2.

1.3 Limite 2: hypothèse nulle

Effect sizes are an important outcome of empirical research. Moving beyond decisions about statistical significance, there is a strong call for researchers to report and interpret effect sizes and associated confidence intervals. This practice is highly endorsed by the American Psychological Association (APA) and the American Educational Research Association (American Educational Research Association, 2006; American Psychological Association, 2010).

En parlant des tailles d'effets, on commence de plus en plus à les utiliser (j'ai une réf qui le dit) mais: - on les calcule sans vraiment les comprendre/interpréter - comme pour le test t de Student et l'ANOVA, on utilise un test qui dépend des mm conditions d'application. Utiliser des tests plus adéquats permettrait d'améliorer les pratiques et à termes, de déterminer des mesures de taille d'effets qui pourront être utilisées a priori dans des tests plus informatifs que ceux visant à détecter l'absence d'effet (cf. tests d'équivalence).

Un paragraphe relatif à la taille d'effet. EN EXPLORATOIRE, ce qui à termes pourrait servir à définir des hypothèses plus informatives pour d'autres chercheurs, qui pourraient être utilisées, soit dans des tests d'effets minimaux, soit pour des tests d'équivalence. Et that's it.

Rem.: "une violation des conditions d'application peut amener à une sous- ou sur-estimation des mesure de taille d'effet (Osborne & Waters, 2002, cités par Hoekstra!)

Le NHST fait l'objet d'énormément de critiques, si bien que certains recommandent de le remplacer par une mesure de taille d'effet accompagnée d'un intervalle de confiance autour de la taille d'effet. Le raisonnement est que si l'IC contient la valeur 0, on ne peut conclure à une différence significative (Counsell and Harlow 2017).

Une des principales critiques des tests d'hypothèse est le fait que l'on compare la différence observée à l'absence totale de différence (= un effet de 0). C'est une question qui est peu intéressante, car peu surprenante. Mais pourquoi comparer à 0 et pas à une autre valeur?

D'après Lakens (2021), un test d'hypothèse (selon l'approche de Neyman-Pearson) vaut la peine à 2 conditions:

- 1) que l'hypothèse nulle soit assez plausible pour que son rejet puisse surprendre au moins certains;
- 2) le chercheur veut appliquer une procédure méthodol qui l'autorise à prendre des décisions quant à la

manière d’agir, tout en contrôlant le taux d’erreur. Agir peut vouloir dire: adopter un traitement, une politique, une intervention, ou abandonner un domaine de recherche, modifier une manipulation, ou de faire un certain type de déclaration ou revendication.

Counsell and Harlow (2017): *the constant calls for reporting effect sizes appears to have had an effect on the Canadian psychology articles as just over 90% of the analyses that used a significance test also included a standardized or unstandardized effect size. Few articles presented an effect size without hypothesis testing, and few of the analyses’ results included a CI.*

Ca se fait apparemment de plus en plus de reporter la taille d’effet (dans leur analyse de 151 études, 90% des analyses incluait une mesure de taille d’effet, standardisée ou non... mais très peu incluait les IC et de plus, ils les donnaient mais sans vraiment en discuter... @Counsell and Harlow (2017) dans la discussion).

Comme déjà mentionné, l’hypothèse nulle est l’absence d’effet. On en reste sur la nil-hypothesis. Du coup, un effet significatif n’a pas vraiment de valeur. En réponse à ce problème, on a écrit deux articles:

- On peut commencer par ajouter une information sur les tailles d’effets (mais du coup ça n’oblige pas à réfléchir à l’avance à l’effet qui nous intéresse)

Dans la revue de Keselman et al. (1998), ils mentionnent que les tailles d’effet ne sont pratiquement jamais reportées malgré les recommandations du manuel de l’APA (1994) (et qu’elles ne sont fournies qu’en cas d’effet significatif).

- On peut aussi faire des tests plus informatifs (tests d’équivalence et/ou tests d’effets minimaux).
*One of the most widely suggested improvements of the use of p values is to replace null-hypothesis tests (where the goal is to reject an effect of exactly 0) with tests of range predictions (where the goal is to reject effects that fall outside of the range of effects that is predicted or considered practically important) (Lakens 2021).

1.4 Pourquoi jusque là la sauce n’a pas pris?

Je suis loin d’être la première à signaler tt ça. Ce qui manque encore dans mon plan d’introduction, c’est que je dois encore trouver le moyen de montrer en quoi mes articles sont une plus-value, ce qu’ils apportent. 2) Parler des packages, des applications Shiny, etc.

D’aucun on fait le constat d’un fossé entre les méthodes inférentielles recommandées dans la littérature scientifique et les techniques réellement utilisées par les chercheurs appliqués [keselman_statistical_1998].

PARLER DES DIFFERENTES REVUES DE LITTERATURE QUI LE DISENT.

Qu'est-ce qui pourrait expliquer cela? 1) Sharpe (2013): lack of awareness (p.573) Manque de conscience des développements dans le domaine?

2) Sharpe (2013): journal editors (p.573) Les éditeurs ne poussent pas assez? → Pas convaincue que ça m'intéresse

3) Sharpe (2013): Publish or perish? (p.574) je ne comprends mm pas en quoi c'est un argument

4) Sharpe (2013): Software (p.574) → aaahh! Certaines pratiques comme les équations structurelles et les analyses de puissance ont été facilitées par des logiciels comme gpower. Cela explique leur popularité. En ce qui concerne les statistiques plus robustes, par contre, elles ont moins de succès car non disponibles dans les logiciels disponibles. Les gens veulent juste qu'on leur dise où cliquer pour avoir le test qu'ils veulent! C'est triste mais faut faire avec (à mon avis).

5) Sharpe (2013): inadequate education (p.574)

6) Sharpe (2013): mindset: facteurs psychologiques t.q. la peur de dévier des pratiques courantes (comme si on n'allait pas être publié si on ne faisait pas comme tout le monde).

Anecdote: les chercheurs font souvent l'erreur de croire qu'il faut vérifier la normalité de la VD en faisant une régression. Dans SPSS, il est assez complexe de le faire car il faut d'abord calculer les résidus, ce qui implique de comprendre que les tests t et ANOVA sont des cas particuliers de régression, puis ensuite a posteriori représenter graphiquement les résidus. C'est chronophage et complexe. Dans Jamovi, par contre, la vérification de la normalité des résidus est automatiquement réalisée lorsqu'on fait un test t. Le rôle des méthodologistes, à mon sens, est de prémâcher le travail, pour permettre à d'autres de créer des outils conçus pour améliorer les pratiques de recherche. À partir du moment où c'est automatiquement fait correctement, il devient moins problématique que les psychologues maîtrisent le détail. Débarassés de ces questions, ils pourront peut-être alors plus se focaliser sur l'important pour mieux comprendre et interpréter les résultats de leurs tests: c'est-à-dire comprendre la distribution d'échantillonnage, dont pratiquement tout découle.

2 Conclusion

The conclusion section should specify the key findings of your study, explain their wider significance in the context of the research field and explain how you have filled the knowledge gap that you have identified in the introduction. This is your chance to present to your reader the major take-home messages of your dissertation research. It should be similar in content to the last sentence of your summary abstract. It should not be a repetition of the first paragraph of the discussion. They can be distinguished in their connection to broader issues. The first paragraph of the discussion will tend to focus on the direct scientific implications of your work (i.e. basic science, fundamental knowledge) while the conclusion will tend to focus more on the implications of the results for society, conservation, etc.

2.0.1 Usage des articles méthodo

Ces derniers semblent assez peu utilisés par les chercheurs. En tout cas s'ils les utilisent, ils les citent très peu dans leurs références pour justifier leurs choix (cf. article de Mills, Abdulla, and Cribbie (2010): le mode du nombre de citation méthodo dans les articles de recherche appliqués est 0, et la médiane vaut 1...). Dans l'autre sens, on constate que les articles méthodologiques sont généralement peu cités, et ils le sont encore 3 fois moins par les chercheurs appliqués que par les autres méthodologistes (Mills, Abdulla, and Cribbie 2010, 56).

On est en droit de questionner l'impact réel des publications méthodologiques, pour 2 raisons, d'après Mills, Abdulla, and Cribbie (2010):

- (1) Les chercheurs appliqués sont noyés sous les articles dans leur domaine d'expertise si bien que cela limite le temps dont ils disposent pour se consacrer aux articles méthodologiques.
- (2) malgré que des nouvelles méthodes sont disponibles, les chercheurs continuent à opter pour des tests traditionnels et familiaux (mais souvent inappropriés).

→ Qu'est-ce qui va pousser les chercheurs à lire des articles méthodologiques? → Si je trouve la réponse à ceci, j'ai mon intro.

2.0.2 Importance des simulations et des logiciels modernes pour enseigner les statistiques fréquentistes

On sait que les chercheurs tendent à privilégier les méthodes qui sont proposées par défaut dans des logiciels de clique bouton (comme SPSS). C'est en tout cas ce que dit Counsell and Harlow (2017) dans le contexte de la gestion des données manquantes (mais je crois que c'est vrai pour tout). Une manière d'améliorer les pratiques serait d'améliorer les options proposées par défaut dans les logiciels de

clic-bouton. C'est à ce genre de choses que j'aspire à travers mes articles.

Malgré tout, un logiciel ne fait pas tout et après avoir utilisé le test adéquat, il est important d'être capable de l'interpréter correctement. Les tests font appel à des notions faussement simples telles que les p-valeurs et les distributions d'échantillonnage. A mon sens, le seul moyen d'enseigner correctement ces notions, c'est à travers des simulations.

2.0.3 Comment écrire/transmettre l'info aux psys

Un consultant doit pouvoir parler de langage des psys, c'est-à-dire décrire et expliquer les méthodes requises d'une manière compréhensible pour les clients (Golinski and Cribbie 2009). Est-ce bien de demander à des mathématiciens/Statisticiens d'enseigner les stat aux psy's? Par forcément, car un psychologue spécialisé en méthodo quanti sera plus à même de comprendre les procédures et méthodes requises par les psys (ex. de la question de la taille d'effet qui n'intéresse pas vraiment les statisticiens; Golinski and Cribbie (2009)).

2.0.4 Recommandations générales

Mills, Abdulla, and Cribbie (2010):

- au moins un reviewer compétant pour analyser le caractère approprié des méthodes stat
- que les éditeurs/reviewers encouragent l'usage d'article de méthodo dans leur recherche → interesting, mais réaliste? La proportion de méthodologistes parmi les psychologues n'est pas assez élevée... Ou alors il faut vraiment de l'interdisciplinarité!

3 Bibliographie

- Aiken, Leona S., Stephen G. West, and Roger E. Millsap. 2008. "Doctoral Training in Statistics, Measurement, and Methodology in Psychology: Replication and Extension of Aiken, West, Sechrest, and Reno's (1990) Survey of PhD Programs in North America." *American Psychologist* 63 (1): 32.
- Byrne, Barbara M. 1996. "The Status and Role of Quantitative Methods in Psychology: Past, Present, and Future Perspectives." *Canadian Psychology/Psychologie Canadienne* 37 (2): 76.
- Counsell, Alyssa, and Lisa Harlow. 2017. "Reporting Practices and Use of Quantitative Methods in Canadian Journal Articles in Psychology." *Canadian Psychology/Psychologie Canadienne* 58 (2): 140.
- Curtis, Deborah A., and Michael Harwell. 1998. "Training Doctoral Students in Educational Statistics in the United States: A National Survey." *Journal of Statistics Education* 6 (1).
- Erceg-Hurn, David M., and Vikki M. Miroseovich. 2008. "Modern Robust Statistical Methods: An Easy Way to Maximize the Accuracy and Power of Your Research." *American Psychologist* 63 (7): 591.
- Everitt, Brian S. 2001. *Statistics for Psychologists: An Intermediate Course*. Psychology Press.
- Golinski, Caroline, and Robert A. Cribbie. 2009. "The Expanding Role of Quantitative Methodologists in Advancing Psychology." *Canadian Psychology/Psychologie Canadienne* 50 (2): 83.
- Haslam, S. Alexander, and Craig McGarty. 2014. *Research Methods and Statistics in Psychology*. Sage.
- Hoekstra, Rink, Henk Kiers, and Addie Johnson. 2012. "Are Assumptions of Well-Known Statistical Techniques Checked, and Why (Not)?" *Frontiers in Psychology* 3: 137.
- Howitt, Dennis, and Duncan Cramer. 2017. *Understanding Statistics in Psychology with SPSS*. Pearson London, UK:
- Keselman, Harvey J., Carl J. Huberty, Lisa M. Lix, Stephen Olejnik, Robert A. Cribbie, Barbara Donahue, Rhonda K. Kowalchuk, Laureen L. Lowman, Martha D. Petoskey, and Joanne C. Keselman. 1998. "Statistical Practices of Educational Researchers: An Analysis of Their ANOVA, MANOVA, and ANCOVA Analyses." *Review of Educational Research* 68 (3): 350–86.
- Lakens, Daniel. 2021. "The Practical Alternative to the p Value Is the Correctly Used p Value." *Perspectives on Psychological Science* 16 (3): 639–48.
- Meehl, Paul E. 1990. "Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles That Warrant It." *Psychological Inquiry* 1 (2): 108–41.

- Mills, Laura, Eva Abdulla, and Robert Cribbie. 2010. "Quantitative Methodology Research: Is It on Psychologists' Reading Lists?"
- Nunnally, Jum. 1960. "The Place of Statistics in Psychology." *Educational and Psychological Measurement* 20 (4): 641–50.
- Osborne, Jason W., and William R. Christianson. 2001. "Educational Psychology from a Statistician's Perspective: A Review of the Quantitative Quality of Our Field."
- Osborne, Jason W., and Elaine Waters. 2002. "Four Assumptions of Multiple Regression That Researchers Should Always Test." *Practical Assessment, Research, and Evaluation* 8 (1): 2.
- Sharpe, Donald. 2013. "Why the Resistance to Statistical Innovations? Bridging the Communication Gap." *Psychological Methods* 18 (4): 572.
- Steyn, H. S. 2000. "Practical Significance of the Difference in Means." *SA Journal of Industrial Psychology* 26 (3): 1–3.
- Student. 1908. "The Probable Error of a Mean." *Biometrika*, 1–25.

4 Annexe(s)

4.1 Annexe A: additional tables

Insert content for additional tables here.

4.2 Annexe B: additional figures

Insert content for additional figures here.

4.3 Annexe C: code

Insert code (if any) used during your dissertation work here.