

Introduction

Pourquoi les stat sont importantes en psycho

Comme dans toute discipline scientifique, les statistiques jouent un rôle très important en psychologie. *"Quantitative methodology [including statistics], broadly defined, occupies a unique and ubiquitous role in the PhD curriculum in psychology [...] quantitative methodology may be the one aspect of doctoral education that continues to unify the discipline of psychology.*

All the conclusions derived from this discipline proceed (or should proceed) from the application of a comprehensive and reliable system called the scientific method. This method is based on the progressive accumulation of evidence using different mathematical resources. Quel que soit le domaine étudié, nous avons besoin des outils mathématiques afin de déterminer si les différences observées dans notre échantillon peuvent être généralisées à une population plus large. @aiken_doctoral_2008

- https://books.google.fr/books?hl=nl&lr=&id=UfGGAwAAQBAJ&oi=fnd&pg=PP1&dq=how+psychologists+use+statistics&ots=gOEL7MeuxL&sig=xtL9KQOfslt_ik6MWefBqu8qOVQ#v=onepage&q=how%20psychologists%20use%20statistics&f=false
- (p.3): https://books.google.fr/books?hl=nl&lr=&id=9p82-ZZeMIsC&oi=fnd&pg=PR20&dq=how+psychologists+use+statistics&ots=hxs_zL6l6f&sig=UziO6G6ncpFZLLvO48cDqMIj3So#v=onepage&q&f=false
- introduction 1.1: https://books.google.fr/books?hl=nl&lr=&id=IoZ5AgAAQBAJ&oi=fnd&pg=PP1&dq=why+psychologists+need+statistics&ots=o5LWHSGE9_&sig=o0DfVUdATAgCdR2SWSPn1_0ENF0#v=onepage&q=why%20psychologists%20need%20statistics&f=false
- file:///C:/Users/Admin/Downloads/Dennis%20Howitt_2017.pdf

Les chercheurs se limitent à des tests de base. Pourquoi?

Au cours des 50 dernières années, on a vu se développer des méthodes statistiques de plus en plus complexes dans le domaine de la psychologie: méta-analyses, équations structurelles, modèles linéaires hiérarchiques... Ces méthodes gagnent en popularité, grâce à des logiciels conviviaux et des textes accessibles pour les psychologues qui facilitent l'usage de ce type d'outils, [@sharp_why_2013; @aiken_doctoral_2008]. A contrario, d'autres formes d'innovation ont beaucoup plus de difficultés à susciter l'intérêt. C'est le cas pour ce qu'on appelle les *tests robustes*, c'est-à-dire les tests qui ne sont pas soumis à des conditions d'application aussi strictes que les tests paramétriques traditionnels.

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].

Très traditionnellement, lorsque les chercheurs en psychologie désirent comparer deux ou plusieurs groupes, ils utilisent la moyenne comme indicateur de tendance central sur base de laquelle se feront les comparaisons. C'est soit un test t soit une ANOVA. Les alternatives ont nettement moins de succès [@keselman_statistical_1998]. Remarque: @Kelman_statistical_1998 ajoute que les post hoc de Tukey sont les préférés aussi.

Ce n'est pas récent: déjà il y a plus de 50 ans, @nunnally_place_1960 soulignait que la plupart des résultats en psycho étaient reportés soit sous forme de comparaison de moyennes, soit sous forme de coefficient de corrélation. Dans une enquête en 2008, @aiken_doctoral_2008 relevait d'ailleurs que les cours d'introduction aux statistiques couvraient principalement les ANOVA et les régressions multiples (aux USA et au Canada pour ce texte).

PARLER DES DIFFERENTES REVUES DE LITTERATURE QUI LE DISENT.

DONNER QUELQUES EXEMPLES. ESSAYER DE TROUVER DES ARTICLES QUI TRAITENT DES PRATIQUES LES PLUS COURAMMENT UTILISEES. Voir s'il n'y aurait pas une enquête ou autre. Regarder aussi dans les livres de stat quels sont les tests les plus énoncés. Très classiquement, à travers ces tests, on postule que tous les échantillons sont extraits de populations dont les moyennes sont identiques

($H_0 : \mu_1 = \mu_2 = \dots = \mu_k$) et si la p -valeur est inférieure au risque alpha, on conclut au rejet de l'hypothèse nulle. Bien qu'il existe plusieurs tests de comparaisons de moyennes, il y a un grand focus sur le test t -de Student et l'ANOVA de Fisher. Il existe pourtant des alternatives plus robustes mais paradoxalement, leur adoption est très compliquée *"Resistance by substantive researchers to statistical innovations is all the more puzzling because it is not universal. Some statistical innovations (e.g., meta-analysis, structural equation modeling) are adopted rapidly even over strong initial objections. Other statistical innovations (e.g., power analysis) are resisted for a long period of time but adopted eventually. And some statistical innovations (e.g., robust statistics) encounter resistance that appears intractable."* [sharpe_why_2013].

Qu'est-ce qui pourrait expliquer cela? 1) @sharpe_why_2013: lack of awareness (p.573) Manque de conscience des développements dans le domaine?

2) @sharpe_why_2013: journal editors (p.573) Les éditeurs ne poussent pas assez? -> Pas convaincue que ça m'intéresse

3) @sharpe_why_2013: Publish or perish? (p.574) je ne comprends mm pas en quoi c'est un argument

4) @sharpe_why_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. 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_why_2013: inadequate education (p.574)

6) @sharpe_why_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).

échec à checker les hypothèses des tests

Les tests t et ANOVA ont la fausse réputation d'être simples et maîtrisés par la plupart. EN 2008, dans une enquête auprès de professeurs de 201 départements conférant des programmes de doctorat en psychologie, @aiken_doctoral_2008 ont mis en évidence que les ANOVA étaient enseignées dans pratiquement tous les programmes (95% des programmes). La plupart (entre 80% et 100%) des personnes interrogées affirmaient que la plupart (si pas tous) leurs étudiants étaient capables de réaliser une ANOVA à un ou plusieurs facteurs (approche a priori ou post hoc).

Ces tests sont faussement simples, car ils reposent sur des notions importantes et complexes telles que la distribution d'échantillonnage et la p -valeur, notions malheureusement encore mal comprises par nombreux chercheurs. Bien comprendre celle-ci requiert du temps et de la patience. C'est un travail de longue haleine. Il m'a fallu des années de simulations et d'études pour réellement comprendre des tests qui semblent aussi simples que le test t , alors que je suis passionnée. Cela montre à quel point il est important pour les psychologues que des personnes prennent le temps de vulgariser le travail pour eux. Des outils comme Jamovi sont mis en place afin de faciliter la vie des chercheurs. Pour continuer à alimenter et mettre à jour ces outils le mieux possible, d'une manière pas trop douloureuse pour le chercheur lambda, il est important que des méthodologistes continuent à jouer le pont entre psychologues et statisticiens comme j'essaye de le faire.

Les tests statistiques conduiront à une interprétation valide à condition de respecter une ou plusieurs conditions d'application. Un non respect des conditions d'application pourra sérieusement affecter le taux d'erreur de types I et II des tests. Pourtant, ces conditions ne sont que peu ou pas vérifiées [@hoekstra_are_2012].

@keselman_statistical_1998: sur 61 articles contenant des designs dans lesquels des échantillons indépendants étaient définis par un seul facteur (between-subjects univariate designs, 13 n'ont pas donné d'information relatives aux écart-types. Parmi les autres, il y avait de l'hétéroscédasticité. En tout, moins d'un article sur 5 se préoccupait des violations possibles des conditions d'application, et lorsqu'ils le faisaient, ils s'inquiétaient surtout de la condition de normalité (moins de l'hétéroscédasticité), et ce malgré le fait que pourtant, des violations de cette conditions sont bien plus dommageables!

-> POURQUOI ELLES NE LE SONT PAS? 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 leur tests: càd comprendre la distribution d'échantillonnage, dont pratiquement tt découle.

Ce constat est loin d'être récent. Par exemple, @keselman_statistical_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 pour cette catégorie de tests, alternatives peu ou pas utilisées

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_are_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, mais ces derniers ne sont que peu ou pas utilisés [@sharpe_why_2013]

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_statistical_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.

- 2) 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_statistical_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).