

# Introduction

## Pourquoi les stat sont importantes en psycho

Dans presque toutes les études en psychologie, les statistiques jouent un rôle très important. *"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](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](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](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

## L'ANOVA et les tests $t$ en psychologie.

Malgré un développement inouï de méthodes statistiques de plus en plus complexes au cours des 50 dernières années (méta-analyses, équations structurelles, modèles linéaires hiérarchiques... @sharp\_why\_2013), les ANOVA et les régressions restent les stars en psychologie. Ces test se retrouvent au coeur des programmes de statistiques au sein des pratiquement toutes les facultés de psychologie (vu pour le Canada et les USA en 2008) [@aiken\_doctoral\_2008; @golinski\_expanding\_2009]. Par ailleurs, la majorité des études en psychologie continuent à reposer sur ces outils. (mettre quelques REVIEWS QUI LE DISENT... IL Y EN A PLEINS)

- ...  
- ...  
- ...

Cela est loin d'être récent. @nunnally\_place\_1960 en faisait déjà le constat il y a plus de 50 ans. De plus, on ne compte plus les livres d'introductions aux statistiques pour psychologues faisant mention de ces tests (en citer quelques uns bien connus). essayer de trouver des livres qui en parlent théoriquement et d'autres qui les mettent en lien avec des logiciels connus.

Compte tenu de tout cela, nous serions en droit de croire que ces tests sont archis connus et parfaitement maîtrisés. C'est une croyance communément partagée. Par exemple, parmi les professeurs interrogés en 2008 par @aiken\_doctoral\_2008 au sein de 201 départements, professeurs conférant des programmes de doctorat en psychologie aux USA et au Canada, presque tous (entre 80% et 100%) affirmaient que la plupart (si pas tous) leurs étudiants étaient capable de réaliser une ANOVA à un ou plusieurs facteurs (que ce soit l'approche a priori ou post hoc). Mais n'est-ce pas un peu trop optimiste?

RELIRE CECI POUR VOIR QUOI FAIRE DE CETTE IDEE: 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.

## La manière dont l'ANOVA et les tests $t$ sont utilisés en psychologie.

Les tests  $t$  et ANOVA ont la fausse réputation d'être simples et maîtrisés par la plupart. Pourtant, encore aujourd'hui, des nombreux manquements dans l'usage de ces tests peuvent être soulignés. C'est ce que nous tentons de faire à travers cette thèse.

Très classiquement, lorsqu'on compare 2 ou plus de 2 groupes définis par les modalités d'un ou plusieurs facteurs catégoriels, on "résume" la distribution des scores au sein de chaque groupe par un indicateur de tendance centrale (la moyenne), et on postule que tous les échantillons sont extraits de populations dont les moyennes sont identiques ( $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ ). On réalise un test  $t$  (de Student) ou une ANOVA (de Fisher) et si la  $p$ -valeur est inférieure au risque alpha, on conclut au rejet de l'hypothèse nulle.

Bien sûr, la moyenne n'est pas le seul indicateur de tendance centrale et des tentatives ont été faites pour introduire d'autres indicateurs dans les tests (citer Wilcoxon par exemple), mais ces alternatives ont eu nettement moins de succès [Keselman\_statistical\_1998]. (Remarque: Keselman\_statistical\_1998 ajoute que les post hoc de Tukey sont les préférés aussi). Et même en se limitant aux comparaisons de moyennes, il existe des alternatives au test  $t$  de Student et à l'ANOVA de Fisher (cf. tests robustes), mais leur adoption est très compliquée.

### Limite 1: conditions d'application

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?

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 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.

## Limite 2: hypothèse nulle

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).

## 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\_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 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\_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 tlm).

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: c-à-d comprendre la distribution d'échantillonnage, dont pratiquement tt découle.