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

1. "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 specifiy 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!)
2. Exemple pour montrer que le caractère « arbitraire » des mesures en psycho n’est pas forcément un souci pour déterminer l’importance d’un effet. le BDI (Beck Depression Inventory): échelle à 21 items (chaque item rapporte entre 0 et 3 points), on parle de dépression sévère quand score entre 30 et 63. Une diminution de 14 points (en passant de 40 à 26) pourrait être interprété par une personne qui connait bien cette échelle (Pek et Flora) 🡪 ça montre qu’une ES non standardisée peut être très parlante ! Ce qui compte surtout, c’est le consensus de la part des chercheurs sur ce qui représente un grand effet pour une échelle en particulier . D’après Pek et Flora.
3. Dire « je m’intéresse au sens d’un effet mais pas forcément à l’amplitude de la différence » n’a pas vraiment de sens. En effet, il existe un lien mathématique très évident entre la taille d’effet et la significativité : dire « j’accepte un effet comme étant significatif s’il est associé à une pvaleur inférieure à .05 » revient exactement au même, mathématiquement parlant, que de dire « j’accepte un effet comme étant significatif s’il est associé à une taille d’effet supérieure ou égale à XXX » (varie en fonction du n). Donc même si c’est sans s’en rendre compte, on s’intéresse à la taille d’effet (Funder et al. 2019).
4. Il peut parfois être argumenté que le sens d’un effet intéresse plus que son amplitude, dans une réplication. C’est le cas par exemple si l’effet a une direction très surprenante (on ne s’attendait pas à observer un effet et donc on tente de répliquer sa significativité et sa directionnalité ; Anderson & Maxwell, 2016).
5. Je ne le retiens jamais pas H0 et la pvaleur, c’et Fisher, et H1, le risque alpha et la puissance, c’est Neyman-Pearson.
6. On constate que les mesures de taille d'effet ne sont toujours pas appréciées à leur juste valeur et sont souvent mal comprises, même par les professionnels [@funder\_evaluating\_2019]. En conséquences, bien qu'on enseigne couramment aux étudiants comment tester la significativité des effets, il est plus rare qu'on leur enseigne comment calculer les tailles d'effets, et encore plus rare qu'on leur enseigne comment évaluer les mesures obtenues. Même lorsque les mesures sont interprétées, elles le souvent souvent sur base de balises qui n'ont aucun sens "dans l'absolu", sans cadre de référence (ex.: r = .10 = petit effet; r=.30 = effet moyenne, etc.). Petit ou moyen par rapport à quoi? Si on veut utiliser des balises, il faut le faire en comparnat aux résultats obtenus dans d'autres études. On peut dire en croisant quelqu'un dans la rue "il est petit" ou "il est grand" parce qu'on le compare à l'ensemble des humains. De la même manière, on pourrait dire si un effet est petit ou grand en comparaison aux autres effets observés. Plusieurs auteurs tels que que Richard et al (2003, cités par Funder et al. 2019) ou plus récemment Gignac & Szodorai (2016, cités par Funder et al. 2019) ont fait de grosses revues méta-analytiques allant dans ce sens. Attention: ils ont trouvé en moyenne un r de .21 par exemple, mais faut pas oublier le biais de publication (donc on sait que si on a un effet de .21, c'est déjà plus grand que bcp d'effet.. Funder et al (2019) ont pris cette information en compte en proposant leur nouvelle benchmark dans leru article.) \*On a tenté d'expliquer la notion d'ES le plus clairement possible. Et bien que nos comparaisons reposent essentiellement sur des critères inférentiels, nous avons tenté de garder la dimension interprétative à l'esprit à travers notre manuscript\*. \*Il y a eu pas mal de discussions pour savoir comment améliorer l'interprétation des mesures (ex.: le binomial effect-size display, ou la propostiion de Benchmark faite par FUnder et al. (2019)).\*
7. D'après @lakens\_practical\_2021, un test d'hypothèse (selon l'approche de Nayman-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 rechercher, modifier une manipulation, ou de faire un certain type de déclaration ou revendication. \*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 ann 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\_practical\_2021].
8. Si on me demande un exemple pratique où étudier la distribution d’échantillonnage peut aider à mieux comprendre des résultats :

Imaginons un cas où on fait un test t pour comparer deux groupes de même moyenne et même écart-type mais avec un échantillon extrait d’une distribution avec asymétrie positive et l’autre avec asymétrie négative. La distribution d’échantillonnage de la moyenne aura une asymétrie de même signe que la distribution de population dont sont extraits chaque échantillon. Du coup :

* La moyenne du groupe avec asymétrie positive sera parfois très très surestimée
* La moyenne du groupe avec asymétrie négative sera parfois très très sous-estimée
* Cela peut fatalement altérer l’estimation de la différence de moyenne. Et donc influencer la puissance du test.

**Echelles de Likert : pistes de degrés de liberté**

1. D'abord, il existe de nombreuses variations des items de type Likert, en fonction du nombre de modalités de réponses proposées[[1]](#footnote-1), du fait d'inclure ou non la possibilité d'opter pour une position neutre ou une absence de position ou encore du fait de proposer un nombre identique de modalités de réponse de part et d'autre de la position neutre (échelles symétriques) ou pas (échelles asymétriques,@joshi\_likert\_2015).
2. Ensuite, il faut réfléchir à la manière de générer les items
   1. En utilisant une loi multinomiale et en spécifiant pour chaque item la distribution de probabilité ? ex. pour un item à 5 niveaux quand on s’attend à un taux d’accord très élevé :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | Pas du tt d’accord | Pas d’accord | Neutre | D’accord | Tout à fait d’accord |
| Proba | .00 | .01 | .03 | .06 | .90 |

* 1. En partant d’une loi continue et en discrétisant les réponses ?

1. Enfin, il faut réfléchir au nombre d’items qu’on combine et à la manière dont on le fait.

Enfin, il est possible de faire varier le nombre d'items constituant l'échelle de Likert (@subedi\_using\_2016 recommande d'en combiner au moins 4 pour utiliser du paramétrique). Et bien sûr, tout cela doit se faire en garantissant le contrôle des paramètres de population de l'échelle finale (forme, indicateurs de tendance centrale et de dispersion). Il faut aussi réfléchir au degré d'homogénéité entre les items. C'est donc extrêmement complexe.

1. A l'origine, les items de type Likert se composaient de 5 modalités sémantiques: "Tout à fait d'accord","D'accord","Sans opinion", "Pas d'accord", "Pas du tout d'accord" [@boone\_analyzing\_2012] [↑](#footnote-ref-1)