

Annexes

Annexe A: erratum de l'article "Why psychologists Should by Default Use Welch's t -test Instead of Student's t -test" (Chapitre 2)

Erreurs conceptuelles Nous spécifions à plusieurs reprises que le test t de Yuen contrôle moins bien le taux d'erreur de type I que le test t de Welch:

- p.14: "*Yuen's t -test is not a good unconditional alternative because we observe an unacceptable departure from the nominal alpha risk of 5 percent for several shapes of distributions [...] particularly when we are studying asymmetric distributions of unequal shapes*";
- p.15: "*As it is explained in the additional file, Yuen's t -test is not a better test than Welch's t -test, since it often suffers high departure from the alpha risk of 5 percent*".

Ceci n'est pas exact d'un point de vue purement statistique. 'A travers le test de Yuen, on ne compare plus les moyennes de chaque groupe, mais les moyennes *trimmées* (soit les moyennes calculées sur les données après avoir écarté les 20% des scores les plus bas ainsi que les 20% des scores les plus élevés). Or, à travers nos simulations, les scénarios créés en vue de tester le taux d'erreur de type I (risque alpha) étaient systématiquement des scénarios dans lesquels les moyennes de chaque groupe étaient identiques. Lorsque les échantillons sont extraits de population parfaitement symétriques, il y aura vraisemblablement peu de différence entre les estimations des moyennes et des moyennes trimmées (puisqu'au niveau de la population, les moyennes et moyennes trimmées seront identiques). Quand les échantillons sont extraits de population asymétriques, par contre, on observera plus vraisemblablement des différences entre les estimations des moyennes et des moyennes trimmées. Au niveau de la population, les moyennes trimmées seront plus proches du mode des distributions que les moyennes et représenteront donc mieux les distributions.

Notons malgré tout que d'un point de vue méthodologique, nous avons déjà relevé que la plupart du temps, les chercheurs définissent l'absence de différence entre les moyennes comme hypothèse nulle et dans ce contexte, nos simulations démontre que le test de Yuen n'est pas approprié. En conclusion, le test de Yuen ne devrait être utilisé que par des chercheurs ayant pleinement conscience du fait que le test t de Student et de Welch ne testent pas la même hypothèse que le test t de Yuen.

Commentaires divers

- p.9: nous décrivons 3 arguments en défaveur de l'usage du test de Levene. En troisième argument, nous mentionnons le manque de puissance du test de Levene. Nous ne mentionnons cependant pas le fait qu'utiliser le test t de Student lorsque le test de Levene est non significatif revient à confondre le non rejet de l'hypothèse d'égalité des variances avec l'acceptation de l'hypothèse d'égalité des variances. Au sein du chapitre 5 sur les tests d'équivalence, il est démontré par simulation que même lorsqu'on s'assure d'avoir une puissance suffisante pour détecter une différence attendue, la stratégie qui consiste à interpréter le non rejet de l'hypothèse nulle comme un soutien en faveur de l'hypothèse nulle n'est pas appropriée.
- p.12: nous mentionnons ceci : "*When both variances and sample sizes are the same in each independent group, the t -values, degrees of freedom, and the p -values in Student's t -test and Welch's t -test are the same (see Table 1).* Par "variances" il faut comprendre "*sample* variances*" ou "*variances estimates*". Nous ne sommes donc pas* en train de dire que les deux statistiques, ainsi que les degrés de liberté et p -valeurs qui leur sont associées seront identiques lorsque la condition d'homogénéité des variances sera respectée au niveau de la population, mais bien lorsque les estimations de chaque variance de population seront identiques.

Mise en forme et Notations Les lettres utilisées pour décrire les statistiques (test- t ou test- F) doivent toujours être inscrites en *italique*. Or, cela a été omis à plusieurs reprises dans l'article. Par exemple, il aurait fallu écrire:

- p.9: "... as the Mann-Whitney U -test..." au lieu de "... as the Mann-Whitney U-test...";
- p.9: " F -ratio test" au lieu de "F-ratio test".

Certaines notations mathématiques auraient également dû être indiquées en italique. Par exemple, à la p.9, il aurait fallu écrire:

- " x_{ij} " au lieu de " x_{ij} ";
- $|x_{ij} - \hat{\theta}_j|$ au lieu de $|x_{ij} - \hat{\theta}_j|$.

Par ailleurs, il est très important d'être consistant dans le choix des notations mathématiques, afin d'éviter d'embrouiller le lecteur. Or, nous n'avons pas toujours respecté cela. Par exemple, nous avons utilisé plusieurs notations différentes pour décrire l'écart-type/la variance. Par exemple:

- p.9: nous utilisons respectivement SD1 et SD2 pour décrire l'écart-type de chaque groupe;
- p.11 (équation 1): nous utilisons respectivement S_1^2 et S_2^2 pour décrire la variance de chaque groupe;
- p.12 (équation 4): nous utilisons respectivement s_1^2 et s_2^2 (lettres minuscules) pour décrire la variance de chaque groupe.

C'est d'autant plus problématique qu'il y a parfois même des inconsistances entre les notations utilisées dans les formules et celles utilisées dans les légendes des formules. Par exemple, nous spécifions p.11 que dans l'équation 1, s_1^2 et s_2^2 (lettres minuscules) représentent les estimations de variance de chaque groupe indépendant, alors qu'en réalité, les estimations des variances sont représentées par S_1^2 et S_2^2 (lettres majuscules) dans l'équation 1.

Faute(s) de frappe

- p.13: nous avons indiqué "see v 2a" au lieu de "see Figure 2a".

Taking parametric assumptions very seriously: Arguments for the Use of Welch's F -test instead of the Classical F -test in One-Way ANOVA (Chapitre 3)

Mise en forme et Notations Une légende est manquante pour certaines notations mathématiques. Par exemple, en ce qui concerne l'équation (1), bien que n_j , k et s_j^2 aient été correctement définis, les définitions pour \bar{x}_j , \bar{x}_\cdot et N ne sont données que plus tard, en référence à d'autres équations. Cela peut rendre la lecture de l'article plus compliquée pour certaines personnes non familières avec ces notations.

Par ailleurs, comme dans l'article précédent sur le test t de Welch, on constate certaines incohérences en termes de notation. Par exemple, si la moyenne de chaque groupe est définie par \bar{x}_j dans l'équation (1), elle est définie par \bar{X}_j dans l'équation (7).

Certaines notations mathématiques auraient également dû être indiquées en italique. Par exemple, à la p.9, il aurait fallu écrire:

- " x_{ij} " au lieu de " x_{ij} ";
- $|x_{ij} - \hat{\theta}_j|$ au lieu de $|x_{ij} - \hat{\theta}_j|$.

Dû à un manque de connaissance de Latex lors de mes premières tentatives d'écritures d'articles via Rmarkdown, certaines majuscules sont manquantes dans les références bibliographiques. S'assurer qu'une lettre apparaisse en majuscule, via latex, implique de l'entourer des symboles $\{\}$, ce qui n'a pas été fait. Par exemple, dans le titre de l'article de Tiku(1971), il aurait fallu indiquer "Power function of the $\{F\}$ -test.." au lieu de "Power function of the F-test...". Cela ne serait pas arrivé, si j'avais utilisé un outil comme Zotero, afin d'exporter directement un fichier au format Bibtex (puisque via ces outils, ce genre de détail est automatiquement inclu), mais je n'ai découvert cette possibilité que très récemment.

Faute(s) de frappe

- p.18: "Although it is important to make sure ~~test~~ **that** assumptions are met";

- p.19: "... we think that a ~~first~~ realistic first step towards progress would be to get researchers...";
- p.20: "Based on mathematical explanations and Monte Carlo simulations";
- p.21: "~~With~~ with $N = \dots$ ";
- p.21: "~~Where~~ where \bar{x}_j and s_j^2 are respectively the group mean and the group variance...";
- p.22: "... negative pairings (the group with the ~~smallest~~ largest sample size is extracted from the population with the smallest SD);
- p.22: "the type I error rate of all tests";
- p.24: "... which is either more liberal or more conservative, depending on the SDs and ~~SD~~ **sample sizes** pairing";

Effect sizes

Equivalence tests

Annexe B

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