

Petite note sur l'ajustement des pvalues.

Rappel sur le petit p.

Avec un petit p à 0.0042 on peut dire que le test est significatif avec un seuil alpha de 5% mais on ne peut pas dire qu'il est significatif au seuil de 1%, même si la pvalue du test vaut 0.0042\$.

En effet, le **risque de première espèce**, noté α , **se fixe a priori** - i.e. avant de réaliser le test. La **significativité** quant à elle, représentée par le **petit p** est une **mesure du degré de confiance avec lequel on peut rejeter H0 avec les données qu'on a recueillies**. Si toutefois on veut afficher dans nos travaux une pvalue $p=0.0042 < 1\%$, alors il faut choisir en amont du travail, avant de réaliser le test, un risque d'erreur de 1% - i.e. un seuil de significativité de 1%.

Dilatation de la pvalue.

Ainsi, si on fixe a priori un seuil de significativité à 5% alors le rejet de l'hypothèse nulle se fera sur la base de ce seuil. Par exemple, si lors d'un test d'homogénéité de deux moyennes on obtient une pvalue de 0.0001, on pourra dire que le test est significatif au seuil de 5%. Autrement dit, on a 5% de chances de rejeter H0 à tort (il s'agit du fameux faux positif ou risque de première espèce ou erreur de type I). Toutefois, ce risque n'est pas seul puisque **sa contrepartie** est qu'on a **95 chances sur 100 de ne pas rejeter H0 à tort** - i.e. risque de type II ou de seconde espèce ou faux négatif. Ces deux risques sont liés.

On a donc la relation suivante,

$\alpha = 0.05$ $\alpha = 1 - (1 - 0.05)$ $\alpha = 1 - \beta$ où α et β représentent, respectivement, le risque de première et de seconde espèce.

Néanmoins, si pour un test on fixe le seuil de significativité à 5% (soit un risque de seconde espèce de 95%) ce seuil ne reste pas fixe à mesure qu'on accroît le nombre de tests à réaliser. En effet, **le risque de ne pas rejeter H0 à tort** (l'erreur de type 2) s'accroît comme suit : si on réalise **3 tests**, dans le cas d'une comparaison de trois moyennes, ce risque sera $(1 - \alpha)^3$. Et puisque **ce risque est lié au risque de première espèce**, le seuil de significativité sera : $1 - (1 - \alpha)^3$. Autrement dit, **à mesure que le nombre de tests à réaliser augmente, le risque alpha augmentera comme suit : $\alpha = 1 - (1 - \alpha)^k$** , avec k le nombre de tests à réaliser. Donc, les chances de conclure à tort vont s'accroître à mesure que le nombre de tests augmente : il faut donc ajuster les pvalues des tests pour nous assurer qu'on ne conclura pas à tort avec une probabilité supérieure à 5% (si tel est le risque d'erreur fixé a priori).

Ajustement de la pvalue.

C'est la toute l'intérêt des procédures d'ajustement des pvalues : nous assurer que le risque d'erreur α ne se dilate pas trop. Une procédure d'ajustement des pvalues est donc nécessaire pour les tests post-hoc après une Anova ou un test de Kruskal-Wallis. Différentes familles de méthodes existent, parmi lesquels la plus connue : la méthode de Bonferroni ; cette méthode n'est pas un choix idéal, il vaut mieux utiliser la méthode de Holm.