Recette de l'Anova

Difficulté : **

Ingrédients :

Data

Materiel :

Jeu de données organisé en lignes (individus) et colonnes (variables) *[csv ou Excel]*

Une VD

Une (ou pls) VI(s)

Un ordinateur quelconque

Logiciel R avec l'interface Rcmdr chargée

Procédure :

- Extraire les données d'une base de données dans un fichier csv organisé en lignes et en colonnes.
- Repérer l'information qui est associé à la variable dépendante repérer la(les) variable(s) indépendante(s).
- Ouvrir R puis charger Rcmdr >> importer les données du jeu de données dans un *data frame (structure qui permet de stoker un jeu de données dans un logiciel)* en suivant bien les options de l'import (type du fichier, séparateur de champs, virgule ou point décimal). Donner un nom explicite à ce *data frame* en fonction du contenu et de son contexte.
- Réaliser un graphe de type boîte à moustaches (*boîte de dispersion, diagramme en boîte, boxplot, diagramme de quartiles*) pour visualiser les variations de la variable dépendante (quantitative) en fonction des modalités de la variable indépendante (qualitative) >> procéder ainsi pour chacune des variables indépendantes identifiées dans le jeu de données.
- Procéder alors au test de l'analyse de variance (ANOVA à un ou 2 facteurs selon les conditions de votre étude), en sélectionnant le menu concernant la comparaison de moyenne puis ANOVA a un facteur *[Statistiques > Moyennes > Anova à un Facteur]* puis sélectionner la variable dépendante (variable réponse) puis le Groupe associé à la variable indépendante et sélectionner si nécessaire les tests post-hoc associés de comparaison par paires (Comparaisons multiples des moyennes, tests non planifiés de Student).
- Définir un nom explicite pour cette analyse dans le cadre en haut (« Entrez un nom pour le Modèle »)
- Lancer le calcul en appuyant sur OK puis analyser les résultats.

Voilà c'est prêt, ce n'est pas plus compliqué que ça ! Il ne vous reste plus qu'à vous régaler.

A noter :

Il faudra sûrement analyser les résidus de cette Anova et au préalable faire un test de normalité de la distribution de la variable dépendante dans chacun des groupes considérés. On préférera le test de Shapiro-Wilk pour les petits échantillons et le test de Kolmogorov-Smirnov pour les échantillons plus importants (n> 50).

Il est également important de vérifier si toutes les données sont complètes, de procéder à une politique de gestion des données manquantes et de chasser les doublons et enfin de regarder si nous avons suffisamment de données pour pouvoir procéder à un test paramétrique (il faudrait idéalement avoir au moins 30 individus dans chacun des groupes considérés pour le test Anova).