

TD n°3

Centralité, dispersion

Comparaison entre groupes

Nous allons calculer la moyenne, la médiane, la variance, l'écart type et les quartiles pour 2 variables. Au final, vous devrez dresser le tableau suivant :

Indice	Age	GraviteCopier
Moyenne
Médiane
Variance
Ecart type
Quartile 0
Quartile 1
Quartile 3
Quartile 4

Mais avant d'utiliser un ordinateur, petit travail manuel...

1 Manuellement

Avant de commencer avec Excel, petit exercice "à la main" (sur une feuille de papier libre avec si besoin l'aide d'une calculatrice) :

1. Les notes du groupe 1 sont 8,9,10 et 13. Calculez la moyenne du groupe.
2. Les notes du groupe 2 sont 4,7,10 et 19. Calculez la moyenne du groupe.
3. Pronostic : Sans faire de calcul, qui du groupe 1 ou du groupe 2 a la plus grande variance ?
4. Vérification : calculez la variance et l'écart type du groupe 1, puis du groupe 2. Cela confirme-t-il ce que vous aviez pronostiqué ?

2 Analyse univariée

2.1 Centralité

On ne peut calculer une moyenne que pour des variables numériques. Pour cela, on utilise la fonction `moyenne()`. Par exemple, `moyenne(H2 :H5)` calculerait la moyenne des cellules comprises entre H2 et H5. Si vous avez donné un nom à une plage de cellule, vous pouvez aussi utiliser `moyenne(nomPlageCellule)`.

Pour calculer la médiane, on utilise la fonction `mediane()`. `mediane(H2 :H5)` calculerait la médiane des cellules comprises entre H2 et H5.

5. Calculez les moyennes des variables `Age` et `GraviteCopier`.
6. Calculez la médiane des variables `Age` et `GraviteCopier`.

2.2 Variance et écart type

La variance se calcule grâce à la fonction `var.p()`. L'écart type se calcule grâce à `ecarttypep()`

7. Calculez la variance des variables `Age` et `GraviteCopier`.
8. Calculez l'écart type des variables `Age` et `GraviteCopier`.

2.3 Quartiles

Les quartiles se calculent grâce à la fonction `quartile()`. Cette fonction prend deux arguments, le premier est la plage de cellule dont on veut calculer le quartile. Le deuxième est le numéro du quartile. Par exemple, pour calculer le troisième quartile de la plage de cellule `maPlageCellule`, on utilise `quartile(maPlageCellule ;3)`

9. Calculez les cinq quartiles des variables `Age` et `GraviteCopier`.

3 Comparaison de groupes

3.1 Introduction

L'analyse univariée permet de connaître une variable. L'analyse bivariée consiste à étudier une variable *relativement* à une autre. Par exemple, on s'intéresse à la moyenne de `[GRAVITECOPIER]` pour les garçons et à la moyenne de `[GRAVITECOPIER]` pour les filles. Cela nous permet ensuite de voir qui fume le plus des garçons ou des filles. La variable qui permet de définir les groupes (ici `[SEXE]`) s'appelle *variable discriminante*.

Les indices *par groupe* se calculent grâce à une série de fonctions qui commence par `BD` (pour Base de Données). Par exemple, pour calculer la `moyenne()` par groupe, on utilise la fonction `bdmoyenne()`. Pour calculer la variance par groupe, on utilise `bdvariance()`. Et ainsi de suite.

3.2 Concrètement

Les fonctions `BD` nécessitent 3 arguments. Pour plus de clarté, nous supposons que nous cherchons à calculer la moyenne de `[GRAVITECOPIER]` pour les hommes. Naturellement, ce qui suit peut s'appliquer à n'importe quel indice et n'importe quelle variable.

- Le premier argument est la plage des cellules où se trouve les données. Elle doit contenir au moins les variables `[SEXE]` et `[GRAVITECOPIER]`, mais peut contenir plus de colonne. Le plus simple est de sélectionner *toutes* les colonnes et de les appeler *baseDonnees*.
- Le deuxième est le nom de la colonne dont on veut calculer la moyenne entre des guillemets. Dans notre cas, c'est `"14A-GraviteCopier"`.
- Le troisième doit contenir deux cellules : la première est le nom de la variable discriminante (ici `[2-SEXE]`) ; la deuxième est la modalité qui nous intéresse (ici `(Homme)`)

Ca sera plus clair avec un petit exemple :

10. Sélectionnez toute la base de donnée et donnez lui un nom (TD 2 nommer une plage de cellules), par exemple `baseDonnees`
11. Sur la feuille 3, créez un petit tableau (cases C2 :C3) qui contiendra dans C2 le nom de la variable discriminante (ici `[2-SEXE]`), juste en dessous dans C3 la modalité qui nous intéresse (ici `(Homme)`) :

2-Sexe
Homme

12. Tout est maintenant prêt pour calculer la moyenne du groupe selon la modalité `(Homme)`. La fonction est `=bdmoyenne()`, le premier argument (la base de données) est `baseDonnees`, le deuxième (colonne dont on cherche la moyenne) est `"14A-GraviteCopier"` (ne pas oublier les guillemets), le troisième (nom et modalité discriminante) est `C2 :C3`.

Au final : `=bdmoyenne(baseDonnees ;"14A-GraviteCopier" ;C2 :C3)` (à calculer dans la case C5).

13. Dans D5, calculez la moyenne de `[GRAVITECOPIER]` pour les femmes.
14. Dans C6 et D6, calculez l'écart type de `[GRAVITECOPIER]` pour les hommes et pour les femmes (fonction `=bdecartype()`)
15. Dans B5 et B6, décrivez le contenu des cases C5 :D6.
16. Qui en moyenne, des hommes ou des femmes, considère que copier est grave ?
17. D'après vous, les personnes qui ont déjà redoublé fraudent-elles plus que les autres ?
18. Pour le savoir, en C8 :D9, calculez la moyenne et l'écart type de `[GRAVITECOPIER]` selon le `[REDOUBLEMENT]`
19. Conclusion ? Ceux qui ont redoublé trichent-ils plus ?