



### Travaux pratiques : Analyse statistiques des données

Pour étudier le niveau de mémoire des sujets d'une population d'étudiants de l'UIE, on leur a demandé de mémoriser les mots d'une liste de 20 mots, de manière à se les rappeler par la suite. Pour chaque sujet, on a noté le nombre de mots correctement mémorisés et obtenu les résultats suivants :

Nombre de mots mémorisés $x$	10	11	12	13	14	15	16	17	18	19	20
Nombre de sujet $n$	13	29	0	18	13	8	6	5	3	2	1

1. Calculer l'effectif total appelé Effect\_total.
2. Pour mieux préparer l'espace de travail, nous commençons par charger pandas, numpy, matplotlib.pyplot
3. Insérer les données sous forme d'un tableau avec deux colonnes ("Nb\_mots", "Nb\_sujets").
4. Créer et calculer une nouvelle colonne appelée "Proportion\_Mots" qui représente la distribution des proportions du nombre de mots mémorisés.
5. Représenter graphiquement la distribution de la variable nombre de mots mémorisés.
6. Créer et calculer une nouvelle colonne appelée "Produit\_Mots\_Sujets" qui est le produit des deux colonnes précédentes.
7. Créer et calculer une nouvelle colonne appelée "Produit\_Mots\_Carres\_Sujets" qui est le produit des de la colonne "Nb\_mots" au carré par la colonne "Nb\_sujets".
8. Calculer le nombre moyen de mots mémorisés, la variance et l'écart-type du nombre de mots mémorisés en utilisant les colonnes du dataset et seulement np.sum et les formule suivantes :

$$\bar{X} = \frac{1}{N} \sum_{i=1}^n x_i * n_i \quad V(X) = \frac{1}{N} \sum_{i=1}^n x_i^2 * n_i - \bar{X}^2 \quad \sigma(X) = \sqrt{V(X)}$$

Soit la serie de données suivante :

```
Serie_exo = [10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10,
11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11,
13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 13,
14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 14, 15, 15, 15, 15, 15, 15, 15, 15,
16, 16, 16, 16, 16, 16, 17, 17, 17, 17, 17, 18, 18, 18, 19, 19, 20]
```

9. Tracer la boites à moustaches de cette série.
10. Calculer la moyenne, la variance, l'écart-type, les quantiles (10%, 25%, 50%, 75%, 90%) du nombre de mots mémorisés en utilisant les modules numpy (mean(), var(), std(), quantile()) pour la serie Serie\_exo.
11. Définir une fonction qui prend en entrée une serie de données et renvoie les valeurs suivantes : la moyenne, la variance, l'écart-type, les quantiles (10%, 25%, 50%, 75%, 90%) du nombre de mots mémorisés
12. Comparer les résultats de la fonction précédente avec celle du module pandas.describe() .

**N.B :** Chaque réponse doit être commentée et expliquée de façon claire et précise.