

Code de partage avec Capytale : a2c3-8736828

On utilisera les bibliothèques suivantes :

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Exercice 1 - un jeu de données pour s'échauffer

On va étudier les deux séries statistiques u et t :

```
u = [1.06, 0.44, 2.25, 3.88, 0.61, 1.97, 3.43, 2.10, 1.50, 1.68, 2.72, 1.35, 2.94, 2.78, 3.43, 3.58]
t = [2.58, 2.25, 2.90, 3.36, 2.41, 2.79, 3.32, 2.81, 2.62, 2.70, 3.17, 2.65, 3.07, 3.13, 3.07, 3.34]
```

1. Déterminer les moyennes et les écarts-type de ces séries statistiques (on pourra utiliser les fonctions prédéfinies dans `numpy`). On cherchera à créer des variables pour conserver les résultats.
2. Ecrire un programme qui permet de déterminer la covariance de u et t (on cherchera à conserver le résultat).
3. Déterminer le coefficient de corrélation linéaire de u et t
4. Déterminer la droite de régression linéaire reliant u et t
5. Représenter graphiquement le nuage de points de la série statistique (u, v) ainsi que la droite de régression linéaire.

Exercice 2 - avec un vrai jeu de données

On va étudier ici une possible corrélation entre le produit intérieur brut et l'espérance de vie à partir de données sur l'année 2020 pour 208 pays (données issues du *World Bank Group*).

Dans un premier temps, il faut importer le jeu de données. On commence donc par exécuter la commande suivante :

```
import pandas as pd
donnees=pd.read_csv('donnees_world_bank.csv', delimiter=';')
```

On rappelle quelques commandes possibles avec pandas : `.head()`, `.describe()`, `.mean()`, `.std()`, `donnees['nom_colonne']`

On procède ensuite comme à l'**exercice 1**.

1. Déterminer les moyennes et les écarts-type de ces séries statistiques (on pourra utiliser les fonctions prédéfinies dans `numpy`). On cherchera à créer des variables pour conserver les résultats.
2. Ecrire un programme qui permet de déterminer la covariance des deux variables à étudier (on cherchera à conserver le résultat).
3. Déterminer le coefficient de corrélation linéaire des deux variables à étudier.
4. Déterminer la droite de régression linéaire reliant les deux variables à étudier.
5. Représenter graphiquement le nuage de points de la série statistique des deux variables à étudier ainsi que la droite de régression linéaire.