

TP3-Statistiques bivariées

Durée : 2-3h

Dans tous le TP, Ω désigne une population et X, Y deux variables quantitatives discrètes sur cette population. On notera x la série statistique $(x_i)_{i \in [1, n]}$ des modalités de X sur un échantillon de taille n et y la série statistique $(y_i)_{i \in [1, n]}$ des modalités de Y sur ce même échantillon.

On suppose importée les bibliothèques `numpy` et `matplotlib.pyplot` grâce aux commandes :

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

1 Moyenne, covariance, coefficient de corrélation linéaire

Moyenne, variance, écart-type

Si les séries statistiques $(x_i)_{i \in [1, n]}$ et $(y_i)_{i \in [1, n]}$ sont représentées en Python par deux vecteurs x et y alors :

- `np.mean(x)` renvoie la moyenne \bar{x} de x ,
- `np.var(x)` renvoie la variance empirique de x ,
- `np.std(x)` renvoie l'écart-type de x .

Remarque 1

Il n'y a pas de formule au programme pour la covariance de (x, y) .

- À l'aide du cours donner la formule de Koenig-Huygens.

- Que renvoie la commande `x*y`?

- En déduire un moyen de calculer la covariance de (x, y) à l'aide `np.mean(x*y)`, `np.mean(x)` et `np.mean(y)`.

- Définir une fonction `corr` qui prend en argument deux vecteurs x, y et qui renvoie le coefficient de corrélation linéaire de (x, y) .

2 Représentation graphique

Nuage de points

En conservant les notations précédentes, la commande :

```
plt.plot(x,y,'.')
```

permet de représenter le nuage de points associé à la série statistique double $(x_i, y_i)_{i \in [1, n]}$.
L'option « '.' » permet d'afficher les points sous forme de ronds.

- À l'aide du cours donner l'équation de la droite de régression linéaire de Y en X.

- Définir une fonction `reg` qui prend en argument deux vecteurs `x`, `y` et qui trace le nuage de points et la droite de régression linéaire de `y` en `x`.

3 Exercices

Résoudre les exercices du TD5 avec Python.