

Les statistiques avec Numpy

Objectif :

Découvrir les fonctions Numpy propres à la statistique ainsi que la normalisation de séries.

Questions

- 1) Récupérer les données de sangliers.csv sous la forme d'un dataframe. Extraire du dataframe les ndarrays `nb_sanglier_preleves` et `nb_permis_chasse` (méthode `to_numpy()`).
- 2) a) Écrire la fonction `moyenne` qui renvoie la moyenne du ndarray à une dimension passé en paramètre, l'appliquer à `nb_sanglier_preleves` et comparer avec `np.mean`.
- b) Écrire la fonction `variance` qui renvoie la variance du ndarray à une dimension passé en paramètre, l'appliquer à `nb_sanglier_preleves` et comparer avec `np.var`.
On ajoutera `ddof=0` comme paramètre à la fonction Numpy.
Delta Degrees of Freedom = 0 impose de diviser par N dans la formule et non par $N - 1$.
La division par $N - 1$ correspond à la correction de Bessel qui élimine le biais s'il on part d'un échantillon et non de la population totale.
- c) Écrire la fonction `ecart_type` qui renvoie l'écart-type du ndarray à une dimension passé en paramètre, l'appliquer à `nb_sanglier_preleves` et comparer avec `np.std` (toujours avec `ddof=0`).
- d) Écrire la fonction `covariance` qui renvoie la covariance des deux ndarrays à une dimension passés en paramètre, l'appliquer à `nb_sanglier_preleves` et `nb_permis_chasse`.
Comparer avec `np.cov` (toujours avec `ddof=0`).
`np.cov(X, Y)` renvoie la matrice de covariance

$$\begin{pmatrix} \text{Cov}(X, X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Cov}(Y, Y) \end{pmatrix} = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & \text{Var}(Y) \end{pmatrix}$$

- e) Écrire la fonction `correlation` qui renvoie le coefficient de corrélation des deux ndarrays à une dimension passés en paramètre, l'appliquer à `nb_sanglier_preleves` et `nb_permis_chasse`.
Comparer avec `np.corrcoef` (`ddof=0` n'est pas nécessaire car N n'intervient plus dans la formule du coefficient de corrélation).
`np.corrcoef(X, Y)` renvoie la matrice de corrélation

$$\begin{pmatrix} \text{Cor}(X, X) & \text{Cor}(X, Y) \\ \text{Cor}(Y, X) & \text{Cor}(Y, Y) \end{pmatrix} = \begin{pmatrix} 1 & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & 1 \end{pmatrix}$$

- 3) a) Afficher le nuage de points correspondant au nombre de sangliers prélevés en fonction du nombre de permis, en ajoutant les légendes des axes et en faisant partir des deux axes de 0. Que peut-on constater ?
- b) **Normalisation** : cela consiste à soustraire la moyenne aux valeurs d'une série puis à les diviser par l'écart-type. Cela permet d'obtenir une série dite « normalisée » avec un écart-type de 1 et une moyenne de 0.

$$X_{norm} = \frac{X - \bar{X}}{\sigma(X)}$$

Écrire la fonction `normaliser` qui renvoie le ndarray passé en paramètre normalisé.
Réafficher le nuage de point avec les séries normalisées. Conclure.

Questions pour approfondir :

- Écrire la fonction `median` et la comparer avec la fonction `np.median`.
- Écrire la fonction `quantile` et la comparer avec la fonction `np.quantil`.