TD15

COMMANDES PANDAS ET COUPLES DE VARIABLES DISCRÈTES.

Exercice 1

Supposons que vous soyez le chez de direction d'une camions ambulants (*Food Trucks*). Vous envisagez différentes villes pour ouvrir un nouveau point de vente. La chaîne a déjà des camions dans différentes villes et vous avez des données pour les bénéfices et les populations des villes. Vous souhaitez utiliser ces données pour vous aider à choisir votre nouvelle implantation.

On dispose d'un fichier data.csv et on utilise la bibliothèque pandas.

1. On exécute les instructions suivantes qui donnent l'affichage ci-après. Que contient le fichier data.csv importé ?

```python
import pandas as pd
import numpy as np
import numpy.random as rd
import matplotlib.pyplot as plt

donnees = pd.csv_read('data.csv', sep=';')

donnees.head()
```

Affichage Python

```
>>>  donnees.head()

    Populations (en 10k)      Profit (en 10k EUR)
0        6.1101                    17.4920
1        5.5277                    9.1302
2        8.5186                    13.6620
3        7.0032                    11.8540
4        5.8598                    6.8233
```

2. On ajoute les commandes suivantes

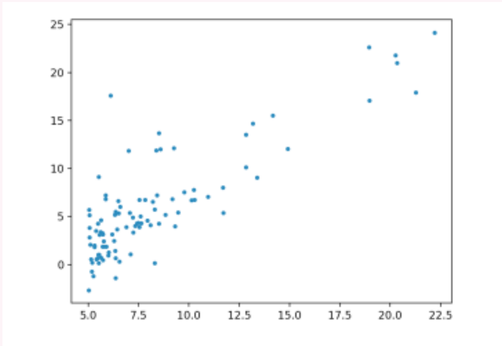
```python
table = donnees.rename(columns={'Population (en 10k)' : 'pop', 'Profit (en 10k EUR)'
     : 'Profit'})

X = table['pop']
Y = table['profit']
plt.grid()
plt.plot(X,Y,'.')
plt.show()
```

qui renvoie le plot suivant

ECG 2 Maths Appliquées, http://louismerlin.fr.

Plot Python



a. Que représente cette figure ?

b. Expliquer pourquoi la figure ci-dessus permet de conjecturer qu'il existe deux réels a et b tels que $ax + b$ où x est le nombre d'habitants (en dizaines de milliers d'habitants) de la ville est une approximation raisonnable du profit (en dizaines de milliers d'euros) d'un *Food Trucks* installé dans cette ville.

c. Quelle quantité pourrait-on calculer pour conforter cette approximation ? Donner une suite d'instructions Python permettant de la calculer.

d. On suppose qu'on a été en mesure de répondre à la question précédente correctement. L'exécution de ces commandes affiche alors

```
0.83788623092365654
```

Est-ce cohérent ?

e. Il y a 182354 habitants à *Légumeville* et pas encore de *Food Truck*. Quelle commande Python permettraient d'estimer raisonnablement le profit d'un camion dans cette localité.

3. Votre société a beau être établie en zone euro, son siège social est dans le Delaware aux États-Unis, et on décide d'exprimer le profit en dollars. Au moment où j'écris ce texte, un euro vaut environ 1.02 dollars. Que devient alors la covariance des séries statistiques habitants / profits ? Même question avec le coefficient de corrélation linéaire.

Exercice 2 d'après **ECRICOME** 2023.

Soit n un entier naturel non nul.

Une urne contient n boules indiscernables au toucher et numérotées de 1 à n. On tire une boule au hasard dans cette urne. Si cette boule porte le numéro k, alors on place dans une seconde urne une boule numérotée 1, deux boules numérotées 2, et plus généralement, pour tout $j \in [\![1, k]\!]$, j boules numérotées j, jusqu'à k boules numérotées k. Les boules de cette 2ème urne sont aussi indiscernables au toucher. On effectue alors un tirage au hasard d'une boule dans cette seconde urne.

Et on note X la variable aléatoire égale au numéro de la première boule tirée et Y la variable aléatoire égale au numéro de la 2ème boule tirée.

1. Reconnaître la loi de X et donner son espérance et sa variance.

2. Déterminer $Y(\Omega)$.

3. Soit $k \in [\![1, n]\!]$.

a. On suppose que l'événement $[X = k]$ est réalisé. Déterminer en fonction de k, le nombre totale de boules présentes dans la seconde urne.

b. Pour tout entier j de $[\![1, n]\!]$, exprimer $\mathbb{P}_{[X=k]}(Y = j)$ en fonction de k et j.
On distinguera les cas $j \leqslant k$ et $j \geqslant k + 1$.

4. **a.** Déterminer deux réels a et b tels que, pour tout entier naturel k non nul,

$$\frac{1}{k(k+1)} = \frac{a}{k} + \frac{b}{k+1}.$$

b. En déduire que, pour tout élément j de $Y(\Omega)$,

$$\mathbb{P}([Y=j]) = \frac{2(n+1-j)}{n(n+1)}.$$

5. Justifier que Y admet une espérance et montrer que $\mathbb{E}(Y) = \frac{n+2}{3}$.

6. Les variables X et Y sont-elles indépendantes ?

7. **a.** Montrer que $\mathbb{E}(XY) = \frac{(n+1)(4n+5)}{18}$.

b. En déduire que $\mathrm{Cov}(X,Y) = \frac{n^2-1}{18}$.

8. **a.** Écrire une fonction en langage Python, nommée seconde_urne, prenant en entrée un entier naturel k non nul, et renvoyant une liste contenant 1 élément valant 1, 2 élément valant 2,..., j éléments valant j,... jusqu'à k éléments valant k.

Par exemple, l'appel de seconde_urne(4) renverra [1,2,2,3,3,3,4,4,4,4].

b. Recopier et compléter la fonction en langage Python suivante pour qu'elle prenne en entrée un entier naturel non nul n et qu'elle renvoie une réalisation du couple de variables aléatoires (X,Y).

```
import numpy.random as rd

def simul_XY(n) :
  X = ......
  urne2 = seconde_urne(......)
  nb = len(urne2)
  i = rd.randint(0,nb)
  Y = .....
  return X,Y
```

c. On considère la fonction en langage Python suivante, prenant en entier un entier naturel n non nul.

```
def fonction(n) :
  liste = [0]*n
  for i in range(10000) :
    j = simul_XY[1]
    liste[j-1] = liste[j-1] + 1/10000
  return liste
```

Quelles valeurs les éléments de la liste renvoyée permettent-ils d'estimer ?

9. Dans toute cette question, on suppose que $n = 20$. On réalise 50 simulations du couple de variables (X,Y) à l'aide de la fonction simul_XY(20). On représente alors les valeurs obtenues sous la forme d'un nuage de points, où les valeurs des réalisations de X dont représentées en abscisse et les valeurs de réalisations de Y en ordonnées. On trace également sur la même figure, la droite de régression linéaire associée à ce nuage de points.

a. Déterminer par un calcul la valeur approchée des coordonnées du point moyen du nuage de points. Quel théorème de probabilités permet de justifier cette approximation ?

b. Parmi les figures représentées ci-dessous, en justifiant soigneusement votre réponse, indiquer celle qui correspond au nuage de points et à la droite de régression étudiés.

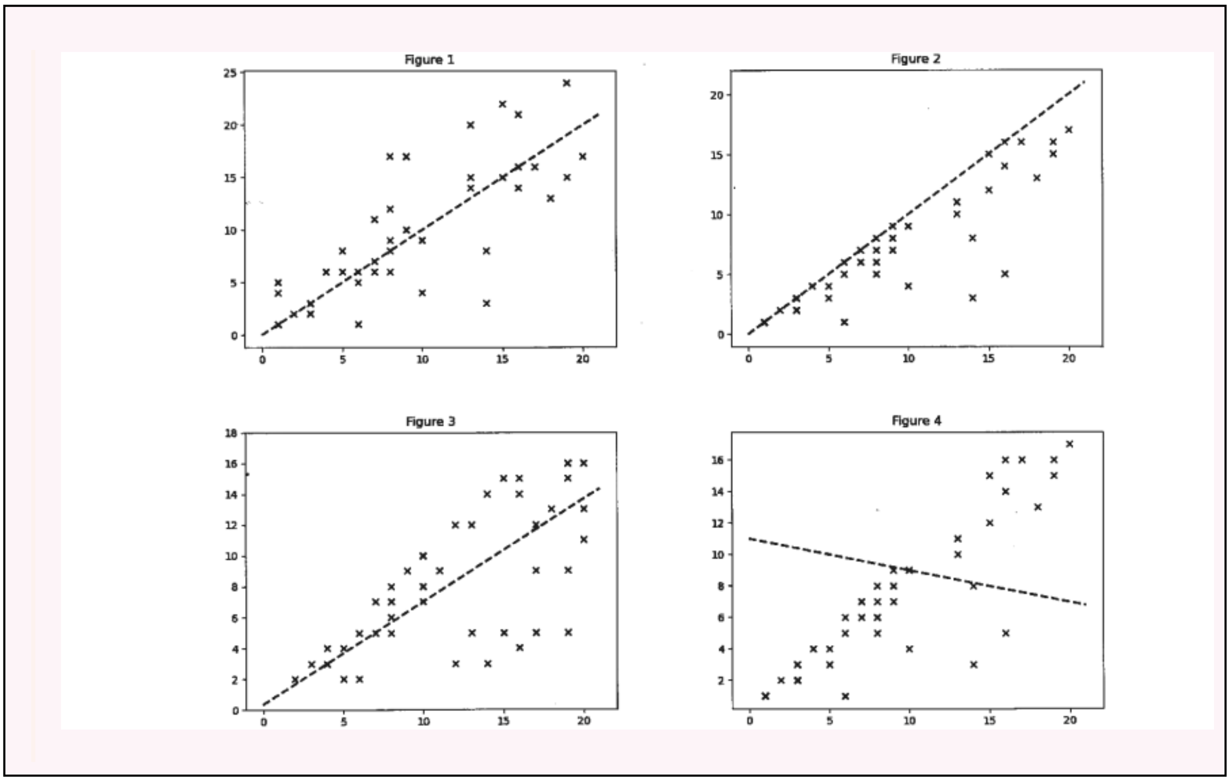


Figure 1
Figure 2
Figure 3
Figure 4