Statistique descriptive

Statistique univariée

Exercice 1

Fait en classe.

Exercice 2 (Correction rapide)

1. Comme les classes sont toutes de même longueur, pas de problème : les hauteurs des rectangles sont proportionnelles aux effectifs.
2. On utilise les centre des classes.

	Groupe A	Groupe B
Q_1	1550	1350
Med	1650	1450
Q3	1750	1550
$\overline{x}$	$\simeq 1633$	$\simeq 1492$
s_x	$\simeq 118$	$\simeq 185$

3. En l'absence d'information supplémentaire, il est impossible de conclure quoi que ce soit.

Exercice 3 (Correction complète)

1. Pour les épinards :
   - Moyenne 2,7
   - Médiane 2,685
   - Ecart-type environ 0,08
   - Q_1 2,62
   - Q_3 2,72
2. Pour les lentilles :
   - Moyenne 8,997
   - Médiane 9,02
   - Ecart-type environ 0,10
   - Q_1 8,94
   - Q_3 9,08
3. Tous les indicateurs prouvent que les lentilles ont une bien meilleure teneur en fer que les épinards.
   On pourrait également noter que le maximum pour le fer est inférieur au minimum pour les lentilles.

Exercice 4

Fait en classe.

Exercice 5 (Correction partielle)

1. On commence par calculer le nombre de total de cas par an :
   - En 2 005 : 1 756
   - En 2 006 : 1 003
   - En 2 007 : 832
   - En 2 008 : 1 151

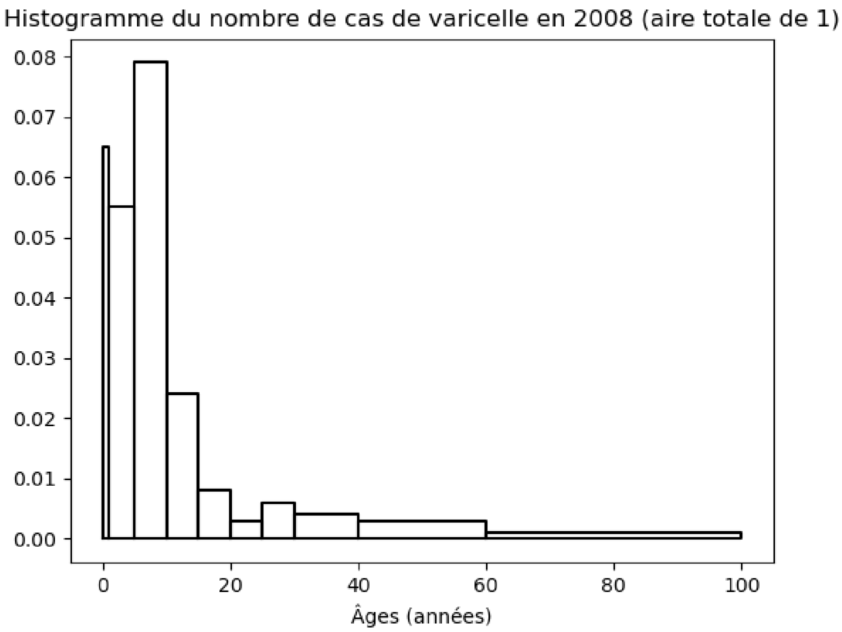
   Le nombre moyen de cas déclarés par an est donc : $\overline{x} = \dfrac{1\,756 + 1\,003 + 832 + 1\,151}{4} = \dfrac{4\,742}{4} = 1\,185,5$.
2. Pour un histogramme, ce sont les *aires* des rectangles qui sont proportionnelles aux effectifs. Or, ici, les classes n'ont pas toutes la même amplitude ; on ne peut donc pas se contenter de placer les effectifs en ordonnées.

   Par convention, on prend la somme des aires des rectangles égale à 1.
   Pour la classe n° i, on note :
   - L_i la longueur de la classe
   - H_i la hauteur du rectangle associé dans l'histogramme
   - f_i la fréquence de la classe

   Pour la classe n° i, on a alors : $\underbrace{L_i\,H_i}_{\text{aire du rectangle}} = f_i$

   Et donc : $H_i = \dfrac{f_i}{L_i}$

Classe d'âge	$[0, 1[$	$[1, 5[$	$[5, 10[$	$[10, 15[$	$[15, 20[$	$[20, 25[$	$[25, 30[$	$[30, 40[$	$[40, 60[$	$[60, 100]$
Longueur L_i	1	4	5	5	5	5	5	10	20	40
Effectif en 2008	75	254	455	139	43	16	35	41	58	35
Fréquence f_i (arrondie à 10^{-3})	$0,065$	$0,221$	$0,395$	$0,121$	$0,037$	$0,014$	$0,030$	$0,036$	$0,050$	$0,030$
Hauteur $H_i = \dfrac{f_i}{L_i}$ (arrondie à 10^{-3})	$0,065$	$0,055$	$0,079$	$0,024$	$0,008$	$0,003$	$0,006$	$0,004$	$0,003$	$0,001$

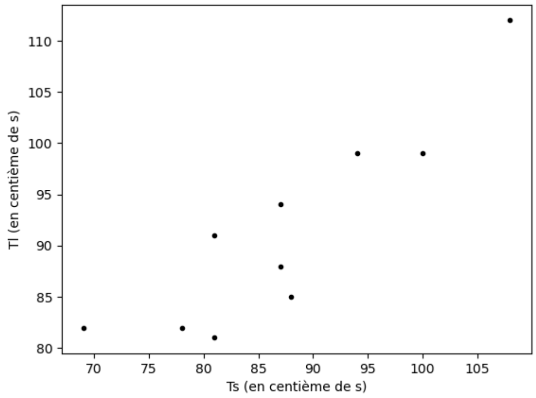


3. On trouve $Q_1 \simeq 4,35$, $Med \simeq 7,71$ et $Q_3 \simeq 12,85$.
4. Passer en argument les effectifs, déterminer l'effectif total, et ensuite méthode usuelle.

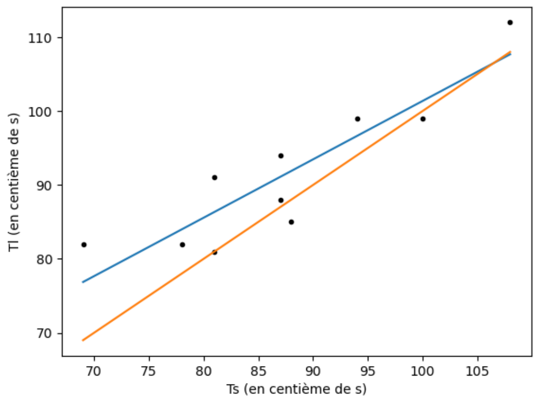
Statistique bivariée

Exercice 6 (Correction complète)

1. Moyenne 87,3 et écart-type environ 10,7
2. Moyenne 91,3 et écart-type environ 9,4
3. Nuage de points :



Le coefficient de corrélation vaut environ 0,895.
Le coefficient de corrélation est suffisamment proche de 1 pour qu'il soit pertinent de proposer un ajustement affine.
La droite qui réalise cet ajustement a pour équation $y = ax + b$ avec $a \simeq 0,790$ et $b \simeq 22,365$.
On ajoute cette droite sur le graphique (celle du haut).
On ajoute également la première bissectrice (droite d'équation $y = x$; celle du bas).



On voit que la droite d'ajustement affine est située au-dessus de la première bissectrice. Dans cette zone du graphique, les temps de réaction T_l sont supérieurs aux temps de réaction T_s.
On en déduit que le temps de réflexe à la lumière est meilleur que le temps de réflexe au son.

4. *La question 4 n'a rien à faire dans cet exercice. J'ai eu un accident de copier/coller. La question 4 aurait dû être dans l'exercice 5...*

Exercice 7

Fait en classe.

Exercice 8 (Correction très rapide)

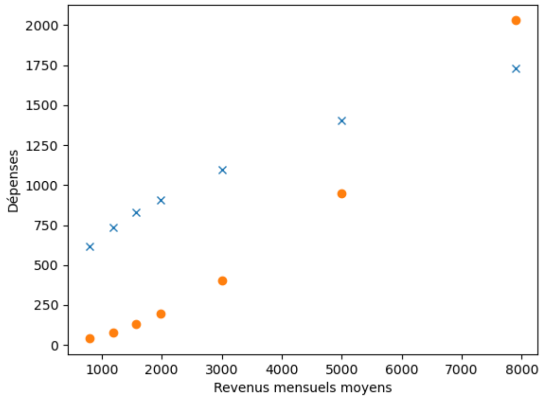
1. $r \simeq 0,026$
2. Oui

Exercice 9

Fait en classe.

Exercice 10 (Correction complète)

1. On propose le graphique suivant, où les dépenses alimentaires sont représentées par des croix et les dépenses habillement-loisir par des disques :



Des ajustements affines seraient probablement pertinents. Mais la forme des deux nuages de points suggère qu'on doit pouvoir trouver de meilleurs ajustements. On anticipe donc un changement de variable.

2. (a) En terme de méthode, le plus simple est de rentrer les données de x dans L_1, de y dans L_2, de z dans L_3.
On utilise alors la calculatrice pour remplir automatiquement $L_4 = \ln(L_1)$, $L_5 = \ln(L_2)$ et $L_6 = \ln(L_3)$.
On utilise ensuite la calculatrice pour effectuer les ajustements affines de L_4 et L_5 d'une part et de L_4 et L_6 d'autre part.

On donne ici plus qu'il n'est demandé, pour pouvoir comparer :
D'une part : $r_{xy} \simeq 0,98988$ et $r_{\ln(x)\ln(y)} \simeq 0,99957$
D'autre part : $r_{xz} \simeq 0,988499$ et $r_{\ln(x\ln(z} \simeq 0,99990$

(b) On prend garde ici à ne pas arrondir *trop tôt*. En effet, si on écrit une expression de la forme $\ln(y) = a\ln(x) + b$ en arrondissant les coefficients a et b à 10^{-1}, alors en passant à l'exponentielle l'erreur d'arrondi se propage exponentiellement et on n'obtient pas un arrondi à 10^{-1} sur les coefficients k et α demandés.

C'est une de ces fois où j'ai un peu perde le fil du changement de programme. Je ne suis pas sûre que les calculs de la fin de l'exercice soient à votre portée.

En valeurs exactes, les calculs sont :

$$\begin{array}{ll} & \ln(y) = a\ln(x) + b \\ \text{Donc} & y = \exp\Big(a\ln(x) + b\Big) \\ \text{Donc} & y = \exp\Big(a\ln(x)\Big)e^b \\ \text{Donc} & y = e^b\, x^a \end{array}$$

Ainsi : $k = e^b$ et $\alpha = a$ et on trouve $k \simeq 30,05$ et $\alpha \simeq 0,45$.

De même, on obtient $h \simeq 44,30$ et $\beta \simeq 1,71$.

3. (a) On utilise les résultats de la question 2b en remplaçant x par 4000. On estime à 1255 euros les dépenses alimentaires et à 640 euros les dépenses habillement-loisirs.

(b) Il s'agit de résoudre :

$$\begin{array}{ll} & kx^\alpha = \dfrac{h}{10^5}x^\beta \\ \Leftrightarrow & \dfrac{k}{h}10^5 = x^{\beta-\alpha} \\ \Leftrightarrow & x = \left(\dfrac{k}{h}10^5\right)^{\frac{1}{\beta-\alpha}} \end{array}$$

On trouve $x \simeq 6831$.

Annexe - Code Python

```
import matplotlib.pyplot as plt

C = [1250,1350,1450,1550,1650,1750,1850,1950]
A = [0,1,2,3,4,2,0]
B = [3,3,3,1,0,1,1]

def histogramme(X,Y):
    n = len(Y)
    for i in range(n):
        abs = [X[i],X[i+1],X[i+1],X[i],X[i]]
        ord = [0,0,Y[i],Y[i],0]
        plt.plot(abs,ord,'k')
    plt.xlabel("Masses (en grammes)")
    plt.ylabel("Effectifs")
    plt.title("Groupe B") #ligne a changer selon le groupe
    plt.show()

TS = [108,100,78,81,69,87,88,81,87,94]
TL = [112,99,82,91,82,94,85,81,88,99]

def ex6():
    plt.plot(TS,TL,'k.')
    plt.xlabel("Ts (en centieme de s)")
    plt.ylabel("Tl (en centieme de s)")
    plt.show()

def ex6_2():
    plt.plot(TS,TL,'k.')
    plt.plot([69,108],[69*0.79+22.365,108*0.79+22.365])
    plt.plot([69,108],[69,108])
    plt.xlabel("Ts (en centieme de s)")
    plt.ylabel("Tl (en centieme de s)")
    plt.show()

X10 = [790,1198,1577,1986,3007,4989,7906]
Y10 = [614,738,828,908,1094,1406,1730]
Z10 = [40,80,131,199,406,951,2030]

def ex10():
    plt.plot(X10,Y10,'x')
    plt.plot(X10,Z10,'o')
    plt.xlabel("Revenus mensuels moyens")
    plt.ylabel("Depenses")
    plt.show()
```