Correction TP6 Python

Exercice 1

1)a)$\overline{x} = \frac{1}{3}(1+2+3) = 2$ et $\overline{y} = \frac{1}{3}(3+5+10) = 6$.

b)$s_x^2 = \frac{1}{3}\left((1-2)^2 + (2-2)^2 + (3-2)^2\right) = \frac{2}{3}$.

$s_y^2 = \frac{1}{3}\left((3-6)^2 + (5-6)^2 + (10-6)^2\right) = \frac{26}{3}$.

2)a)$s_{xy} = \frac{1}{3}\big((1-2)(3-6) + (2-2)(5-6) + (3-2)(10-6)\big) = \frac{7}{3}$.

2)b)programme :

```
def covariance(x,y,n):
    liste=[]
    mx=np.mean(x)
    my=np.mean(y)
    for i in range(n):
        liste.append((x[i]-mx)*(y[i]-my))
    return(np.mean(liste))
```

2)c)La commande covariance(x,y,3) renvoie la valeur de s_{xy}.

3)$\rho_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\frac{7}{3}}{\sqrt{\frac{2}{3}}\sqrt{\frac{26}{3}}} = \frac{7}{\sqrt{52}} \approx 0,97$.

Exercice 2

1)exemple : 0,10,290.

2)exemple : 0,8,12,380.

3)cela est dû au fait qu'une petite partie de la population perçoit un salaire très élevé, ce qui tire la moyenne vers le haut.

Exercice 3

1)programme :

```
import numpy as np
import numpy.random as rd
import matplotlib.pyplot as plt
x=[rd.randint(1,1001)for k in range(100)]
print(x)
print(np.mean(x))
print(np.median(x))
print(np.std(x))
plt.boxplot(x)
```

2)La boîte à moustache donne la médiane, le premier et troisième quartile, ainsi que l'étendue de la série statistique.

Exercice 4

1)Soit x le nombre de cadres et y le nombre d'ouvriers.
Les données de l'énoncé mènent au système :

$$\left\{\begin{array}{lcl} x+y & = & 360 \\ \dfrac{4000x+2400y}{x+y} & = & 2800 \end{array}\right. \Longleftrightarrow \left\{\begin{array}{lcl} x+y & = & 360 \\ y=3x & & \end{array}\right. \Longleftrightarrow \left\{\begin{array}{l} x=90 \\ y=270 \end{array}\right.$$

2)Le salaire d'un ouvrier est de $2400+2400\times\dfrac{5}{100}=2520$.

Le salaire moyen est $\dfrac{4000\times 90+2520\times 270}{360}=2890$.

Exercice 5

1)programme :

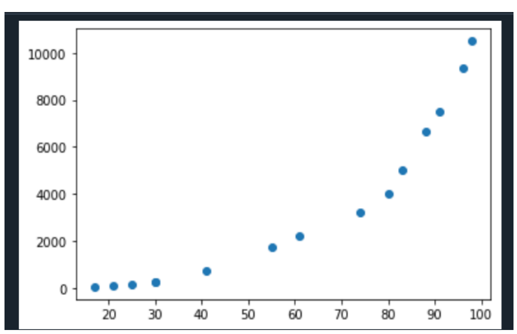
```
import numpy as np
x=np.array([[55,74,80,91,30,88,17,96,83,30,21,41,61,25,98]])
y=np.array([[1750,3200,4000,7500,250,6640,50,9360,5000,270,90,720,2200,150,10500]])
print(x)
print(y)
```

2)a)Les commandes np.mean(x) et np.mean(y) donnent :
taille moyenne $=\overline{x}=59,3\ cm$ et poids moyen $=\overline{y}=3445\ g$.

2)b)Les commandes np.median(x) et np.median(y) donnent :
taille médiane $=\overline{x}=61\ cm$ et poids médian $=\overline{y}=2200\ g$.

3)a)programme :

```
import numpy as np
import matplotlib.pyplot as plt
x=np.array([[55,74,80,91,30,88,17,96,83,30,21,41,61,25,98]])
y=np.array([[1750,3200,4000,7500,250,6640,50,9360,5000,270,90,720,2200,150,10500]])
plt.scatter(x,y)
```



Les points du nuage ne sont pas alignés. Il y a une dépendance entre x et y (quand x augmente, y a tendance a augmenter), mais cette dépendance n'est pas pas affine.

b)programme :

```
import numpy as np
def covariance(x,y):
    prod=x*y
    mx=np.mean(x)
    my=np.mean(y)
    return np.mean(prod)-mx*my
```

c)On complète le programme précédent avec :

```
import matplotlib.pyplot as plt
x=np.array([[55,74,80,91,30,88,17,96,83,30,21,41,61,25,98]])
y=np.array([[1750,3200,4000,7500,250,6640,50,9360,5000,270,90,720,2200,150,10500]])
rho=covariance(np.log(x),np.log(y))/(np.std(np.log(x))*np.std(np.log(y)))
print(rho)
plt.scatter(np.log(x),np.log(y))
```

Le coefficient de corrélation linéaire de la série $(\ln x, \ln y)$ vaut 0,998.
Il est très proche de 1 donc $\ln y$ est une fonction presque affine de $\ln x$.

d)Il existe des réels a et b telles que $\forall x > 0, \forall y > 0,\ \ln y = a \ln x + b$.

On déduit : $y = e^{a\ln x + b} = e^{a \ln x} e^b = x^a e^b$. Donc $y = \beta x^\alpha$.

<u>Exercice 6</u>

1)L'équation de D est de la forme $y = ax + b$.
Le point moyen appartient à D, ses coordonnées vérifient l'équation de D.
On a donc $\overline{y} = a\overline{x} + b$, puis $b = \overline{y} - a\overline{x}$.
En reportant, l'équation de D est : $y = ax + \overline{y} - a\overline{x}$, soit $y = \overline{y} + a(x - \overline{x})$.

2)$d_i = \left|y_i - (\overline{y} + a(x_i - \overline{x}))\right| = \left|y_i - \overline{y} - a(x_i - \overline{x})\right|$.

Donc $d_i^2 = \left(y_i - \overline{y} - a(x_i - \overline{x})\right)^2$.

3)a) $\forall a \in \mathbf{R},\ f'(a) = \displaystyle\sum_{i=1}^{n} \left(d_i^2\right)'$

$$\begin{aligned}
&= \sum_{i=1}^{n} -2(x_i - \overline{x})\left(y_i - \overline{y} - a(x_i - \overline{x})\right)\\
&= \sum_{i=1}^{n} 2a(x_i - \overline{x})^2 - 2(y_i - \overline{y})(x_i - \overline{x})\\
&= 2a\sum_{i=1}^{n} (x_i - \overline{x})^2 - 2\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})\\
&= 2ans_x^2 - 2ns_{xy}.
\end{aligned}$$

3)b) $f'(a) \geq 0 \iff 2ans_x^2 - 2ns_{xy} \geq 0 \iff a \geq \dfrac{s_{xy}}{s_x^2}$.

a	$-\infty$		$\frac{s_{xy}}{s_x^2}$		$+\infty$
$f'(a)$		$-$	0	$+$	
$f(a)$		$\searrow$		$\nearrow$	

$f(a)$ atteint bien son minimum pour $a = \frac{s_{xy}}{s_x^2}$.

3)c)L'équation de D est donc $y = \overline{y} + \frac{s_{xy}}{s_x^2}(x - \overline{x})$.