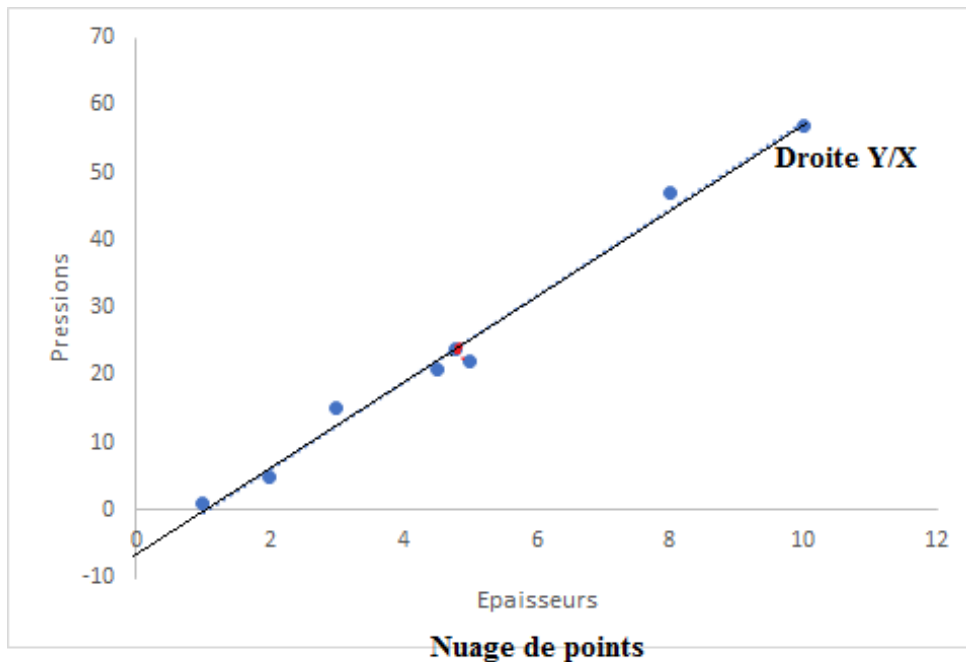


Corrigé série 3

Exercice 1 1. Nuage de points



(On est dans le cas d'une série double injective c.à.d. tous les $n_{ij} = 1$.)

2. On complète le tableau pour faciliter le calcul des différentes caractéristiques

x_i	1	2	3	4.5	5	8	10	$\sum x_i = 33.5$
y_j	1	5	15	21	22	47	57	$\sum y_j = 168$
x_i^2	1	4	9	20.25	25	64	100	$\sum x_i^2 = 223.25$
y_j^2	1	25	225	441	484	2209	3249	$\sum y_j^2 = 6634$
$x_i y_j$	1	10	45	94.5	110	376	570	$\sum x_i y_j = 1206.5$

Calcul des moyennes

$$\bar{X} = \frac{1}{7} \sum_{i=1}^7 x_i = \frac{33.5}{7} \simeq 4.7857.$$

$$\bar{Y} = \frac{1}{7} \sum_{j=1}^7 y_j = \frac{168}{7} = 24.$$

Calcul des écarts-type

$$\begin{aligned} \sigma_X &= \sqrt{\frac{1}{7} \sum_{i=1}^7 x_i^2 - \bar{X}^2} = \sqrt{\frac{223.25}{7} - 22.9441} = \sqrt{8.9899} \\ &\implies \sigma_X \simeq 2.9983. \end{aligned}$$

$$\begin{aligned} \sigma_Y &= \sqrt{\frac{1}{7} \sum_{j=1}^7 y_j^2 - \bar{Y}^2} = \sqrt{\frac{6634}{7} - 576} = \sqrt{371.7143} \\ &\implies \sigma_Y \simeq 19.2799. \end{aligned}$$

3. Covariance de X et Y

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{7} \sum_{i=1}^7 \sum_{j=1}^7 x_i y_j - \bar{X}\bar{Y} = \frac{1206.5}{7} - 4.7857 \times 24 \\ &\implies \text{Cov}(X, Y) = 57.5003. \end{aligned}$$

Coefficient de corrélation linéaire entre X et Y

$$\begin{aligned} \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{57.5003}{2.9983 \times 19.2799} \\ &\implies \rho(X, Y) \simeq 0.9947; \end{aligned}$$

Comme $\rho^2(X, Y) \simeq 0.9947 > 0.9$ alors il existe une très forte corrélation entre la pression déchirure et l'épaisseur de la feuille d'aluminium et la relation est linéaire.

4. La droite de régression $D_{Y/X}$:

$$\begin{aligned} Y &= aX + b \text{ où } a = \frac{\text{Cov}(X, Y)}{\sigma_X^2} \text{ et } b = \bar{Y} - a\bar{X} \\ a &= \frac{57.5003}{2.9983^2} \simeq 6.3962. \\ b &= 24 - 6.3962 \times 4.7857 = -6.6102. \end{aligned}$$

D'où

$$(D_{Y/X}) : Y = 6.3962X - 6.6102.$$

5. Pour $X = 12$ on a $Y = 6.3962 \times 12 - 6.6102 = 70.1442$.

Exercice 2 1. On complète le tableau pour faciliter le calcul des différentes caractéristiques

$Y \setminus X$	1	2	4	$\sum n_j$
1	8	7	5	20
3	10	9	11	30
$\sum n_i$	18	16	16	50

Calcul des moyennes

$$\bar{X} = \frac{1}{50} \sum_{i=1}^3 n_i x_i = \frac{18 + 32 + 64}{50} = \frac{114}{50} = 2.28.$$

$$\bar{Y} = \frac{1}{50} \sum_{j=1}^2 n_j y_j = \frac{20 + 90}{50} = \frac{110}{50} = 2.2.$$

Calcul des variances

$$\begin{aligned} \sigma_X^2 &= \frac{1}{50} \sum_{i=1}^3 n_i x_i^2 - \bar{X}^2 = \frac{18 + 64 + 256}{50} - 2.28^2 \\ &\implies \sigma_X^2 = 1.5616. \end{aligned}$$

$$\begin{aligned} \sigma_Y^2 &= \frac{1}{50} \sum_{j=1}^2 n_j y_j^2 - \bar{Y}^2 = \frac{20 + 270}{50} - 2.2^2 \\ &\implies \sigma_Y^2 = 0.96. \end{aligned}$$

2. La droite de régression de X en Y , $D_{X|Y}$ a pour équation :

$$X = aY + b \text{ où } a = \frac{\text{Cov}(X, Y)}{\sigma_Y^2} \text{ et } b = \bar{X} - a\bar{Y}$$

on a

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{50} \sum_{i=1}^3 \sum_{j=1}^2 n_{ij} x_i y_j - \bar{X}\bar{Y} \\ &= \frac{(1 \cdot 1 \cdot 8 + 2 \cdot 1 \cdot 7 + 4 \cdot 1 \cdot 5 + 1 \cdot 3 \cdot 10 + 2 \cdot 3 \cdot 9 + 4 \cdot 3 \cdot 11)}{50} - 2.28 \cdot 2.2 \\ &= \frac{258}{50} - 5.016 = 0.144, \end{aligned}$$

alors

$$X = \frac{0.144}{0.96}Y + \left(2.28 - \frac{0.144}{0.96}2.2\right)$$

$$\Rightarrow (D_{X/Y}) : X = 0.15Y + 1.95.$$

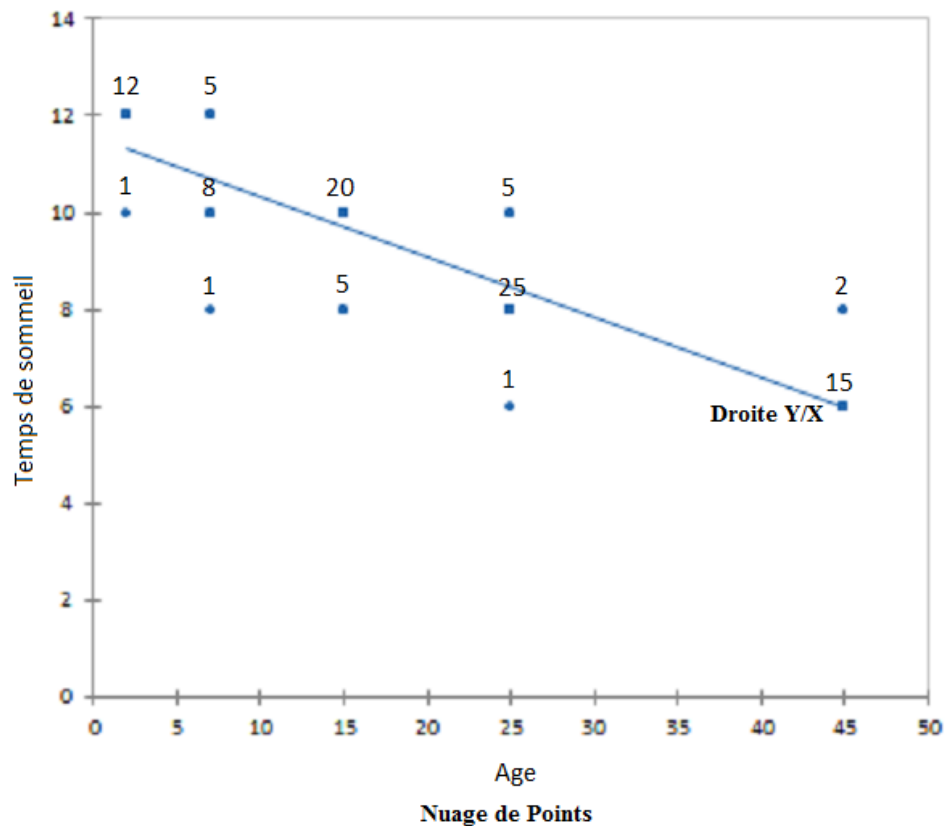
3. Pour répondre à cette question on calcule d'abord le coefficient de corrélation linéaire

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{0.144}{\sqrt{1.5616} \cdot \sqrt{0.96}}$$

$$\Rightarrow \rho(X, Y) \simeq 0.1176.$$

Comme $\rho^2(X, Y) \simeq 0.0139 < 0.1$ alors l'ajustement est très mauvais. X et Y sont non corrélées.

Exercice 3 1. Tracer le nuage de points du couple (X, Y) .



2. Pour vérifier si les variables X et Y sont statistiquement indépendantes il faut que

$$\forall i = 1, \dots, 5; \forall j = 1, \dots, 4 \quad f_{ij} = f_{i.} \times f_{.j}$$

Il suffit qu'il y ait un seul cas non vérifié pour qu'elles ne soient pas statistiquement indépendantes, on a

$$f_{33} = \frac{n_{33}}{n} = \frac{20}{100} \text{ et } f_{3.} \times f_{.3} = \frac{n_{3.}}{n} \times \frac{n_{.3}}{n} = \frac{25}{100} \times \frac{34}{100} = \frac{850}{10000}$$

d'où $f_{33} \neq f_{3.} \times f_{.3}$ alors les deux variables sont dépendantes.

3. On complète le tableau pour faciliter le calcul des différentes caractéristiques

$Y \setminus X$	Centre]1; 3]]3; 11]]11; 19]]19; 31]]31; 59]	$n_{.j}$	$n_{.j}y_j$	$n_{.j}y_j^2$
Centre		2	7	15	25	45	—	—	—
]5; 7]	6				1	15	16	96	576
]7; 9]	8		1	5	25	2	33	264	2112
]9; 11]	10	1	8	20	5		34	340	3400
]11; 13]	12	12	5				17	204	2448
$n_{i.}$	—	13	14	25	31	17	100	904	8536
$n_{i.}x_i$	—	26	98	375	775	765	2039		
$n_{i.}x_i^2$	—	52	686	5625	19375	34425	60163		
$n_{ij}x_iy_j$	—	308	1036	3600	6400	4770		16114	

Coefficient de corrélation linéaire de X et Y

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

$$Cov(X, Y) = \frac{1}{100} \sum_{i=1}^5 \sum_{j=1}^4 n_{ij}x_iy_j - \bar{X}\bar{Y}$$

on a

$$\bar{X} = \frac{1}{100} \sum_{i=1}^5 n_{i.}x_i = \frac{2039}{100} = 20,39$$

$$\bar{Y} = \frac{1}{100} \sum_{j=1}^4 n_{.j}y_j = \frac{904}{100} = 9,04$$

alors

$$Cov(X, Y) = \frac{16114}{100} - 20,39 \cdot 9,04 = -23,1856.$$

En plus,

$$\sigma_X = \sqrt{\frac{1}{100} \sum_{i=1}^5 n_i x_i^2 - \bar{X}^2} = \sqrt{\frac{60163}{100} - 20,39^2} \simeq 13,6337.$$

$$\sigma_Y = \sqrt{\frac{1}{10} \sum_{j=1}^4 n_j y_j^2 - \bar{Y}^2} = \sqrt{\frac{8536}{100} - 9,04^2} \simeq 1,9075.$$

D'où

$$\rho(X, Y) = \frac{-23,1856}{13,6337 \cdot 1,9075} \simeq -0,8917.$$

Comme $0.1 < \rho^2(X, Y) = 0,7951 < 0,9$ on peut dire que le temps de sommeil est lié à l'âge et évoluent dans le sens contraire c'est-à-dire que plus on avance dans l'âge plus le temps de sommeil diminue, on a une relation fonctionnelle non pas linéaire.

4. Droite de régression de Y en X

$$Y = aX + b \text{ où } a = \frac{Cov(X, Y)}{\sigma_X^2} \text{ et } b = \bar{Y} - a\bar{X}$$

$$a = \frac{-23,1856}{20,39^2} \simeq -0,0559$$

$$b = 9,04 + 0,0559 \cdot 20,39 = 10,1798.$$

D'où

$$(D_{Y/X}) : Y = -0,0559X + 10,1798.$$

5. Droite de régression de X en Y

$$X = a'Y + b' \text{ où } a' = \frac{Cov(X, Y)}{\sigma_Y^2} \text{ et } b' = \bar{X} - a'\bar{Y}$$

$$a' = \frac{-23,1856}{9,04^2} \simeq -0,2837$$

$$b' = 20,39 + 0,2837 \cdot 9,04 = 22,9546.$$

D'où

$$(D_{X/Y}) : X = -0,2837Y + 22,9546.$$

Pour déterminer le point d'intersection des deux droites de régression on résout le système suivant

$$\begin{cases} Y = -0,0559X + 10,1798 \\ X = -0,2837Y + 22,9546 \end{cases}$$

d'où

$$\begin{aligned} Y &= -0,0559(-0,2837Y + 22,9546) + 10,1798 \\ &= 0,0159Y + 8,8966 \\ \implies Y &= \frac{8,8966}{0,9841} \simeq 9,04 = \bar{Y}. \end{aligned}$$

et

$$X = -0,2837 \times 9,04 + 22,9546 \simeq 20,39 = \bar{X},$$

on appelle le point $G(\bar{X}, \bar{Y})$ d'intersection des deux droites de régression : le **point moyen** ou bien le **centre de gravité** du nuage de point.