

**Exercice n° 01 :**

Un laboratoire fabrique depuis 10 ans un vaccin destiné à immuniser contre une certaine maladie, la production en millions de doses étalée sur dix ans est fournie par le tableau suivant :

$x_i =$ rang de l'année	0	1	2	3	4	5	6	7	8	9
$y_i =$ production en $10^6$ doses	48	45	39	33	29	26	23	20	18	15

- 1) Ajuster graphiquement cette distribution à deux variables. Quels sont les inconvénients de cette méthode d'ajustement ?
- 2) Utiliser la calculatrice pour trouver les droites de régression  $Dy(x)$  et  $Dx(y)$ . Laquelle de ces deux droites vous paraît présenter le plus d'intérêt ?
- 3) Quelle serait la production de la 12<sup>ème</sup> année ?

**Solution :**

- 1) Ajustement aux jugés de cette distribution : C'est une droite de la forme  $y = a * x + b$  qui passe par deux points par exemple le point A (1 ; 45) et le point B (8 ; 18).

Calculons donc les deux coefficients a et b à partir des coordonnées des points A et B.

$$\begin{cases} y_1 = a x_1 + b \\ y_2 = a x_2 + b \end{cases} \Rightarrow \begin{cases} 45 = a + b \\ 18 = a * 8 + b \end{cases} \Rightarrow \begin{cases} a = -3.857142857 \\ b = 48.857142857 \end{cases}$$

D'où la droite d'ajustement aux jugés :  $y = -3.857 x + 48.857$

On pourrait prendre deux autres points A'(2 ; 39) et B'(7 ; 20). On aurait pu avoir une autre droite d'ajustement aux jugés...L'inconvénient de cette méthode d'ajustement est que le résultat obtenu n'est pas unique.

- 2) En utilisant la calculatrice :  $Dy(x) = 46.3455 - 3.721 x$  ;  $Dx(y) = 12.2509 - 0.262 y$

**Remarque :** ces 2 droites passent par le point moyen G ( $\bar{x} = 4.5$  ;  $\bar{y} = 29.6$ ).

- 3) La production de la 12<sup>ème</sup> année :  $\hat{y}(11) = 5.412$

**Exercice n° 02 :**

Pour étudier les mécanismes hormonaux de la puberté on a mesuré les concentrations de deux hormones : l'œstradiol et l'œstrone pour un groupe de 8 adolescentes. Les résultats sont :

$x_i =$ concentration œstradiol pg/ml	7.5	16.5	22	30	39	54	69	77
$y_i =$ concentration œstrone pg/ml	9	18.5	21.5	27	32.5	48.5	57	58

On note par H le point moyen des quatre premiers points du nuage et par K le point moyen des quatre autres points.

- 1) Calculer les coordonnées des points H et K et déterminer la droite d'ajustement Y.
- 2) Utiliser la méthode des moindres carrés ordinaires pour déterminer Y(X).
- 3) Calculer la covariance et le coefficient de corrélation linéaire.

**Solution :**

- 1) Calcul des coordonnées des points H et K : l'abscisse de H est la moyenne de  $x_1$  à  $x_4$  et son ordonnée est la moyenne de  $y_1$  à  $y_4$  d'où H (19 ; 19). l'abscisse de K est la moyenne de  $x_5$  à  $x_8$  et son ordonnée est la moyenne de  $y_5$  à  $y_8$  d'où K (59.75 ; 49).

$$\begin{cases} y_1 = a x_1 + b \\ y_2 = a x_2 + b \end{cases} \Rightarrow \begin{cases} 19 = a * 19 + b \\ 49 = a * 59.75 + b \end{cases} \Rightarrow \begin{cases} a = 0.736196319 \\ b = 5.012269939 \end{cases}$$

D'où la droite d'ajustement de Mayer :  $y = 0.736 x + 5.012$

Soit le point moyen G ( $\bar{x} = 39.375$  ;  $\bar{y} = 34$ ), la droite de Mayer passe par le point G.

- 2) La droite des moindres carrés est  $y = 0.728 x + 5.3211$

- 3)  $Cov(x ; y) = \frac{13941.25}{8} - 39.375 * 34 = 403.90625$  ;  $r = \frac{403.90625}{23.5488 * 17.27721} = 0.99274858$

**Exercice n° 3 :**

- 1) Si le coefficient de corrélation entre x et y est égal à 0.66,  $\sigma_x^2 = 34.18$  et  $\sigma_y^2 = 121.36$  ;  $\bar{x} = 30$  et  $\bar{y} = 57.55$  Trouver les deux droites de régression.
- 2) Les droites de régression relatives à un ensemble donné sont :  $Dy(x) \rightarrow 45.414 = 2.993x - y$  et  $Dx(y) \rightarrow x - 0.228y = 31.576$  Calculer le coefficient de corrélation r.

**Solution :**

- 1) La pente  $a = \frac{0.66 * \sigma_y}{\sigma_x} = 1.243642887$        $b = \bar{y} - a \bar{x} = 20.24071339$   
 $Dy(x) = 1.243642887 x + 20.24071339$        $Dx(y) = a'y + b'$  avec  $aa' = 0.66^2$  et  $b' = \bar{x} - a'\bar{y}$   
 $Dx(y) = 0.35026132 y + 9.842460997$       Ces 2 droites passent par G ( $\bar{x} = 30$  ;  $\bar{y} = 57.55$ )
- 2) On a :  $a a' = r^2$  d'où  $r = \sqrt{2.993 * 0.228} = 0.826077478 \approx 82.61 \%$

**Exercice n° 4 :**

Le tableau suivant concerne les âges auxquels 100 couples se sont mariés :

Classes	Femmes Y	[17 ; 22[	[22 ; 27[	[27 ; 32[	[32 ; 37[	$\Sigma$
Maris X	Centres					
[20 ; 25[		14	9	1	0	
[25 ; 30[		18	7	2	1	
[30 ; 35[		4	13	3	1	
[35 ; 40[		1	9	10	2	
[40 ; 45[		0	1	2	2	
$\Sigma$						

- 1) Compléter le tableau. Calculer le tableau de contingence des fréquences.
- 2) Calculer les distributions, les moyennes et les variances marginales de X et de Y.
- 3) Calculer la covariance entre X et Y ainsi que le coefficient de corrélation linéaire.
- 4) Trouver par la méthode des moindres carrés, les deux droites de régression  $Dy(x)$  et  $Dx(y)$ .

**Solution :**

- 1) Les centres et les fréquences sont figurés dans le tableau suivant :

Classes	Femmes Y	[17 ; 22[	[22 ; 27[	[27 ; 32[	[32 ; 37[	$\Sigma$
Maris X	Centres	19.5	24.5	29.5	34.5	
[20 ; 25[	22.5	0.14	0.09	0.01	0	0.24
[25 ; 30[	27.5	0.18	0.07	0.02	0.01	0.28
[30 ; 35[	32.5	0.04	0.13	0.03	0.01	0.21
[35 ; 40[	37.5	0.01	0.09	0.10	0.02	0.22
[40 ; 45[	42.5	0	0.01	0.02	0.02	0.05
$\Sigma$		0.37	0.39	0.18	0.06	1

- 2) La distribution marginale de x :

$x_i$	22.5	27.5	32.5	37.5	42.5
$n_i$	24	28	21	22	5

La distribution marginale de y :

$y_j$	19.5	24.5	29.5	34.5
$n_j$	37	39	18	6

Les moyennes et variances marginales sont :

$$\bar{x} = 30.3 \quad \bar{y} = 24.15 \quad \sigma_x^2 = 36.66 \quad \sigma_y^2 = 19.6275$$

- 3) La covariance (x ; y) = 15.48      Le coefficient de corrélation  $r = 0.577088251$

- 4)  $Dy(x) = 11.35556465 + 0.422258592 x$        $a a' = r^2$  d'où :  $a' = \frac{r^2}{a} = 0.788689338$

$$\bar{x} = a' \bar{y} + b' \quad \text{d'où : } b' = 11.25315246 \quad Dx(y) = 11.25315246 + 0.788689338 y$$

Ces 2 droites passent par le point moyen G ( $\bar{x} = 30.3$  ;  $\bar{y} = 24.15$ ).