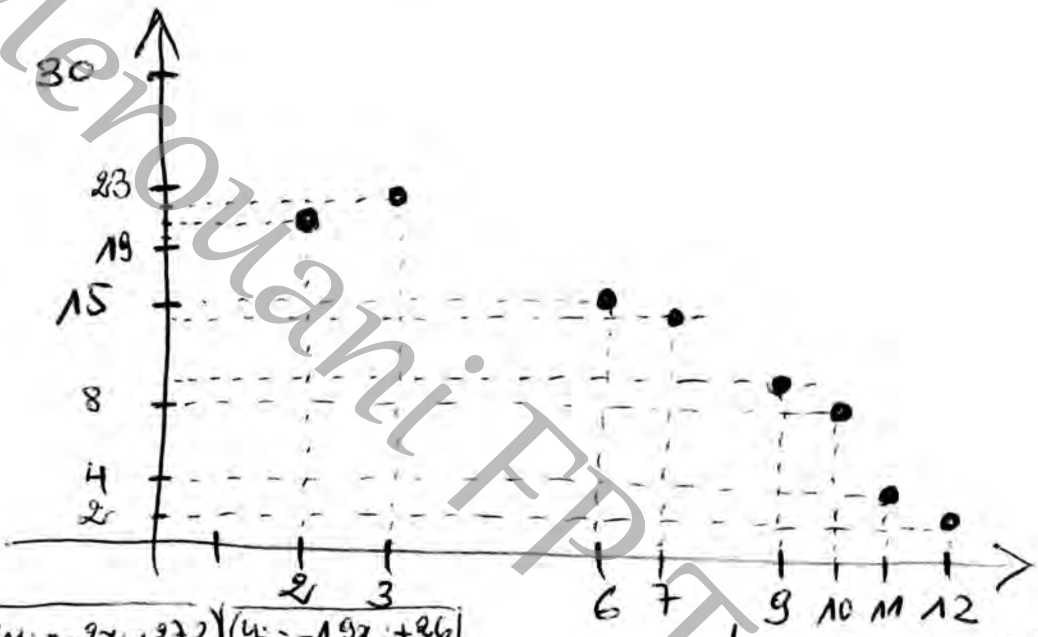
Correction des exercices de T.D.
Statistique descriptive (Ajustement et correlation)

Exercice n° 1 page 121

1°) Représentation graphique du nuage de points Y_i en fonction X_i :



2°)

x_i	y_i	($y_i = -2x_i + 27,2$) y_i (du 1er)	($y_i = -1,9x_i + 26$) y_i (2^{eme})	$(y_i - y_i(1^{er}))^2$	$(y_i - y_i(2^{eme}))^2$
2	21	23,2	22,2	4,84	1,44
3	22	21,2	20,3	0,64	2,89
6	15	15,2	14,6	0,04	0,16
7	14	13,2	12,7	0,64	1,69
9	10	9,2	8,9	0,64	1,21
10	8	7,2	7	0,64	1
11	4	5,2	5,1	1,44	1,21
12	2	3,2	3,2	1,44	1,44
60	96			10,32	11,04

Comme $\sum (y_i - y_i^{(1^{er})})^2 < \sum (y_i - y_i^{(2^{ème})})^2$,

alors d'après le principe de la méthode des moindres carrées, le 1^{er} statisticien a réalisé le meilleur ajustement.

3) Ajustement par la méthode des moindres carrées:

$$\bar{x} = \frac{60}{8} = 7{,}5 \qquad \bar{y} = \frac{96}{8} = 12$$

$$a = \frac{Cov(X,Y)}{Var(X)} = \frac{\sum_{i=1}^{8} x_i y_i - \bar{y} \sum_{i=1}^{8} x_i}{\sum_{i=1}^{8} x_i^2 - \bar{x} \sum_{i=1}^{8} x_i}$$

x_i	y_i	x_i^2	$x_i y_i$
2	21	4	42
3	22	9	66
6	15	36	90
7	14	49	98
9	10	81	90
10	8	100	80
11	4	121	44
12	2	144	24
60	96	544	534

Alors $a = \frac{534 - (12 \times 60)}{544 - (7{,}5 \times 60)}$

$$a = \frac{534 - 720}{544 - 450} = \frac{-186}{94} \simeq -1{,}98$$

et $b = \bar{y} - a\bar{x}$

$= 12 - (-1{,}98) \times 7{,}5$

$= 12 + 14{,}85 = 26{,}85$

Donc, l'équation de la droite d'ajustement par la méthode des moindres carrées sera:

$$\boxed{y = -1{,}98x + 26{,}85}$$

$y_{MMC} = -1{,}98x_i + 26{,}85$

x_i	y_i	$y_i(M.M.C)$	$(y_i - y_i(M.M.C))^2$	$y_i - y_i(M.M.C)$
2	21	22,89	3,5721	−1,89
3	22	20,91	1,1881	1,09
6	15	14,97	0,0009	0,03
7	14	12,99	1,0201	1,01
9	10	9,03	0,9409	0,97
10	8	7,05	0,9025	0,95
11	4	5,07	1,1449	−1,07
12	2	3,09	1,1881	−1,09
			9,9576	0

D'abord dans la question 4), corrigeons :

4) Vérifier, à partir ... que le total $(y_i - y_i \text{ ajustés})$ ~~^{2}~~ est égal à zéro.

Ce que l'on voit clairement dans la dernière colonne du tableau ci-dessus.

5) D'après la colonne avant dernière du tableau ci-dessus on voit que $\sum (y_i - y_i(M.M.C))^2 \simeq 9{,}96$ est plus faible que $\sum (y_i - y_i(1^{er}))^2 = 10{,}32$ et que

$$\sum (y_i - y_i(2^{ème}))^2 = 11{,}04.$$

Exercice n° 2 :

x_i	y_i	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y_i-\bar{y})$
2,4	38	−0,6	−2	0,36	4	1,2
3	42	0	2	0	4	0
3	42	0	2	0	4	0
2,5	39	−0,5	−1	0,25	1	0,5
3,2	40	0,2	0	0,04	0	0
3,5	45	0,5	5	0,25	25	2,5
2	35	−1	−5	1	25	5
1,8	24	−1,2	−16	1,44	256	19,2
3	38	0	−2	0	4	0
3,2	40	0,2	0	0,04	0	0
3,8	44	0,8	4	0,64	16	3,2
4,6	53	1,6	13	2,56	169	20,8
36	480	0	0	6,58	508	52,4

1) Coefficient de corrélation linéaire entre X et Y :

$$r = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} = \frac{\sum_i (x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})^2 \sum_i (y_i-\bar{y})^2}}$$

$$\bar{x} = \frac{36}{12} = 3 \qquad \bar{y} = \frac{480}{12} = 40$$

$$r = \frac{52,4}{\sqrt{6,58 \times 508}} = \frac{52,4}{\sqrt{3342,64}} = \frac{52,4}{57,81} \simeq 0,91$$

Donc, on a une forte corrélation positive (les deux croissent ou varient dans le même sens) entre les dépenses en publicité et les ventes.

2) La droite d'estimation de y à partir de x :

Appliquons la méthode des moindres carrées pour déterminer cette droite ; on a :

$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{52,4}{6,58} = 7,96$$

et $b = \bar{y} - a\bar{x} = 40 - (7,96 \times 3) = 40 - 23,88$

$b = 16,12$

D'où la droite cherchée est d'équation :

$$\boxed{y = 7,96x + 16,12}$$

Exercice n°3 :

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
−5	33	−9	18	81	324	−162
−1	25	−5	10	25	100	−50
3	17	−1	2	1	4	−2
10	3	6	−12	36	144	−72
13	−3	9	−18	81	324	−162
20	75	0	0	224	896	−448

1°) Coefficient de corrélation linéaire entre X et Y :

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

$\bar{x} = \frac{20}{5} = 4$ $\qquad \bar{y} = \frac{75}{5} = 15$

$$r = \frac{-448}{\sqrt{224 \times 896}} = \frac{-448}{\sqrt{200704}} = \frac{-448}{448} = -1$$

$\Rightarrow$ une forte correlation négative

2) L'équation de la fonction (linéaire) de régression de y par rapport à x ; de la forme $y = ax + b$, par la méthode de moindres carrées :

$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{-448}{224} = -2$$

et $b = \bar{y} - a\bar{x} = 15 + 2 \times 4 = 15 + 8 = 23$

D'où $\boxed{y = -2x + 23}$

3) Représentation du nuage de points et de la droite de régression :