

CH I – RÉGRESSION LINÉAIRE SUR DONNÉES INDIVIDUELLES

OBJECTIFS :

- Régression linéaire et ajustement linéaire
- La méthode des moindres carrés

I – Le principe de l'ajustement linéaire

1 Les données

Soient deux variables numériques x et y .

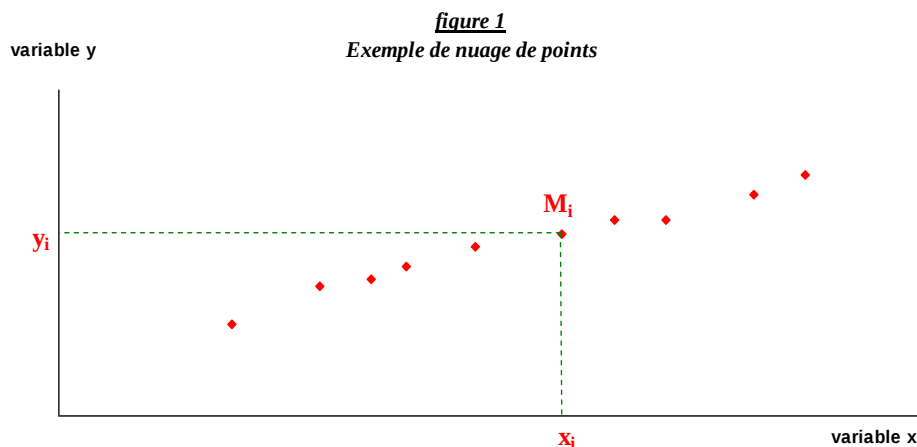
- Si x est la **variable explicative** et y la **variable à expliquer**, pour chaque valeur de x , on relève la valeur correspondante de y .
- Si y est la **variable explicative** et x la **variable à expliquer**, pour chaque valeur de y , on relève la valeur correspondante de x .

On dispose ainsi, concernant ces deux variables x et y , de N observations dites **individuelles** : le couple $(x_i ; y_i)$ est observé à une seule reprise, donc l'effectif associé vaut : $n_i = 1$.

Observation n°	Variable x	Variable y
1	x_1	y_1
$\vdots$	$\vdots$	$\vdots$
i	x_i	y_i
$\vdots$	$\vdots$	$\vdots$
N	x_N	y_N

2 La représentation graphique des données

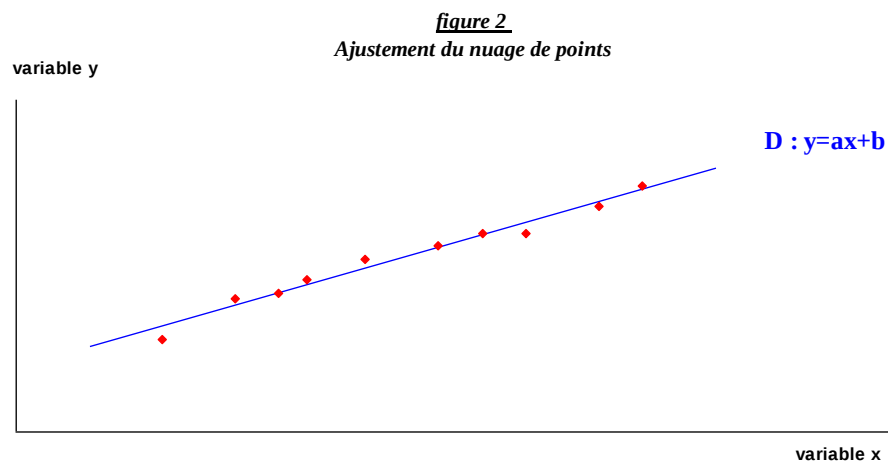
On obtient un « **nuage de points** » comportant N points $M_i (x_i, y_i)$ avec : $1 \leq i \leq N$



3 L'ajustement linéaire

Afin d'expliquer la démarche, on se placera dans le cas où x est la variable explicative et y la variable à expliquer. Cela signifie donc que l'on tente d'expliquer les variations de y à l'aide des variations de x . L'observation de ce nuage de points permet de constater, bien que les N points M_i ne soient pas alignés, une **tendance linéaire** évidente : cela signifie que les variations de y sont (à peu près) proportionnelles aux variations de x .

On recherche alors une droite D d'équation : $y = ax + b$ (car x est la variable explicative), qui représente « **au mieux** » le nuage de points. On dit que l'on ajuste le nuage de points par la droite D .



La droite D d'équation : $y = ax + b$ se nomme : **droite de régression ou droite d'ajustement de y en x** .

L'opération consistant à ajuster le nuage de points par la droite D porte le nom d'ajustement linéaire.

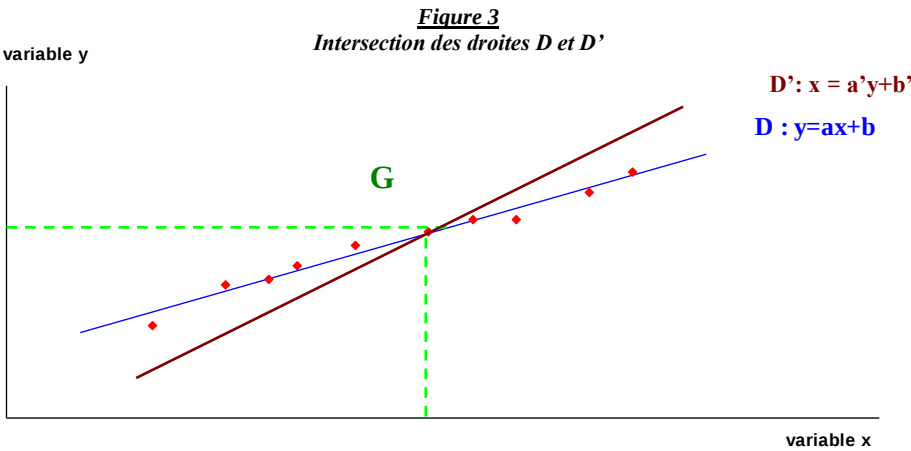
Remarque : Dans le cas où y est la variable explicative et x la variable à expliquer, on ajustera le nuage de points par la droite D' d'équation : $x = a'y + b'$ appelée **droite de régression ou droite d'ajustement de x en y** .

4 **Point moyen**

Les deux droites de régression D : $y = a x + b$ et D' : $x = a' y + b'$ s'intersectent en un point situé à l'intérieur du nuage de points : **Le point moyen** G de coordonnées : $(\bar{x}; \bar{y})$

$$\bar{x} = \frac{1}{N} \left[\sum_i x_i \right] \text{ (Moyenne des valeurs de x) et } \bar{y} = \frac{1}{N} \left[\sum_i y_i \right] \text{ (Moyenne des valeurs de y)}$$

La lettre grecque sigma majuscule : Σ représente un total, une somme (en français : S)



II – La méthode de MAYER

1 **Le principe**

- On décompose le nuage de points en deux parties selon les valeurs croissantes de x pour la droite d’ajustement de y en x (selon les valeurs croissantes de y pour la droite d’ajustement de x en y)
- On recherche G₁ point moyen de la première partie du nuage de points
- On recherche G₂ point moyen de la seconde partie du nuage de points
- Pour la droite d’ajustement de $y = ax + b$: $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}}$ et $b = y_{G_1} - ax_{G_1}$ (ou $y_{G_2} - ax_{G_2}$)
- Pour la droite d’ajustement de $x = ay + b$: $a = \frac{x_{G_2} - x_{G_1}}{y_{G_2} - y_{G_1}}$ et $b = x_{G_1} - ay_{G_1}$ (ou $x_{G_2} - ay_{G_2}$)

Remarque : Cette méthode, rapide, donne des résultats satisfaisants mais souvent peu précis.

2 **Exemple**

La société Mercure vend des machines agricoles. Suite à sa restructuration en 2009, ses bénéfices annuels en k€ (milliers d’euros) ont évolué selon le tableau suivant :

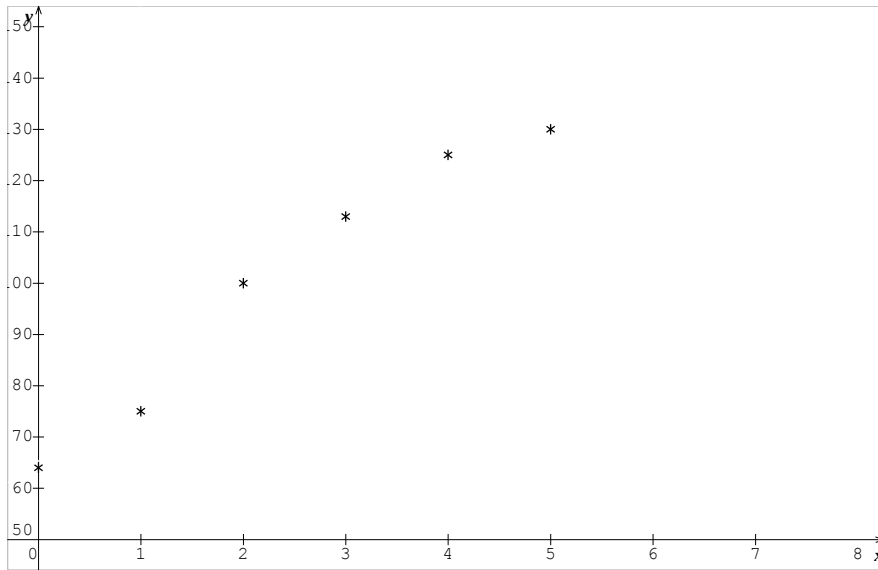
Année	2009	2010	2011	2012	2013	2014
Rang x_i	0	1	2	3	4	5
Bénéfices y_i	64	75	100	113	125	130

1. Représenter sur le graphique ci-dessous le nuage de points
2. Déterminer G₁ le point moyen correspondant aux 3 premiers points et G₂ le point moyen correspondant aux trois derniers points
3. Déterminer l’équation de D droite de Mayer de y en x.
4. Déterminer l’équation de D' droite de Mayer de x en y.

5. En déduire une estimation du bénéfice réalisé par cette société en 2015.
6. Déterminer une estimation de la première année où le bénéfice dépassera 200 k€

Solution :

1. nuage de points



2. $x_{G_1} = \frac{x_1+x_2+x_3}{3} = 1$ et $y_{G_1} = \frac{y_1+y_2+y_3}{3} = 79,67$ donc $G_1(1 ; 79,67)$.

De même $G_2(4 ; 122,67)$

3. $y = a x + b$. $a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} = 14,33$ et $b = y_{G_1} - a x_{G_1} = 65,33$.

D : $y = 14,33 x + 65,33$

4. $x = a y + b$. $a = \frac{x_{G_2} - x_{G_1}}{y_{G_2} - y_{G_1}} = 0,07$ et $b = x_{G_1} - a y_{G_1} = -4,56$

D' : $x = 0,07 y - 4,56$

5. On connaît $x = 6$ et on cherche y donc on utilise la droite D

$y = 14,33 \times 6 + 65,33 = 151,31$ donc le bénéfice en 2015 sera environ de 151 310 euros

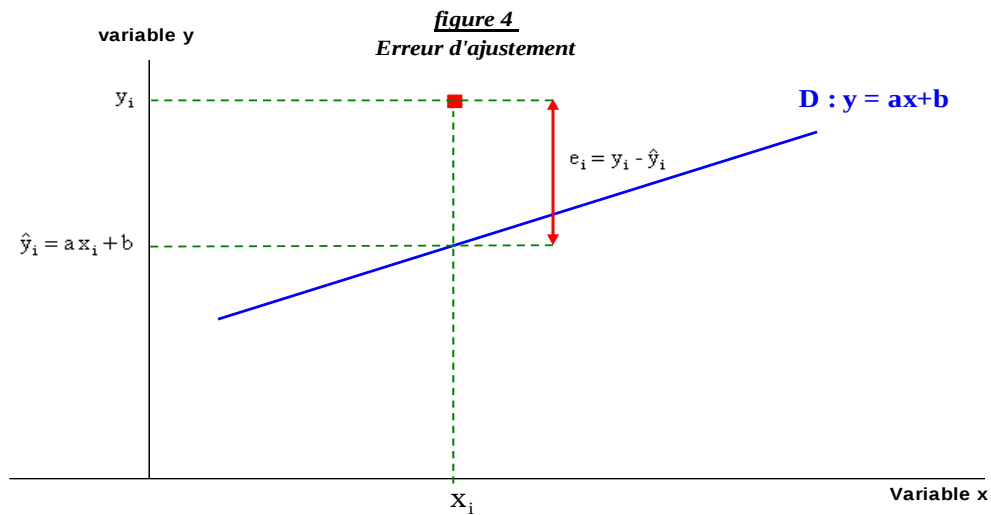
6. On connaît $y = 200$ et on cherche x donc on utilise la droite D'

$x = 0,07 \times 200 - 4,56 = 9,44$ donc il faut attendre 10 ans soit en 2019

III – La méthode des moindres carrés

1 Le principe

Sur la figure 4, on a représenté, pour simplifier, un seul point du nuage de points de coordonnées (x_i, y_i) ($1 \leq i \leq N$) et la droite D d'ajustement de y en x du nuage de points.



En remplaçant y_i par $\hat{y}_i$ dans l'ajustement du nuage de points par la droite D, on commet une erreur :

$$e_i = y_i - \hat{y}_i \quad \text{Avec : } \hat{y}_i = ax_i + b$$

Comme la droite D traverse le nuage de points, il y a des erreurs e_i **positives** et des erreurs e_i **négatives** de telle sorte que :

$$\sum e_i = 0 \quad \text{La somme des erreurs } e_i \text{ vaut 0 : il y a compensation entre les erreurs}$$

Pour tenir compte de toutes les erreurs, **on totalise les carrés des erreurs**, c'est-à-dire que l'on considère l'expression :

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (ax_i + b)]^2$$

La **méthode des moindres carrés** consiste à calculer les réels a et b, donc les paramètres de l'équation de la droite D, de sorte que :

$$\sum e_i^2 \text{ soit minimum} \Leftrightarrow \boxed{\sum [y_i - (ax_i + b)]^2 \text{ soit minimum}}$$

Il s'agit d'un problème d'optimisation sans contrainte à deux variables a et b (Cf. cours de maths).

Remarque : Cette méthode demande plus de calculs mais permet de trouver l'ajustement affine la plus précise possible. Sauf précision dans l'énoncé, c'est cette méthode qui sera toujours utilisée.

2 Formulaire pratique

N.B. : Les données sont **individuelles**, donc l'effectif associé au couple $(x_i ; y_i)$ vaut : $n_i = 1$. Par suite, l'effectif n_i n'apparaît pas dans le formulaire car, par exemple :

$$\bar{x} = \frac{1}{N} \left[\sum_i n_i x_i \right] = \frac{1}{N} \left[\sum_i 1 x_i \right] = \frac{1}{N} \left[\sum_i x_i \right]$$

➔ Calculs préalables

■ Variable x :

Moyenne : $\bar{x} = \frac{1}{N} \left[\sum_i x_i \right]$

Variance : $V(x) = \frac{1}{N} \left[\sum_i x_i^2 \right] - \bar{x}^2$

■ **Variable y :**

Moyenne : $\bar{y} = \frac{1}{N} \left[\sum_i y_i \right]$

Variance : $V(y) = \frac{1}{N} \left[\sum_i y_i^2 \right] - \bar{y}^2$

■ **Covariance de x et de y :**

Covariance : $\text{Cov}(x, y) = \frac{1}{N} \left[\sum_i x_i y_i \right] - \bar{x} \bar{y}$

➔ **Droites de régression D et D' et coefficient de corrélation linéaire**

■ **Droite D de régression de y en x :** **D: $y = ax + b$**

Calcul de a : $a = \frac{\text{Cov}(x, y)}{V(x)}$

Calcul de b : $b = \bar{y} - a \bar{x}$

■ **Droite D' de régression de x en y :** **D': $x = a'y + b'$**

Calcul de a' : $a' = \frac{\text{Cov}(x, y)}{V(y)}$

Calcul de b' : $b' = \bar{x} - a' \bar{y}$

3 Le tableau de calculs

Afin d'effectuer les calculs préalables, il est nécessaire de compléter, dans le cas de données individuelles un tableau statistique de la forme :

		$\bar{y}$				
		Calcul des moyennes		Calcul des variances		Covariance
N modalités		x_i	y_i	x_i^2	y_i^2	$x_i y_i$
		x_1	y_1	x_1^2	y_1^2	$x_1 y_1$
		...	...	...	...	...
		x_i	y_i	x_i^2	y_i^2	$x_i y_i$
		...	...	...	...	...
		x_N	y_N	x_N^2	y_N^2	$x_N y_N$
Les totaux		$\sum x_i$	$\sum y_i$	$\sum x_i^2$	$\sum y_i^2$	$\sum x_i y_i$

4 Exemple

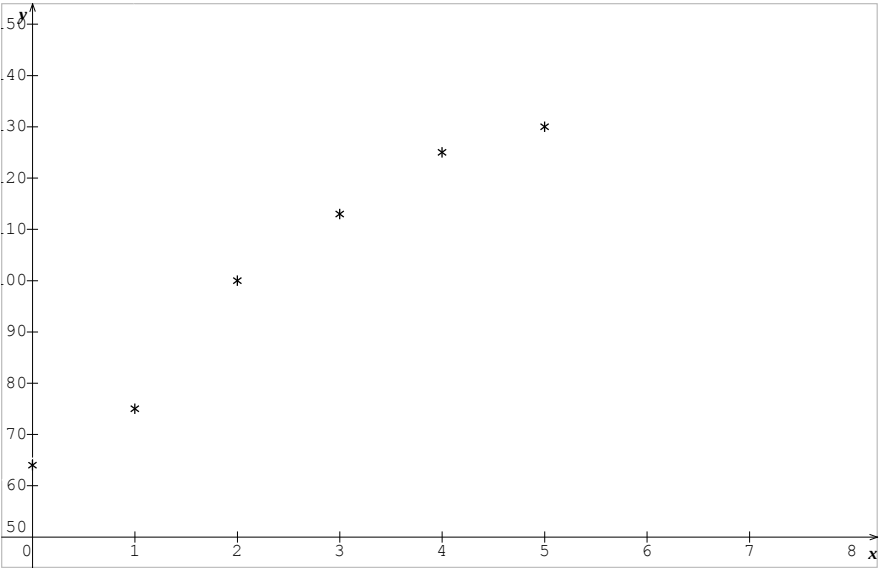
La société Mercure vend des machines agricoles. Suite à sa restructuration en 2009, ses bénéfices annuels en k€ (milliers d’euros) ont évolué selon le tableau suivant :

Année	2009	2010	2011	2012	2013	2014
Rang x_i	0	1	2	3	4	5
Bénéfices y_i	64	75	100	113	125	130

- 1. Représenter sur le graphique ci-dessous le nuage de points
- 2. Compléter le tableau statistique
- 2. Déterminer G le point moyen correspondant au nuage de points
- 3. Déterminer l’équation de D droite de régression de y en x.
- 4. Déterminer l’équation de D’ droite de régression de x en y.
- 5. En déduire une estimation du bénéfice réalisé par cette société en 2015.
- 6. Déterminer une estimation de la première année où le bénéfice dépassera 200 k€

Solution :

1. nuage de points



2. Complétons le tableau statistique (?) :

N = 6 modalités	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	0	64	? 0	? 4096	? 0
	1	75	? 1	? 5625	? 75
	2	100	4	10000	200
	3	113	9	12769	339
	4	125	16	15625	500
Total	5	130	25	16900	650
	15	607	? 55	? 65015	? 1764

3. $x_G = \frac{15}{6} = 2,5$ et $y_{G_1} = \frac{607}{6} = 101,17$ donc $G(2,5 ; 101,17)$.

4. **D : $y = a x + b$**

Variance : $V(x) = \frac{1}{N} \left(\sum_i x_i^2 \right) - \bar{x}^2 = \frac{1}{6} \times 55 - 2,5^2 = \boxed{2,92}$

Covariance : $\text{Cov}(x, y) = \frac{1}{N} \left(\sum_i x_i y_i \right) - \bar{x} \bar{y} = \frac{1}{6} \times 1764 - 2,5 \times 101,17 = \boxed{41,08}$

Calcul de a :

$$a = \frac{\text{Cov}(x, y)}{V(x)} = \frac{41,08}{2,92} = \boxed{14,07}$$

Calcul de b :

$$b = \bar{y} - a \bar{x} = 101,17 - 14,07 \times 2,5 = \boxed{66}$$

Conclusion :

La droite D de régression de y en x a pour équation : $\boxed{y = 14,07 x + 66}$

5. **D' : $x = a y + b$**

Variance : $V(y) = \frac{1}{N} \left(\sum_i y_i^2 \right) - \bar{y}^2 = \frac{1}{6} \times 65015 - 101,17^2 = 600,46$

Calcul de a :

$$a = \frac{\text{Cov}(x, y)}{V(y)} = 0,068$$

Calcul de b

$$b = \bar{x} - a \bar{y} = -4,41$$

Conclusion :

La droite D' de régression de x en y a pour équation $x = 0,068 y - 4,41$

5. On connaît $x = 6$ et on cherche y donc on utilise la droite D

$y = 14,07 \times 6 + 66 = 150,42$ donc le bénéfice en 2015 sera environ de 150 420 euros

6. On connaît $y = 200$ et on cherche x donc on utilise la droite D'

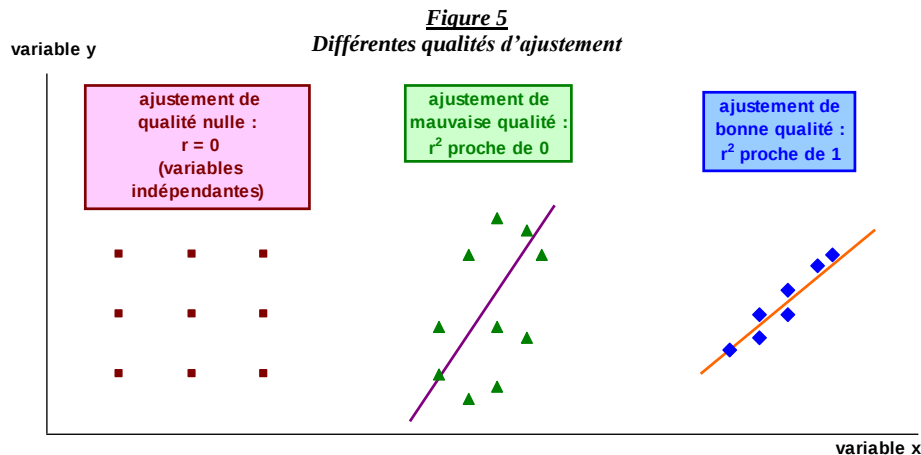
$x = 0,068 \times 200 - 4,41 = 9,19$ donc il faut attendre 10 ans soit en 2019

Remarque : On obtient sensiblement les mêmes résultats que ceux obtenus avec la droite de MAYER.

IV – CORRELATION

1 La qualité de la régression (ou de l'ajustement) linéaire

On peut, *dans tous les cas*, déterminer l'équation des droites D et D'. Cependant, l'ajustement du nuage de points par la droite D ou par la droite D' n'est pas toujours justifié. C'est pourquoi, on détermine, pour estimer la qualité de cet ajustement **le coefficient de corrélation linéaire** r ou mieux encore **le coefficient de détermination** r^2 . Sa valeur indique immédiatement si l'ajustement linéaire obtenu est ou non de bonne qualité.



2 Calculs pratiques

La lettre grecque sigma minuscule : σ représente l'écart-type (en français : s)

Ecart-type : $\sigma(x) = \sqrt{V(x)}$

Ecart-type : $\sigma(y) = \sqrt{V(y)}$

Coefficient de corrélation linéaire et coefficient de détermination :

Coefficient de corrélation linéaire : $r = \frac{\text{Cov}(x,y)}{\sigma(x) \sigma(y)}$

Coefficient de détermination : $r^2 = \frac{\text{Cov}^2(x,y)}{V(x) \times V(y)}$

3 Interprétations :

On a toujours : $0 \leq r^2 \leq 1$ et donc : $-1 \leq r \leq 1$

- Plus le **coefficient de détermination** r^2 est proche de 0 (ou de 0 %), moins bonne est la qualité de l'ajustement linéaire.

Lorsque $r = 0$, on dit que les variables x et y sont **non corrélées linéairement**.

- Plus le **coefficient de détermination** r^2 est proche de 1 (ou de 100 %), meilleure est la qualité de l'ajustement linéaire.

L'ajustement linéaire est de **bonne qualité** lorsque : $75 \% \leq r^2 \leq 85 \%$

L'ajustement linéaire est de **très bonne qualité** lorsque : $85 \% \leq r^2 \leq 95 \%$

L'ajustement linéaire est de **d'excellente qualité** lorsque : $95 \% \leq r^2 \leq 100 \%$

- Si de plus : $r > 0$, alors la **corrélation** est **positive**

$r < 0$, alors la *corrélation* est *négative*

➡ On a également : $r^2 = a a'$

4 **Indicateur de la qualité de l’ajustement linéaire**

Outre la détermination de a et de b, la *méthode des moindres carrés* permet de définir un indicateur de la qualité de l’ajustement linéaire. Celui-ci sera justifié à condition que l’ensemble des valeurs $\hat{y}_i$ avec : $\hat{y}_i = a x_i + b$ soient « proches » des valeurs y_i correspondantes.

La notion de « proximité » est associée en Statistique à la notion de « dispersion », donc au calcul de la variance, d’où l’indicateur suivant :

$$r^2 = \frac{\text{Variance expliquée par la régression linéaire de y en x}}{\text{Variance totale de y}} \Leftrightarrow r^2 = \frac{V(\hat{y}_i)}{V(y_i)}$$

r^2 : *coefficient de détermination* (en %) r : *coefficient de corrélation linéaire* (jamais en %)
Remarque : Le raisonnement précédent s’applique de même à la droite D’ d’équation : $x = a' y + b'$ de régression de x en y.

V – Exemples

1 **Exemple 1**

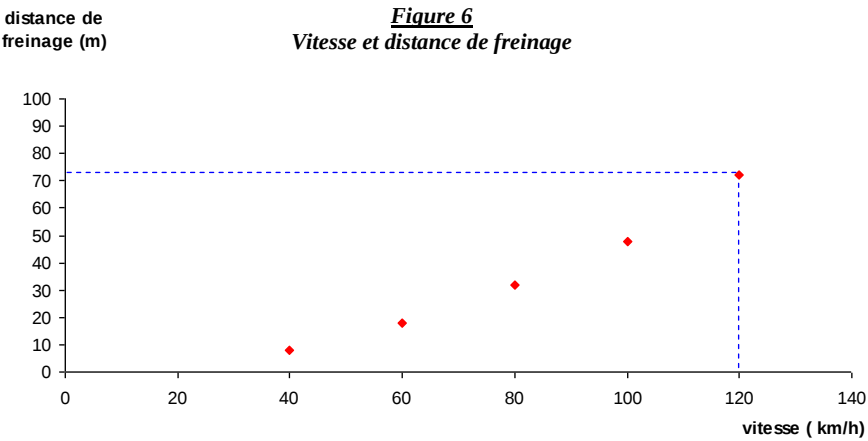
Le tableau suivant retrace la distance de freinage d'un véhicule automobile roulant sur route sèche, en fonction de sa vitesse.

vitesse x (en km/h)	40	60	80	100	120
distance y (en m)	8	18	32	48	72

- 1. Sur la représentation graphique ci-dessous, placer le dernier point de coordonnées : $x = 120$ km/h et $y = 72$ m.
- 2. Compléter le tableau statistique (?).
- 3. a) Déterminer l'équation de la droite D de régression de y en x.
b) Déterminer l'équation de la droite D' de régression de x en y.
c) Calculer le coefficient de corrélation linéaire r. Interpréter.
- 4. Estimer la distance de freinage d'un véhicule roulant à 130 km/h.
- 5. Estimer la vitesse d'un véhicule dont la distance de freinage est de 25 mètres.

Solution

- 1. Représentation graphique :



2. Complétons le tableau statistique (?) :

N = 5 modalités	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	40	8	? 1 600	? 64	? 320
	60	18	? 3 600	? 324	? 1 080
	80	32	6 400	1 024	2 560
	100	48	10 000	2 304	4 800
	120	72	14 400	5 184	8 640
Total	400	178	? 36 000	? 8 900	? 17 400

Remarque : on a développé, sur cet exemple, un tableau de calculs complet car les questions portent sur la totalité de ce qu'il est possible de demander par la méthode des moindres carrés :

- la droite D de régression de y en x
- la droite D' de régression de x en y
- le coefficient de corrélation linéaire r

3. Nous devons d'abord établir les calculs préalables.

➡ Calculs préalables :

■ Vitesse x :

Moyenne :
$$\bar{x} = \frac{1}{N} \left(\sum_i x_i \right) = \frac{1}{5} \times 400 = \boxed{80 \text{ km/h}}$$

Variance :
$$V(x) = \frac{1}{N} \left(\sum_i x_i^2 \right) - \bar{x}^2 = \frac{1}{5} \times 36\,000 - 80^2 = \boxed{800}$$

Ecart-type :
$$\sigma(x) = \sqrt{V(x)} \approx \boxed{28,28 \text{ km/h}} \quad (28,28427125 \dots)$$

■ Distance de freinage y :

Moyenne :
$$\bar{y} = \frac{1}{N} \left(\sum_i y_i \right) = \frac{1}{5} \times 178 = \boxed{35,6 \text{ m}}$$

Variance :
$$V(y) = \frac{1}{N} \left(\sum_i y_i^2 \right) - \bar{y}^2 = \frac{1}{5} \times 8\,900 - 35,6^2 = \boxed{512,64}$$

Ecart-type :
$$\sigma(y) = \sqrt{V(y)} \approx \boxed{22,64 \text{ m}} \quad (22,64155472 \dots)$$

■ Covariance :
$$\text{Cov}(x, y) = \frac{1}{N} \left(\sum_i x_i y_i \right) - \bar{x} \bar{y} = \frac{1}{5} \times 17\,400 - 80 \times 35,6 = \boxed{632}$$

a) Nous pouvons dès lors répondre facilement à la première question posée.

➡ Equation de la droite de régression de y en x : $\boxed{D : y = a x + b}$

■ Calcul de a :

$$a = \frac{\text{Cov}(x, y)}{V(x)} = \frac{632}{800} = \boxed{0,79}$$

■ Calcul de b :

$$b = \bar{y} - a \bar{x} = 35,6 - 0,79 \times 80 = \boxed{-27,60}$$

■ Conclusion :

La droite D de régression de y en x a pour équation : $\boxed{y = 0,79x - 27,60}$

b) De même pour la seconde équation de droite.

➡ Equation de la droite de régression de x en y : $\boxed{D' : x = a' y + b'}$

■ Calcul de a' :

$$a' = \frac{\text{Cov}(x, y)}{V(y)} = \frac{632}{512,64} \approx \boxed{1,23} \quad (1,23283 \dots)$$

☑ Calcul de b' :

$$b' = \bar{x} - a'\bar{y} \approx 80 - 1,23 \times 35,6 \approx \boxed{36,11}$$

☑ Conclusion :

La droite D' de régression de x en y a pour équation : $\boxed{x = 1,23y + 36,11}$

c) On peut, pour terminer, s'intéresser à la qualité de l'ajustement linéaire.

➡ Qualité de l'ajustement linéaire :

☑ Coefficient de corrélation linéaire :

$$r = \frac{\text{Cov}(x,y)}{\sigma(x)\sigma(y)} = \frac{632}{\sqrt{800 \times 512,64}} \approx \boxed{0,9869}$$

☑ Coefficient de détermination :

$$r^2 \approx \boxed{0,9739} \quad \text{ou} \quad r^2 \approx \boxed{97\%}$$

☑ Interprétation :

r^2 est très proche de 1, donc l'ajustement linéaire est de qualité excellente.

4. Estimation de la distance de freinage d'un véhicule roulant à 130 km/h :

On utilise l'équation de la droite D de régression de la distance de freinage y (en mètres) en fonction de la vitesse x (en km/h) : $\boxed{y = 0,79x - 27,60}$

Pour $x = 130 \text{ km/h}$: $y = 0,79 \times 130 - 27,60 = 75,1 \text{ m} \approx \boxed{75 \text{ m}}$

5. Estimation de la vitesse d'un véhicule dont la distance de freinage est de 25 mètres :

On utilise l'équation de la droite D' de régression de la vitesse x (en km/h) en fonction de la distance de freinage y (en mètres) : $\boxed{x = 1,23y + 36,11}$

Pour $y = 25 \text{ m}$: $x = 1,23 \times 25 + 36,11 = 66,86 \text{ km/h} \approx \boxed{67 \text{ km/h}}$

2 Exemple 2

On a observé, durant 8 années, les frais x de publicité (en milliers d'euros) et le chiffre d'affaires annuel y (en millions d'euros) d'une grosse entreprise.

année n°	1	2	3	4	5	6	7	8
Frais de publicité x (en k€)	45	42	38	49	41	44	50	52
Chiffre d'affaires y (en M€)	31	33	20	37	25	30	40	43

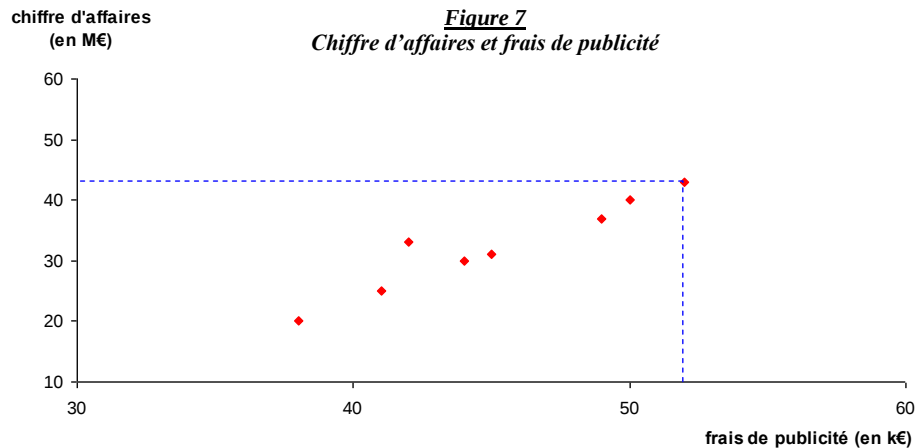
On donne les résultats suivants obtenus à l'aide d'une calculatrice scientifique :

$$\sum_i x_i = 361 \text{ k€}, \quad \sum_i y_i = 259 \text{ M€}, \quad \sum_i x_i^2 = 16\,455, \quad \sum_i y_i^2 = 8\,793 \quad \text{et} \quad \sum_i x_i y_i = 11\,935$$

1. Sur la représentation graphique ci-dessous, placer le dernier point de coordonnées :
 $x = 52 \text{ k€}$ et $y = 43 \text{ M€}$.
2. Déterminer l'équation de la droite D de régression de y en x ainsi que la qualité de l'ajustement linéaire.
3. Pour l'année à venir, cette entreprise prévoit de dépenser 55 milliers d'euros pour sa publicité.
A combien peut-on estimer son chiffre d'affaires à venir ?

Solution

1. Représentation graphique :



2. Nous devons d'abord établir les calculs préalables.

➡ Calculs préalables :

❑ Frais de publicité x :

Moyenne : $\bar{x} = \frac{1}{N} \left(\sum_i x_i \right) = \frac{1}{8} \times 361 = \boxed{45,125 \text{ k€}}$

Variance : $V(x) = \frac{1}{N} \left(\sum_i x_i^2 \right) - \bar{x}^2 = \frac{1}{8} \times 16\,455 - 45,125^2 = \boxed{20,609375}$

Ecart-type : $\sigma(x) = \sqrt{V(x)} \approx \boxed{4,54 \text{ k€}}$

❑ Chiffre d'affaires y :

Moyenne : $\bar{y} = \frac{1}{N} \left(\sum_i y_i \right) = \frac{1}{8} \times 259 = \boxed{32,375 \text{ M€}}$

Variance : $V(y) = \frac{1}{N} \left(\sum_i y_i^2 \right) - \bar{y}^2 = \frac{1}{8} \times 8\,793 - 32,375^2 = \boxed{50,9844}$

Ecart-type : $\sigma(y) = \sqrt{V(y)} \approx \boxed{7,14 \text{ M€}}$

❑ Covariance :

$$\text{Cov}(x, y) = \frac{1}{N} \left(\sum_i x_i y_i \right) - \bar{x} \bar{y}$$

$$\text{Cov}(x, y) = \frac{1}{8} \times 11\,935 - 45,125 \times 32,375$$

$$\text{Cov}(x, y) = \boxed{30,953125}$$

a) Nous pouvons dès lors répondre facilement à la première question posée.

➡ Equation de la droite de régression de y en x : $D : y = a x + b$

❑ Calcul de a :

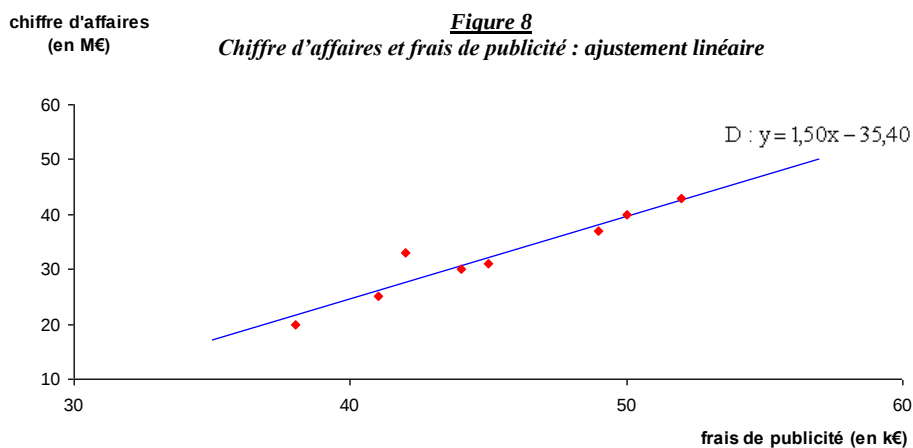
$$a = \frac{\text{Cov}(x, y)}{V(x)} = \frac{30,953125}{20,609375} \approx \boxed{1,50}$$

❑ Calcul de b :

$$b = \bar{y} - a \bar{x} = 32,375 - 1,50 \times 45,125 \approx \boxed{-35,40}$$

❑ Conclusion :

La droite D de régression de y en x a pour équation : $D : y = 1,50x - 35,40$



b) On peut, pour terminer, s'intéresser à la qualité de l'ajustement linéaire.

➡ Qualité de l'ajustement linéaire :

❑ Coefficient de corrélation linéaire :

$$r = \frac{\text{Cov}(x,y)}{\sigma(x) \sigma(y)} = \frac{30,953125}{\sqrt{20,609375 \times 50,9844}} \approx \boxed{0,9549}$$

❑ Coefficient de détermination :

$$r^2 \approx \boxed{0,9118} \quad \text{ou} \quad r^2 \approx \boxed{91\%}$$

❑ Interprétation :

r^2 est proche de 1, donc l'ajustement linéaire est de qualité excellente.

3. Prévision du chiffre d'affaires pour l'année à venir :

On utilise l'équation de la droite D de régression du chiffre d'affaires y en fonction des frais de publicité x :

$$D : y = 1,50x - 35,40$$

Pour $x = 55 \text{ k€}$: $y = 1,50 \times 55 - 35,40 = \boxed{47,1 \text{ M€}}$