Correction TD : Statistiques

Exercice 1(points)(BAC 2005)

1. Calcul du coefficient de corrélation linéaire

$$r = \frac{\operatorname{cov}(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$$

$$\bar{x} = \frac{1}{8}\sum_{i=1}^{8} x_i = \frac{60+80+100+120+140+160+180+200}{8} = \frac{1040}{8} = 130$$

$$\bar{y} = \frac{1}{8}\sum_{i=1}^{8} y_i = \frac{952+805+630+522+510+324+205+84}{8} = \frac{4032}{8} = 504$$

$$V(x) = \frac{1}{8}\sum_{i=1}^{8} x_i^2 - \bar{x}^2 = \frac{152000}{8} - (130)^2 = 2100$$

$$V(y) = \frac{1}{8}\sum_{i=1}^{8} y_i^2 - \bar{y}^2 = \frac{2637870}{8} - (504)^2 = 75717{,}75$$

$$\operatorname{cov}(x,y) = \frac{1}{8}\sum_{i=1}^{8} x_i y_i - \bar{x}\bar{y} = \frac{424100}{8} - 130\times 504 = -12507{,}5$$

$$r = \frac{-12507{,}5}{\sqrt{2100}\times\sqrt{75717{,}75}} = \frac{-12507{,}5}{45{,}83\times 275{,}17} \approx -0{,}99$$

$r \approx -1$, donc la valeur trouvée justifie la recherche d'un ajustement linéaire.

2. Équation de la droite de régression de y en x :

$$y = ax + b, \text{ avec :}$$

$$a = \frac{\operatorname{cov}(x,y)}{V(x)} = \frac{-12507{,}5}{2100} = -5{,}95$$

$$b = \bar{y} - a\bar{x} = 504 - (-5{,}95)(130) = 1277{,}5$$

$$\Rightarrow y = -5{,}95x + 1277{,}5$$

3. Les frais de conception sont de 28 000 000 F. Le prix de fabrication de chaque produit est de 25 000 F.

a x est le prix de vente, donc y est le nombre d'exemplaires du produit.

le prix de vente est

$$yx = (-5{,}95x + 1277{,}5)x \quad \text{en milliers de francs}$$

le prix de revient est

$$25000y + 28000000 = 25y + 28000 \quad \text{en milliers de francs.}$$

Donc

$$z = (-5{,}95x + 1277{,}5)x - 25y - 28000$$

$$z = (-5{,}95x + 1277{,}5)x - 25(-5{,}95x + 1277{,}5) - 28000$$

$$z = -5{,}95x^2 + 1426{,}25x - 59937{,}5$$

b Déterminons le prix de vente x permettant de réaliser un bénéfice maximum.

On a :

$$z(x) = -5{,}95x^2 + 1426{,}25x - 59937{,}5$$

z est une fonction continue et dérivable en x sur $\mathbb{R}$ et :

$$z'(x) = -11{,}9x + 1426{,}25 \qquad \Rightarrow \qquad z'(x) = 0 \quad \text{si} \quad x = 119{,}85$$

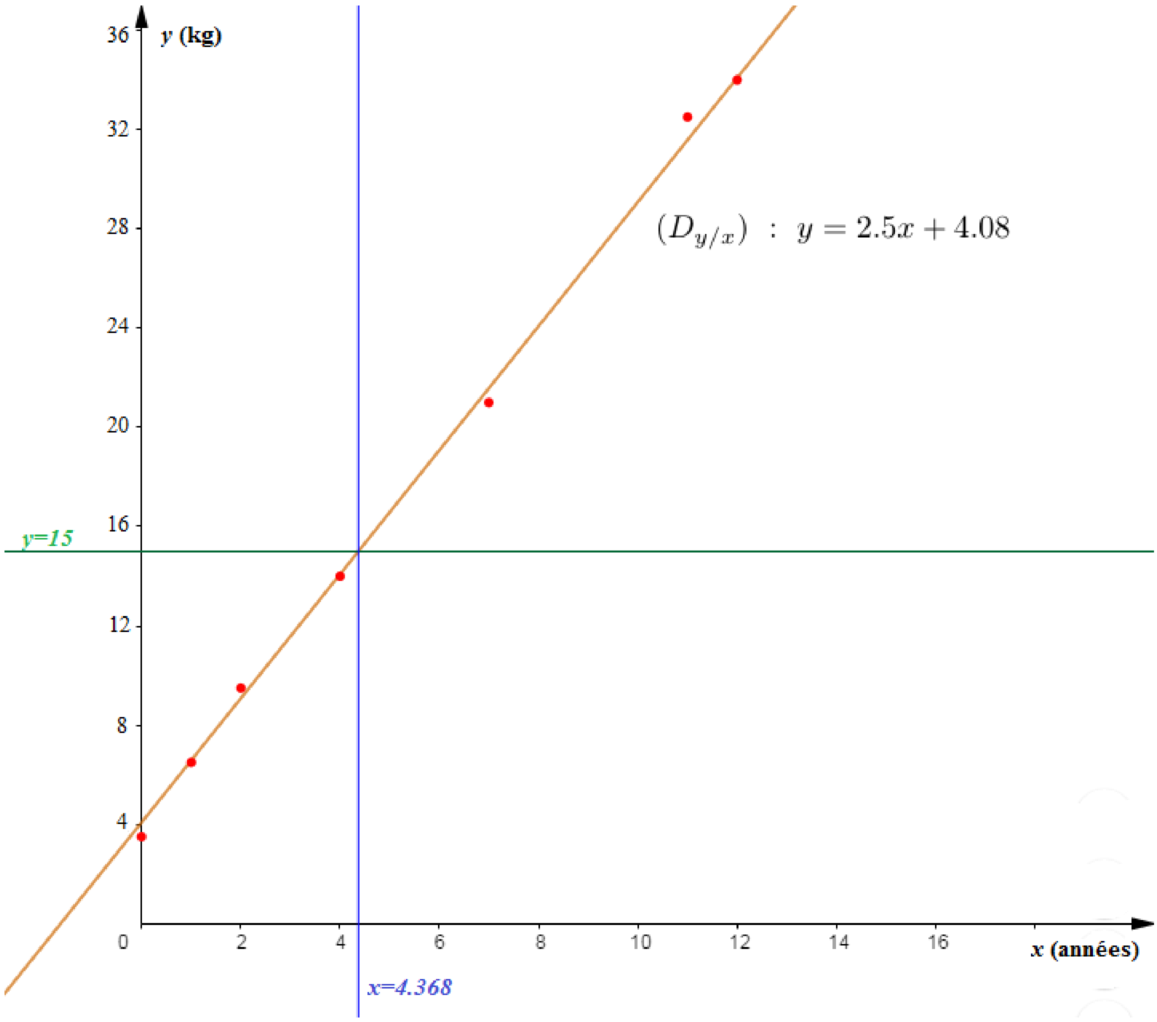
On voit ainsi que z atteint son maximum pour $x = 119{,}85$ en milliers de francs.
Donc le prix de vente permettant de réaliser un bénéfice maximum est $x = 119{,}850$ F.

$$z(119{,}85) = -5{,}95 \times (119{,}85)^2 + 1426{,}25 \times 119{,}85 - 59937{,}5 = 25532{,}628 \quad \text{en milliers de francs}$$

D'où le bénéfice maximum est **25.532.628 F**

Exercice 2(04,5 points)(BAC 2008)

1 Le nuage de point



2 G a pour coordonnées $(\bar{x},\ \bar{y})$ avec :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{7} x_i \simeq 5{,}28 \quad \text{et} \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{7} y_i \simeq 17{,}28$$

3 a

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{7} x_i y_i - \bar{x}\times\bar{y} \simeq 50{,}63$$

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{7} x_i^2 - \bar{x}^2} \simeq 4{,}66 \quad \text{et} \quad \sigma_y = \sqrt{\frac{1}{N}\sum_{i=1}^{7} y_i^2 - \bar{y}^2} \simeq 11{,}39$$

Le coefficient de corrélation linéaire est donné par la formule :

$$r = \frac{\sigma_{xy}}{\sigma_x \times \sigma_y} \simeq 0{,}99$$

b r est très proche de 1, donc on a une très bonne corrélation.

4 $D_{y/x}$ a pour équation :

$$y - \bar{y} = a(x - \bar{x}) \quad \text{avec} \quad a = \frac{\sigma_{xy}}{\sigma_x^2}$$

Après calculs, on trouve qu'une équation de $D_{y/x}$ est :

$$y = 2{,}5x + 4{,}08$$

5 a D'après le graphique, on constate que les valeurs de y supérieures à 15 correspondent aux valeurs de x supérieures à 4,368. On ainsi dire qu'à partir de 5 ans le poids de l'enfant sera supérieur à 15 kg.

b Pour retrouver ce résultat par le calcul, on considère l'équation $D_{y/x}$ de la droite de régression de y en x.

Soit $(D_{y/x}) : y = 2{,}5x + 4{,}08$, alors on a :

$$\begin{aligned} y > 15 \Leftrightarrow &\quad 2{,}5x + 4{,}08 > 15 \\ \Leftrightarrow &\quad 2{,}5x > 15 - 4{,}08 \\ \Leftrightarrow &\quad x > \frac{10{,}92}{2{,}5} \\ \Leftrightarrow &\quad x > 4{,}368 \end{aligned}$$

D'où, $y > 15 \;\Leftrightarrow\; x > 4{,}368$

Exercice 3(03 points)(BAC 2009)

1 (D_1) droite de régression de Y en X ayant pour équation :

$$y = ax + b, \text{ on a :}$$

$$a = \frac{\operatorname{cov}(X,Y)}{V(X)} \qquad \text{et} \qquad b = \bar{y} - a\bar{x}$$

(D_2) droite de régression de X en Y ayant pour équation :

$$x = a'y + b', \text{ on a :}$$

$$a' = \frac{\operatorname{cov}(Y,X)}{V(Y)} \qquad \text{et} \qquad b' = \bar{x} - a'\bar{y}$$

On en déduit que :

$$\begin{aligned} aa' &= \frac{\operatorname{cov}(X,Y)}{V(X)} \cdot \frac{\operatorname{cov}(Y,X)}{V(Y)} \\ &= \frac{(\operatorname{cov}(X,Y))^2}{V(X)\cdot V(Y)} \\ &= \left(\frac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}\right)^2 \\ &= r^2 \end{aligned}$$

D'où, $\boxed{aa' = r^2}$

(D_1) droite de régression de Y en X ayant pour équation réduite :

$$y = 2{,}4x, \quad \text{on a :} \quad a = 2{,}4 \quad \text{et} \quad b = 0$$

(D_2) droite de régression de X en Y ayant pour équation réduite :

$$x = \frac{3{,}5}{9}y + \frac{24}{9}, \quad \text{on a :} \quad a' = \frac{3{,}5}{9} \quad \text{et} \quad b' = \frac{24}{9}$$

D'après la question précédente, le coefficient de corrélation vérifie :

$$\begin{aligned} r^2 &= aa' \\ &= 2{,}4 \times \frac{3{,}5}{9} \\ &= \frac{14}{15} \end{aligned}$$

Puisque $r = \dfrac{\operatorname{cov}(X,Y)}{\sigma(X)\sigma(Y)}$, que $\sigma(X)$ et $\sigma(Y)$ sont positifs par définition et que $\operatorname{cov}(X,Y)$ est positif par hypothèse, alors r est positif.

Donc,

$$r = \sqrt{\frac{14}{15}}$$

2 On a :

$$\begin{cases} -a\bar{x} + \bar{y} = b & (1) \\ \bar{x} - a'\bar{y} = b' & (2) \end{cases}$$

Je garde l'équation 1. Je multiplie l'équation 2 par a pour éliminer $\bar{x}$:

$$\begin{cases} -a\bar{x} + \bar{y} = b \\ a\bar{x} - aa'\bar{y} = ab' \end{cases}$$

J'additionne membre à membre : $(1 - aa')\bar{y} = b + ab'$, c'est-à-dire :

$$\bar{y} = \frac{b + ab'}{1 - r^2}$$

Pour trouver $\bar{x}$, j'utilise l'équation la plus simple ; ici c'est la 2 : $\bar{x} = b' + a'\bar{y}$, c'est-à-dire :

$$\begin{aligned} \bar{x} &= b' + a' \cdot \frac{b + ab'}{1 - r^2} \\ &= \frac{b'(1 - r^2) + a'(b + ab')}{1 - r^2} \\ &= \frac{b' + a'b}{1 - r^2} \end{aligned}$$

Donc,

$$\bar{x} = \frac{b' + a'b}{1 - r^2}$$

Application numérique : Comme $\dfrac{1}{1 - r^2} = 15$, on a :

$$\bar{y} = 15 \times 2{,}4 \times \frac{24}{9} \qquad \text{et} \qquad \bar{x} = 15 \times \frac{24}{9}$$

$$\boxed{\bar{y} = 96 \quad \text{et} \quad \bar{x} = 40}$$

Exercice 4(03 points)(BAC 2010)

On considère le tableau ci-dessous indiquant les résultats d'une étude sur le nombre d'années x en service des ouvriers d'une entreprise et de leur salaire y en milliers de francs.

Notons x_i les modalités de x et n_i l'effectif de x_i, avec $1 \leq i \leq 6$.

Et notons y_j les modalités de y et n_j l'effectif de y_j, avec $1 \leq j \leq 4$.

Soit N l'effectif total.

$x\backslash y$	2	6	10	14	18	22	n_j
75	a	5	0	0	0	0	$a+5$
125	0	7	1	0	2	0	10
175	2	0	9	8	15	4	38
225	0	1	0	3	b	1	$b+5$
n_i	$a+2$	13	10	11	$b+17$	5	$N=a+b+58$

1. Déterminons a et b pour que $\bar{x} = \dfrac{596}{59}$ et $\bar{y} = \dfrac{8450}{59}$.

On sait que :

$$\bar{x} = \frac{\sum_{i=1}^{6} n_i x_i}{N} \qquad \text{et} \qquad \bar{y} = \frac{\sum_{j=1}^{4} n_j y_j}{N}$$

Alors :

$$\bar{x} = \frac{2(a+2) + 6\times 13 + 10\times 10 + 14\times 11 + 18(b+17) + 22\times 5}{a+b+58}$$

$$\bar{y} = \frac{75(a+5) + 10\times 125 + 38\times 175 + 225(b+5)}{a+b+58}$$

Or $\bar{x} = \dfrac{596}{59}$ et $\bar{y} = \dfrac{8450}{59}$, d'où :

$$\frac{2(a+2) + 6\times 13 + 10\times 10 + 14\times 11 + 18(b+17) + 22\times 5}{a+b+58} = \frac{596}{59}$$

$$\frac{75(a+5) + 10\times 125 + 38\times 175 + 225(b+5)}{a+b+58} = \frac{8450}{59}$$

On obtient ainsi le système suivant :

$$\begin{cases} 239a - 233b = 4900 \\ 161a - 193b = 2580 \end{cases}$$

D'où $a = 40 \quad$ et $\quad b = 20$

On suppose $a = 40 \quad$ et $\quad b = 20$ dans la suite.

En associant à chaque valeur x_i la moyenne m_i de la série conditionnelle $(y/x = x_i)$, on a obtenu le tableau suivant :

x	2	6	10	14	18	22
m	80	113	170	189	199	185

2 a Calculons le coefficient de corrélation linéaire.

b Déterminons d'abord les moyennes $\bar{x}$ et $\bar{m}$, les variances V_x, V_m, les écarts-types σ_x, σ_m, et la covariance de x et m.
Notons que la série statistique double (x, m) est injective.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{6} x_i \quad , \quad \bar{m} = \frac{1}{N}\sum_{i=1}^{6} m_i$$

$$V_x = \frac{1}{N}\sum_{i=1}^{6} x_i^2 - \bar{x}^2 \quad , \quad V_m = \frac{1}{N}\sum_{i=1}^{6} m_i^2 - \bar{m}^2$$

$$\sigma_x = \sqrt{V_x}, \qquad \sigma_m = \sqrt{V_m}, \qquad \text{et} \quad \operatorname{cov}(x,m) = \frac{1}{N}\sum_{i=1}^{6} x_i m_i - \bar{x}\bar{m}$$

$$\bar{x} = 12, \quad \bar{m} = 156, \quad V_x \simeq 46{,}66, \quad V_m \simeq 1933{,}33$$

$$\sigma_x \simeq 6{,}83, \quad \sigma_m \simeq 43{,}96, \quad \operatorname{cov}(x,m) \simeq 267{,}33$$

c Calculons maintenant le coefficient de corrélation linéaire r :

$$r = \frac{\operatorname{cov}(x,m)}{\sigma_x \sigma_m}$$

D'où $r \simeq 0{,}89$
Puisque r est proche de 1, il y a alors une forte corrélation entre x et m.

d La droite de régression de m en x, notée $D_{m/x}$, a pour équation $m = ax + b$ avec :

$$a = \frac{\operatorname{cov}(x,m)}{V_x} \qquad \text{et} \qquad b = \bar{m} - a\bar{x}$$

$$\Rightarrow \quad D_{m/x} : \quad m = 5{,}73x + 87{,}25$$

e Si $x = 30$, alors $m \simeq 259{,}128$. D'où le salaire moyen d'un ouvrier ayant 30 ans d'ancienneté est sensiblement égal à **259130 F**.

Exercice 5(05 points)(BAC 2013)

1 a Soit :

$$r = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Donc, $r = -0{,}973$
Ce qui signifie qu'il y a une forte corrélation.

b) La droite de régression de Y en X est :

$$y = ax + b \quad \text{avec} \quad a = \frac{\text{cov}(X,Y)}{V(X)} \quad \text{et} \quad b = \bar{Y} - a\bar{X}$$

$$\Rightarrow \quad y = -0{,}874x + 4{,}12$$

Ainsi, $\boxed{y = -0{,}874x + 4{,}12}$

b Si $x = 6$ alors $y \simeq -1{,}124$
Cette équation ne permet pas d'estimer le degré de salinité car au $6^{\text{ième}}$ mois de pluie, le degré de salinité ne peut être négatif.

2 $Z = \ln(Y - 1)$

a Le tableau correspondant à la série (X, Z) est donné par :

X_i	0	1	2	3	4
Z_i	1,182	0,875	0,010	−1,830	−4,610

b Le coefficient de corrélation linéaire de cette série (X, Z) est :

$$r = \frac{\text{cov}(X,Z)}{\sigma_X \sigma_Z} = -0{,}944$$

La droite de régression de Z en X est :

$$z = ax + b \quad \text{avec} \quad a = \frac{\text{cov}(X,Z)}{V(X)} \quad \text{et} \quad b = \bar{Z} - a\bar{X}$$

$$\Rightarrow z = -1{,}428x + 1{,}982$$

D'où, $\boxed{z = -1{,}428x + 1{,}982}$

c Exprimons Y en fonction de X. On a : $z = \ln(y-1)$ et $z = -1{,}428x + 1{,}982$ d'où,
$$\ln(y-1) = -1{,}428x + 1{,}982 \quad \Rightarrow \quad y - 1 = e^{-1{,}428x+1{,}982}$$
$$\Rightarrow \quad y = e^{-1{,}428x+1{,}982} + 1$$

Ainsi, $\boxed{y = e^{-1{,}428x+1{,}982} + 1}$

d Si $x = 6$ alors, $y = 1{,}001$. Le degré de salinité estimé au $6^{\text{ième}}$ mois est positif, il est très proche de celui du quatrième mois et lui est inférieur.
Donc, l'équation $y = e^{-1{,}428x+1{,}982} + 1$ nous permet de faire cette estimation.

Exercice 6(02,5 points)(BAC 2015)

1 Le coefficient de corrélation linéaire r est défini par

$$r = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

D'où : $r \approx$ **0,69**.

2 a La droite de régression de Y en X, $(D_{Y/X})$, a pour équation $y = \mathbf{92{,}59x - 4{,}35}$.

b Il faut investir 3,29 milliards de *FCFA* si l'on désire un chiffre d'affaire de 300 milliards.