Exercices corrigés de statistiques

Exercice n°1

Lors d'une période de sécheresse, un agriculteur relève la quantité totale (en m^3) utilisée par son exploitation depuis le premier jour et donne le résultat suivant :

Nombre de jours écoulés x_i	1	3	5	8	10
Volume utilisé (en m^3) y_i	2,25	4,3	8	17,5	27

Le plan est muni d'un repère orthogonal.

On prendra pour unité sur l'axe des abscisses 1 cm pour 1 jour et sur l'axe des ordonnées 0,5 cm pour 1 m^3.

1) Représentez alors la série $(x_i;y_i)$.

2) Donnez l'équation de la droite Δ des moindres carrés sous la forme $y=ax+b$ (a et b sont les arrondis à 10^{-2} près des valeurs lues sur la calculatrice).

Représentez la droite Δ sur le graphique.

3) Le nuage de points permet d'envisager un ajustement par la parabole P qui passe par des points A(1;2,25) et B(10;27), et qui a pour équation $y=ax^2+b$ où a et b sont deux nombres réels

a) Déterminez a et b et donnez l'équation de la parabole P.

b) Représentez la parabole P sur le graphique.

4) Dans cette question, on compare les deux ajustements à l'aide du tableau suivant :

x_i	1	3	5	8	10		
y_i	2,25	4,3	8	17,5	27		
$\lvert y_i-(ax_i+b)\rvert$	2,54	0,91	2,71			Total T_1 :	
$\lvert y_i-(ax_i^2+b)\rvert$	0	0,05	0,25			Total T_2 :	

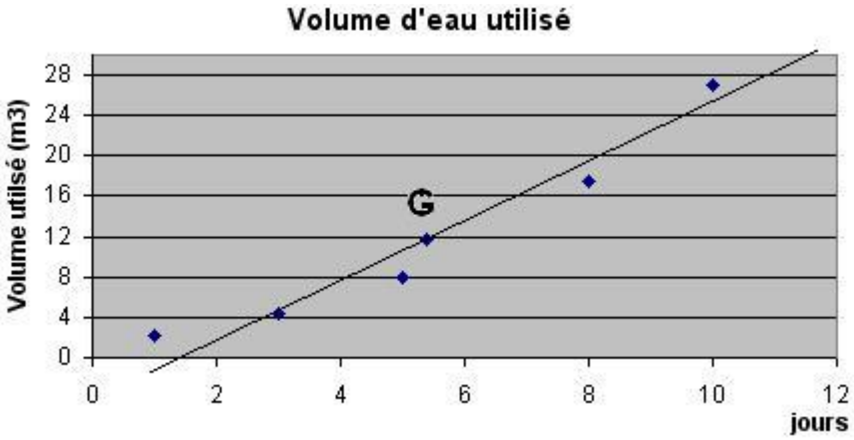
Les deux totaux calculés évaluent, pour chaque ajustement, la somme des écarts entre les ordonnées des points du nuage et les ordonnées des points de même abscisse de l'ajustement.

Donnez les arrondis à 10^{-1} près des deux totaux T_1 et T_2 calculés ci-dessus.

Déduisez l'ajustement qui paraît le mieux adapté.

Correction :

1) Les coordonnées du G sont données par
$$G\begin{vmatrix} x_G = \bar{x} = \dfrac{\sum x_i}{5} = 5,4 \\ y_G = \bar{y} = \dfrac{\sum y_i}{5} = 11,81 \end{vmatrix}$$



2) L »'équation de la droite Δ des moindres carrés est de la forme $y=ax+b$ avec $a = \dfrac{Cov(x;y)}{V(x)}$ et $G \in \Delta$

On obtient :en utilisant une calculatrice :

L'équation de la droite d'ajustement de y en x est donc $y=2,74x-3,04$

3) a) Si le point A(1;2,25) appartient à la parabole P d'équation $y=ax^2+b$, alors les coordonnées de ce point vérifient l'équation de la parabole, c'est-à-dire

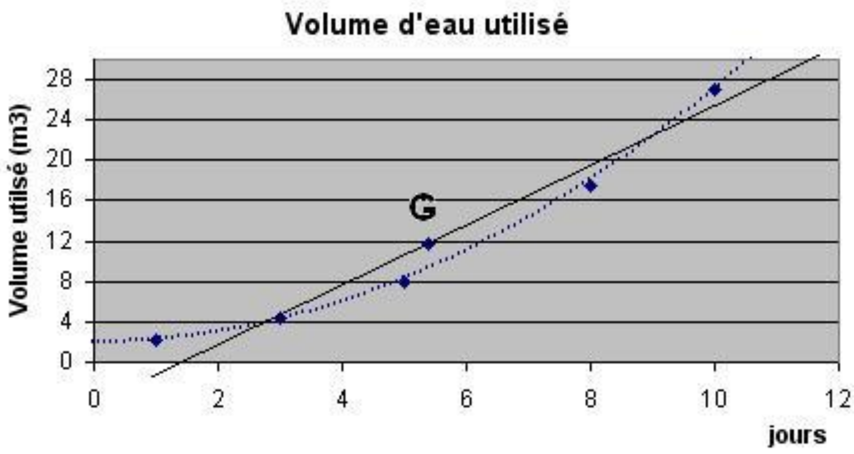
$2,25=a\text{x}1^2+b \Leftrightarrow a+b=2,25$

De même pour le point B : $27=a\text{x}10^2+b \Leftrightarrow 100a+b=27$

On résout donc le
$$\begin{cases} a+b=2,25 & L_1 \\ 100a+b=27 & L_2 \end{cases} \Leftrightarrow \begin{cases} b=2,25-a & L_1 \\ 99a=24,75 & L_2-L_1 \end{cases} \Leftrightarrow \begin{cases} b=2,25-0,25=2 & L_1 \\ a=0,25 & L_2-L_1 \end{cases}$$

L'équation de la parabole est donc $y=0,25x^2+2$

b)



4)

x_i	1	3	5	8	10	
y_i	2,25	4,3	8	17,5	27	
$\lvert y_i-(ax_i+b)\rvert$	2,54	0,91	2,71	$\lvert 17,5-(2,74\times 8+3,04)\rvert=7,46$	$\lvert 27-(2,74\times 10+3,04)\rvert=3,44$	**Total T_1 : 17,06**
$\lvert y_i-(ax_i^2+b)\rvert$	**0**	0,05	0,25	$\lvert 17,5-(0,25\times 8+2)\rvert=3,5$	**0**	**Total T_2 : 3,8**

L'ajustement qui paraît le mieux adapté est celui pour lequel la somme des écarts est la plus faible. Il s'agit donc de l'ajustement à l'aide de la parabole *P*.

Exercice n°2. l'exercice qui résume tout

Le tableau suivant donne la dépense, en millions d'euros, des ménages en produits informatiques (matériels, logiciels, réparations) de 1990 à 1999 :

Année	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Rang x_i de l'année	0	1	2	3	4	5	6	7	8	9
Dépense y_i	398	451	423	501	673	956	1077	1285	1427	1490

1) a) Dessiner le nuage de points $M_i(x_i\ ;y_i)$ dans le plan muni d'un repère orthogonal avec, pour unités graphiques 1 cm pour un rang en abscisse, 1cm pour 200 millions d'euros en ordonnée.

b) Déterminez les coordonnées de G, point moyen de nuage. Placez le point G.

2) Le modèle étudié dans cette question sera appelé « droite de Mayer »

a) G_1 désigne le point moyen des 5 premiers points du nuage et G_2 celui des 5 derniers points. Déterminer les coordonnées de G_1 et G_2.

Placez ces points sur le graphique précédent et tracez la droite (G_1G_2).Le point G appartient-il à cette droite ?

b) Donnez l'équation de la droite (G_1G_2) sous la forme $y=ax+b$ (on arrondira les coefficients à 0,1 prés)

c) Calculez la somme des carrés des résidus pour cet ajustement : $S_1 = \sum_{i=0}^{9} \left[y_i - (ax_i + b) \right]^2$

d) En utilisant cet ajustement, effectuer une prévision sur les dépenses de l'année 2005.

3) Ajustement des moindres carrés

a) Donner, à l'aide de la calculatrice, une équation de la droite d'ajustement affine de y en x , sous la forme $y=mx+p$ par la méthode des moindres carrés (*les coefficients seront arrondis à 0,1 près*).

b) Représenter D dans le repère précédent.

c) Calculez la somme des carrés des résidus pour cet ajustement : . Conclusion ?

d) En utilisant cet ajustement, effectuer une prévision sur les dépenses de l'année 2005.

4) Ajustement logarithmique

La croissance des dépenses semblant « ralentir » entre 1997 et 1999, on envisage un ajustement logarithmique entre 1994 et 1999.

On pose $t_i=\ln(x_i)$

a) Recopier et compléter le tableau suivant où t_i est arrondi est arrondi à 10^{-3}

t_i						
y_i	673	956	1077	1285	1427	1490

b) Ecrire une équation de la droite d'ajustement affine de y en t par la méthode des moindres carrés (*les coefficients seront arrondis à* 10^{-3} *près*).

c) En utilisant cet ajustement, effectuer une prévision sur les dépenses de l'année 2005.

d) Tracer la courbe d'équation $y=f(x)$, où on explicitera l'expression de f

5) Ajustement exponentiel

Si, au contraire de la question 4, on ne s'intéresse qu'aux années 1990 à 1996, la forme du nuage suggère plutôt un ajustement exponentiel.

Pour $0 \leq i \leq 6$, on pose $z_i=\ln(y_i)$

a) Recopier et compléter le tableau suivant où z_i est arrondi est arrondi à 10^{-3}

x_i	0	1	2	3	4	5	6
z_i							

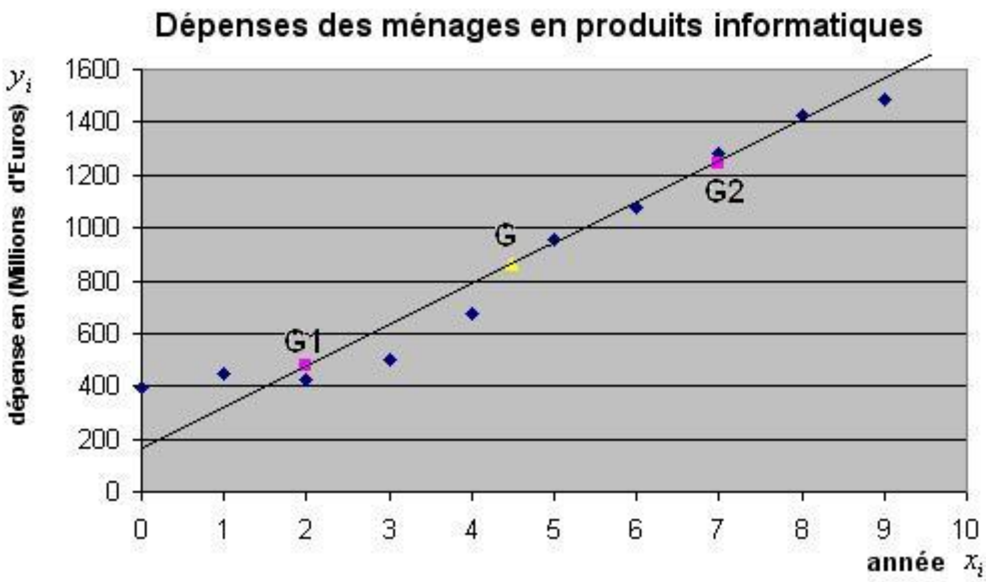
b) Ecrire une équation de la droite d'ajustement affine de z en x par la méthode des moindres carrés (*les coefficients seront arrondis à* 10^{-3} *près*).

c) En utilisant cet ajustement, effectuer une prévision sur les dépenses de l'année 2000.

d) Tracer la courbe d'équation $z=g(x)$, où on explicitera l'expression de g

Correction :

1) a) Le nuage de points $M_i(x_i\,;y_i)$ représentant la série statistique double est :



b) Les coordonnées de G sont :

$$\begin{cases} x_G = \dfrac{0+1+2+...8+9}{10} = 4,5 \\ y_G = \dfrac{398+451+...+1490}{10} = 868,1 \end{cases}$$ (représenté par un triangle sur le graphique)

2) a) Les coordonnées de G_1 et G_2 sont :

$$\begin{cases} x_{G1} = \dfrac{0+1+2+3+4}{5} = 2 \\ y_{G1} = \dfrac{398+451+...+673}{5} = 489,2 \end{cases}$$ et $$\begin{cases} x_{G2} = \dfrac{5+6+7+8+9}{5} = 7 \\ y_{G2} = \dfrac{956+1077+1490}{5} = 1247 \end{cases}$$ (représentés par deux carrés sur le graphique)

Le point G appartient à (G_1G_2) car puisque $x_G = \frac{x_{G1} + x_{G2}}{2}$ et , le point G est le milieu de $[G_1G_2]$

b) l'équation de la droite (G_1G_2) est de la forme $y=ax+b$ avec $a = \frac{y_{G2} - y_{G1}}{x_{G2} - x_{G1}} = \frac{1247 - 489,2}{7-2} = 151,56$.

Puisque le point G appartient à la droite (G_1G_2), ses coordonnées vérifient l'équation de cette droite , donc $b=\bar{y}-a\bar{x}$ =868,1-151,56x4,5=186,08

L'équation de (G_1G_2) est donc y=151,6x+186,1, où les coefficients ont été arrondisà 0,1 près

Remarque : Les coefficients ne sont arrondis qu'au moment de la présentation des résultats et surtout pas pendant les étapes intermédiaires de calcul !

c) Pour cet ajustement,

$$S_1 = \sum_{i=0}^{9}\left[y_i - (ax_i + b)\right]^2 = \sum_{i=0}^{9}\left[y_i - (151,6x_i + 186,1)\right]^2 = 102348,9$$.

Détail du calcul :

x_i	0	1	2	3	4	5	6	7	8	9
y_i	398	451	423	501	673	956	1077	1285	1427	1490
$[y_i-(151,6x_i+186,1)]^2$	44901,61	12836,89	4395,69	19572,01	14280,25	141,61	349,69	1421,29	789,61	3660,25

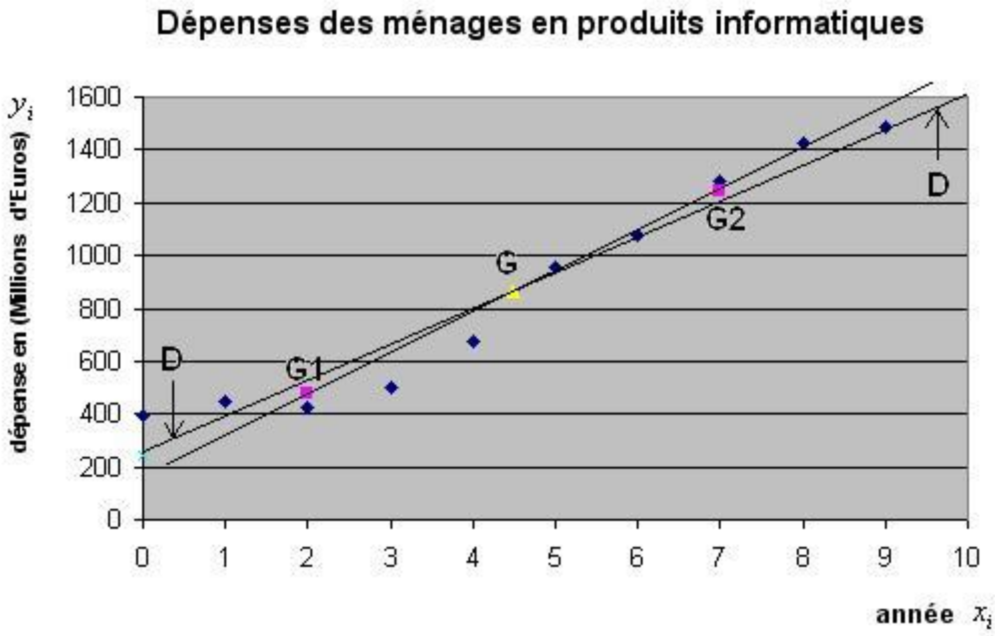
d) L'année 2005 correspond à x=15, donc d'après l'ajustement, à une dépense égale à y=151,6x15+186,1=2460,1 millions d'euros

3) a)

L'équation de y en x est donc y=139,3x+241,3 (les coefficients ont été arrondis à 0,1 près)

b) Cette droite D passe par le point G

Pour la tracer, un deuxième point est suffisant (prenons l'ordonnée à l'origine : x=0=>y=241,3)



c) Pour cet ajustement,

$$S_2 = \sum_{i=0}^{9}\left[y_i - (ax_i + b)\right]^2 = \sum_{i=0}^{9}\left[y_i - (139,3x_i + 241,3)\right]^2 = 89824,05$$

détail du calcul :

x_i	0	1	2	3	4	5	6	7	8	9
y_i	398	451	423	501	673	956	1077	1285	1427	1490
$[y_i-(139,3x_i+241,3)]^2$	24554,89	4956,16	9389,61	25027,24	15750,25	331,24	0,01	4705,96	5083,69	25

La somme de cet ajustement étant nettement inférieure à celle du modèle mayer, on peut en déduire que pour la période 1990-1999, le modèle des moindres carrées est plus pertinent

d) L'année 2005 correspond à x=15, donc d'après l'ajustement, à une dépense égale à y=139,3x15+241,3=2330,8 millions d'euros

4) a)

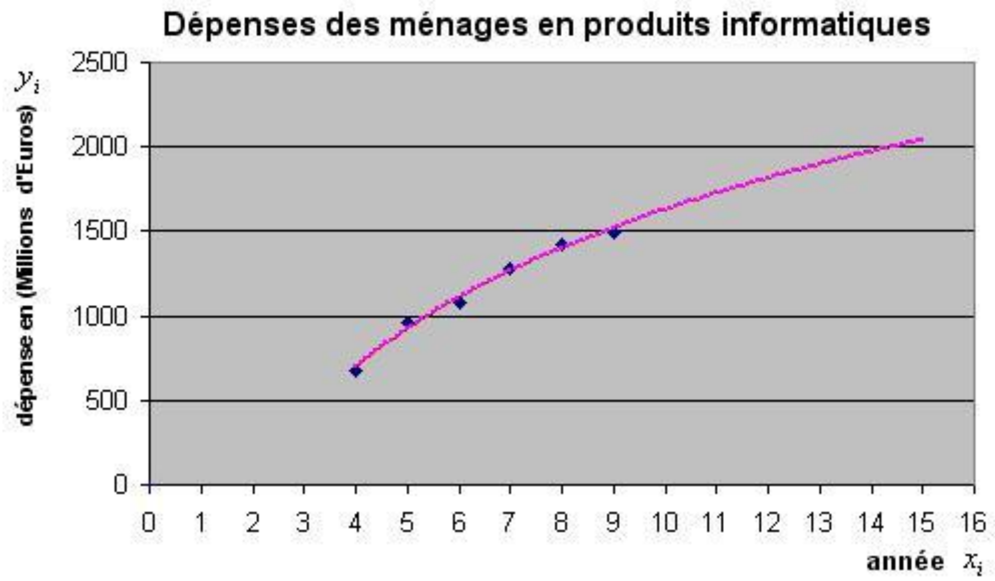
t_i	1,386	1,609	1,792	1,946	2,079	2,197
y_i	673	956	1077	1285	1427	1490

b) Une équation de la droite de régression de y en t est donc y=1020,514t-721,139

c) L'année 2005 correspond à x=15 donc à t=2,708, donc d'après l'ajustement, à y=1020,514x2,708-721,139=2042,413 millions d'euros

d) Puisque t=ln$x$, on en déduit y=1020,514ln(x)-721,139.

La courbe de f définie par $f(x)$=1020,514ln(x)-721,139 est donnée ci-dessous :



5) a)

x_i	0	1	2	3	4	5	6
z_i	5,986	6,111	6,047	6,217	6,512	6,863	6,981

b) Une équation de la Une équation de la droite de régression de z en x est donc z=0,177x+5,857

c) L'année 2000 correspond à x=10 donc à z=0,177x10+5,857=7,627 et puisque z = ln$y$ on a donc $y = e^z = e^{7,627} = 2052,882$ millions d'euros

d) Puisque z = ln$y$, on en déduit

$$y = e^z = e^{0,177x+5,857}$$