Statistiques à deux variables : les exercices

Exercice 1

(*Dans tout cet exercice, les résultats concernant la population seront arrondis au million*).

Le tableau suivant donne l'évolution de la population de l'Inde de 1951 à 1991.

année	1951	1961	1971	1981	1991
Rang x_i	1	2	3	4	5
Population y_i (en millions)	361	439	548	683	846
z_i					

On cherche à étudier l'évolution de la population y en fonction du rang x de l'année.

1. Représenter graphiquement le nuage de points $(x_i\ ;\ y_i)$ dans le plan muni d'un repère orthonormal d'unités graphiques 2 sur l'axe des abscisses et 1 cm pour 100 millions sur l'axe des ordonnées.
2. Le modèle étudié dans cette question sera appelée « droite de Mayer ».
   (a) G_1 désigne le point moyen des trois premiers points du nuage et G_2 celui des deux derniers points. Déterminer les coordonnées de G_1 et de G_2.
   (b) Déterminer l'équation réduite de (G_1G_2) sous la forme $y = ax + b$.
   (c) Tracer la droite (G_1G_2) sur le graphique précédent.
   (d) En utilisant cet ajustement, calculer la population de l'Inde que l'on pouvait prévoir pour 2001.
3. (a) À l'aide de la calculatrice, déterminer un ajustement affine de y en x par la méthode des moindres carrés. Déterminer le coefficient de corrélation linéaire.
   (b) Tracer cette droite $\mathcal{D}$ sur le graphique précédent.
   (c) En utilisant cet ajustement, calculer la population de l'Inde que l'on pouvait prévoir pour 2001.
4. On cherche un autre ajustement et on se propose d'utiliser le changement de variable suivant : $z = \ln y$.
   (a) Recopier le tableau ci-dessus et compléter la dernière ligne.
   (b) À l'aide de la calculatrice, déterminer un ajustement affine de z en fonction de x par la méthode des moindres carrés (*les coefficients seront arrondis au millième*).
   (c) Déterminer le coefficient de corrélation linéaire, comparer avec celui trouvé dans la question 3. et conclure.
   (d) En déduire qu'une approximation de la popu1ation y, exprimée en millions d'habitants, en fonction du rang x de l'année est donnée par : $y \approx 289\ e^{0,215x}$.
   (e) Représenter graphiquement cette fonction $\mathcal{C}$ sur le graphique précédent.
   (f) En utilisant cet ajustement, calculer la population que l'on pouvait prévoir pour 2001.

5. Les résultats obtenus en 2001 ont révélé que la population comptait 1027 millions d'habitants. Déterminer une estimation de la population en 2011 en choisissant le modèle qui semble le plus approprié. Justifier ce choix.

Exercice 2

Une entreprise fabrique des chaudières de deux types :

– des chaudières dites « à cheminée »,

– des chaudières dites « à ventouse ».

Les deux parties de cet exercice peuvent être traitées de façon indépendante.

A. Ajustement affine

Le nombre de chaudières fabriquées lors des années précédentes est donné par le tableau suivant :

Rang de l'année : x_i	0	1	2	3	4	5
Nombre de chaudières fabriquées par milliers : y_i	15,35	15,81	16,44	16,75	17,19	17,30

1. À l'aide d'une calculatrice, déterminer :
   (a) le coefficient de corrélation linéaire de la série statistique double de variables x et y ; arrondir à 10^{-2} ;
   (b) déterminer une équation de la droite de régression de y en x, sous la forme $y = ax + b$, où a sera arrondi à 10^{-3} et b sera arrondi à l'unité.
2. En supposant que la tendance observée se poursuive pendant deux années, estimer le nombre de chaudières qui seront fabriquées l'année de rang 7.

B. Probabilités conditionnelles

L'entreprise a fabriqué en un mois 900 chaudières à cheminée et 600 chaudières à ventouse.
Dans ce lot, 1 % des chaudières à cheminée sont défectueuses et 5 % des chaudières à ventouse sont défectueuses.
On prélève au hasard une chaudière dans la production de ce mois.
Toutes les chaudières ont la même probabilité d'être prélevées.

On considère les évènements suivants :

A : « La chaudière est à cheminée » ;

B : « La chaudière est à ventouse » ;

D : « La chaudière présente un défaut ».

1. Déterminer $P(A)$, $P(B)$, $P(D/A)$ et $P(D/B)$.
2. Calculer $P(D \cap A)$ et $P(D \cap B)$.
3. En remarquant que $D = (D \cap A) \cup (D \cap B)$ et que les événements $D \cap A$ et $D \cap B$ sont incompatibles, calculer $P(D)$ et $P(\bar{D})$.

Exercice 3

On donne ci-dessous la proportion, en pourcentage, du nombre d'enfants nés hors mariage en France métropolitaine.

Année a_i	1980	1985	1990	1995	2000	2003
Proportion y_i	11,4	19,6	30,1	37,6	42,6	45,2

On souhaite effectuer un ajustement de cette série statistique de la proportion en fonction de l'année.

1. (a) Construire le nuage de points de coordonnées $(a_i,\ y_i)$ dans le plan muni du repère orthogonal suivant
   - ▷ sur l'axe des abscisses, on placera 1980 à l'origine et on prendra comme unité $0,5$ cm,
   - ▷ sur l'axe des ordonnées, on placera 10 à l'origine et on prendra comme unité $0,5$ cm.

   (b) Un ajustement affine semble-t-il adapté ?
2. On note a l'année et y la proportion, on pose $x = a - 1950$ et $t = \ln x$.

   (a) Compléter le tableau suivant :

Année a_i	1980	1985	1990	1995	2000	2003
$x_i = a_i - 1950$	30					
$t_i = \ln x_i$	3,401					
y_i	11,4					

   On donnera pour t des valeurs arrondies au millième.

   (b) Exprimer y en fonction de t par une régression linéaire en utilisant la méthode des moindres carrés. On arrondira les coefficients au dixième.

   (c) En déduire la relation : $y = 61,3 \ln x - 197$.

   (d) Quel pourcentage du nombre d'enfants nés hors mariage (arrondi à 1 %), peut-on prévoir en 2010 en utilisant cet ajustement ?

   (e) À partir de quelle année peut-on prévoir que la proportion du nombre d'enfants nés hors mariage sera-t-elle supérieure à 60 % ?

Exercice 4

Le tableau suivant donne, en milliers, le nombre de Pactes civils de solidarité (PACS) signés chaque année en France :

Années	2000	2001	2002	2003	2004
Rang de l'année, x_i	0	1	2	3	4
Nombres de PACS en milliers, y_i	22,1	19,4	25	31,1	39,6

Source INSEE.

1. Calculer, à 0,1 près, le pourcentage d'augmentation du nombre de milliers de Pactes civils de solidarité entre 2000 et 2004.

2. **On envisage un ajustement affine**

   (a) À l'aide de la calculatrice, donner l'équation de la droite d'ajustement de y en x par la méthode des moindres carrés, sous la forme $y = ax + b$. Par la suite, on pose $f(x) = ax + b$.

   (b) En supposant que cet ajustement affine est valable jusqu'en 2007, donner une estimation du nombre de milliers de Pactes civils de solidarité signés en 2007.

3. **On envisage un autre type d'ajustement**
   On modélise le nombre de milliers de Pactes civils de solidarité signés durant l'année $2000 + x$ (x entier) à l'aide de la fonction g définie par

   $$g(x) = 1,6x^2 - 1,8x + 21,4.$$

   (a) En utilisant ce second modèle, calculer le nombre de milliers de Pactes civils de solidarité signés en 2007.

   (b) On suppose que l'évolution se poursuit selon ce modèlejusqu'en 2015. Le nombre de milliers de Pactes civils de solidarité signés en 2010 sera-t-il supérieur à 100 000 ? Justifier.

4. **Comparaison des deux ajustements**
   Pour chacun des deux modèles, on calcule ci-dessous le tableau des carrés des écarts entre les valeurs réelles et les valeurs calculées à l'aide de chacun des deux ajustements.

x_i	0	1	2	3	4
$[(y_i - f(x_i)]^2$	16	11,36	5,95	1,02	7,95

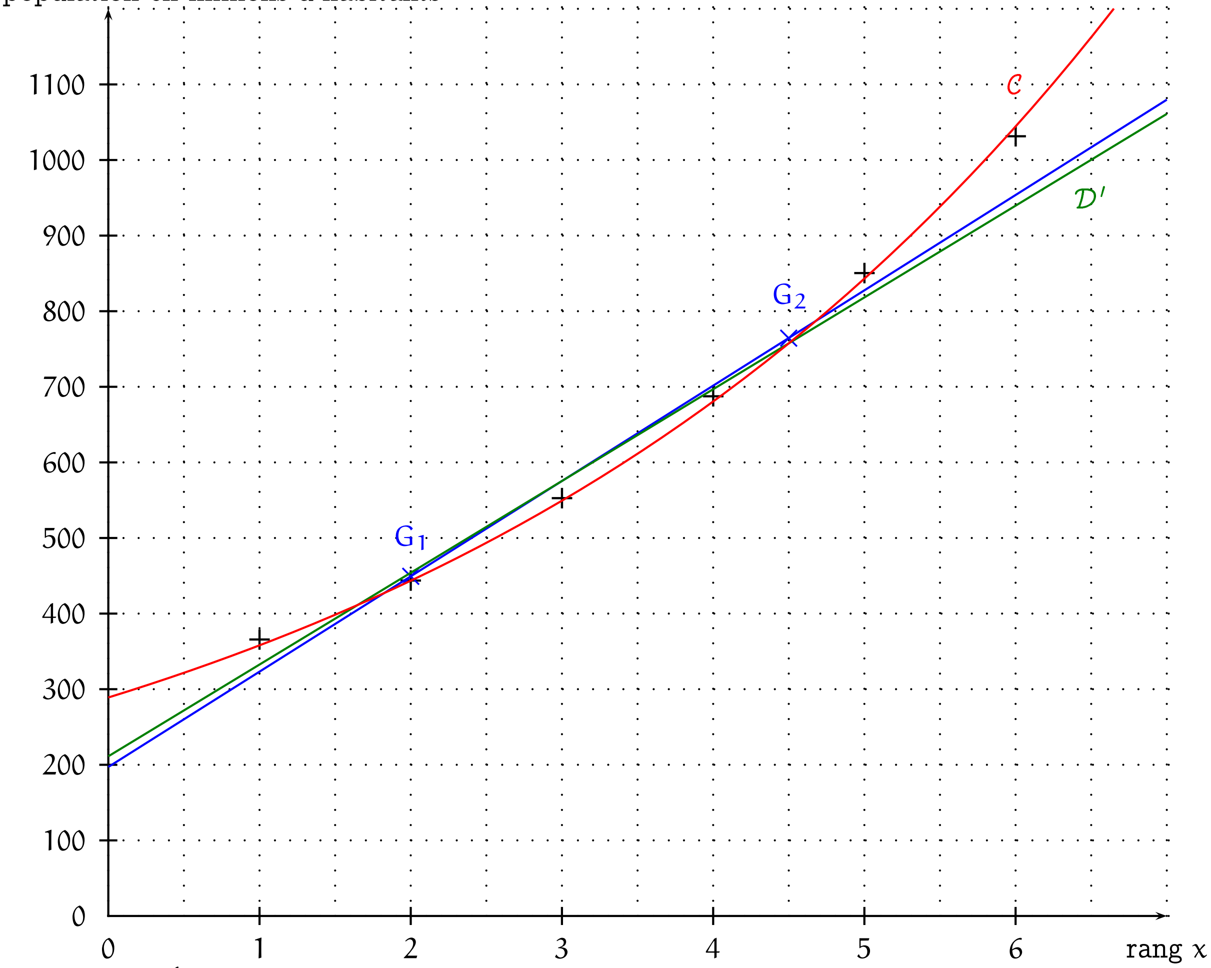
x_i	0	1	2	3	4
$[(y_i - g(x_i)]^2$	0,49				

   (a) Recopier et compléter le deuxième tableau, les valeurs étant arrondies au centième.

   (b) Lequel de ces deux ajustements semble le plus proche de la réalité ? Justifier.

Exercice 1

1. Nuage de points :



2. (a) $G_1 : \begin{cases} x_{G_1} = \frac{1+2+3}{3} = 2 \\ y_{G_1} = \frac{361+439+548}{3} = 449,3 \end{cases}$ donc, $\boxed{G_1(\ 2\ ;\ 449,3\ )}$

   $G_2 : \begin{cases} x_{G_2} = \frac{4+5}{2} = 4,5 \\ y_{G_2} = \frac{683+846}{2} = 764,5 \end{cases}$ donc, $\boxed{G_2(\ 4,5\ ;\ 764,5\ )}$

   (b) La droite (G_1G_2) n'est pas parallèle à l'axe des ordonnées, elle a donc pour équation $y = ax + b$ avec :

   $a = \dfrac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_2}} = \frac{764,5-449,3}{4,5-2} = 126,1.$

   De plus, elle passe par le point $G_1(2;449,3)$ d'où :

   $y_{G_1} = ax_{G_1} + b \quad \Rightarrow \quad 449,3 = 126,1 \times 2 + b \quad \Rightarrow \quad b = 197,1.$

   Conclusion : $\boxed{(G_1G_2) : y = 126,1x + 197,1}$.

   (c) Voir graphique.

   (d) 2001 correspond au rang $x = 6$ donc : $y = 126,1 \times 6 + 197,1 = 953,7$.

   $\boxed{\text{En 2001, on pouvait prévoir 954 millions d'habitants}}$

3. (a) La calculatrice donne $\mathcal{D} : y = ax + b$ avec $a = 121,4,\ b = 211,2$ et $r = 0,9908$.

   Donc $\boxed{\mathcal{D} : y = 121,4x + 211,2 \text{ avec } r = 0,9908}$

   (b) Voir graphique.

(c) 2001 correspond au rang $x = 6$ donc : $y = 121,4 \times 6 + 211,2 = 939,6$.

$\boxed{\text{En 2001, on pouvait prévoir 940 millions d'habitants}}$

4. (a) Tableau :

année	1951	1961	1971	1981	1991
Rang x_i	1	2	3	4	5
Population y_i (en millions)	361	439	548	683	846
z_i	$5,889$	$6,084$	$6,306$	$6,526$	$6,741$

(b) La calculatrice donne $\mathcal{D}' : z = ax + b$ avec $a = 0,216,\ b = 5,665$.

Donc $\boxed{\mathcal{D}' : z = 0,215x + 5,666}$

(c) $r' = 0,9998$. r' est plus proche de 1 que r, donc, cette approximation doit être plus appropriée car les points seront plus près de la droite.

(d) On a $\ln y = 0,215x + 5,666 \iff y = e^{0,215x+5,666} \iff y = (e^{0,215})^x \times e^{5,666}$

Soit $\boxed{y \approx 289\ e^{0,215x}}$

(e) Voir graphique.

(f) 2001 correspond au rang $x = 6$ donc : $y = 289\ e^{0,215\times 6} = 1049,9$.

$\boxed{\text{En 2001, on pouvait prévoir 1050 millions d'habitants suivant cet ajustement}}$

5. Le troisième ajustement semble le plus approprié car le plus proche des résultats réels.

$\boxed{\text{Dans ce cas, on peut prévoir en 2011 une population de } y = 289\ e^{0,215\times 7} = 1302 \text{ habitants}}$