Corrélation et causalité
Correction partielle

∽

Exercice 1

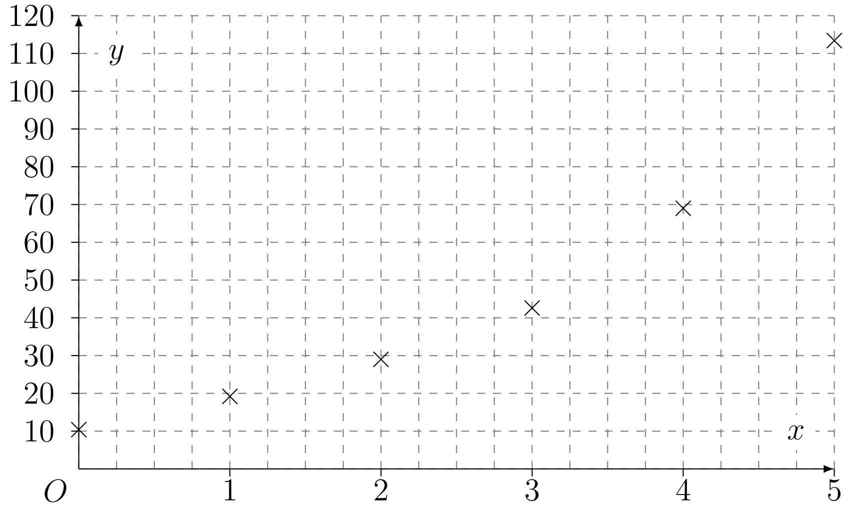
1. Les couples $(x_i; y_i)$ sont les coordonnées de points dans le repère. Ainsi :
   - Série A : Nuage 4
   - Série B : Nuage 3
   - Série C : Nuage 1
   - Série D : Nuage 2
2. Deux nuages semblent permettre un ajustement affine : les 1 et 2. On « voit » presque une droite qui pourrait s'approcher de ces ensembles de points.

   Le nuage 3 ressemblerait plutôt à une parabole (fonction du second degré).

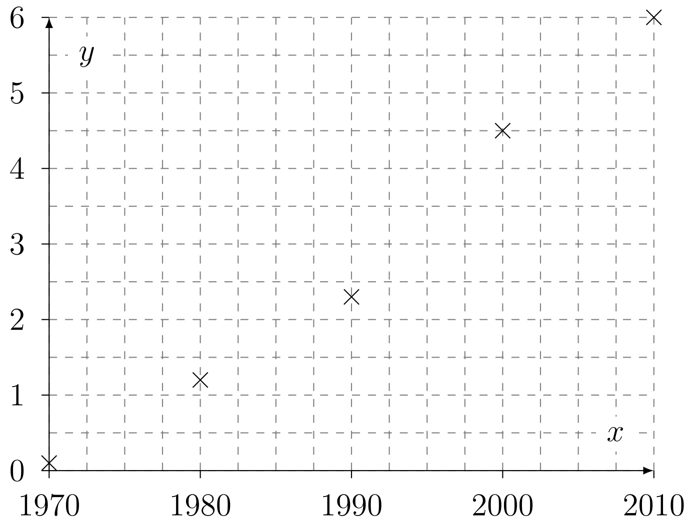
   Le nuage 4 semble avoir des points trop dispersés pour un ajustement quel qu'il soit.
3. - Série A : $\overline{x} = \dfrac{1+2+3+4+5+6}{6} = 3{,}5 = \dfrac{7}{2}$, et $\overline{y} = \dfrac{2+1+(-1)+0{,}5+2+1}{6} = \dfrac{11}{12}$.
     Ainsi, $G\left(\dfrac{7}{2}; \dfrac{11}{12}\right)$.
   - Série B on obtient de même $G(40; 32)$.
     Pour vérifier, on peut utiliser la calculatrice.
     * Avec Numworks, il faut aller dans la partie Régression. On remplit les deux listes de valeurs, puis on demande les valeurs statistiques.
     * Avec TI, on édite les deux listes dans STAT, puis dans STAT on demande les 2-Var Stats en renseignant les deux listes utilisées.
     * Avec Casio, on entre les données dans la partie Stats. On indique dans 'set' les deux listes utilisées, puis on demande les '2Var'.
   - Pour la série C on obtient $G\left(0{,}45; \dfrac{7}{12}\right)$.
   - Pour la série D on obtient $G\left(\dfrac{8}{3}; \dfrac{11}{4}\right)$.

Exercice 2

1. Pour la série 1 :



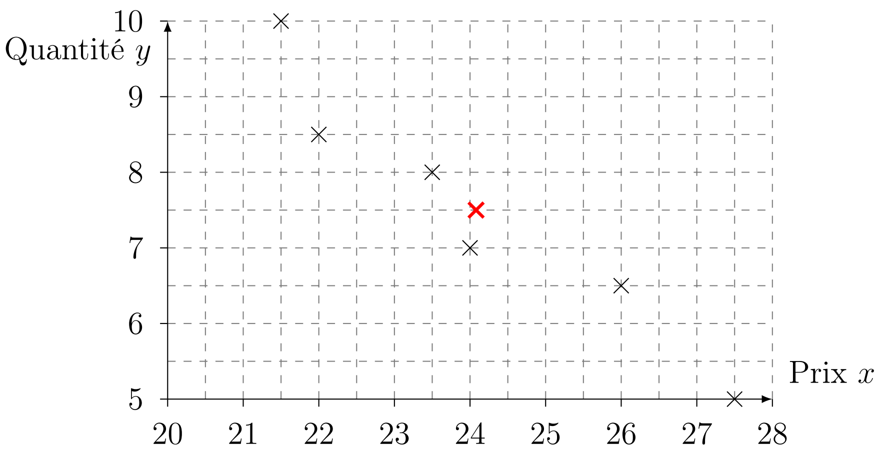
Pour la série 2 :



2. Pour la série 1, un ajustement par une fonction polynomiale de degré 2 semble convenir. Pour la série 2, un ajustement affine semble convenir.
3.

Exercice 3

1. Le nuage de points est le suivant :



2. Le point moyen a pour coordonnées $G(24{,}08; 7{,}5)$ (arrondies au centième pour l'abscisse). Le point est placé en rouge sur la figure ci-dessus.
3. Cela signifie qu'en moyenne, le marchand vend 7,5 kg de cèpes à 24,88€/kg.

Exercice 4

Nous ne détaillerons les calculs que pour le premier tableau.
Pour les autres nous donnerons tout de même les résultats de la calculatrice.

1. Il faut d'abord calculer les moyennes : $\overline{x} = \dfrac{1+5+8+10}{4} = 6$ et $\overline{y} = \dfrac{3+6+10+15}{4} = 8{,}5$
   Ensuite,

$$\begin{aligned}
\Sigma(x_i - \overline{x})(y_i - \overline{y}) &= (1-6)(3-8{,}5) + (5-6)(6-8{,}5) \\
&\quad + (8-6)(10-8{,}5) + (10-6)(15-8{,}5) \\
&= (-5)(-5{,}5) + (-1)(-2{,}5) \\
&\quad + 2(1{,}5) + 4(6{,}5) \\
&= 59
\end{aligned}$$

et

$$\begin{aligned}\Sigma(x_i-\overline{x})^2 &= (1-6)^2+(5-6)^2+(8-6)^2+(10-6)^2\\ &= 25+1+4+16\\ &= 46\end{aligned}$$

Alors $a=\dfrac{59}{46}\simeq 1{,}2826$
Par suite, $b=\overline{y}-a\overline{x}\simeq 8{,}5-1{,}2826\times 6\simeq 0{,}8044$.
L'équation est donc $y=1{,}2826x+0{,}8044$, ce que donne à peu près la calculatrice.
Pour calculer r, nous avons besoin en plus de $\Sigma(y_i-\overline{y})^2$:

$$\begin{aligned}\Sigma(y_i-\overline{y})^2 &= (3-8{,}5)^2+(6-8{,}5)^2+(10-8{,}5)^2+(15-8{,}5)^2\\ &= 5{,}5^2+2{,}5^2+1{,}5^2+6{,}5^2\\ &= 81\end{aligned}$$

Par suite, $r\simeq\dfrac{59}{\sqrt{46*81}}\simeq 0{,}9666$ (ce que donne aussi la calculatrice).
Cela qui indique, puisque sa valeur absolue est très proche de 1, que l'ajustement affine est adapté.

2. La calculatrice donne $\overline{x}=12$, $\overline{y}=182$, $a\simeq 10{,}909$, $b\simeq 51{,}091$ et $r\simeq 0{,}948$ (là aussi l'ajustement est adapté, mais un peu moins).
3. La calculatrice donne $\overline{x}=180$, $\overline{y}=35$, $a\simeq -0{,}138$, $b\simeq 59{,}880$ et $r\simeq -0{,}976$ (là aussi l'ajustement est adapté).
   On remarque que r est négatif, mais ce qui compte c'est la valeur absolue de r.

Exercice 5

1. On utilise simplement la calculatrice, qui donne les valeurs de a et b.
   L'équation de la droite des moindres carrés est $y=0{,}48x+1{,}94$.
2. La valeur de r, donnée par la calculatrice, est $r\simeq 0{,}992$.
   Comme $|r|$ est très proche de 1, on en déduit que cet ajustement est adapté.
3. (a) Pour 2022, cela fera $2022-2015=7$ années écoulées depuis 2015.
   Or pour $x=7$, on obtient $y=0{,}48\times 7+1{,}94=5{,}3$.
   Autrement dit, selon le modèle, la production devrait être de 5,3 tonnes en 2022.
   (b) On souhaite savoir quand la production dépassera 7 tonnes.
   On résout donc :

$$\begin{aligned}0{,}48x+1{,}94\geqslant 7 &\Leftrightarrow 0{,}48x\geqslant 5{,}06\\ &\Leftrightarrow x\geqslant\frac{5{,}06}{0{,}48}\\ &\Leftrightarrow x\geqslant 10{,}54\end{aligned}$$

   Ainsi, durant la onzième année après 2015, autrement dit en 2026, la production dépassera 7 tonnes.

Exercice 6

1. L'équation de la droite des moindres carrés est $y=0{,}61x+1{,}54$.
2. La valeur de r est $r\simeq 0{,}996$.
   Comme $|r|$ est très proche de 1, on en déduit que cet ajustement est adapté.

3. (a) Pour 2022, cela fera $2022 - 2015 = 7$ années écoulées depuis 2015.
Or pour $x = 7$, on obtient $y = 0{,}61 \times 7 + 1{,}54 = 5{,}81$.
Autrement dit, selon le modèle, la production devrait être de 5,81 tonnes en 2022.

   (b) On souhaite savoir quand la production dépassera 7 tonnes.
On résout donc :

$$\begin{aligned} 0{,}61x + 1{,}54 \geqslant 7 &\Leftrightarrow 0{,}61x \geqslant 5{,}46 \\ &\Leftrightarrow x \geqslant \frac{5{,}46}{0{,}61} \\ &\Leftrightarrow x \geqslant 8{,}95 \end{aligned}$$

Ainsi, durant la neuvième année après 2015, autrement dit en 2024, la production dépassera 7 tonnes.

Exercice 7

1. La droite d_3 passe par les points extrêmes, c'est donc la droite des points extrêmes.
La droite de Mayer passe par définition par le point moyens des trois premiers points et celui des trois derniers points. On peut vérifier par exemple que le point moyen des trois premiers points a pour coordonnées (2,4), point par lequel la droite d_1 est la seule à passer.
La droite de Mayer est donc la droite d_1
La droite des moindres carrés est nécessairement celle qui reste, donc d_2. On peut remarquer qu'elle passe, comme la droite d_1 de Mayer, par le point moyen, dont les coordonnées sont $(3{,}5; 3)$.
2. (a) On a $A(1;3)$ et $B(6;2)$ deux points de d_3. Donc le coefficient directeur de d_3 est : $a = \dfrac{y_B - y_A}{x_B - x_A} = \dfrac{2-3}{6-1} = -\dfrac{1}{5} = -0{,}2$.
Par suite, l'équation est donc de la forme $y = -0{,}2x + b$.
Or la droite passe (entre autres) par $A(1;3)$, donc on a $3 = -0{,}2 \times 1 + b$.
Autrement dit $b = 3 + 0{,}2 = 3{,}2$.
On a donc bien $d_3 : y = -0{,}2x + 3{,}2$.

   (b) La droite de Mayer passe par $G_1(2;4)$ et $G_2(5;2)$ (calculer les coordonnées de ces points moyens).
Alors son coefficient directeur est $a = \dfrac{2-4}{5-2} = -\dfrac{2}{3}$.
Par suite, l'équation est de la forme $y = -\dfrac{2}{3}x + b$.
Or G_1 appartient à la droite, donc $4 = -\dfrac{2}{3} \times 2 + b$.
Ainsi, $b = 4 + \dfrac{4}{3} = \dfrac{16}{3}$.
On a bien $d_2 : y = -\dfrac{2}{3}x + \dfrac{16}{3}$.

   (c) Pour la droite des moindres carrés nous nous contenterons ici de la calculatrice, qui donne effectivement l'équation annoncée : $y = -0{,}4x + 4{,}4$.
Sinon il s'agit de refaire les calculs.
À noter : le coefficient de corrélation vaut $r = -0{,}316$ et indique, puisque sa valeur absolue est plus proche de 0 que de 1, que cette régression linéaire n'est pas justifiée (ce qui n'est pas très étonnant à la vue du nuage de points).
3. Nous avons déjà indiqué plus haut que les droites de Mayer et des moindres carrés sont sécantes au point moyen.
D'autre part, ajustement affine semble a priori peu pertinent (la calculatrice donne d'ailleurs $r \simeq -032$).

Exercice 8

Exercice 9

1. (a) D'après la calculatrice, l'équation de la droite des moindres carrés est :
   $y = -1970{,}091x + 22682{,}91$
   Le coefficient de corrélation vaut $r = -0{,}9832$ (plutôt bon).

   (b) On obtient les valeurs suivantes grâce à l'équation (et la calculatrice) :

Année	2020	2021	2022	2023	2024	2025
Rang x	10	11	12	13	14	15
Cote y	2982	1011,9	-958,18	-2928,3	-4898,4	-6868,6

   Nous voyons là des valeurs absurdes (car négatives...).

2. (a) La calculatrice (ou un tableur) permet d'obtenir des listes par calcul à partir d'autres listes.

x_i	0	1	2	3	4	5	6	7	8	9
$z_i = \ln(y_i)$	10,093	10,007	9,757	9,66	9,535	9,41	9,225	9,129	8,908	8,708

   (b) Pour z, on trouve l'équation suivante de la droite des moindres carrés :
   $z = -0{,}14985x + 10{,}11767$, avec $r = -0{,}9958$ (meilleur que le précédent).

   (c) On a donc $z = \ln(y) = -0{,}14985x + 10{,}11767$, puis en appliquant l'exponentielle on obtient :
   $\mathrm{e}^z = \mathrm{e}^{\ln(y)} = \mathrm{e}^{-0{,}14985x+10{,}11767}$.
   Autrement dit, $y = \mathrm{e}^{10{,}11767}\,\mathrm{e}^{-0{,}14985x} \simeq 24776{,}88\,\mathrm{e}^{-0{,}14985x}$ (ce que peut donner la calculatrice si elle peut faire la régression exponentielle).

   (d) Avec cet ajustement, le tableau pour les années 2020 à 2025 devient :

Année	2020	2021	2022	2023	2024	2025
Rang x	10	11	12	13	14	15
Cote y	5536,7	4766,2	4102,9	3532	3040,4	2617,3

3. Bien évidemment, les valeurs les plus réalistes sont celles données par l'ajustement exponentiel. Un argument permet de le justifier : le coefficient de corrélation était meilleur.