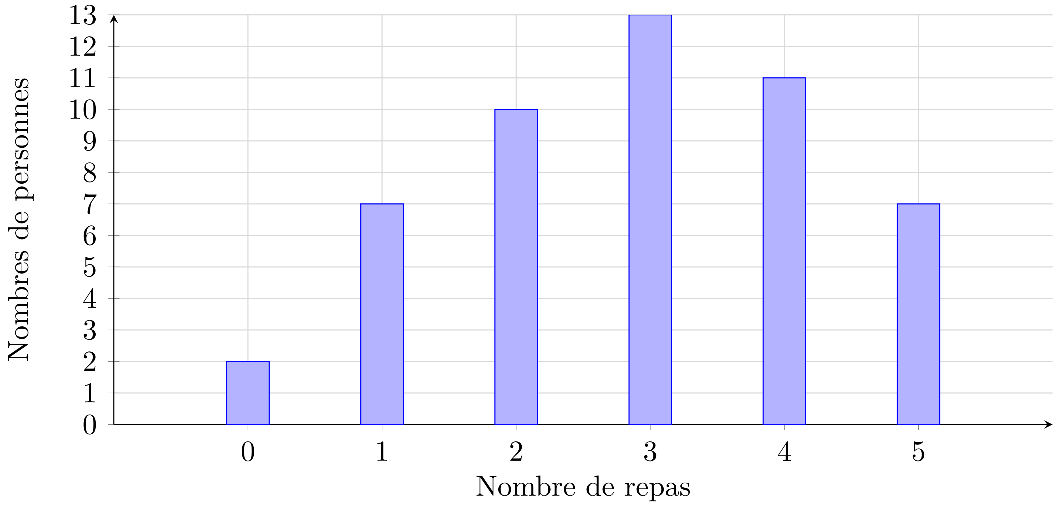
Corrigé de la liste d'exercices n°14 — Statistique descriptive

Exercice 1. On interroge 50 élèves sur le nombre de repas pris à la cantine dans la semaine. Les données sont représentées dans le diagramme ci-dessous.



1.
- **Étendue :** $\max(x) - \min(x) = 5 - 0 = 5$.
- **Mode :** la valeur la plus fréquente est 3 (effectif 13) $\Rightarrow$ mode $= 3$.
- **Cumul des effectifs :**

x	0	1	2	3	4	5
effectif n_x	2	7	10	13	11	7
cumul	2	9	19	32	43	50

- **Médiane :** pour $N = 50$, la médiane est la moyenne des 25$^{\text{e}}$ (3) et 26$^{\text{e}}$ valeurs (3).

$$\Rightarrow \text{médiane} = 3.$$

- **Quartiles :** $N/4 = 12{,}5$ et $3N/4 = 37{,}5$.

$$Q_1 = \text{première valeur dont le rang cumulé } \geq 12{,}5 \Rightarrow Q_1 = 2 \quad (13^{\text{e}} valeur),$$

$$Q_3 = \text{première valeur dont le rang cumulé } \geq 37{,}5 \Rightarrow Q_3 = 4 \quad (38^{\text{e}} valeur).$$

2.

$$\overline{x} = \frac{1}{N}\sum_x x\, n_x = \frac{0\cdot 2 + 1\cdot 7 + 2\cdot 10 + 3\cdot 13 + 4\cdot 11 + 5\cdot 7}{50} = \frac{145}{50} = 2{,}9.$$

On peut calculer la variance par $V(X) = \frac{1}{N}\sum_x x^2 n_x - \overline{x}^2$. Ici,

$$\sum_x x^2 n_x = 0^2\cdot 2 + 1^2\cdot 7 + 2^2\cdot 10 + 3^2\cdot 13 + 4^2\cdot 11 + 5^2\cdot 7 = 515,$$

donc

$$V = \frac{515}{50} - 2{,}9^2 = 10{,}3 - 8{,}41 = 1{,}89, \qquad \sigma = \sqrt{V} = \sqrt{1{,}89} \approx 1{,}375 \text{ (soit } \approx 1{,}38 \text{ au centième).}$$

Exercice 2. Taux de cholestérol (en g/L) d'un échantillon de 100 personnes :

Classe	[1.2,1.4[	[1.4,1.6[	[1.6,1.8[	[1.8,2.0[	[2.0,2.2[	[2.2,2.4[	[2.4,2.6[	[2.6,2.8[	[2.8,3.0[	[3.0,3.2[
Effectif	5	14	15	21	18	10	7	5	3	2

1. On considère les classes de largeur $h = 0{,}2$ de $[1{,}2; 3{,}2[$ et les effectifs

$$(5,\ 14,\ 15,\ 21,\ 18,\ 10,\ 7,\ 5,\ 3,\ 2), \qquad N = 100.$$

Les centres de classes sont $c_i \in \{1{,}3;\ 1{,}5;\ 1{,}7;\ 1{,}9;\ 2{,}1;\ 2{,}3;\ 2{,}5;\ 2{,}7;\ 2{,}9;\ 3{,}1\}$. Les effectifs cumulés sont : $5,\ 19,\ 34,\ 55,\ 73,\ 83,\ 90,\ 95,\ 98,\ 100$.
On place les positions $N/4 = 25{,}0$ et $3N/4 = 75{,}0$ dans la table cumulée.

— Q_1 **:** 25^e valeur donc $Q_1 \in [1{,}6; 1{,}8[$. Avec

$$Q_1 \approx 1{,}7 \text{ g/L}.$$

— **Médiane :** moyenne de la 50^e et de la 51^e valeur donc Med $\in [1{,}8; 2{,}0[$. Avec

$$\text{Med} \approx 1{,}9 \text{ g/L}.$$

— Q_3 **:** 75^e valeur donc $Q_3 \in [2{,}2; 2{,}4[$. Avec

$$Q_3 \approx 2{,}3 \text{ g/L}.$$

2. On assimile chaque valeur de la classe à son centre c_i :

$$\overline{x} = \frac{1}{N} \sum f_i c_i = \frac{199{,}6}{100} = \boxed{1{,}996 \text{ g/L} \approx 2{,}00 \text{ g/L}}.$$

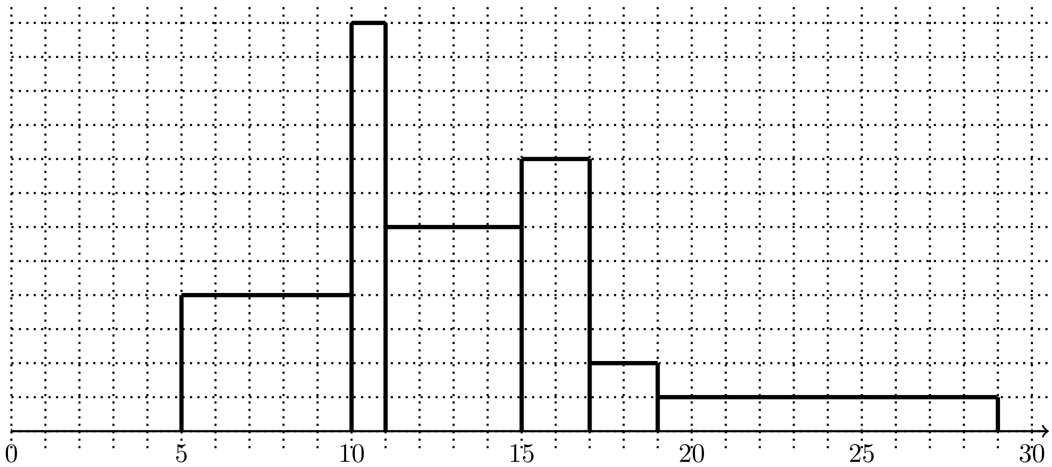
Pour la variance (au sens « population ») :

$$\sigma^2 = \frac{1}{N} \sum f_i (c_i - \overline{x})^2 = \frac{1}{N} \sum f_i c_i^2 - \overline{x}^2 = \frac{416{,}04}{100} - 1{,}996^2 = 0{,}176384.$$

Donc

$$\sigma = \sqrt{0{,}176384} \approx \boxed{0{,}420 \text{ g/L}}.$$

Exercice 3.
Dans l'histogramme suivant, l'effectif de la classe $[17; 19[$ est égal à 2.



1. L'aire de chaque rectangle est proportionnelle à l'effectif de la classe :

$$\text{Aire} = \text{largeur} \times \text{hauteur}.$$

On nous indique que l'effectif de la classe $[17;19[$ est 2. Son aire vaut $2 \times 2 = 4$, donc le coefficient de proportionnalité est :

$$k = \frac{2}{4} = \frac{1}{2}.$$

Ainsi, pour chaque classe :

$$\text{Effectif} = \frac{\text{Aire}}{2} = \frac{\text{Largeur} \times \text{Hauteur}}{2}.$$

Classe	Largeur	Hauteur	Aire	Effectif
[5 ;10[	5	4	20	10
[10 ;11[	1	12	12	6
[11 ;15[	4	6	24	12
[15 ;17[	2	8	16	8
[17 ;19[	2	2	4	2
[19 ;29[	10	1	10	5

Total : $N = 10 + 6 + 12 + 8 + 2 + 5 = 43$.

2. La classe modale est $[11;15[$.

3. On assimile chaque classe à son **centre** :

Classe	Centre c_i
[5 ;10[	7,5
[10 ;11[	10,5
[11 ;15[	13
[15 ;17[	16
[17 ;19[	18
[19 ;29[	24

La moyenne est :

$$\overline{x} = \frac{1}{N}\sum_i n_i c_i = \frac{10 \times 7{,}5 + 6 \times 10{,}5 + 12 \times 13 + 8 \times 16 + 2 \times 18 + 5 \times 24}{43} = \frac{578}{43} \approx 13{,}44.$$

$$\boxed{\overline{x} \approx 13{,}4}$$

Exercice 4. Partie A.

1. **Interprétation du point** $A(1400; 250)$**.** À $x = 1400$ h, l'effectif cumulé vaut 250 : <u>250 ampoules ont une durée de vie inférieure ou égale à 1400 h</u>, soit $\dfrac{250}{600} \approx 41{,}7\,\%$ du lot.
2. **Estimation de la médiane et de** Q_1**.**
   — La **médiane** correspond au rang cumulé $N/2 = 300$ (soit $y = 6$). On lit l'abscisse du point de la courbe d'ordonnée 300 : $M \approx 1470$ h.
   — Le **premier quartile** Q_1 correspond au rang $N/4 = 150$ (soit $y = 3$).

$$Q_1 \approx 1300.$$

3. **Ampoules défectueuses (durée** < 1000 **h).**
   (a) À $x = 1000$ h, l'effectif cumulé lu est 50. Proportion de défectueuses : $\dfrac{50}{600} \approx \boxed{8{,}3\,\%}$.
   (b) <u>Affirmation</u> : « Parmi les défectueuses, une majorité a une durée de vie < 400 h. » À $x = 400$ h, l'effectif cumulé est 40. Parmi les 50 défectueuses (< 1000 h), 40 ont < 400 h, soit $\dfrac{40}{50} = 80\%$. $\Rightarrow$ **Affirmation vraie** (une large majorité).

Partie B.

Durée (h)	200	250	380	500	660	750	870	1260	1800	2200
Effectif	3	11	18	30	19	17	14	15	14	9

Total $N = 3 + 11 + 18 + 30 + 19 + 17 + 14 + 15 + 14 + 9 = 150$.

1. **Durée de vie moyenne du lot Y.**

$$\bar{x} = \frac{1}{150}\Big(200\cdot3+250\cdot11+380\cdot18+500\cdot30+660\cdot19+750\cdot17+870\cdot14+1260\cdot15+1800\cdot14+2200\cdot9\Big).$$

On a donc

$$\bar{x} = \frac{126\,560}{150} \approx \boxed{843{,}7 \text{ h}}.$$

2. **Après augmentation de +10%, la moyenne devient :**

$$\bar{x}_{\text{nouvelle}} = 1{,}10 \times 843{,}7 \approx \boxed{928{,}1 \text{ h}}.$$

Exercice 5.

1. Lorsque la moyenne est nettement supérieure à la médiane, la série présente une asymétrie à droite. Quelques valeurs élevées tirent la moyenne vers le haut alors que la moitié centrale des données reste plus basse. On parle d'une distribution dissymétrique positive, typique de la présence de valeurs extrêmes élevées.
2. Constructions demandées.
   (a) Série où la moyenne est 100 fois plus grande que la médiane.
   Exemple simple (3 valeurs, triées) :
   $$\mathcal{S}_1 = \{0,\ 1,\ a\}.$$
   La médiane vaut 1 (deuxième valeur) et la moyenne vaut
   $$\overline{x} = \frac{0+1+a}{3} = 100 \times \text{médiane} = 100 \text{ donc } a = 299$$
   Donc $\mathcal{S}_1 = \{0,\ 1,\ 299\}$.
   Cette construction illustre un effet de valeur extrême à droite.
   (b) Série où la médiane est 100 fois plus grande que la moyenne.
   On peut utiliser des valeurs négatives pour maintenir une moyenne très petite tout en gardant une médiane élevée. Avec trois valeurs :
   $$\mathcal{S}_2 = \{-197,\ \boxed{100},\ 100\}.$$
   La médiane vaut 100 (valeur centrale) et la moyenne vaut
   $$\overline{x} = \frac{-197+100+100}{3} = 1,$$
   donc la médiane est 100 fois la moyenne. (Les signes négatifs rendent cela possible avec un petit effectif.)
   (c) Série où la moyenne est le double de l'écart-type.
   On précise la convention : on utilise l'écart-type empirique avec facteur $1/n$,
   $$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}.$$
   Considérons un ensemble de 5 valeurs contenant quatre fois la même valeur $a > 0$ et un zéro :
   $$\mathcal{S}_3 = \{0,\ a,\ a,\ a,\ a\}.$$
   Alors
   $$\overline{x} = \frac{4a}{5}, \qquad \sigma = \sqrt{\frac{(0-\overline{x})^2 + 4(a-\overline{x})^2}{5}} = \sqrt{\frac{\left(\frac{4a}{5}\right)^2 + 4\left(\frac{a}{5}\right)^2}{5}} = \frac{2a}{5}.$$
   On obtient
   $$\overline{x} = 2\sigma \qquad \text{pour tout } a > 0.$$
   Par exemple, avec $a = 5$: $\overline{x} = 4$ et $\sigma = 2$, donc $\overline{x} = 2\sigma$.

Exercice 6. On note X la variable aléatoire correspondant aux notes actuelles des élèves.

$$\text{Moyenne actuelle : } \mu_X = 8, \qquad \text{Écart-type actuel : } \sigma_X = 3.$$

On souhaite obtenir une nouvelle série de notes Y telle que :

$$\mu_Y = 10, \qquad \sigma_Y = 4.$$

On décide de prendre une transformation entre les deux séries est **affine** :

$$Y = aX + b.$$

Alors :

$$\begin{cases} \mu_Y = a\mu_X + b, \\ \sigma_Y = |a|\sigma_X. \end{cases}$$

Calcul de a et b :

$$\begin{cases} a \times 8 + b = 10, \\ a \times 3 = 4. \end{cases}$$

D'où :

$$a = \frac{4}{3}, \qquad b = 10 - \frac{4}{3} \times 8 = 10 - \frac{32}{3} = -\frac{2}{3}.$$

Ainsi :

$$\boxed{Y = \frac{4}{3}X - \frac{2}{3}}$$

Interprétation :

— Le professeur doit multiplier toutes les notes par $\frac{4}{3}$ (c'est-à-dire les augmenter de 33% environ),

— puis retirer $\dfrac{2}{3}$ (0,67 environ) point à chacune.

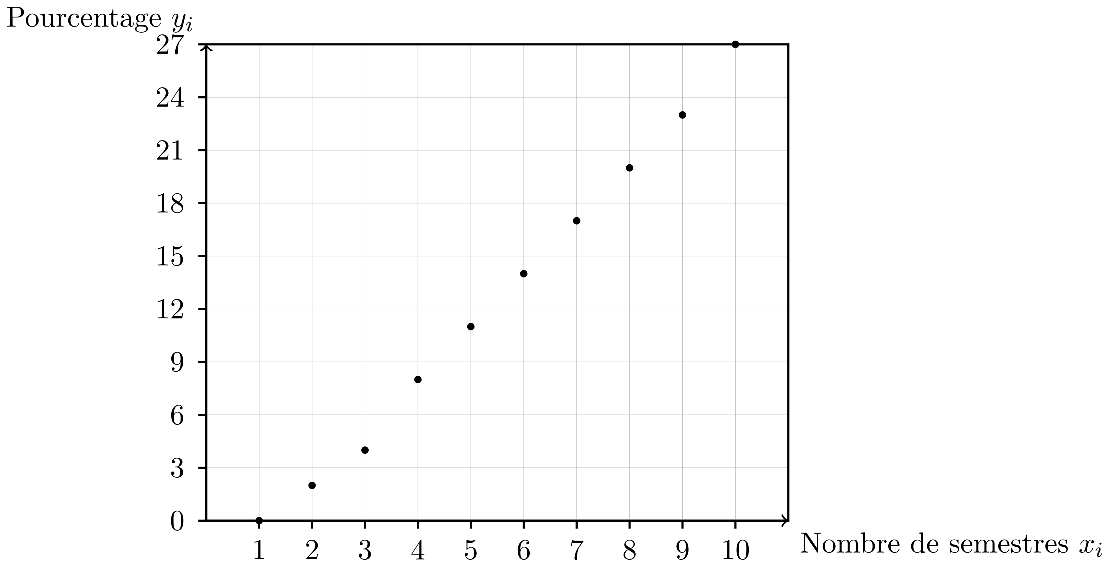
Conclusion :

$$\boxed{\text{Nouvelle note } = \frac{4}{3} \times \text{ancienne note} - \frac{2}{3}}$$

Le professeur doit donc « gonfler » les notes d'un facteur $\frac{4}{3}$ tout en les recentrant légèrement pour satisfaire le Proviseur.

Exercice 7.

1. Nuage de points



2. On note $\bar{x} = \dfrac{1 + \cdots + 10}{10} = \dfrac{11}{2} = 5{,}5$ et $\bar{y} = \dfrac{0 + 2 + \cdots + 27}{10} = \dfrac{126}{10} = 12{,}6$. On calcule

$$S_{xx} = \sum (x_i - \bar{x})^2 = \frac{165}{2} = 82{,}5, \qquad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 250.$$

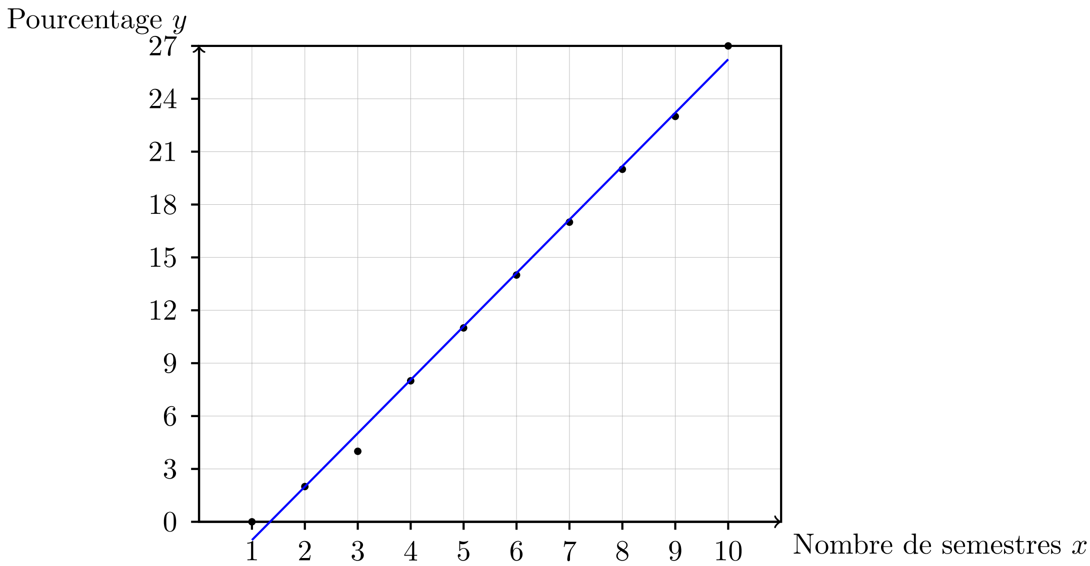
Ainsi la pente et l'ordonnée à l'origine valent

$$a = \frac{S_{xy}}{S_{xx}} = \frac{250}{165/2} = \frac{500}{165} = \frac{100}{33} \approx 3{,}03, \qquad b = \bar{y} - a\bar{x} = 12{,}6 - \frac{100}{33} \cdot 5{,}5 = -\frac{61}{15} \approx -4{,}07.$$

$$\boxed{\hat{y} = ax + b = \frac{100}{33}\, x - \frac{61}{15} \quad \approx \; 3{,}03\, x - 4{,}07}$$

3. Dans le repère où 1 unité verticale $= 3\,\%$, on trace la courbe $y = \hat{y}$ en posant l'ordonnée TikZ à $\hat{y}/3$. Par exemple, pour $x = 0$, $\hat{y}(0) = b \approx -4{,}07$ (hors cadre) ; on peut utiliser deux points dans le cadre, par exemple $x = 5$ et $x = 10$:

$$\hat{y}(5) \approx 3{,}03 \times 5 - 4{,}07 \approx 11{,}1, \quad \hat{y}(10) \approx 26{,}2.$$

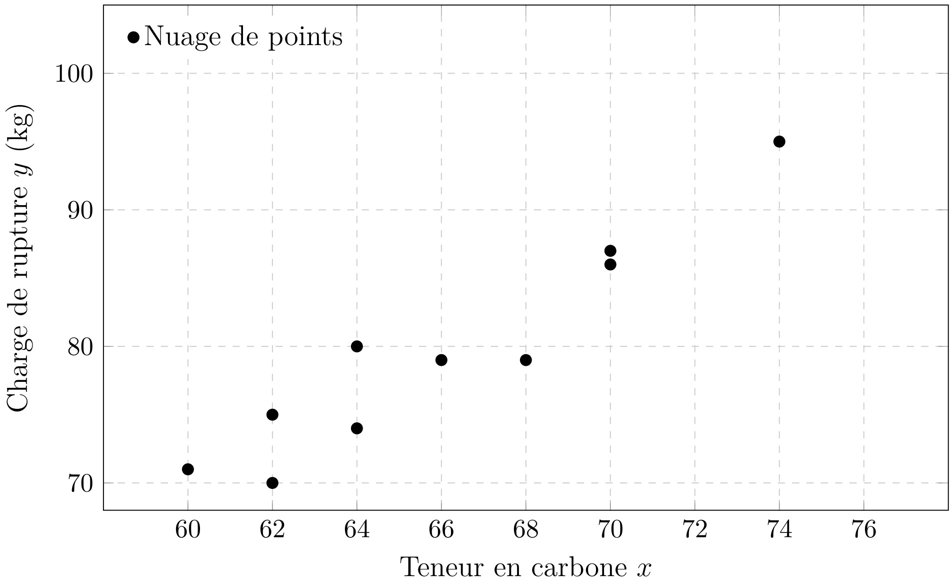


4.

$$\hat{y}(12) = \frac{100}{33} \times 12 - \frac{61}{15} = \frac{5329}{165} \approx \boxed{32{,}3\,\%}.$$

Exercice 8.

1. Représentation graphique du nuage de points.



2. Coordonnées du point moyen du nuage.
   On calcule

$$\bar{x} = \frac{1}{10}\sum x_i = \frac{70+60+68+64+66+64+62+70+74+62}{10} = 66{,}0,$$

$$\bar{y} = \frac{1}{10}\sum y_i = \frac{87+71+79+74+79+80+75+86+95+70}{10} = 79{,}6.$$

   Le point moyen est donc

$$G(\bar{x}, \bar{y}) = (66{,}0\,;\,79{,}6).$$

3. Coefficient de corrélation linéaire et interprétation.
   Les sommes de carrés et produits centrés sont

$$S_{xx} = \sum(x_i - \bar{x})^2 = 176{,}0, \qquad S_{yy} = \sum(y_i - \bar{y})^2 = 552{,}4, \qquad S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = 296{,}0.$$

   Le coefficient de corrélation vaut

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{296}{\sqrt{176 \times 552{,}4}} \approx 0{,}949.$$

   Il s'agit d'une corrélation forte et positive ; un ajustement linéaire est pertinent.

4. Points moyens G_1 et G_2 et droite de Mayer.
   Pour la méthode de Mayer, on retient les cinq points les plus à gauche (plus petites abscisses) et les cinq points les plus à droite (plus grandes abscisses).
   Cinq plus à gauche : $(60, 71), (62, 75), (62, 70), (64, 74), (64, 80)$.

$$G_1(\overline{x}_1, \overline{y}_1) = \left(\frac{60+62+62+64+64}{5}, \frac{71+75+70+74+80}{5}\right) = (62{,}4\,;\,74{,}0).$$

   Cinq plus à droite : $(66, 79), (68, 79), (70, 87), (70, 86), (74, 95)$.

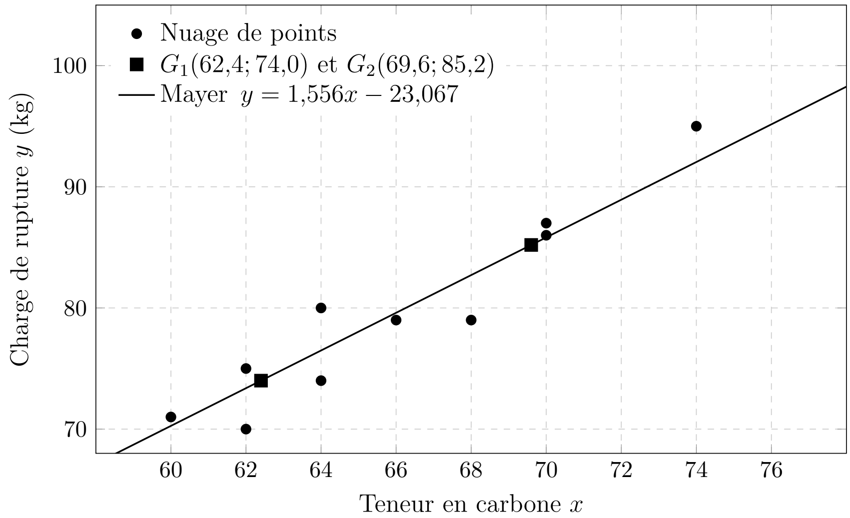
$$G_2(\overline{x}_2, \overline{y}_2) = \left(\frac{66+68+70+70+74}{5}, \frac{79+79+87+86+95}{5}\right) = (69{,}6\,;\,85{,}2).$$

La pente et l'ordonnée à l'origine de la droite (G_1G_2) sont

$$a_M = \frac{85{,}2 - 74{,}0}{69{,}6 - 62{,}4} = \frac{11{,}2}{7{,}2} = \frac{14}{9} \approx 1{,}556, \qquad b_M = 74{,}0 - a_M \times 62{,}4 = 74{,}0 - \frac{14}{9} \times 62{,}4 \approx -23{,}067.$$

Une équation de la droite de Mayer est $\boxed{y = 1{,}556\,x - 23{,}067.}$

Représentation de G_1, G_2 et de la droite de Mayer (avec le nuage) :



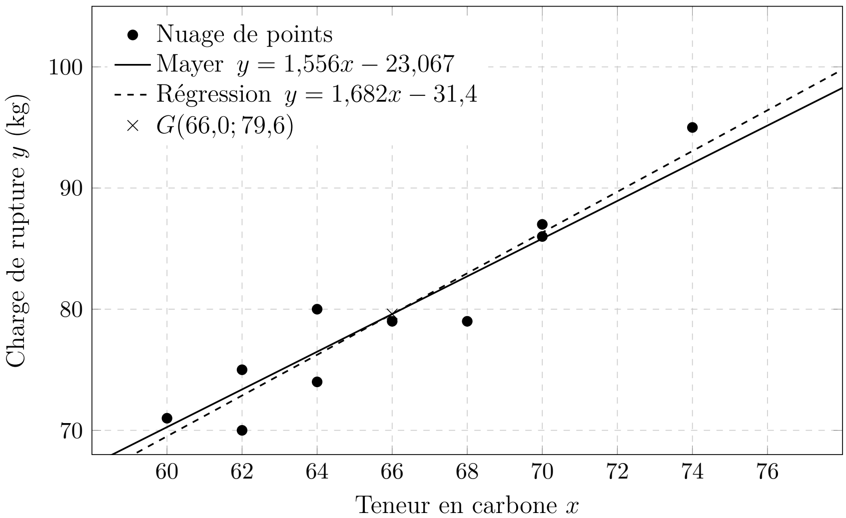
5. (a) Droite de régression de y en x par la méthode des moindres carrés.
Les coefficients de la droite $y = ax + b$ sont

$$a = \frac{S_{xy}}{S_{xx}} = \frac{296{,}0}{176{,}0} \approx 1{,}681818, \qquad b = \bar{y} - a\,\bar{x} \approx 79{,}6 - 1{,}681818 \times 66 \approx -31{,}4.$$

Une équation réduite est

$$\boxed{y = 1{,}682\,x - 31{,}4} \quad \text{(coefficients arrondis à } 10^{-3} \text{ près).}$$

(b) Représentation de la droite de régression, avec rappel de la droite de Mayer et du point moyen global G.



6. Estimations pour une teneur en carbone égale à $x = 77$.
Par la droite de Mayer :

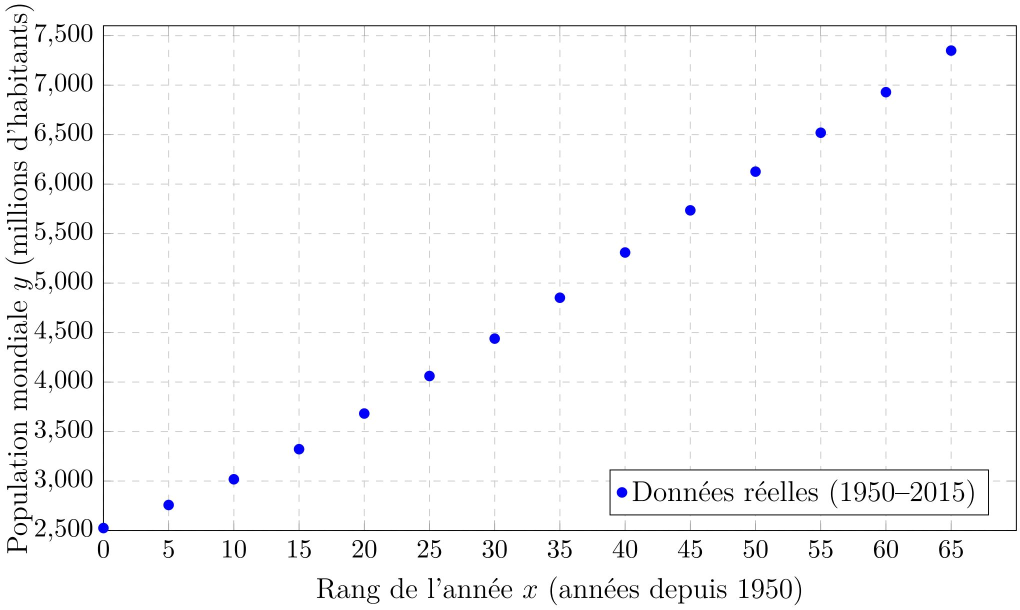
$$y_M(77) = 1{,}556 \times 77 - 23{,}067 \approx 96{,}7 \text{ kg}.$$

Par la droite de régression :

$$y_R(77) = 1{,}682 \times 77 - 31{,}4 \approx 98{,}1 \text{ kg}.$$

Exercice 9.

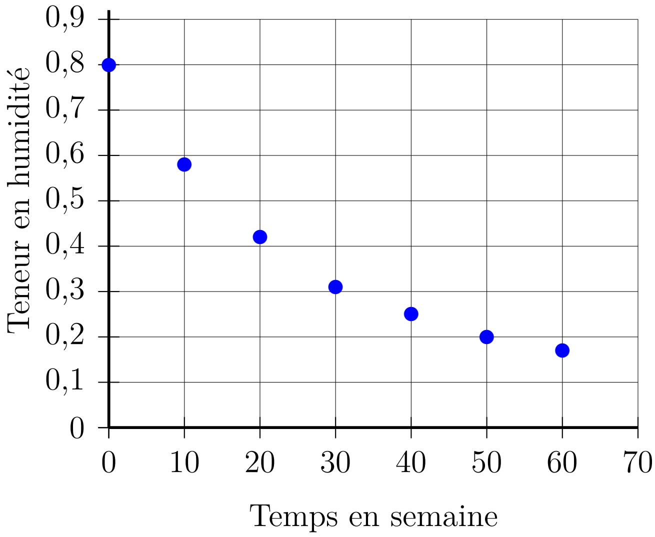
1. Nuage de points.



2. Ajustement d'une loi de la forme $y = 2525 + \lambda x^a$.
On pose, pour $x > 0$,
$$Y := y - 2525 \quad \text{et} \quad T := \ln x.$$
Le modèle devient
$$Y = \lambda x^a \iff \ln Y = \ln \lambda + a \ln x \iff Z = \alpha + a\,T,$$
avec $Z := \ln(y - 2525)$ et $\alpha := \ln \lambda$. On effectue alors une régression linéaire de Z en fonction de T (en omettant le point $x = 0$ car $\ln 0$ n'est pas défini).
Les estimateurs des moindres carrés sont
$$a = \frac{n \sum T_i Z_i - \left(\sum T_i\right)\left(\sum Z_i\right)}{n \sum T_i^2 - \left(\sum T_i\right)^2}, \qquad \alpha = \overline{Z} - a\,\overline{T}, \qquad \lambda = e^{\alpha},$$
où la somme porte sur les $n = 13$ valeurs $x \in \{5, 10, \dots, 65\}$.
Résultat numérique (arrondi à 10^{-3}) :
$$a \approx 1{,}205, \qquad \lambda \approx 31{,}828.$$
On obtient donc le modèle ajusté
$$\boxed{\widehat{y}(x) = 2525 + 31{,}828\, x^{1{,}205}} \qquad (x > 0).$$
Remarque. L'adéquation du modèle est très bonne sur l'intervalle observé (coefficient de détermination sur les y : $R^2 \approx 0{,}999$).

3. Prévision pour 2100.
L'année 2100 correspond à $x = 2100 - 1950 = 150$. La prévision (en millions) est
$$\widehat{y}(150) = 2525 + 31{,}828 \times 150^{1{,}205} \approx 15\,889{,}9.$$
$$\boxed{\widehat{y}(2100) \approx 15{,}9 \text{ milliards d'habitants}}$$
Attention : cette extrapolation est très éloignée de l'intervalle de calibration (1950–2015) et doit être interprétée avec prudence.

Exercice 10. Le bois d'épicéa est couramment utilisé en France pour la construction. Avant son utilisation, il est nécessaire de le faire sécher. La teneur en humidité du bois d'épicéa correspond au pourcentage d'eau contenu dans le bois. La teneur en humidité, en pourcentage, du bois d'épicéa est une fonction f du temps t, exprimé en semaine.
On a effectué un relevé de la teneur en humidité d'une poutre en épicéa en fonction du temps, exprimé en semaine. Les données sont représentées sur le graphique ci- dessous.



1. Les points ne sont pas suffisamment alignés pour qu'on puisse envisager un ajustement affine.
2. On désigne par H la teneur en humidité dans le bois, en pourcentage, et on pose : $y = \ln(H - 0,1)$.

t	0	10	20	30	40	50	60
H	0,80	0,58	0,42	0,32	0,25	0,20	0,17
y	$-0,36$	$-0,73$	$-1,14$	$-1,51$	$-1,90$	$-2,30$	$-2,66$

   (a) Dans le tableau, la valeur manquante y, arrondie au centième, correspondant au temps $t = 30$ est $\ln(0,32 - 0,1)$ soit $-1,51$.
   (b) À la calculatrice, on trouve une équation de la droite d'ajustement de y en t, par la méthode des moindres carrés : $y = -0,039t - 0,357$.
   (c) $y = \ln(H - 0,1) \iff -0,039t - 0,357 = \ln(H - 0,1) \iff \mathrm{e}^{-0,039t-0,357} = H - 0,1$
   $\iff H = \mathrm{e}^{-0,039t-0,357} + 0,1 \iff H = \mathrm{e}^{-0,357}\,\mathrm{e}^{-0,039t} + 0,1$
   $\iff H = 0,7\,\mathrm{e}^{-0,039} + 0,1$ car $\mathrm{e}^{-0,357} \approx 0,7$
3. On admet dans la suite que l'évolution de la teneur en humidité de la poutre, en fonction du temps, est donnée par l'expression : $H(t) = 0,7\,\mathrm{e}^{-0,04t} + 0,1$.
   (a) La teneur en humidité de la poutre après 70 semaines serait :
   $H(20) = 0,7\,\mathrm{e}^{-0,04\times 70} + 0,1 \approx 0,143$, soit $14,3\ \%$.
   (b) La teneur en humidité est inférieure à $5\,\%$ si $H(t) < 0,05$.
   $H(t) < 0,05 \iff 0,7\,\mathrm{e}^{-0,04t} + 0,1 < 0,05 \iff 0,7\,\mathrm{e}^{-0,04t} < -0,05$
   Mais, pour tout réel x, $\mathrm{e}^{x} > 0$ donc pour tout t, $\mathrm{e}^{-0,04t} > 0$ donc l'inéquation n'a pas de solution.
   La teneur en humidité ne pourra jamais être inférieure à $5\,\%$

Exercice 11.

1. (a) Nuage de points.

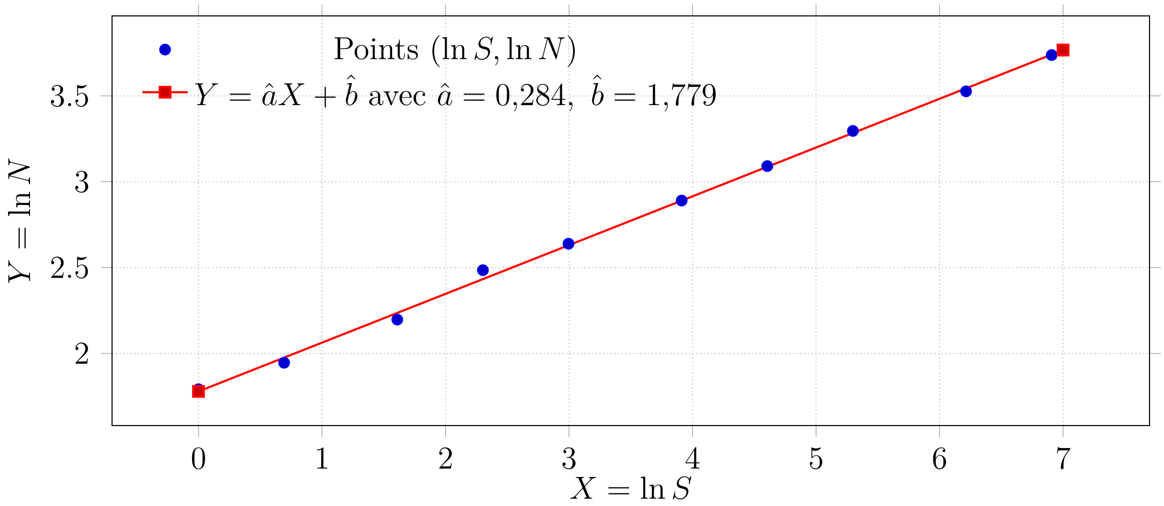


FIGURE 1 – Linéarisation : régression de $\ln N$ sur $\ln S$.

(b) La relation $N = cS^a$ devient $\ln N = a\ \ln S + \ln c$. On ajuste donc Y sur X. Le nuage (X, Y) est quasi aligné croissant.

2. (a)

$$S_{xx} = \sum X^2 - \frac{(\sum X)^2}{n} = 168{,}295 - \frac{34{,}539^2}{10} \approx 49{,}000.$$

$$S_{xy} = \sum XY - \frac{(\sum X)(\sum Y)}{n} = 109{,}263 - \frac{34{,}539 \times 27{,}600}{10} \approx 13{,}935.$$

(b)

$$\hat{a} = \frac{S_{xy}}{S_{xx}} \approx \frac{13{,}935}{49{,}000} \approx 0{,}284.$$

$$\hat{b} = \overline{Y} - \hat{a}\,\overline{X} = \frac{27{,}600}{10} - 0{,}284 \times \frac{34{,}539}{10} \approx 1{,}779.$$

D'où $\hat{c} = \exp(\hat{b}) \approx \exp(1{,}779) \approx 5{,}92$.

(c) Loi SPAR estimée :

$$\boxed{\hat{N} \;=\; 5{,}92\, S^{\,0{,}284}}.$$

3. (a)

$$S_{yy} = \sum Y^2 - \frac{(\sum Y)^2}{n} = 80{,}146 - \frac{27{,}600^2}{10} \approx 3{,}970.$$

$$R^2 = \frac{S_{xy}^2}{S_{xx}\, S_{yy}} \approx \frac{13{,}935^2}{49{,}000 \times 3{,}970} \approx 0{,}999.$$

(b) Cette valeur $R^2 \approx 0{,}999$. indique qu'à l'échelle log–log, la surface explique quasi parfaitement le nombre d'espèces observé, ce qui est cohérent avec une loi SPAR, tout en nécessitant les précautions ci-dessus pour l'écologie réelle

4. (a) Pour $S = 300\,\text{ha}$, $\hat{N} \approx 5{,}92 \times 300^{0{,}284} \approx 29$ espèces.
Pour $S = 5000\,\text{ha}$, $\hat{N} \approx 5{,}92 \times 5000^{0{,}284} \approx 59$ espèces.

(b) **Interprétation de a.** Si on multiplie l'aire par k, le nombre d'espèces est multiplié par k^a. Ici, doubler la surface ($k = 2$) augmente N d'un facteur $2^{0{,}284} \approx 1{,}22$ ($\approx +22\%$).