Correction de certains exercices de Statistiques

Cette correction vous permet de comparer vos résultats et de vous assurer de la justesse de vos raisonnements.
Dans la majorité des cas les calculs numériques ne sont pas détaillés.
Si vous repérez des erreurs, n'hésitez pas à m'en faire part.

Statistiques

Exercices du cours (11-statistiques descriptives cours & exos)

Exercice 1

(a) Pour qu'il y ait *progression*, il faut que le taux de réussite augmente. Hors le taux de réussite *global* passe de $\frac{15}{25}$ à $\frac{17}{25}$. Il y a donc une progression globale. L'affirmation est correcte.

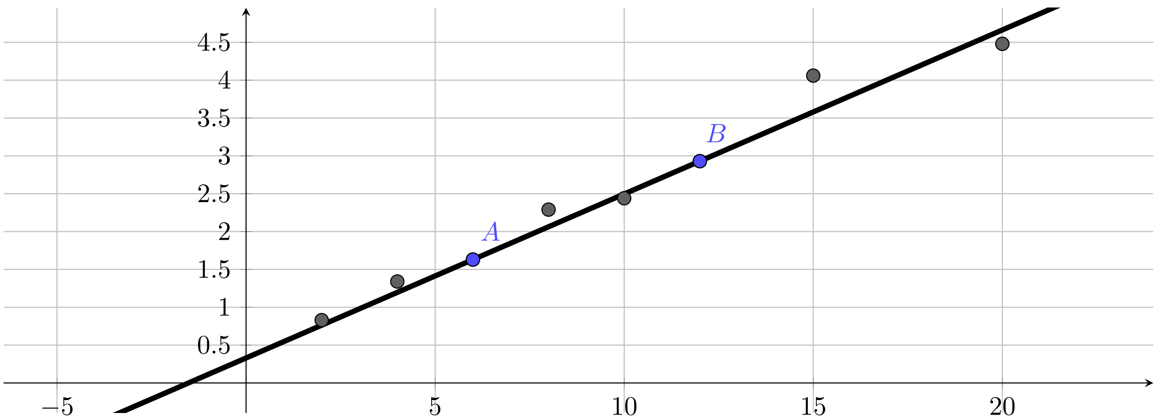
(b) Examinés en détail, les résultats font apparaître une régression à la fois chez les non-redoublants et chez les doublants. Néanmoins, cette affirmation est incorrecte, car imprécise. Elle contredit la précédente, qui est juste.

1. Le paradoxe tient à l'effet de structure. Si le taux de réussite des deux catégories diminue, l'effectif varie aussi considérablement.

   Sans cette variation d'effectif, il ne serait pas possible de diminuer les deux taux de réussite et d'obtenir une augmentation du taux de réussite.
2. La réponse précédente permet de résoudre le problème.

Exercice 2

1. Voici le graphe obtenu après la question 2.a)



2. (a) voir figure

   (b) On lit graphiquement pour $x = 18$ la valeur $y = 4.2$ On peut espérer une resistance de 4.2 m^2/w pour une épaisseur de 18 cm.
3. (a) La droite de Mayer est obtenue en séparant selon les x croissant le nuage en deux parties égales (à une valeur près.)

   Pour les 4 premières valeurs sur les x croissant on trouve le point moyen G_1 en calculant la moyenne des abscisses et la moyenne des ordonnées.

   On a $G_1(5, 1.53)$. On fait de même avec les quatre plus grandes valeurs pour x croissant et on trouve $G_2(14.25; 3.48)$.

   Tracer la droite de Mayer revient à construire ces deux points et les relier.

   Je m'abstiens pour ne pas surcharger le graphique.

   On détermine ensuite la droite de Mayer en donnant l'équation de cette droite $(G_1; G_2)$

   On utilise pour cela les formules $y = ax + b$ avec $a = \dfrac{y_b - y_a}{x_b - x_a}$ et $b = y_a - ax_a$.

   Il vient $a = \dfrac{3.48 - 1.53}{14.25 - 5} = 0.21$ et $b = 0.49$: $y = 0.21x + 0.49$

   (b) On applique le même principe avec $x = 18$ on obtient $y = 4.27$

   La différence est modeste.

   Remarque : Ces méthodes « graphiques » sont inutiles quand une machine fait les calculs à notre place. Ce sera toujours le cas dorénavant.

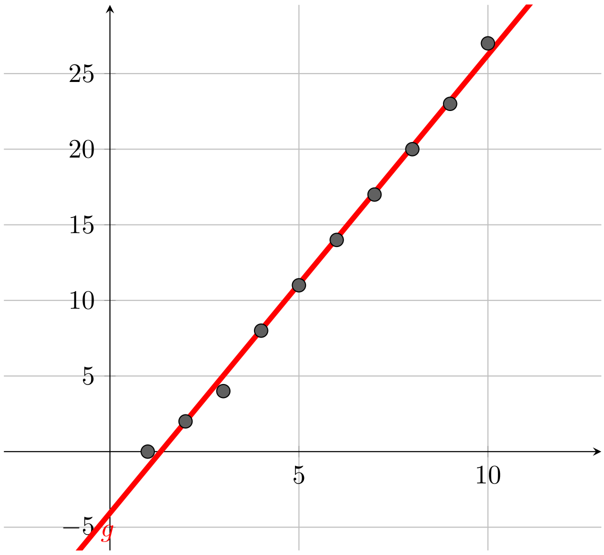
Exercice 3

1. A : 7, 8, 11, 12, 13, 13, 13 et B : 4, 7, 9, 12, 13, 13, 19
   mode : série A, la valeur la plus fréquente est 13. Série B le mode est 13.
   moyenne : série A : $\dfrac{7+8+11+12+13+13+13}{7} = 11$, série B : 11
   médiane : série A : 12 (3ème valeur), série B 12 (3ème valeur)
2. écart absolu moyen à la moyenne :
   On calcule les écarts à la moyenne, sans leurs signes et on fait la moyenne.
   Pour la série A ils valent : 4, 3, 0, 1, 2, 2, 2 et la moyenne est $\dfrac{14}{7} = 2$
   Inutile pour la série B, c'est stupide.
3. Variance on utilise la seule formule humainement acceptable : $V(X) = \overline{x^2} - \overline{x}^2$
   La méthode consiste à calculer les écarts précédents, les élever au carré et en prendre la moyenne.
   Pour la série A les carrés des écarts sont : 16, 9, 0, 1, 4, 4, 4 et la moyenne vaut $\dfrac{38}{7} = 5.43$.
   L'écart-type est la racine de la variance et vaut $\sigma(X) = \sqrt{V(X)} = 2.33$
   Pour la série B on trouve $V(X) = 11.22$ et $\sigma(X) = 3.35$

Par manque de temps, on passe directement à l'exercice 9.

Exercice 9

1. Voici le graphe obtenu après avoir tracé la droite de régression linéaire par la méthode des moindres carrés.



2. obtenir une droite de regression linéaire par la méthode des moindres carrés
   sur TI 83 : chercher (regression linéaire ti83), première vidéo :
   https ://www.youtube.com/watch ?v=uzMAL7uPE44
   sur casio : chercher (regression linéaire casio), première vidéo :
   https ://www.youtube.com/watch ?v=AYfsjiwsW00
   On obtient : $y = 3.03x - 4.07$
   Par manque de temps je zappe toute la recherche des coefficients...
   Seule la méthode à la calculatrice est au programme. Relisez le cours pour les détails.
3. Comme dans l'exercice précédent, on remplace x par 12 dans l'équation et
   $y = 3.03 \times 12 - 4.07 = 32.33$
   Au cours de 12 semestre d'utilisation, on estime qu'il y a 32.33% des lots qui vont subir une panne.

Exercices demandés jeudi 19 mars : 10, 12 et 13

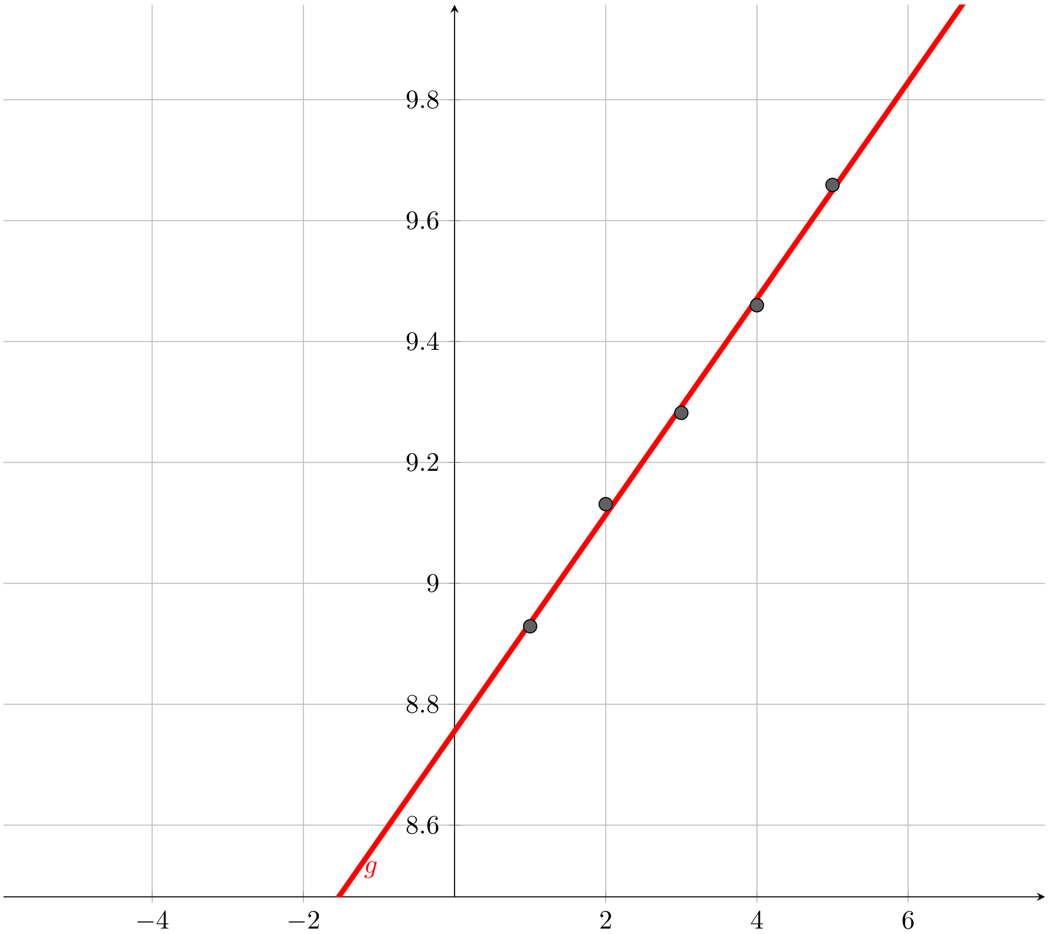
Exercice 10

(a)

(b) On calcule les valeurs suivantes :

Année	x_i	$y_i = \ln p_i$
1992	1	$\ln 7550 = 8.929$
1993	2	9.131
1994	3	9.282
1995	4	9.460
1996	5	9.659

(c) Voici le nuage de points obtenu, ainsi que la droite de régression linéaire.
Avant de tracer la droite, on peut remarquer l'alignement quasi parfait des points.
Un ajustement linéaire de y en x est donc envisageable.



1. (a) La méthode est indiquée dans l'exercice précédent.
On obtient $y = 0.179x + 8.756$.

(b) La calculatrice donne directement le coefficient de correlation linéaire $r = 0.9989$
Ce nombre étant proche de sa valeur maximale (1), il est important de donner des décimales jusqu'à ce qu'on trouve autre chose que 9. En l'occurence il est excellent et les points sont presque parfaitement alignés. Ce résultat confirme la remarque faite à la question 1.b)

(c) on a $y = 0.179x + 8.756 \iff \ln p = 0.179x + 8.756 \iff p = \exp(0.179x + 8.756)$

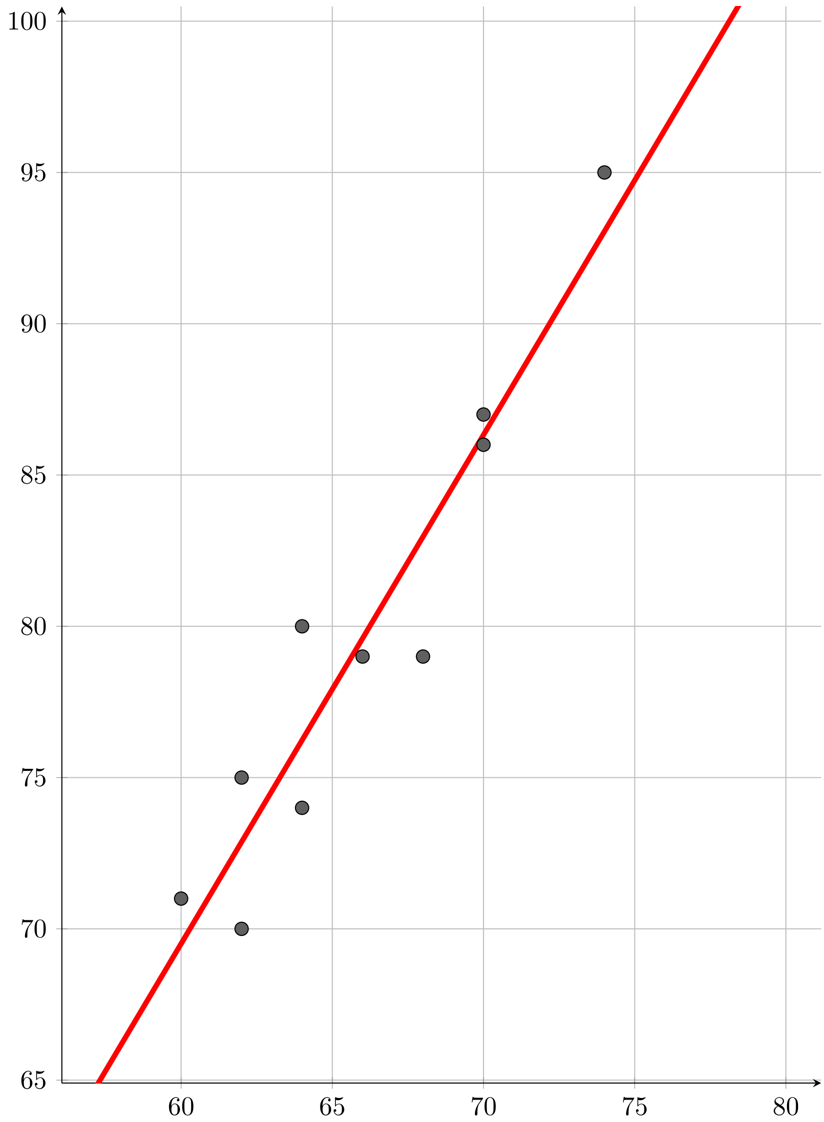
$$p = \exp(8.756) \times \exp(0.179x) = 6348.666\mathrm{e}^{0.179x}$$

(d) 1998 correspond au rang $x = 7$ on substitue dans l'expression précédente pour trouver $p_7 = 6348.666\mathrm{e}^{0.179\times 7} = 22226$ passagers.

Remarque finale. La situation précédente mérite un commentaire : le tableau de départ montre une explosion du nombre de passagers. La progression est exponentielle.
Si vous calculez les valeurs successives de $\dfrac{p_{i+1}}{p_i}$: $\dfrac{9235}{7550}, \dfrac{10741}{9235}$ etc. vous trouvez des valeurs très proches.
Cela signifie que, d'une année à l'autre, le nombre de passagers est multiplié par une constante (ou presque).
C'est ce qu'on appelle une progression *exponentielle* et c'est pour cela que les logarithmes des nombres de passagers conduisent à un ajustement affine d'excellente qualité.
Aurait-on essayé un ajustement affine avec les valeurs brutes (les formules donnent des réponses !) on aurait trouvé un coefficient de corrélation linéaire bien moins proche de 1.
Donc la démarche de l'exercice est la bonne dans ce genre de contexte.

Exerice 12

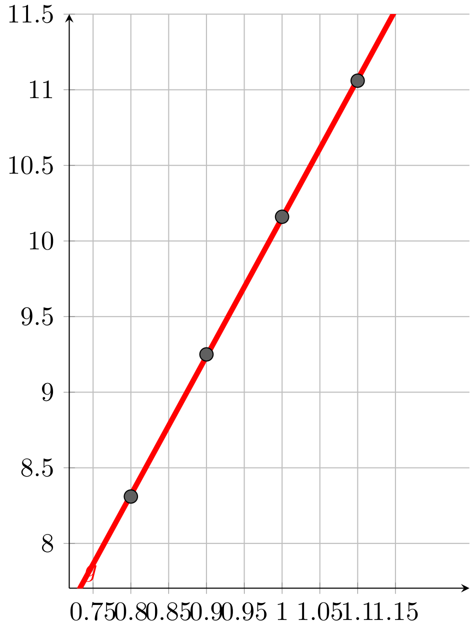
1. Voici le nuage de points et la droite de régression de y en x



2. On calcule la moyenne $\overline{x}$ des x_i et la moyenne $\overline{y}$ des y_i, le point moyen $G = (\overline{x}, \overline{y}) = (66, 79.6)$
3. Le coefficient de corrélation linéaire r vaut $r = 0.949$.
4. (a) La droite de régression de y en x est : $y = 1.68x - 31.4$.
   (b) Voir figure
5. On remplace x par 77 dans l'équation et il vient : 98.1
   La charge de rupture d'un acier ayant une teneur en carbone de 77 est 98.1 kg

Exercice 13

1. Voici le nuage de points et la droite de regression linéaire de y en x.



On calcule la moyenne $\overline{x}$ des x_i et la moyenne $\overline{y}$ des y_i, le point moyen $G = (\overline{x}, \overline{y}) = (0.95, 9.695)$

Le coefficient de corrélation linéaire r vaut $r = 0.99995$.

1. La droite de régression de y en x est : $y = 9.16x + 0.993$.
2. Voir figure

$Y = -\log D \Longleftrightarrow y = -\dfrac{\ln D}{\ln 10} \Longleftrightarrow D = \exp(-y \times \ln(10))$
on a $y = 9.16x + 0.993 \Longrightarrow D = \exp((-9.16x - 0.993) \times \ln(10))$

$$D = \exp((-9.16\frac{1000}{T} - 0.993) \times \ln(10)) = 0.1016\mathrm{e}^{\dfrac{-21091}{T}}$$

On applique la formule précédente avec les valeurs de T fournies et on obtient :

T	D
900	0.00792
1000	0.0102
1100	0.0126
1200	0.0150
1300	0.0174