Université Paris I, Panthéon - Sorbonne

Licence M.I.A.S.H.S. première année 2022 – 2023

Feuilles de TD, cours de L1 Probabilités

Jean-Marc Bardet (Université Paris 1, SAMM)

Email: bardet@univ-paris1.fr
Page oueb: http://samm.univ-paris1.fr/-Jean-Marc-Bardet-

U.F.R. 27 et Equipe SAMM (Statistique, Analyse et Modélisation Multidisiplinaire)
Université Panthéon-Sorbonne, 90 rue de Tolbiac, 75013 Paris.

Feuille n^o 1:

Un peu de statistiques descriptives

1. **(*)** Calculer la moyenne et l'écart type du nombre de lettres dans les mots de cette phrase.

2. **(*)** On considère les longueurs de 30 oeufs de coucous (en mm):

22.5	20.1	23.3	22.9	23.1	22.0	22.3	23.6	24.7	23.7
24.0	20.4	21.3	22.0	24.2	21.7	21.0	20.1	21.9	21.9
21.7	22.6	20.9	21.6	22.2	22.5	22.2	24.3	22.3	22.6

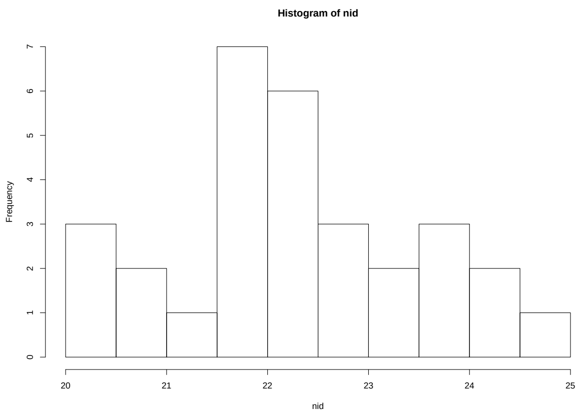
- Construire un tableau de fréquences pour des classes de longueur 0.5 mm. Illustrer graphiquement les données à l'aide d'un histogramme, puis d'un polygone des fréquences cumulées. En déduire la médiane.
- Calculer la moyenne et l'écart-type à partir des 30 données, puis à partir des données par classes. Différence?

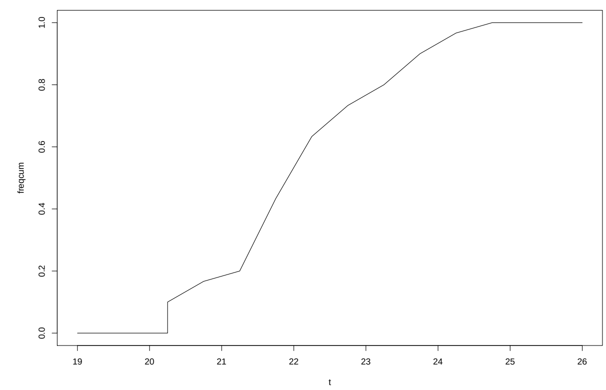
Proof.

- Le tableau des fréquences est le suivant:

Taille	]20,20.5]	]20.5,21]	]21,21.5]	]21.5,22]	]22,22.5]	]22.5,23]	]23,23.5]	]23.5,24]	]24,24.5][	]24.5,25]
Effectifs	3	2	1	7	6	3	2	3	2	1
Fréquences	0.1	0.066	0.033	0.233	0.2	0.1	0.066	0.1	0.066	0.033
Fréq Cum	0.1	0.166	0.2	0.433	0.633	0.733	0.8	0.9	0.966	1

Voici l'histogramme demandé ainsi que le polygône des fréquences cumulées: Sur le polygône on peut déduire





la médiane qui est l'antécédent de 0.5, soit $21.75 + 0.5/3 \simeq 21.92\,mm$.

- En travaillant directement sur les 30 mesures, on obtient une moyenne de $22.32\,mm$, une variance de $1.394267\,mm^2$ et donc un écart-type de $1.180791\,mm$.
Si on travaille sur les données par classes, on obtient une moyenne calculée par:

$$\frac{1}{30}\left(3*20.25+2*20.75+21.25+7*21.75+6*22.25+3*22.75+2*23.25+3*23.75+2*24.25+24.75\right) \simeq 22.283$$

Pour la variance, on obtient:

$$\frac{1}{30}\left(3*20.25^2+2*20.75^2+21.25^2+7*21.75^2+6*22.25^2+3*22.75^2+2*23.25^2+3*23.75^2+2*24.25^2+24.75^2\right)-22.283^2 \simeq 1.432.$$

Ainsi, l'écart-type obtenu à partir des classes est de $\simeq 1.197$. Il y a donc une petite différence par rapport à la moyenne et à l'écart-type obtenus directectement à partir des données.

□

3. **(*)** Les salaires hebdomadaires (en dollars) d'un échantillon de femmes d'une même entreprise en 2014 étaient:

Salaires hebdomadaires (Dollars)	Nombre de femmes
moins de 150	1
entre 150 et 200	4
entre 200 et 250	28
entre 250 et 300	42
entre 300 et 350	33
entre 350 et 400	18
entre 400 et 450	13
entre 450 et 600	9
600 et plus	2

   (a) Représenter par un histogramme la répartition des salaires. Construire le polygône des fréquences cumulées et en déduire la médiane et les quantiles à 10% et 90% de cette distribution. Quel est le ratio entre ces deux quantiles (mesure de la répartition des salaires)?

   (b) Calculer la moyenne et l'écart type de cette distribution.

   (c) Que se passe t-il pour la médiane, la moyenne et l'écart-type si chaque femme est augmentée de 50 dollars par semaine? Est augmentée de 10%?

4. **(*)** On a demandé à 1000 étudiants et à 1500 étudiantes la couleur de leurs yeux:

Couleur des yeux	Nombre d'étudiants	Nombre d'étudiantes
Bleu	420	650
Vert	120	150
Marron	450	680
Divers	10	20

   Représenter sur un même graphe les deux diagrammes à bâtons des fréquences. Peut-on conclure que les hommes ont des yeux de couleur différente de ceux des femmes?

5. **(*)** 60 % de la population gagne moins que le salaire moyen a dit un politicien. Est-ce posssible? Illustrer votre réponse.

   Proof. Oui cela est tout à fait possible. Prenons par exemple la grille de salaires en euros: 1000, 1200, 1200, 2000, 2600. Alors la moyenne est de 1600 euros, et 60% gagnent moins. □

6. **(*)** Un ensemble d'observations a pour moyenne 20 et pour écart type 4. Quel est l'effet sur la moyenne et sur l'écart-type de:

   - On ajoute 3 à toutes les observations.
   - On multiplie toutes les observations par 10.
   - On soustrait 12 à toutes les observations, puis on les divise par 2.

   Proof. Soit $(x_1, \ldots, x_n)$ tel que $\overline{x}_n = \frac{1}{n}(x_1 + \ldots + x_n) = 20$ et $\frac{1}{n}\left((x_1 - \overline{x}_n)^2 + \ldots + (x_n - \overline{x}_n)^2\right) = 16$.
   • Si on ajoute 3 à toutes les observations, on travaille sur les $x_i + 3$, et ainsi la moyenne devient 23, mais l'écart-type est inchangé et vaut toujours 4 puisque les $x_i - \overline{x}_n$ conservent les mêmes valeurs.
   • Si on multiplie tous les x_i par 10, la moyenne devient 200, et l'écart-type devient 40 (la variance est elle multipliée par 100).
   • Si on considère les $(x_i - 12)/2$, alors la moyenne devient $8/2 = 4$ et l'écart-type vaut 2.

   □

7. **(**)** On constate qu'en France il y une personne sur 6 qui possède un ordinateur personnel et une personne sur 25 qui est chauve. On constate aussi que une personne sur 75 est chauve et possède un ordinateur personnel. Quel est le pourcentage de Français chevelus? De Français chauves sans ordinateur? De Français chevelus avec ordinateur? De Français chevelus sans ordinateur?

Proof. On note n le nombre de français, C l'ensemble des français chauves, P l'ensemble des français possédant un ordinateur personnel. On a donc $\mathrm{Card}(C) = n/25$, $\mathrm{Card}(P) = n/6$ et $\mathrm{Card}(C \cap P) = n/75$.
Il est clair que $\mathrm{Card}(Chevelus) = n - \mathrm{Card}(C) = n - n/25 = \frac{24}{25}\,n$: le pourcentage de français chevelus est de $24/25 = 96\%$.
On cherche ensuite $\mathrm{Card}(C \cap \overline{P}) = \mathrm{Card}(C) - \mathrm{Card}(C \cap P) = n/25 - n/75 = \frac{2}{75}\,n$: le pourcentage de français chauves sans ordinateurs est de $2/75 \simeq 2.66\%$
Puis, $\mathrm{Card}(\overline{C} \cap P) = \mathrm{Card}(P) - \mathrm{Card}(C \cap P) = n/6 - n/75 = \frac{23}{150}\,n$: le pourcentage de français chevelus avec ordinateur est de $23/150 \simeq 15.33\%$
Et enfin $\mathrm{Card}(\overline{C} \cap \overline{P}) = \mathrm{Card}(\overline{C}) - \mathrm{Card}(\overline{C} \cap P) = \frac{24}{25}\,n - \frac{23}{150}\,n = \frac{121}{150}\,n$: le pourcentage de français chevelus sans ordinateur est de $\frac{121}{150} \simeq 80.66\%$. □

8. **(***)** Soit $(x_1, \ldots, x_n)$, n valeurs réelles observées avec $n \in \mathbf{N}^*$.

   (a) Montrer que $\overline{x}$ est l'unique minimum de la fonction $f(x) = \sum_{i=1}^n (x_i - x)^2$.

   (b) Montrer que la fonction $g(x) = \sum_{i=1}^n |x_i - x|$ est minimisée par la médiane de $(x_1, \ldots, x_n)$. Est-ce l'unique minimum?

Proof. (a) On peut écrire que $f(x) = \sum_{i=1}^n \big((x_i - \overline{x}) + (\overline{x} - x)\big)^2 = \sum_{i=1}^n (x_i - \overline{x})^2 + \sum_{i=1}^n (\overline{x} - x)^2 + 2\sum_{i=1}^n (x_i - \overline{x})(\overline{x} - x)$. Dans les deux dernières sommes $(\overline{x} - x)$ ne dépend pas de i et peut donc être sorti de la somme. On obtient ainsi: $f(x) = \sum_{i=1}^n (x_i - \overline{x})^2 + n\,(\overline{x} - x)^2 + 2(\overline{x} - x)\sum_{i=1}^n (x_i - \overline{x})$. Mais $\sum_{i=1}^n (x_i - \overline{x}) = \big(\sum_{i=1}^n x_i\big) - n\,\overline{x} = n\,\overline{x} - n\,\overline{x} = 0$, donc: $f(x) = \sum_{i=1}^n (x_i - \overline{x})^2 + n\,(\overline{x} - x)^2$. La somme étant une constante ne dépendant pas de x, $f(x)$ est donc minimum quand $(\overline{x} - x)^2$ est minimum, ce qui est le cas pour $x = \overline{x}$. Ainsi f atteint un unique minimum en $x = \overline{x}$.

(b) Dans toute la suite, on ordonnera la famille $(x_1, \ldots, x_n)$ et ainsi on notera $\min_i(x_i) = x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)} = \max_i(x_i)$. On va travailler par la suite suivant la parité de x.

1. pour $n = 2p + 1$, on montre que le minimum est atteint en $x_{(p+1)}$. En effet, dans un tel cas

$$g(x_{(p+1)}) = \sum_{i=1}^{p} (x_{(p+1)} - x_{(i)}) + \sum_{i=p+2}^{2p+1} (x_{(i)} - x_{(p+1)}) = \sum_{i=p+2}^{2p+1} x_{(i)} - \sum_{i=1}^{p} x_{(i)}.$$

Pour $m \in]x_{(j)}, x_{(j+1)}]$ et $j = p+1, \ldots, 2p+1$ (avec, par convention, $x_{(2p+2)} = +\infty$ et les sommes vides nulles),

$$\begin{aligned} g(m) &= \sum_{i=1}^{j} (m - x_{(i)}) + \sum_{i=j+1}^{2p+1} (x_{(i)} - m) \\ &= \sum_{i=1}^{p} (m - x_{(i)}) + \sum_{i=p+2}^{2p+1} (x_{(i)} - m) + \sum_{i=p+1}^{j} (x_{(i)} - m) - \sum_{i=p+2}^{j} (x_{(i)} - m) \\ &= \sum_{i=p+2}^{2p+1} x_{(i)} - \sum_{i=1}^{p} x_{(i)} + 2\sum_{i=p+2}^{j} (m - x_{(i)}) + (m - x_{(p+1)}) \\ &> g(x_{(p+1)}). \end{aligned}$$

Par symétrie, il en est de même si $m \in [x_{(j)}, x_{(j+1)}[$ et $j = 0, \ldots, p$. On en déduit que le minimum de $g(.)$ est atteint en $x_{(p+1)}$, qui est la médiane des (x_j).

2. Si $n = 2p$, on déduit de ce qui précède que le minimum de g doit se trouver entre $x_{(p)}$ et $x_{(p+1)}$. En effet, pour $m \in]x_{(p)}, x_{(p+1)}]$,

$$\begin{aligned} g(m) &= \sum_{i=1}^{p} (m - x_{(i)}) + \sum_{i=p+1}^{2p} (x_{(i)} - m) \\ &= \sum_{i=p+1}^{2p} x_{(i)} - \sum_{i=1}^{p} x_{(i)}, \end{aligned}$$

qui est indépendant de m. Maintenant, pour $m \in [x_{(j)}, x_{(j+1)}[$ et $j = p+1, \ldots, 2p+1$, on a :

$$\begin{aligned}
g(m) &= \sum_{i=1}^{j}(m - x_{(i)}) + \sum_{i=j+1}^{2p+1}(x_{(i)} - m) \\
&= \sum_{i=1}^{p}(m - x_{(i)}) + \sum_{i=p+1}^{2p}(x_{(i)} - m) + 2\sum_{i=p+1}^{j}(m - x_{(i)}) \\
&> \sum_{i=p+1}^{2p} x_{(i)} - \sum_{i=1}^{p} x_{(i)}.
\end{aligned}$$

On retrouve bien que le minimum, qui est unique, est atteint par toute valeur entre $x_{(p)}$ et $x_{(p+1)}$. Si $x_{(p)} = x_{(p+1)}$, alors le minimum est atteint en une unique valeur $xx_{(p)} = x_{(p+1)}$, sinon il ne l'est pas. On note que la médiane, qui est la moyenne de $x_{(p)}$ et $x_{(p+1)}$, soit $\frac{1}{2}(x_{(p)} + x_{(p+1)})$ est telle que g atteint son minimum en la médiane.

□

9. **(***)** (Paradoxe de Simpson) Dans l'état de Floride en 1974, on a décompté les condamnations prononcées pour meurtres. Comme aux USA les statistiques ethniques sont autorisées, on a réparti celles-ci en fonction de la couleur de peau et on obtient les résultats suivants:

   - Parmi les "blancs" meurtriers, 72 ont été condamnés à mort et 2185 ont été condamnés à une autre peine;
   - Parmi les "noirs" meurtriers, 59 ont été condamnés à mort et 2448 ont été condamnés à une autre peine.

   (a) Quelles sont les fréquences de verdict de condamnation à mort suivant la couleur de peau du meurtrier? Quelle population vous semblent être privilégiée?

   (b) On peut rendre un peu plus précis ces résultats en distinguant la couleur de peau de la victime. Ainsi lorsque la victime était blanche, les meurtriers blancs ont écopé 72 fois de la peine de mort et 2074 fois d'une autre peine et les meurtriers noirs ont écopés 48 fois de la peine de mort et 239 fois d'une autre peine. En déduire les résultats lorsque la victime était noire (on pourra notamment utiliser des dessins d'ensembles). Qu'en déduisez vous finalement quant à la justice en Floride? (on dira que la couleur de peau de la victime est une variable de confusion).

Proof. (a) On a donc $f_B = 72/(2185 + 72) \simeq 3.19\%$ de condamnations à mort pour les meurtriers blancs et $f_N = 59/(2448 + 59) \simeq 2.35\%$ de condamnations à mort pour les meurtriers noirs. Il apparaît donc en première analyse que les meurtriers noirs sont moins souvent condamnés à mort que les meurtriers blancs.

(b) On déduit que si la victime est noire, il y a eu $72 - 72 = 0$ condamnation à mort et $2185 - 2074 = 111$ autres condamnations si le meurtrier est blanc, et il y a eu $59 - 48 = 11$ condamnations à mort et $2448 - 239 = 2209$ autres condamnations si le meurtrier est noir.
On s'aperçoit que la fréquence de verdict de mort quand la victime est blanche est de $f_{BB} = 72/(2074+72) \simeq 3.36\%$ de condamnations à mort pour les meurtriers blancs et $f_{NB} = 48/(239 + 48) \simeq 16.72\%$ de condamnations à mort pour les meurtriers noirs. Et quand la victime est noire la fréquence de verdict de mort est de $f_{BN} = 0/111 = 0\%$ de condamnations à mort pour les meurtriers blancs et $f_{NN} = 11/(2209 + 11) \simeq 0.48\%$ de condamnations à mort pour les meurtriers noirs.
La justice en Floride dépend donc totalement de la couleur de peau de la victime et de celle du meurtrier: la mort d'un blanc est démesurément plus punie que celle d'un noir, avec en plus 5 fois plus de chances d'être condamné à mort pour un meurtrier noir plutôt que pour un meurtrier blanc!

□

10. **(**)** Soit $(x_1, \ldots, x_{2n+1})$, $2n+1$ nombres réels inclus dans $[0, 1]$. Notons respectivement $\overline{x}$ et $\overline{m}$ la moyenne empirique et la médiane empirique de $(x_1, \ldots, x_{2n+1})$.

    (a) Montrer que $\sum_{i=1}^{2n+1} x_i \geq (n+1)\,\overline{m}$. En déduire que $\overline{x} \geq \frac{1}{2}\,\overline{m}$.

    (b) Montrer que $\overline{x} \leq \frac{1}{2}\big(1 + \overline{m}\big)$.

    (c) Soit $\overline{\sigma^2}$ la variance empirique de $(x_1, \ldots, x_{2n+1})$. Montrer que $\max\Big(0\,,\,\frac{1}{2}\,\big(\overline{m}\big)^2 - \big(\overline{x}\big)^2\Big) \leq \overline{\sigma^2} \leq \overline{x}(1 - \overline{x})$.

Proof. (a) Puisque $2n+1$ est impair, on sait que $\overline{m} = x_{(n+1)}$, avec la notation $x_{(i)}$ désignant le i-éme plus petit élément de $(x_1, \dots, x_{2n+1})$. On sait également que $\overline{x} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} x_i$.
On a $x_1 + \cdots + x_{2n+1} = x_{(1)} + \cdots + x_{(2n+1)}$, et $x_{(1)} < x_{(2)} < \cdots < x_{(2n+1)}$. Ainsi, comme les x_i sont positifs, on a $x_1 + \cdots + x_{2n+1}\ geq x_{(n+1)} + x_{(n+2)} + \cdots + x_{(2n+1)}$. Et comme $x_{(n+i)} > x_{(n+1)}$ pour tous $i = 1, \dots, n$, on a: $x_1 + \cdots + x_{2n+1} \geq x_{(n+1)} + x_{(n+1)} + \cdots + x_{(n+1)} \geq (n+1) x_{(n+1)} = (n+1)\overline{m}$. On déduit que $\overline{x} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} x_i \geq \frac{(n+1)}{2n+1}\, x_{(n+1)} \geq \frac{1}{2}\, \overline{m}$.
Comme pour tous i on a $x_{(i)} \leq 1$, on a également

$$x_1 + \cdots + x_{2n+1} \leq x_{(1)} + \cdots + x_{(n)} + \frac{1}{2}\,(x_{(n+1)} + 1) + 1 + 1 + \cdots + 1 \leq (n + 1/2)\, x_{(n+1)} + (n + 1/2).$$

En renormalisant par $2n+1$, on obtient le résultat.

(b) La variance empirique est définie par: $\overline{\sigma^2} = \frac{1}{2n+1}, \sum_{i=1}^{2n+1} x_i^2 - (\overline{x})^2$.
Comme $x_i \in [0,1]$ pour tous i, alors $x_i^2 \leq x_i$. Ainsi $\sum_{i=1}^{2n+1} x_i^2 \leq \sum_{i=1}^{2n+1} x_i$ et donc $\overline{\sigma^2} \leq \overline{x} - (\overline{x})^2 = \overline{x}(1 - \overline{x})$.
On a également $\overline{\sigma^2} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} \big(x_i - (\overline{x})\big)^2$, donc $\overline{\sigma^2} \geq 0$ comme une somme de carrés. De plus, on a

$$\sum_{i=1}^{2n+1} x_i^2 = \sum_{i=1}^{2n+1} x_{(i)}^2 \geq (n+1)\, x_{(n+1)}^2 \geq (n + 1/2) \big(\overline{m}\big)^2.$$

On en déduit que $\frac{1}{2n+1} \sum_{i=1}^{2n+1} x_i^2 \geq \frac{1}{2} \big(\overline{m}\big)^2$, et donc $\overline{\sigma^2} \geq \frac{1}{2} \big(\overline{m}\big)^2 - (\overline{x})^2$. On a ce résultat et également $\overline{\sigma^2} \geq 0$, donc $\overline{\sigma^2} \geq \max\big(0\,,\, \frac{1}{2} \big(\overline{m}\big)^2 - (\overline{x})^2\big)$.

□

11. **(*)** Soit $n \geq 2$ et $(y_i)_{1 \leq i \leq n}$ une famille de réels non tous égaux. Soit le nuage de points $(x_i, y_i)_{1 \leq i \leq n}$, où les $x_i = 1$ pour $1 \leq i \leq n-1$ et $x_n = 2$. Déterminer l'équation de la droite de régression par moindres carrés de ce nuage de points et le coefficient de détermination R^2.

Proof. Dans ce cas, la moyenne empirique des (x_i) est $\overline{x} = 1 + \frac{1}{n}$ et donc la variance empirique des (x_i) est $\overline{\sigma}_x^2 = \frac{1}{n} \Big(\sum_{i=1}^{n-1} \frac{1}{n^2} + \big(1 - \frac{1}{n}\big)^2 \Big) = \frac{n-1}{n^2}$. La covariance empirique entre les (x_i) et les (y_i) est $\overline{\sigma}_{xy} = \frac{1}{n} \Big(\sum_{i=1}^{n} y_i + y_n \Big) - \overline{y}\,\big(1 + \frac{1}{n}\big) = \frac{1}{n}\,(y_n - \overline{y})$. Donc $\widehat{a} = \frac{\overline{\sigma}_{xy}}{\overline{\sigma}_x^2} = \frac{n}{n-1}\,(y_n - \overline{y})$ et $\widehat{b} = \overline{y} - \widehat{a}\,\overline{x} = \Big(\frac{2n}{n-1}\Big)\,\overline{y} - \Big(\frac{n+1}{n-1}\Big)\, y_n$. L'équation de la droite de régression par moindres carrés de ce nuage de points est donc

$$y = \frac{n}{n-1}\,(y_n - \overline{y})\, x + \Big(\frac{2n}{n-1}\Big)\, \overline{y} - \Big(\frac{n+1}{n-1}\Big)\, y_n.$$

Le coefficient de détermination R^2 est $R^2 = \frac{\overline{\sigma}_{xy}^2}{\overline{\sigma}_x^2\, \overline{\sigma}_y^2} = \frac{1}{(n-1)\, \overline{\sigma}_y^2}\,(y_n - \overline{y})^2$. □

12. **(***)** Pour $n \geq 2$, on a observé la famille de couples de nombres réels $\big((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\big)$, où les x_i ne sont pas tous égaux. Supposons qu'il existe $a^* \in \mathbf{R}$ et $b^* \in \mathbf{R}$, des paramètres inconnus, tels que $y_i = a^*\, x_i + b^* + \varepsilon_i$ pour tout $1 \leq i \leq n$, où $(\varepsilon_1, \dots, \varepsilon_n)$ est une famille de réels non observés.

(a) Supposons, seulement dans cette question, que $\varepsilon_i = 0$ pour tout $i \in \{1, \dots, n\}$. Déterminer l'équation de la droite de régression par moindres carrés passant du nuage de points $(x_i, y_i)_{1 \leq i \leq n}$.

(b) Désormais les ε_i sont des nombres réels quelconques non observés. On note $y = \widehat{a}\, x + \widehat{b}$ la droite de régression par moindres carrés du nuage de points $(x_i, y_i)_{1 \leq i \leq n}$ et $\widehat{\varepsilon}_i = y_i - \widehat{a}\, x_i - \widehat{b}$ pour tout $i \in \{1, \dots, n\}$. Montrer que $\sum_{i=1}^{n} \widehat{\varepsilon}_i = 0$.

(c) On note respectivement $\overline{x}$ et $\overline{\varepsilon}$ les moyennes empiriques des (x_i) et des (ε_i), $\overline{\sigma}_x^2$ la variance empirique des (x_i) et $\overline{\sigma}_{x,\varepsilon}$ la covariance empirique entre (x_i) et (ε_i). Montrer que $\widehat{a} = a^* + \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}$, puis montrer que $\widehat{\varepsilon}_i = (\varepsilon_i - \overline{\varepsilon}) - \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\,(x_i - \overline{x})$.

(d) En déduire que $\frac{1}{n} \sum_{i=1}^{n} \big(\widehat{\varepsilon}_i\big)^2 \leq \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2$. Ne pouvait-on pas montrer ce résultat plus simplement?

Proof. (a) Notons $\Delta(a,b) = \sum_{i=1}^{n} \big(y_i - (a\, x_i + b)\big)^2$. Dans le cas où $\varepsilon_i = 0$ pour tout i, $\Delta(a,b) = \sum_{i=1}^{n} \big(a^*\, x_i + b^*\_(a\, x_i + b)\big)^2 = \sum_{i=1}^{n} \big((a^* - a)\, x_i + (b^*\_b)\big)^2$. Donc, pour tout (a,b), $\Delta(a,b) \geq 0$ et $\Delta(a^*, b^*) = 0$ implique que (a^*, b^*) est l'unique minimum de $\Delta(a,b)$: la droite de régression par MC est $y = a^*\, x + b^*$.

(b) On a $\sum_{i=1}^n \widehat{\varepsilon}_i = \sum_{i=1}^n (y_i - \widehat{a}\, x_i - \widehat{b}) = \sum_{i=1}^n \big((y_i - \overline{y}) - \widehat{a}\,(x_i - \overline{x})\big)$ puisque $\widehat{b} = \overline{y} - \widehat{a}\,\overline{x}$. En conséquence

$$\sum_{i=1}^n \widehat{\varepsilon}_i = \sum_{i=1}^n \big(y_i - \overline{y}\big) - \widehat{a} \sum_{i=1}^n (x_i - \overline{x}) = 0.$$

(c) On a $\widehat{a} = \frac{\overline{\sigma}_{x,y}}{\overline{\sigma}_x^2} = \frac{\overline{\sigma}_{x,a^*x+b^*+\varepsilon}}{\overline{\sigma}_x^2} = \frac{a^*\,\overline{\sigma}_{x,x} + \overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2} = a^* + \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}$.
On a également

$$\widehat{\varepsilon}_i = (y_i - \overline{y}) - \widehat{a}\,(x_i - \overline{x}) = (a^* x_i + b^* + \varepsilon_i - \overline{a^* x + b^* + \varepsilon}) - \Big(a^* + \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\Big)\,(x_i - \overline{x}) = (\varepsilon_i - \overline{\varepsilon}) - \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\,(x_i - \overline{x}).$$

(d) En utilisant une formule précédente,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \big(\widehat{\varepsilon}_i\big)^2 &= \frac{1}{n}\sum_{i=1}^n \Big((\varepsilon_i - \overline{\varepsilon}) - \frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\,(x_i - \overline{x})\Big)^2 \\
&= \frac{1}{n}\sum_{i=1}^n (\varepsilon_i - \overline{\varepsilon})^2 - 2\,\frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\,\frac{1}{n}\sum_{i=1}^n (\varepsilon_i - \overline{\varepsilon})(x_i - \overline{x}) + \Big(\frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\Big)\,\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 \\
&= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - (\overline{\varepsilon})^2 - 2\,\frac{\overline{\sigma}_{x,\varepsilon}^2}{\overline{\sigma}_x^2} + \Big(\frac{\overline{\sigma}_{x,\varepsilon}}{\overline{\sigma}_x^2}\Big)\,\overline{\sigma}_x^2 \\
&= \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2 - (\overline{\varepsilon})^2 - \frac{\overline{\sigma}_{x,\varepsilon}^2}{\overline{\sigma}_x^2} \\
&\leq \frac{1}{n}\sum_{i=1}^n \varepsilon_i^2.
\end{aligned}$$

Une preuve directe de ce résultat aurait pu être: puisque $(\widehat{a}, \widehat{b})$ est l'unique minimum de $\Delta(a,b)$, alors $\Delta(\widehat{a}, \widehat{b}) \leq \Delta(a,b)$ pour tout $(a,b) \in \mathbf{R}^2$, et donc $\sum_{i=1}^n \big(\widehat{\varepsilon}_i\big)^2 = \Delta(\widehat{a}, \widehat{b}) \leq \Delta(a^*, b^*) = \sum_{i=1}^n \varepsilon_i^2$.

$\square$

Feuille n^o 2:

Espace de probabilité, événements et dénombrement

1. **(*)** On choisit une carte au hasard dans un jeu de 52 cartes. Quelle est la probabilité que ce soit une dame (décrire l'ensemble fondamental et la probabilité considérée) ? On en choisit deux. Quelle est la probabilité que ce soit deux dames (décrire l'ensemble fondamental et la probabilité considérée)?

2. **(*)** En informatique, on utilise le système binaire pour coder les caractères. Un bit (binary digit : chiffre binaire) est un élément qui prend la valeur 0 ou la valeur 1. Avec 8 chiffres binaires (un octet), combien de caractères peut-on coder?

3. **(*)** Soit A l'ensemble des nombres de quatre chiffres, le premier étant non nul.
   (a) Calculer le nombre d'éléments de A.
   (b) Dénombrer les éléments de A:
   a) composés de quatre chiffres distincts;
   b) composés d'au moins deux chiffres identiques;
   c) composés de quatre chiffres distincts autres que 5 et 7.

4. **(*)** Une boîte contient 3 jetons, un rouge, un vert et un bleu. On considère l'expérience consistant à tirer au hasard un jeton dans la boîte, à l'y remettre puis à en tirer un second. Décrire l'ensemble fondamental et la mesure de probabilité. Déterminer la probabilité de ne pas avoir un jeton rouge. Même question si le second jeton est tiré sans qu'on ait remis le premier.

Proof. Dans le premier cas, un tirage possible est (R, V) ou (V, V). On a ainsi $\Omega = \{R, V, B\}^2$, $\mathcal{A} = \mathcal{P}\big(\{R, V, B\}^2\big)$ et $\mathbb{P}$ est la mesure de probabilité uniforme.
Soit A l'événement "Ne pas tirer un jeton rouge". Alors $A = \big\{(B, V), (V, B), (V, V), (B, B)\big\}$. Comme $\mathrm{Card}(\Omega) = 3^2$, on en déduit que $\mathbb{P}(A) = 4/9$.

Dans le second cas, un tirage possible est (R, V) mais plus (V, V). On a ainsi Ω qui est composé de l'ensemble des arrangements de 2 éléments choisis dans $\{R, V, B\}$, $\mathcal{A} = \mathcal{P}\big(\Omega\big)$ et $\mathbb{P}$ est la mesure de probabilité uniforme.
Soit A l'événement "Ne pas tirer un jeton rouge". Alors $A = \big\{(B, V), (V, B)\big\}$. Comme $\mathrm{Card}(\Omega) = A_3^2 = 6$, on en déduit que $\mathbb{P}(A) = 2/6 = 1/3$.

$\square$

5. **(*)** Soit $(\Omega, \mathcal{A})$ un espace probabilisable et soit E, F et G trois événements de $\mathcal{A}$. Trouver des expressions pour les événements suivants que l'on dira réalisés lorsque:

   - E seul l'est;
   - E et G le sont mais pas F;
   - au moins l'un des trois l'est;
   - au moins deux d'entre eux le sont;
   - les trois le sont;
   - aucun ne l'est;
   - au plus deux d'entre eux le sont;
   - exactement deux le sont;
   - au plus trois le sont.

Proof.

- E seul l'est: cela s'écrit $E \cap \overline{F} \cap \overline{G}$.
- E et G le sont mais pas F: cela s'écrit $E \cap \overline{F} \cap G$.
- au moins l'un des trois l'est: cela s'écrit $\overline{\overline{E} \cap \overline{F} \cap \overline{G}}$.
- au moins deux d'entre eux le sont: cela s'écrit $E \cap F \cap \overline{G} \bigcup E \cap G \cap \overline{F} \bigcup F \cap G \cap \overline{E} \bigcup E \cap F \cap G = E \cap F \bigcup E \cap G \bigcup F \cap G$.
- les trois le sont: cela s'écrit $E \cap F \cap G$.
- aucun ne l'est: cela s'écrit $\overline{E} \cap \overline{F} \cap \overline{G}$.
- au plus deux d'entre eux le sont: cela s'écrit $\overline{E \cap F \cap G}$.
- exactement deux le sont: $E \cap F \cap \overline{G} \bigcup E \cap G \cap \overline{F} \bigcup F \cap G \cap \overline{E}$.
- au plus trois le sont: cela s'écrit Ω.

$\square$

6. **(**)** Au Scrabble, on a un tirage avec 7 lettres distinctes. Combien de mots de 7 lettres peut-on constituer au maximum à partir de ce tirage? Combien de mots de 3 lettres? Répondre aux mêmes questions si deux des lettres (et uniquement 2) du tirage initial sont identiques?

Proof. Dans le cas de lettres distinctes, comme un mot de 7 lettres est un 7-uplet de ces lettres (il y a de l'ordre), donc un arrangement de 7 lettres parmi les 7 lettres distinctes choisies, le nombre de mots de 7 lettres correspond à $A_7^7 = 7!/0! = 7! = 5040$ mots. Cela correspond aussi au nombre de permutations dans un ensemble taille 7.
Pour déterminer le nombre de mots de 3 lettres, qui correspondent aux arrangements de taille 3 parmi les 7 lettres distinctes du tirage. Le nombre de mots de 3 lettres correspond au nombre de ces arrangements, c'est-à-dire à $A_7^3 = 7!/(7-3)! = 7 * 6 * 5 = 210$ mots.
On suppose désormais que 2 lettres sont identiques. Pour le cas des mots de 7 lettres, on peut reprendre le résultat précédent, sauf que l'on doit diviser par 2! le résultat du fait que pour les 2 lettres identiques, chaque mot de 7 lettres peut s'écrire de 2 manières différentes (en gardant un ordre entre les 2 lettres identiques). Ainsi il y a $7!/2! = 2520$ mots possibles.
Pour les mots de 3 lettres, on doit considérer 3 cas: 1/ le nombre de mots où la lettre double n'apparaît pas, soit A_5^3 mots différents, 2/ ceux où la lettre double n'apparaît qu'une fois et cela peut être à 3 places différentes, soit $3 * A_5^2$ mots différents et enfin, 3/ ceux où la lettre double apparaît 2 fois, et cela peut se faire à 3 places différentes, soit $3 * A_5^1$ mots différents. Au total on aura donc $5 * 4 * 3 + 3 * 5 * 4 + 3 * 5 = 135$ mots différents. $\square$

7. **(**)** On demande à 5 personnes d'écrire leur nom sur un papier, puis on mélange les papiers et on les redistribue au hasard. Quelle est la probabilité qu'au moins une personne ait le papier avec son nom (décrire l'ensemble fondamental, la tribu et l'événement considéré)? Que tous aient le papier avec leur nom?

Proof. Si on numérote $\{1, 2, 3, 4, 5\}$ les 5 personnes (non, je ne suis pas un numéro!), un résultat de l'expérience aléatoire est un 5-uplet issu de ces 5 numéros sans répétition, donc un arrangement. Ainsi on en déduit l'ensemble fondamental $\Omega = \big\{\text{Arrangements de longueur 5 issus de } \{1, 2, 3, 4, 5\}\big\}$. Ce qui conduit à $\text{Card}(\Omega) = A_5^5 = 5!/0! = 120$. Naturellement, comme il n'y a aucune autre indication, la tribu est $\mathcal{A} = \mathcal{P}(\Omega)$. L'événement A_1 qui appartient nécessairement à $\mathcal{A}$ est $A_1 = \big\{(i_1, i_2, i_3, i_4, i_5) \in \Omega,\ \exists 1 \le k \le 5,\ i_k = k\big\}$. On peut également définir A_1 par son événement contraire, soit $A_1 = \overline{\big\{(i_1, i_2, i_3, i_4, i_5) \in \Omega,\ \forall 1 \le k \le 5,\ i_k \neq k\big\}}$. Pour trouver la probabilité de A_1, il faut préciser que tous les papiers semblent être tirés avec la même probabilité, donc on est en situation d'équiprobabilité et ainsi $\mathbb{P}(A_1) = \text{Card}(A_1)/\text{Card}(\Omega) = 1 - \text{Card}(\overline{A}_1)/\text{Card}(\Omega)$. Pour trouver $\text{Card}(\overline{A}_1)$ qui correspond à la situation où aucun n'a un papier à son nom, considérons le numéro 1. Il a 4 places possibles, et à chaque place prise on peut décompter 11 possibilités. Au total, $\text{Card}(\overline{A}_1) = 44$ et ainsi $\mathbb{P}(A_1) = \text{Card}(A_1)/\text{Card}(\Omega) = 1 - 44/120 = 19/30$.
L'événement "tous ont le papier avec leur nom" est l'événement $A_2 = \big\{(1, 2, 3, 4, 5)\big\}$. On a clairement $\text{card}(A_2) = 1$ et ainsi $\mathbb{P}(A_2) = 1/120$. $\square$

8. **(**)** On constitue un groupe de 6 personnes choisies au hasard parmi 25 femmes et 32 hommes.
1) De combien de façons peut-on constituer ce groupe de 6 personnes?
2) Après avoir précisé l'ensemble fondamental, la tribu et l'événement considéré, dans chacun des cas suivants, quelle est la probabilité d'obtenir pour ces 6 personnes: a) uniquement des hommes; b) des personnes de même sexe; c) au moins une femme et au moins un homme.

Proof. On dispose d'une population de 57 personnes différentes et on en choisit 6 parmi ces 57 sans faire attention à l'ordre. Aussi un événement élémentaire est une combinaison (un sous-groupe) de 6 éléments parmi ces 57.
1/ Il y a donc $\mathrm{Card}(\Omega) = C_{57}^6 = \frac{57!}{51!\,6!} = 36288252$ possibilités de groupes. 2/ On associe naturellement à Ω la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Tous les tirages sont équiprobables, et donc on associe la mesure de probabilité uniforme sur $(\Omega, \mathcal{A})$.
a/ L'événement $E_a \in \mathcal{A}$, l'union des combinaisons contenant uniquement 6 hommes. On a $\mathrm{Card}(E_a) = C_{32}^6$ et ainsi $\mathbb{P}(E_a) = C_{32}^6/C_{57}^6 = \frac{32!\,51!}{26!\,57!} \simeq 0.0249$. b/ L'événement $E_b \in \mathcal{A}$, l'union des combinaisons contenant uniquement 6 femmes et des combinaisons contenant uniquement 6 hommes. On a $\mathrm{Card}(E_b) = C_{32}^6 + C_{25}^6$ et ainsi $\mathbb{P}(E_b) = (C_{32}^6 + C_{25}^6)/C_{57}^6 = \frac{32!\,51!}{26!\,57!} + \frac{25!\,51!}{19!\,57!} \simeq 0.0298$. c/ L'événement $E_c \in \mathcal{A}$, l'union des combinaisons contenant au moins 1 femme au moins 1 homme. Il est clair que l'événement contraire de E_c est E_b. On en déduit donc que On a $\mathrm{Card}(E_b) = C_{32}^6 + C_{25}^6$ et ainsi $\mathbb{P}(E_c) = 1 - \mathbb{P}(E_b) \simeq 0.9702$. □

9. **(**)** Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit A un événement de $\mathcal{A}$ non réduit à l'ensemble vide. On définit pour tout $B \in \mathcal{A}$, $\mu(B) = \mathbb{P}(B \cap A)/\mathbb{P}(A)$. Montrer que μ est bien une mesure de probabilité sur $(\Omega, \mathcal{A})$. Serait-il possible de définir μ sur un autre espace probabilisable? Mêmes questions avec ν telle que pour tout $B \in \mathcal{A}$, $\nu(B) = \mathbb{P}(B \cup A)$.

Proof. Il est clair que $\mu(\Omega) = \mathbb{P}(A)/\mathbb{P}(A) = 1$, que $0 \leq \mathbb{P}(B \cap A) \leq \mathbb{P}(A)$ pour tout $B \in \mathcal{A}$ car $B \cap A \subset A$ et comme $\mathbb{P}$ est à valeur dans $[0,1]$ alors μ l'est aussi. Par ailleurs, pour toute famille $(A_i)_{i \in I}$, où $I \subset \mathbf{N}$, d'événements disjoints de $\mathcal{A}$, alors $\mathbb{P}\Big(A \cap \bigcup_{i \in I} A_i\Big) = \mathbb{P}\Big(\bigcup_{i \in I} A \cap A_i\Big)$. Mais les événements $A \cap A_i$ forment également une famille dénombrable d'événements disjoints (car pour $i \neq j$, $(A \cap A_i) \cap (A \cap A_j) = A \cap (A_i \cap A_j) = \emptyset$ car $A_i \cap A_j = \emptyset$). Donc comme $\mathbb{P}$ est une mesure de probabilités, $\mathbb{P}\Big(\bigcup_{i \in I} A \cap A_i\Big) = \sum_{i \in I} \mathbb{P}\big(A \cap A_i\big)$ et ainsi $\mu(\bigcup_{i \in I} A_i) = \sum_{i \in I} \mathbb{P}\big(A \cap A_i\big)/\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}\big(A \cap A_i\big)/\mathbb{P}(A) = \sum_{i \in I} \mu\big(A_i\big)$. La propriété est bien vérifiée et μ est bien une probabilité.
Il serait possible de définir μ sur $(A, \mathcal{A}')$, où $\mathcal{A}' = \big\{E \in \mathcal{A},\, E \subset A\big\}$, qui est une sous-tribu de $\mathcal{A}$.
Si on considère ν, alors $\nu(\emptyset) = \mathbb{P}(A) > 0$ donc ν ne peut pas être une mesure de probabilités. □

10. **(***)** Dans une assemblée de n personnes, à partir de quelle valeur de n a-t-on au moins une chance sur deux que deux personnes soient nées le même jour? Préciser les hypothèses faites...

Proof. On va supposer pour simplifier que les années ont exactement 365 jours et qu'il y a situation d'équiprobabilité, c'est-à-dire qu'il y a autant de chance de naître n'importe quel jour de l'année (ce qui n'est pas vrai stricto sensu, il y a plus de naissances par exemple en mai que décembre en France). On chosit donc n numéros avec remise dans $E = \{1, \cdots, 365\}$ et le résultat de cette expérience aléatoire est donc un n-uplet de E^n. D'où $\Omega = E^n$ et il y a donc $\mathrm{Card}(\Omega) = 365^n$ tirages possibles.
On considérera donc $\mathcal{A} = \mathcal{P}(\Omega)$ et l'événement $A \in \mathcal{A}$ "il y a au moins égalité de 2 coordonnées du tirage". On préfera travailler avec $\overline{A}$, toutes les coordonnées sont distinctes. Calculons le cardinal de $\overline{A}$. Pour la première coordonnée, on peut choisir tout numéro dans E, pour la seconde, il n'y plus que 364 possibilités, ... Aussi a-t-on $\mathrm{Card}(\overline{A}) = A_{365}^n$ et $\mathbb{P}(A) = 1 - A_{365}^n/365^n$.
On cherche n tel que $\mathbb{P}(A) \geq 1/2$ ou encore $A_{365}^n/365^n \leq 1/2$. En passant au logarithme, on a l'équation:

$$\sum_{i=0}^{n-1} \log(365 - i) - \log(365) = \sum_{i=0}^{n-1} \log(1 - i/365) =\leq -\log(2).$$

Mais en première approximation $\log(1 - i/365) \simeq -i/365$. Donc si n reste petit par rapport à 365, cela donne

$$\sum_{i=0}^{n-1} i/365 \geq \log 2 \quad \text{soit} \quad n(n-1) \geq 365\, 2 \log 2.$$

En première approximation cela donne $n^2 \geq 506$, d'où $n \geq 22.5$ et donc n doit être plus grand que 23.
Si on vérifie avec un logiciel numérique pour $n = 22$, on trouve $\mathbb{P}(A) \simeq 0.4757$ et pour $n = 23$, $\mathbb{P}(A) \simeq 0.5073$.

□