

Chapitre 2. Statistiques à 1 variable

Yann Barsamian

École Européenne de Bruxelles 1

Année scolaire 2020–2021



- Caractéristiques d'une série (moyenne, écart-type, mode, médiane, quartiles)
- Représentations graphiques (histogramme, boîte à moustaches)
- Propriétés

Soit une série statistique prenant p différentes valeurs x_1, x_2, \dots, x_p (les x_i) et où chaque valeur x_i (pour $1 \leq i \leq p$) a pour effectif n_i . Par exemple, une étude sur 30 élèves concernant le temps de travail journalier (en minutes) à la maison donne les résultats suivants :

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1

Dans ce tableau, on lit par exemple que 4 élèves travaillent 20 minutes par jour à la maison.

1) La moyenne¹ :

La moyenne est notée \bar{x} . C'est la somme des valeurs divisée par le nombre de valeurs. Lorsque l'on a des effectifs pour les valeurs, il faut pondérer la moyenne : on multiplie chaque valeur par l'effectif, et on divise par l'effectif total (qui se calcule par $n_1 + n_2 + \dots + n_p$ si l'énoncé ne le donne pas).

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_p \cdot x_p}{n_1 + n_2 + \dots + n_p}$$

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1

$$\begin{aligned}\bar{x} &= \frac{1 \cdot 5 + 4 \cdot 10 + 4 \cdot 15 + 3 \cdot 20 + 7 \cdot 30 + 7 \cdot 40 + 3 \cdot 50 + 1 \cdot 70}{30} \\ &= \frac{875 \div 5}{30 \div 5} = \frac{175}{6} \approx 29,17. \text{ Ainsi, les élèves travaillent en moyenne} \\ &\underline{29,17 \text{ minutes par jour.}}\end{aligned}$$

1. https://www.youtube.com/watch?v=88_16UbkdZM

2) L'écart-type² :

L'écart-type est noté $\sigma(x)$. Il représente une mesure de dispersion autour de la moyenne. Plus $\sigma(x)$ est grand, plus les x_i sont dispersées.

$$\sigma(x) = \sqrt{\frac{n_1 \cdot (x_1 - \bar{x})^2 + n_2 \cdot (x_2 - \bar{x})^2 + \dots + n_p \cdot (x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}}$$

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1

On avait calculé la moyenne qui valait $\frac{175}{6}$. Le calcul donne $\sigma(x) = \sqrt{\frac{1 \cdot (5 - \frac{175}{6})^2 + 4 \cdot (10 - \frac{175}{6})^2 + \dots + 1 \cdot (70 - \frac{175}{6})^2}{30}} \approx 15,17$.

Remarque : on définit également la variance $V(x)$, c'est le carré de l'écart-type ($V(x) = \sigma(x)^2$ ou $\sigma(x) = \sqrt{V(x)}$).

2. <https://www.youtube.com/watch?v=CiFoBkipJQk>.

I/ Caractéristiques d'une série

Remarque importante : quand on a une série donnée par classes d'intervalles, on prend pour valeur d'une classe la valeur centrale de la classe pour faire les calculs.

Exemple similaire à précédemment, où on a demandé aux élèves dans quelle tranche leur travail journalier se situait (toujours 30 élèves) :

Temps	[0 ; 15[[15 ; 30[[30 ; 45[[45 ; 60[[60 ; 75[
Effectif	5	7	14	3	1

Pour la tranche [0; 15[on utiliserait la valeur 7,5, pour la tranche [15; 30[on utiliserait la valeur 22,5, etc.

$\bar{x} = \frac{5 \cdot 7,5 + 7 \cdot 22,5 + 14 \cdot 37,5 + 3 \cdot 52,5 + 1 \cdot 67,5}{30} = 31,5$. Ainsi, les élèves travaillent en moyenne 31,5 minutes par jour.

$$\sigma(x) = \sqrt{\frac{5 \cdot (7,5 - 31,5)^2 + 7 \cdot (22,5 - 31,5)^2 + \dots + 1 \cdot (67,5 - 31,5)^2}{30}}$$

$\approx 14,80$.



Moyenne et écart-type

Pour une série statistique prenant p différentes valeurs x_1, x_2, \dots, x_p (les x_i) et où chaque valeur x_i a pour effectif n_i , les formules pour la moyenne et l'écart-type sont les suivantes :

$$\bar{x} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_p \cdot x_p}{n_1 + n_2 + \dots + n_p}$$

$$\sigma(x) = \sqrt{\frac{n_1 \cdot (x_1 - \bar{x})^2 + n_2 \cdot (x_2 - \bar{x})^2 + \dots + n_p \cdot (x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}}$$

Calculatrice : tableur ; menu statistiques, calcul statistique, statistiques à 1 variable.

3) La médiane³ :

La médiane sépare la série en deux parties de même effectif : au moins 50% des valeurs qui sont inférieures ou égales à la médiane, et au moins 50% des valeurs qui sont supérieures ou égales à la médiane.

Pour calculer la médiane d'une série de n nombres, on commence par ordonner les valeurs de manière croissante :

$$u_1 \leq u_2 \leq \dots \leq u_n$$

- si n est impair, c'est la valeur centrale : celle de rang $\frac{n+1}{2}$
Ex. : 2, 5, 7, 8, 9 : la valeur de rang $\frac{5+1}{2} = 3$: c'est 7
- si n est pair, c'est le nombre au milieu des deux valeurs centrales : la demi-somme des valeurs de rang $\frac{n}{2}$ et $\frac{n}{2} + 1$
Ex. : 2, 5, 7, 8, 9, 12 : la demi-somme des valeurs de rang $\frac{6}{2} = 3$ et $\frac{6}{2} + 1 = 4$: c'est $\frac{7+8}{2} = 7,5$

3. <https://www.youtube.com/watch?v=g1OCTw--VYQ>

Médiane avec effectifs :

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1
n_i cum.	1	5	9	12	19	26	29	30

Ici on a 30 valeurs, la médiane est donc la demi-somme des valeurs $\frac{30}{2} = 15$ et $\frac{30}{2} + 1 = 16$. Où sont ces valeurs ? Pour le savoir, on peut construire le tableau des effectifs cumulés.

On lit que les valeurs de rang 15 et 16 sont toutes les deux égales à 30, donc la médiane vaut $\frac{30+30}{2} = 30$. Il y a au moins 50% des élèves qui travaillent 30 minutes ou moins, et il y a au moins 50% des élèves qui travaillent 30 minutes ou plus.

4) Les quartiles⁴ :

Le **1^{er} quartile Q1** (**3^{ème} quartile Q3**) : la plus petite valeur de la série supérieure ou égale à au moins **25%** (**75%**) des valeurs.

Le rang de **Q1** (**Q3**) est le premier entier supérieur ou égal à $\frac{n}{4}$ ($\frac{3n}{4}$), c'est-à-dire **25%** (**75%**) de p .

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1
n_i cum.	1	5	9	12	19	26	29	30

- $\frac{30}{4} = 7,5$ donc Q1 est la 8e valeur.
- $\frac{3 \times 30}{4} = 22,5$ donc Q3 est la 23e valeur.

4. <https://www.youtube.com/watch?v=Yjh-9nMVmEw>,
<https://www.youtube.com/watch?v=IjsDK00Dwlw>

4) Les quartiles⁴ :

Le **1^{er} quartile Q1** (**3^{ème} quartile Q3**) : la plus petite valeur de la série supérieure ou égale à au moins **25%** (**75%**) des valeurs.

Le rang de **Q1** (**Q3**) est le premier entier supérieur ou égal à $\frac{n}{4}$ ($\frac{3n}{4}$), c'est-à-dire **25%** (**75%**) de n .

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1
n_i cum.	1	5	9	12	19	26	29	30

- $\frac{30}{4} = 7,5$ donc Q1 est la 8e valeur. C'est 15.
- $\frac{3 \times 30}{4} = 22,5$ donc Q3 est la 23e valeur.

4. <https://www.youtube.com/watch?v=Yjh-9nMVmEw>,
<https://www.youtube.com/watch?v=IjsDK00Dwlw>

4) Les quartiles⁴ :

Le **1^{er} quartile Q1** (**3^{ème} quartile Q3**) : la plus petite valeur de la série supérieure ou égale à au moins **25%** (**75%**) des valeurs.

Le rang de **Q1** (**Q3**) est le premier entier supérieur ou égal à $\frac{n}{4}$ ($\frac{3n}{4}$), c'est-à-dire **25%** (**75%**) de p .

x_i	5	10	15	20	30	40	50	70
n_i	1	4	4	3	7	7	3	1
n_i cum.	1	5	9	12	19	26	29	30

- $\frac{30}{4} = 7,5$ donc Q1 est la 8e valeur. C'est 15.
- $\frac{3 \times 30}{4} = 22,5$ donc Q3 est la 23e valeur. C'est 40.

4. <https://www.youtube.com/watch?v=Yjh-9nMVmEw>,
<https://www.youtube.com/watch?v=IjsDK00Dwlw>



Médiane et quartiles

Pour une série statistique comprenant n valeurs (pas forcément différentes), on calcule la médiane et les quartiles de la manière suivante :

- Médiane :
 - n impair : valeur centrale ; rang $\frac{n+1}{2}$
 - n pair : au milieu des deux valeurs centrales ; rang $\frac{n}{2}$ et $\frac{n+1}{2}$
(dans ce cas, ce n'est pas forcément une valeur de la série)
- Q1 : son rang est le premier entier $\geq \frac{n}{4}$
- Q3 : son rang est le premier entier $\geq \frac{3n}{4}$

Calculatrice : tableur ; menu statistiques, calcul statistique, statistiques à 1 variable.

1) Le diagramme en bâton :

Pour chaque valeur du critère, on dessine un bâton dont la hauteur est l'effectif de cette valeur.

(exemples dans l'exercice 11)

2) L'histogramme :

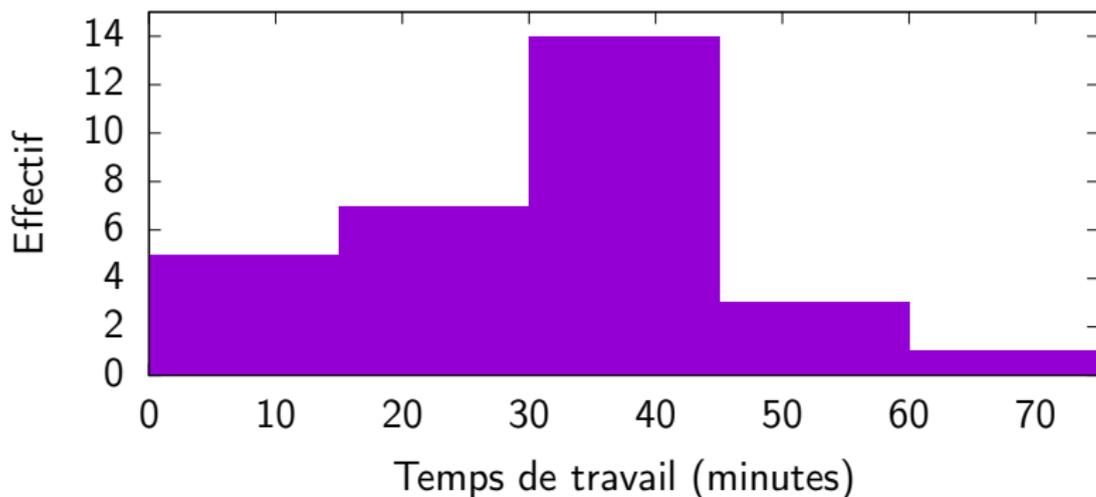
On regroupe les valeurs par intervalle, et on peut alors tracer un histogramme. Si les intervalles sont de même amplitude (ça sera toujours le cas pour nous), c'est comme un diagramme en bâtons, sauf que les bâtons sont collés.

(exemples dans les exercices 5 et 12, et à la diapositive suivante)

II/ Représentations graphiques

Dans l'exemple précédent (où on a demandé aux 30 élèves dans quelle tranche leur travail journalier se situait), voici l'histogramme associé :

Temps	[0 ; 15[[15 ; 30[[30 ; 45[[45 ; 60[[60 ; 75[
Effectif	5	7	14	3	1



3) La boîte à moustaches⁵ :

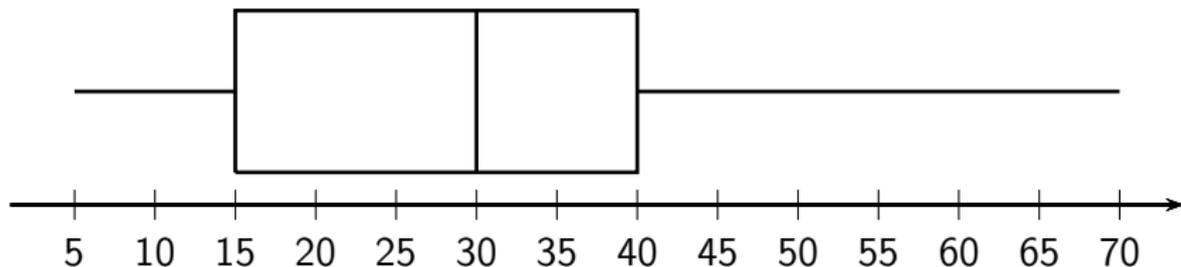
C'est le diagramme qu'il s'agit de bien maîtriser pour cette année. La boîte à moustache est un diagramme représentant les différentes caractéristiques que nous avons déjà calculées :

- il faut une échelle (un axe gradué horizontal, qui couvre les valeurs minimale et maximale de la série)
- ensuite, on trace une boîte rectangulaire qui va de $Q1$ à $Q3$, avec une ligne verticale à la médiane
- enfin, on trace deux traits horizontaux : le premier de la gauche du rectangle jusqu'à la valeur minimale, l'autre de la droite du rectangle jusqu'à la valeur maximale

5. <https://www.youtube.com/watch?v=1a7c0Yf8VyM>

Exemple sur notre série à 30 valeurs, où on avait calculé précédemment (voir diapos 9 et 10) :

- valeur minimale : 5
- valeur maximale : 70
- Q1 : 15
- Q3 : 40
- médiane : 30



1) Ajout d'un nombre :

On ajoute à toutes les valeurs d'une série $x_i = x_1, x_2, \dots, x_p$ un nombre a , et on nomme la nouvelle série y :

$$y_1 = x_1 + a, y_2 = x_2 + a, \dots, y_p = x_p + a$$

- $\bar{y} = \bar{x} + a$ (la médiane et les quartiles augmentent de a également)
- $\sigma(y) = \sigma(x)$ (l'écart inter-quartile ne change pas non plus)

2) Multiplication par un nombre :

On multiplie toutes les valeurs d'une série $x = x_1, x_2, \dots, x_p$ par un nombre m , et on nomme la nouvelle série z :

$$z_1 = x_1 \times m, z_2 = x_2 \times m, \dots, z_p = x_p \times m$$

- $\bar{z} = \bar{x} \times m$ (la médiane et les quartiles sont multipliés par m également)
- $\sigma(z) = \sigma(x) \times m$ (l'écart inter-quartile est multiplié par m également)