Statistiques

1. Moyenne et écart-type

Compléments sur la moyenne

1.1. Moyenne pondérée

Exemple avec des notes :
Dans le tableau suivant sont regroupées les notes obtenues par les élèves d'une seconde lors d'un contrôle

4	5	6	6	6	8	8	9	10	11	11	11	12	12	12	12	13
13	14	14	16	16	16	16	16	16	16	17	17	17	17	19	19	19

La série statistique est l'ensemble des notes collectées. La population est l'ensemble des élèves de seconde .

L'effectif total est le nombre d'élèves de la classe, à savoir 34. Les valeurs extrêmes sont 4 et 19.

Pour une meilleure lisibilité et pour simplifier l'étude, on peut commencer par compter le nombre d'individus ayant obtenu chaque note :

Note	4	5	6	8	9	10	11	12	13	14	16	17	19
Effectif	1	1	3	2	1	1	3	4	2	2	7	4	3
Fréquence à 10^{-2} près	0,03	0,03	0,09	0,06	0,03	0,03	0,09	0,12	0,06	0,06	0,21	0,12	0,09

On considère une série statistique donnée par le tableau suivant :

Valeur	x_1	x_2	x_3	$\dots$	x_{p-1}	x_p
Effectif	n_1	n_2	n_3	$\dots$	n_{p-1}	n_p
Fréquence	f_1	f_2	f_3	$\dots$	f_{p-1}	f_p

Définition La **moyenne** de cette série statistique est le réel noté $\overline{x}$ défini par

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{n_1 + n_2 + \ldots n_p} = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$$

en notant $N = n_1 + n_2 + \ldots + n_p$ l'effectif total de la série.

Propriété On peut également calculer la moyenne à l'aide des fréquences :

$$\overline{x} = x_1f_1 + x_2f_2 + \ldots + x_pf_p.$$

Exemple avec les notes
Avec les données de l'exemple , la moyenne de la classe est :

$$\overline{x} = \frac{1\times4+1\times5+3\times6+2\times8+1\times9+1\times10+3\times11+4\times12+2\times13+2\times14+7\times16+4\times17+3\times19}{34} = \frac{434}{34}$$

$$\overline{x} = \frac{1}{34}\times4+\frac{1}{34}\times5+\frac{3}{34}\times6+\frac{2}{34}\times8+\frac{1}{34}\times9+\frac{1}{34}\times10+\frac{3}{34}\times11+\frac{4}{34}\times12+\frac{2}{34}\times13+\frac{2}{34}\times14+\frac{7}{34}\times16+\frac{4}{34}\times17+\frac{3}{34}\times19$$

Une valeur approchée de $\overline{x}$ à 10^{-2} près est 12,76.

1.2. Linéarité de la moyenne

Exemple avec des notes :

Multiplier les notes de chaque élève par 1,5, que devient la moyenne de la classe ?

Ajouter 2 points à la moyenne de chaque élève, que devient la moyenne de la classe ?

Propriété Soit S une série statistique de données x_i de moyenne $\overline{x}$. Pour tout réel m et p, la nouvelle série $y_i = mx_i + p$ (obtenue en multipliant chaque valeur par m et en ajoutant p à chacune des valeurs) a pour moyenne $\overline{y} = m\overline{x} + p$.

Démonstration page 315.

1.3. Ecart-type

On considère une série statistique S, de moyenne $\overline{x}$, donnée par le tableau suivant :

Valeur	x_1	x_2	x_3	$\dots$	x_{p-1}	x_p
Effectif	n_1	n_2	n_3	$\dots$	n_{p-1}	n_p
Fréquence	f_1	f_2	f_3	$\dots$	f_{p-1}	f_p

Définition

- La **variance**, notée V, est la moyenne des carrés des écarts des valeurs avec la moyenne soit : $V = \frac{\sum_{i=1}^{p} n_i(x_i-\overline{x})^2}{\sum_{i=1}^{p} n_i} = \frac{n_1(x_1-\overline{x})^2+\ldots+n_p(x_p-\overline{x})^2}{n_1+\ldots+n_p} = f_1(x_1-\overline{x})^2+\ldots+f_p(x_p-\overline{x})^2$
- L'**écart-type** de S, noté σ est la racine carrée de la variance : $\sigma = \sqrt{V}$. (de même unité que les x_i.)

Exemple avec les notes

La variance vaut :

$$V = \frac{1\times\left(4-\frac{434}{34}\right)^2+1\times\left(5-\frac{434}{34}\right)^2+3\times\left(6-\frac{434}{34}\right)^2+2\times\left(8-\frac{434}{34}\right)^2+1\times\left(9-\frac{434}{34}\right)^2+1\times\left(10-\frac{434}{34}\right)^2}{34}$$

$$+\frac{3\times\left(11-\frac{434}{34}\right)^2+4\times\left(12-\frac{434}{34}\right)^2+2\times\left(13-\frac{434}{34}\right)^2+2\times\left(14-\frac{434}{34}\right)^2+7\times\left(16-\frac{434}{34}\right)^2+4\times\left(17-\frac{434}{34}\right)^2}{34}$$

$$+\frac{3\times\left(19-\frac{434}{34}\right)^2}{34} = \frac{5254}{289} \approx 18,18.$$

L'écart-type vaut :

$\sigma = \sqrt{\frac{5254}{289}} \approx 4,26$

Propriété L'écart-type mesure un écart moyen à la moyenne d'une série : pour deux séries de même moyenne, celle dont l'écart-type est le plus faible a ses données plus homogènes.

Remarque Le couple $(\overline{x}, \sigma)$ permet de résumer la série. Il utilise toutes les valeurs de la série et est donc sensible aux valeurs extrêmes.

2. Médiane et écart interquartile

2.1. Médiane

Rappel : Définition La **médiane** M d'une série statistique est un réel qui partage cette série en deux parties telles que :

- Au moins 50 % des valeurs sont inférieures ou égales à la médiane ;
- Au moins 50 % des valeurs sont supérieures ou égales à la médiane.

Propriété En pratique, on adopte la démarche suivante pour déterminer la médiane M d'une série statistiques d'effectif total N :

- On range d'abord les N valeurs du caractère par ordre croissant.
- Si N est pair, $N = 2k$, M est la moyenne des deux valeurs « centrales » de la série, c'est-à-dire $M = \frac{x_k + x_{k+1}}{2}$.
- Si N est impair, $N = 2k+1$, M est la valeur centrale de la série, c'est-à-dire x_{k+1}.

Exemple

Note	4	5	6	8	9	10	11	12	13	14	16	17	19
Effectif	1	1	3	2	1	1	3	4	2	2	7	4	3
Effectifs Cumulés Croissants	1	2	5	7	8	9	12	16	18	20	27	31	34

Dans l'exemple, l'effectif total est 34, c'est-à -dire pair. $34 = 2 \times 17$.

La médiane est donc la moyenne des deux valeurs centrales de la série, à savoir les 17$^{\text{e}}$ et 18$^{\text{e}}$ valeurs. Donc $M = \dfrac{13+13}{2} = 13$, ce qui signifie qu'au moins la moitié des notes est inférieure ou égale à 13, et qu'au moins la moitié des notes est supérieure ou égale à 13 .

Définition On considère une série statistique.

- Le premier **quartile** Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des données soient inférieures ou égales à Q_1.
- Le troisième **quartile** Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des données soient inférieures ou égales à Q_3.

Exemple
On considère toujours les données de l'exemple des notes.

- $34 \times \dfrac{25}{100} = 8,5$ donc le premier quartile Q_1 de la série est la 9$^{\text{e}}$valeur, d'où $Q_1 = 10$, ce qui signifie qu'au moins un quart des notes sont inférieures ou égales à 10 (en réalité 9 notes, soit environ 26 %).
- $34 \times \dfrac{75}{100} = 25,5$ donc le troisième quartile Q_3 de la série est la 26$^{\text{e}}$ valeur, d'où $Q_3 = 16$, ce qui signifie qu'au moins trois quarts des notes sont inférieures ou égales à 16 (en réalité 27 notes, soit environ 79 %).

Remarque
Le fait que le partage théorique en 25 %, 50 % et 75 % de la série statistique à l'aide des indicateurs Q_1, M et Q_3 ne soit pas tout à fait exact provient du fait que la série comporte des valeurs identiques. Ce phénomène a tendance à s'amoindrir lors d'une étude sur une population plus importante avec un caractère dont les modalités sont plus disparates.

2.11 Écart interquartile

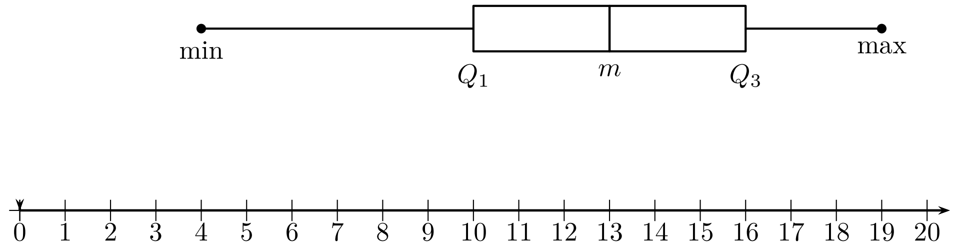
Définition On considère une série statistique de premier quartile Q_1 et de troisième quartile Q_3.

- On appelle **intervalle interquartile** l'intervalle $[Q_1\ ;\ Q_3]$.
- On appelle **écart interquartile** la différence $Q_3 - Q_1$.

Exemple
Dans l'exemple, l'intervalle interquartile est $[10\ ;\ 16]$ et l'écart interquartile est $Q_3 - Q_1 = 16 - 10 = 6$.

On peut représenter graphiquement les valeurs extrêmes, les quartiles et la médiane par un diagramme en boite, appelé aussi boite à moustaches, conçue de la manière suivante :

- au centre une boite allant du premier au troisième quartile, séparée en deux par la médiane ;
- de chaque côté une moustache allant du minimum au premier quartile pour l'une, et du troisième quartile au maximum pour l'autre.



Les diagrammes en boîte permettent une interprétation visuelle et rapide de la dispersion des séries statistiques. Ils permettent également d'apprécier des différences entre des séries (lorsqu'elles ont des ordres de grandeurs comparables).