Chapitre 11 : Statistiques

Table des matières

1 Contenu

- Indicateurs de tendance centrale d'une série statistique : moyenne pondérée.
- Linéarité de la moyenne.
- Indicateurs de dispersion : écart interquartile, écart type.

Dans tout ce chapitre, on considèrera les 3 séries statistiques suivantes :

Série A :
Notes obtenues à un contrôle dans une classe de 30 élèves :

2 – 3 – 3 – 4 – 5 – 6 – 6 – 7 – 7 – 7 – 8 – 8 – 8 – 8 – 8 – 9 – 9 – 9 – 9 – 9 – 9 – 10 – 10 – 11 – 11 – 11 – 13 – 13 – 15 – 16

Série B :
Salaires en euros des employés d'une entreprise :

Salaires	[900; 1200]	[1200; 1400]	[1400; 1600]	[1600; 1800]	[1800; 2000]	[2000; 2400]	**TOTAL**
Effectif	30	30	60	80	40	40	280

Série C :
Proportion d'adhérents à un club sportif dans différentes sections :

- 17% jouent au handball,
- 25% houent au rugby,
- 58% jouent au tennis.

2 Méthodes de représentation

2.1 Vocabulaire

La **population** est l'ensemble des individus sur lesquels portent l'étude statistique. (Par exemple la classe de BTS domotique, la popolation féminine, les fonctionnaires . . .) dont chaque élément est appelé **individu**.

Un **échantillon** est une partie de la population considérée.

Le **caractère** (ou **variable**) d'une série statistique est une propriété étudiée sur chaque individu :

- Lorsque le caractère ne prend que des valeurs (ou **modalités**) numériques, il est **quantitatif** :
  - **discret** s'il ne peut prendre que des valeurs isolées (notes, âge . . .)
  - **continu** dans le cas contraire (poids, taille . . .). Dans ce cas on effectue souvent un regroupement des valeurs par **classes**.
- Sinon, on dit qu'il est **qualitatif** (couleur des yeux, sport pratiqué . . .): les modalités ne sont pas des nombres.

A chaque valeur (ou classe) est associée un **effectif** n: c'est le nombre d'individus associés à cette valeur.
Faire des **statistiques**, c'est recueillir, organiser, synthétiser, représenter et exploiter des données, numériques ou non, dans un but de comparaison, de prévision, de constat . . .
Les plus gros "consommateurs" de statistiques sont les assureurs (risques d'accidents, de maladie des assurés), les médecins (épidémiologie), les démographes (populations et leur dynamique), les économistes (emploi, conjoncture économique), les météorologues . . .

2.2 Tableaux

Définition:
On considère une série statistique X à caractère quantitatif, dont les p valeurs sont données par $x_1, x_2, \ldots, x_p$ d'effectifs associés $n_1, n_2, \ldots, n_p$ avec $n_1 + n_2 + ... + n_p = N$.

- A chaque valeur (ou classe) est associée une fréquence f_i: c'est la proportion d'individus associés à cette valeur.
- $f_i = \dfrac{n_i}{N}$ est un nombre compris entre 0 et 1, que l'on peut écrire sous forme de pourcentage.

Exemple:
On peut représenter la **série A** par un tableau d'effectifs, et le compléter par la distribution des fréquences :

Notes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Eff.	0	1	2	1	1	2	3	5	6	2	3	0	2	0	1	1	0	0	0
Fréq. en %	0	3	7	3	3	7	10	17	20	7	10	0	7	0	3	3	0	0	0

! Remarque

On peut vérifier que la somme des fréquences est égale à 1 (ou à 100 si on les exprime en pourcentages).

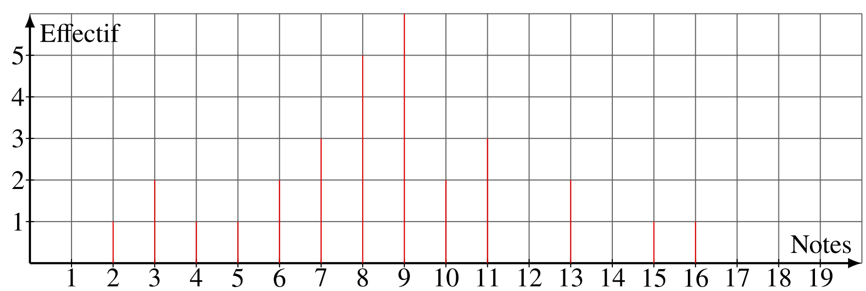
On peut aussi faire un regroupement par classe, ce qui rend l'étude moins précise, mais qui permet d'avoir une vision plus globale.

Exemple:
Toujours pour la **série A**, si on regroupe les données par classes d'amplitude 5 points, on obtient :

Notes	[0 ; 5 [	[5 ; 10 [	[10 ; 15 [	[15 ; 20 [	**total**
Effectif	4	17	7	2	30
Fréquence	$0,13$	$0,57$	$0,23$	$0,07$	1

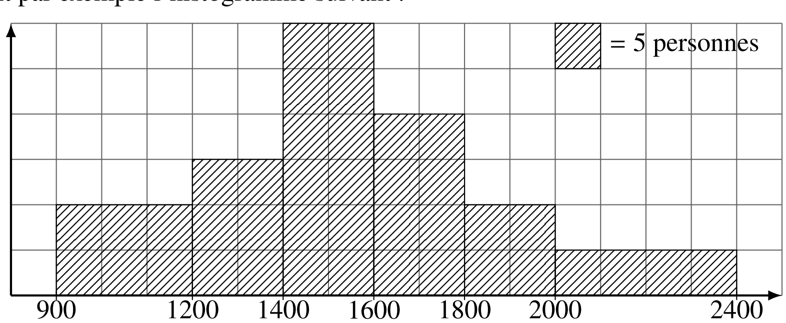
2.3 Graphiques

Lorsque le caractère étudié est **quantitatif et discret**, on peut représenter la série statistique étudiée par un **diagramme en bâtons**: la hauteur de chaque bâton est alors proportionnelle à l'effectif (ou à la fréquence) associé à chaque valeur.
Exemple:
Voici le diagramme en bâtons représentant la série des notes de la **série A** :



Lorsque le caractère étudié est **quantitatif et continu**, et lorsque les modalités sont regroupées en classes, on peut représenter la série par un **histogramme**: l'aire de chaque rectangle est alors proportionnelle à l'effectif (ou à la fréquence) associée à chaque classe.
Lorsque les classes ont la même **amplitude**, c'est la hauteur qui est proportionnelle à l'effectif.

Exemple:
Pour la **série B**, on obtient par exemple l'histogramme suivant :



Enfin, lorsque le caractère est **qualitatif**, on peut représenter la série par :

- **Un diagramme circulaire** (" camemberts") :
  La mesure de chaque secteur angulaire est proportionnelle à l'effectif associé.

- **Un diagramme en tuyaux d'orgue :**
  Chaque classe est représentée par un rectangle de même largeur et de longueur proportionnelle à l'effectif, donc à la fréquence.

- **Un diagramme en bandes :**
  Chaque classe est représentée par un rectangle de même largeur et de longueur proportionnelle à l'effectif, donc à la fréquence.

Exemple:
Diagrammes de la **série C**.

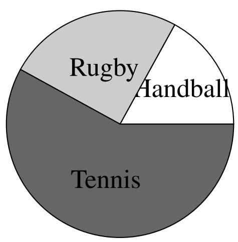


Diagramme circulaire

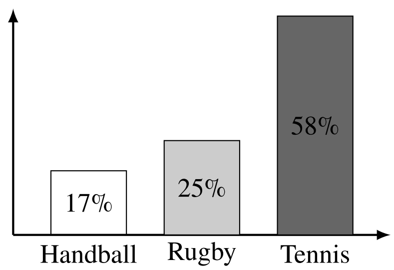


Diagramme en tuyau d'orgue

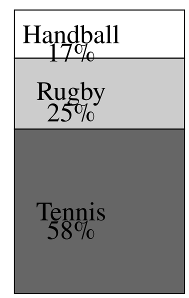


Diagramme en bande

3 Caractéristiques de positions

3.1 Moyenne

Définition:
Si une série statistique x_1 ; ... ; x_p prend n_i fois la valeur x_i, pour $i \in \{1, ..., p\}$, alors on a la moyenne pondérée qui vaut

$$\bar{x} = \frac{n_1 \times x_1 + \ldots + n_p \times x_p}{n_1 + \ldots + n_p}$$

Exemple:

- Dans la **série A**, la moyenne du contrôle est égale à $\overline{m} = \dfrac{254}{30} \approx 8,47$.
- Dans la **série B**, une estimation du salaire moyen est donné par : $\overline{S} = \dfrac{460500}{280} \approx 1644,64$.

! Remarque

On peut aussi calculer une moyenne à partir de la distribution de fréquences : $\overline{x} = f_1x_1 + f_2x_2 + \cdots + f_px_p$

Propriété: *linéarité de la moyenne*

- Si on ajoute (ou soustrait) un même nombre k à toutes les valeurs d'une série, alors la moyenne de cette série se trouve augmentée (resp. diminuée) de k.
- Si on multiplie (ou divise) par un même nombre non nul k toutes les valeurs d'une série, alors la moyenne de cette série se trouve multipliée (resp. divisée) par k.

Démonstration:

$$\overline{x+k} = \frac{n_1 \times (x_1 + k) + \ldots + n_p \times (x_p + k)}{n_1 + \ldots + n_p} = \frac{n_1 \times x_1 + \ldots + n_p \times x_p}{n_1 + \ldots + n_p} + \frac{n_1 \times k + \ldots + n_p \times k}{n_1 + \ldots + n_p} = \bar{x} + k$$

Exemple:

- Si on ajoute $1,5$ points à chaque note du contrôle, alors la moyenne de classe devient $\overline{m} = 8,47 + 1,5 = 9,97$.
- Si on augmente chaque note de 10%, cela revient à multiplier chaque note par $1,1$, ce qui donne $\overline{m} = 8,47 \times 1,1 = 9,32$.

3.2 Médiane et quartiles

Définition:
On appelle médiane Me d'une série statistique à n éléments, la plus petite valeur de la série qui est supérieure ou égale à 50% des valeurs de la série.

- Si n est impair alors la médiane est la valeur centrale.
- Si n est pair alors la médiane est la moyenne des deux valeurs centrales.

! Remarque

La médiane peut ne pas être égale à une des valeurs de la série statistique.

Exemple:
On relève les pointures de 10 personnes et on obtient :
45 ; 39 ; 42 ; 48 ; 41 ; 42 ; 43 ; 42 ; 38 ; 49
On réoganise ces données dans l'ordre croissant et on a donc la médiane : 42.

! Remarque

La médiane n'est pas affectée par les valeurs extrêmes.
Si on remplace le 49 par 54 dans l'exemple précédent le résultat ne change pas.

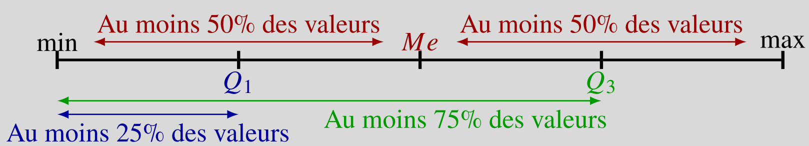
Exemple:
On souhaite calculer la médiane de la **série A**.

- Pour cela, on commence par remplir le tableau des effectifs cumulés croissants :

Notes	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
Eff.	0	1	2	1	1	2	3	5	6	2	3	0	2	0	1	1	0	0	0
ECC.	0	1	3	4	5	7	10	15	21	23	26	26	28	28	29	30	30	30	30

- Ensuite, l'effectif étant de 30, on chosit la moyenne entre la $15^{ième}$ et la $16^{ième}$ note.
  On obtient $Med = \dfrac{8+9}{2} = 8,5$.
- Ce qui signifie que la moitié des notes est inférieure ou égale à $8,5$, et que l'autre moitié des notes est supérieure ou égale à $8,5$.

Définition:
On appelle 1° quartile Q_1 (respectivement 3° quatile Q_3) la plus petite valeur de la série qui est supérieure ou égale à 25% (respectivement 75%) des valeurs de la série.



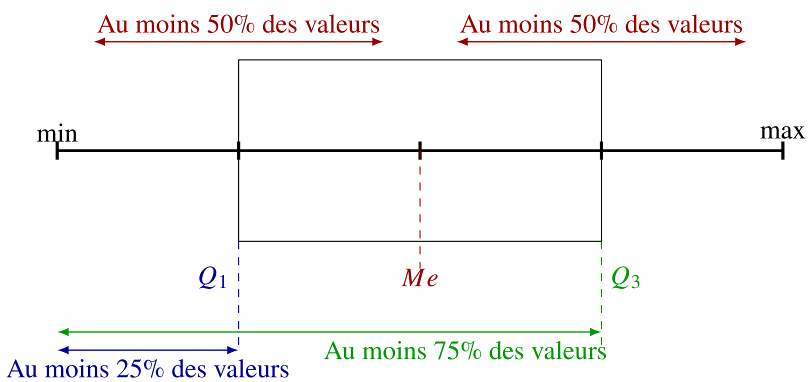
Exemple:
On reprends l'exemple avec les pointures des 10 personnes
On obtient donc $Q_1 = 40$ et $Q_3 = 46,5$.

Exemple:

- Pour la **série A**, la calculatrice nous donne $Q_1 = 7$, $Méd = 8,5$ et $Q_3 = 10$.
- Pour la **série B**, on trouve $Q_1 = 1500$, $Med = 1700$ et $Q_3 = 1900$.

Il est possible de représenter graphiquement les résultats de médiane et de quartiles d'une série statistique par un diagramme en boîte.

Les diagrammes en boîte, ou "boîte à moustaches", permettent de visualiser rapidement des caractéristiques de position (voir figure ci-dessous).



4 Caractéristiques de dispersion

4.1 Ecart interquartiles

Définition:
On appelle écart interquartile la donnée $EQ = Q_3 - Q_1$.

Exemple:
On reprends l'exemple avec les pointures des 10 personnes
On obtient donc $EQ = Q_3 - Q_1 = 4$.

! Remarque

Plus EQ est petit, plus la médiane est représentative de la série statistique.

4.2 Ecart-type d'une série statistique

Définition:
Si une série statistique x_1 ; ... ; x_p prend n_i fois la valeur x_i, pour $i \in \{1, \ldots, p\}$, alors on définit respectivement la variance et l'écart-type par:

$$V = \frac{n_1(x_1 - \bar{x})^2 + \ldots + n_p(x_p - \bar{x})^2}{n_1 + \ldots + n_p} \quad \text{et} \quad \sigma = \sqrt{V}$$

! Remarque

L'écart type représente la distance moyenne entre les valeurs de la série et la moyenne.
Au plus il est petit, au plus la moyenne est représentative de la série statistique.

Exercice:
Calculer la variance et l'écart type de la série statistique des pointures des 10 personnes.

5 Exercice bilan

Une classe de 15 élèves a obtenu les résultats suivants (sur 20) au dernier devoir de mathématiques.

14 – 6 – 19 – 4 – 15 – 17 – 18 – 17 – 13 – 10 – 11 – 4 – 6 – 17 – 12

1. Calculer la moyenne et l'écart-type de cette série statistique.
2. Calculer la médiane et l'écart interquartile de cette série statistique.
3. Interpréter ce résultat.