

*Outils mathématiques de gestion*

*Statistiques*

- 1 Vocabulaire
- 2 Caractéristiques de position
- 3 Caractéristiques de dispersion
- 4 Boîtes à moustaches, box-plot
  - Quartiles
  - Écart interquartile

Une **unité statistique**, ou **individu statistique**, est un membre d'une **population** sur laquelle porte une **enquête statistique** (un recueil d'informations).

Le nombre d'individus ou d'unités statistiques appartenant à la population est appelé **taille** de la population ou **effectif total**.

On observe (ou on mesure) sur chaque individu une **variable** (ou **caractère**, ou encore **attribut**).

Les valeurs (numériques ou non) que peut prendre une variable statistique s'appellent des **modalités**.

Le nombre d'individus qui présentent une modalité de la variable statistique est appelé **effectif** de cette modalité.

La **fréquence** d'une modalité est l'effectif de cette modalité divisé par l'effectif total (c'est une proportion!).

Une enquête peut porter sur toute la population (on peut alors parler de **recensement**) ou sur une partie de la population. Cette partie s'appelle **échantillon** ou **groupe observé**.

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est l'effectif totale des élèves de la classe ?

Quel est le caractère (la variable) étudié ?

Quelles sont les modalités et leur effectif ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est l'effectif totale des élèves de la classe ?

Il y a 21 élèves au total.

Quel est le caractère (la variable) étudié ?

Quelles sont les modalités et leur effectif ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est l'effectif totale des élèves de la classe ?

Il y a 21 élèves au total.

Quel est le caractère (la variable) étudié ?

La note.

Quelles sont les modalités et leur effectif ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est l'effectif totale des élèves de la classe ?

Il y a 21 élèves au total.

Quel est le caractère (la variable) étudié ?

La note.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1



## Variables quantitatives

Une variable est **quantitative** si ses modalités sont des nombres.

Parmi les variables quantitatives, il y a les variables

- **discrètes** (nombre d'enfants dans un foyer, nombre de pièces d'un logement, âge d'une personne, ...) → ensemble dénombrable
- **continues** (taille d'une personne, température, ...) → intervalle

## Variables qualitatives

Une variable est **qualitative** si ses modalités ne sont pas des nombres.

Parmi les variables qualitatives, il y a les variables

- **ordinales** (qualité d'un vin, ...) → classement entre les modalités
- **nominales** (sexe d'une personne, votre groupe de TD, ...) → pas de classement entre les modalités

# Description de données univariées

Univarié = une variable.

## Variable qualitative, exemple 1

*Individus* : tous les mots d'un livre

*Variable* : classe grammaticale

*Modalités* : nom, pronom, adjectif, verbe, adverbe, déterminants, prépositions, conjonctions, interjections

## Variable qualitative, exemple 2

*Individus* : les t-shirts d'un magasin

*Variable* : taille

*Modalités* : XXXS, XXS, XS, S, M, L, XL, XXL, XXXL

On peut représenter les effectifs et les fréquences à l'aide de diagrammes en barres, diagrammes en secteurs, etc.

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est le caractère (la variable) étudié ?

La note.

Quel est le type de la variable ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8

Quel est le caractère (la variable) étudié ?

La note.

Quel est le type de la variable ?

C'est une variable quantitative (les modalités sont des nombres) et discrète (les valeurs possibles sont dénombrables).

# Description de données univariées

On observe un caractère quantitatif  $x$  sur  $n$  individus. Soient  $x_1, x_2, \dots, x_n$  les valeurs observées.

L'ensemble des nombres  $\{x_1, \dots, x_n\}$  est appelé **série statistique simple**.

On va noter  $v_1, v_2, \dots, v_k$  les différentes modalités,  $n_1, n_2, \dots, n_k$  les effectifs correspondants et  $f_1, f_2, \dots, f_k$  les fréquences correspondantes.

## Définitions

Pour  $i = 1, \dots, k$ , l'**effectif cumulé** de la modalité  $v_i$  est

$$N_i = n_1 + n_2 + \dots + n_i$$

et la **fréquence cumulée** est

$$F_i = f_1 + f_2 + \dots + f_i.$$

# Caractéristiques de position

Les caractéristiques de position nous renseignent sur l'ordre de grandeur des données.

## Moyenne

La **moyenne** de la série statistique est

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \cdots + x_n).$$

On peut aussi la calculer à l'aide des modalités et de leur effectif,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i v_i = \frac{n_1 v_1 + n_2 v_2 + \cdots + n_k v_k}{n_1 + \cdots + n_k}.$$

La moyenne est un *indicateur de tendance centrale*, elle permet de « résumer » la série statistique en une seule valeur.

Inconvénient : sensible aux valeurs extrêmes qui peuvent être aberrantes !

# Caractéristiques de position

## Médiane

On range nos  $n$  observations  $x_1, \dots, x_n$  dans l'ordre croissant.

- Si l'effectif  $n$  est impair, la **médiane** est l'observation centrale, celle qui occupe la position  $\frac{n+1}{2}$ .
- Si l'effectif est pair, on appelle **médiane** la demi-somme des deux observations centrales, celles qui occupent les rangs  $\frac{n}{2}$  et  $\frac{n}{2} + 1$ .

La médiane est un indicateur du **robuste**, de nombreuses observations peuvent varier sans que la médiane soit modifiée.

Interprétation de la médiane : 50% des observations occupent un rang qui précède celui de la médiane (mais cela ne signifie pas que 50% des observations ont une valeur inférieures ou égales à celle de la médiane!).



## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1
Effectifs cumulés	2	5	6	9	12	15	16	18	20	21

Quel est la moyenne ?

Quel est la médiane ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1
Effectifs cumulés	2	5	6	9	12	15	16	18	20	21

Quel est la moyenne ?

On a  $\bar{x} \simeq 10,3$ .

Quel est la médiane ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1
Effectifs cumulés	2	5	6	9	12	15	16	18	20	21

Quel est la moyenne ?

On a  $\bar{x} \simeq 10,3$ .

Quel est la médiane ?

L'effectif  $n = 21$  est impair, la médiane est donc l'observation au rang  $\frac{n+1}{2} = 11$  qui est 10.

# Caractéristiques de position

Les caractéristiques de position ne suffisent pas pour synthétiser les informations contenues dans un ensemble de données.

Elles doivent être complétées par des paramètres de dispersion qui quantifient globalement l'importance des écarts entre les observations ou encore leur dispersion autour d'une valeur centrale (moyenne ou médiane).

# Caractéristiques de dispersion

## Étendue

L'**étendue** d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la série.

$$e = x_{\max} - x_{\min}$$

## Variance et écart-type

La **variance** d'une série statistique est

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2,$$

l'**écart-type** est

$$\sigma = \sqrt{V}.$$

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1

Quel est l'étendue ?

Quel est la variance ?

Quel est l'écart-type ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1

Quel est l'étendue ?

On a  $e = 16 - 5 = 11$ .

Quel est la variance ?

Quel est l'écart-type ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1

Quel est l'étendue ?

On a  $e = 16 - 5 = 11$ .

Quel est la variance ?

On a  $V \simeq 10,1$

Quel est l'écart-type ?



## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1

Quel est l'étendue ?

On a  $e = 16 - 5 = 11$ .

Quel est la variance ?

On a  $V \simeq 10,1$

Quel est l'écart-type ?

On a  $\sigma = \sqrt{V} \simeq 3,18$

# Quartiles

La médiane est une valeur qui sépare les observations en deux groupes de même effectif (50% et 50%).

On peut de même chercher des valeurs telles que 25% ou 75% des observations soient inférieures à notre valeur. Ce sont les **quartiles**.

## Premier quartile

On range nos  $n$  observations  $x_1, \dots, x_n$  dans l'ordre croissant.

- Si  $\frac{1}{4} \times n$  n'est pas un entier, le **premier quartile** est l'observation de rang  $\left\lfloor \frac{n}{4} \right\rfloor + 1$  (où  $\lfloor x \rfloor$  désigne la partie entière inférieure du nombre  $x$ ).
- Si  $\frac{1}{4} \times n$  est un entier, le **premier quartile** est la demi-somme des observations de rang  $\frac{n}{4}$  et  $\frac{n}{4} + 1$ .

# Quartiles

La médiane est une valeur qui sépare les observations en deux groupes de même effectif (50% et 50%).

On peut de même chercher des valeurs telles que 25% ou 75% des observations soient inférieures à notre valeur. Ce sont les **quartiles**.

## Troisième quartile

On range nos  $n$  observations  $x_1, \dots, x_n$  dans l'ordre croissant.

- Si  $\frac{3}{4} \times n$  n'est pas un entier, le **troisième quartile** est l'observation de rang  $\left\lfloor \frac{3n}{4} \right\rfloor + 1$  (où  $\lfloor x \rfloor$  désigne la partie entière inférieure).
- Si  $\frac{3}{4} \times n$  est un entier, le **troisième quartile** est la demi-somme des observations de rang  $\frac{3n}{4}$  et  $\frac{3n}{4} + 1$ .

# Écart interquartile

## Écart interquartile

On note  $q_1$  et  $q_3$  le premier troisième quartile d'une série statistique.  
L'**écart interquartile** est

$$EIQ = q_3 - q_1.$$

C'est un indicateur de dispersion.

On peut l'utiliser pour déterminer si des observations sont atypiques ou non.

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1
Effectifs cumulés	2	5	6	9	12	15	16	18	20	21

Quel est la médiane ?

On a  $m = 10$

Quels sont les premier et troisième quartiles ?

## Exemple

À la suite d'une interrogation, les élèves d'une classe ont obtenu les notes suivantes :

5, 10, 7, 13, 5, 9, 14, 15, 11, 16, 11, 10, 14, 7, 9, 10, 15, 9, 11, 7, 8.

Quelles sont les modalités et leur effectif ?

Notes	5	7	8	9	10	11	13	14	15	16
Effectifs	2	3	1	3	3	3	1	2	2	1
Effectifs cumulés	2	5	6	9	12	15	16	18	20	21

Quel est la médiane ?

On a  $m = 10$

Quels sont les premier et troisième quartiles ?

$\frac{21}{4} = 5,25$  et  $\frac{3 \times 21}{4} = 15,75$  ne sont pas entiers, donc  $q_1$  et  $q_3$  sont les observations de rang 6 et 16 respectivement,  $q_1 = 8$  et  $q_3 = 13$

# Boîte à moustaches

Une manière de représentation graphiquement les informations : la boîte à moustaches.

## Boîte à moustaches

La **boîte à moustaches** se construit de la manière suivante :

- Les limites de la boîte sont les premier et troisième quartiles.
- On indique la médiane à l'intérieur de la boîte.
- Les « moustaches » sont tracées jusqu'aux valeurs extrêmes.
- (On les arrête parfois à concurrence de 1,5 fois l'écart interquartile afin de prendre en compte les éventuelles « données aberrantes ». Ces points au-delà des moustaches sont parfois marqués par des astérisques.)