# SERIES STATISTIQUES A UNE VARIABLE

Dans tout le chapitre, on considère une *population*, c'est-à-dire un ensemble d'éléments dont chacun sera appelé *individu*. On cherche à étudier une partie de cette population (éventuellement cette population dans son ensemble), ce qu'on appelle un *échantillon*.
On notera *N* le nombre d'individus dans cet échantillon.
Les problèmes considérés dans ce chapitre sont les suivants :

(1) de quelles manières peut-on étudier une certaine propriété concernant les individus de cet échantillon (propriété que l'on appellera un *caractère* ou encore une *variable statistique*), et comment représenter cette étude selon le type de variable étudiée ?
(2) Comment caractériser la répartition de la variable statistique étudiée dans l'échantillon au moyen de *paramètres pertinents* permettant notamment d'en *mesurer la tendance centrale et la dispersion* autour de cette tendance ?

## I. Types de variables statistiques et modes de représentation

### 1. Les deux principaux types de variables statistiques

#### a) Variables qualitatives

Une variable est dite qualitative lorsqu'elle n'est pas numérique, c'est-à-dire lorsqu'elle ne peut pas être mesurée par des nombres. Exemples : couleur des yeux, sexe…
Remarque : affecter un numéro à chacune des *modalités* d'une variable qualitative n'en fait pas pour autant une variable numérique, car les opérations n'y ont aucune signification.
(Par exemple : 1 pour yeux bleus ; 2 pour yeux marrons ; 3 pour yeux verts. $1 + 2 = 3$ ?)

#### b) Variables quantitatives

Une variable est dite quantitative lorsque ses modalités sont numériques, c'est-à-dire lorsqu'elle peut être mesurée par des nombres. Il en existe deux catégories :

**Variables discrètes**
Une variable quantitative est dite discrète lorsqu'elle ne peut prendre qu'un ensemble fini ou infini dénombrable de valeurs ponctuelles, isolées.
Exemple : nombre d'enfants par femme.

**Variables continues**
Une variable quantitative est continue dans le cas contraire, c'est-à-dire lorsqu'elle prend ses valeurs dans des intervalles. Exemples : poids, taille d'un individu…

Remarque : la mesure elle-même discrétise une variable continue. On dit qu'on discrétise une variable continue lorsqu'on regroupe les valeurs qu'elle prend selon des intervalles déterminés, appelés *classes*.
Inversement, on peut regrouper par intervalles les valeurs d'une variable discrète.

### 2. Effectifs, fréquences

Soit $X$ une variable statistique (qualitative ou quantitative) se déclinant en $n$ modalités $(x_i)_{1\leq i\leq n}$ .
A tout $i \in [\![1; n]\!]$ correspond un nombre $n_i$ d'individus, appelé effectif associé à la modalité $x_i$ .

La somme de ces effectifs correspond au nombre d'individus dans l'échantillon : $\sum_{i=1}^{i=n} n_i = N$

On peut calculer la fréquence $f_i$ associée à chaque modalité $x_i$ : $f_i = \frac{n_i}{N}$

Par définition, $f_i \in [0; 1]$ et la somme des fréquences est égale à 1 : $\sum_{i=1}^{i=n} f_i = 1$

# 3. Modes de représentation d'une variable statistique

Selon le type de variable étudiée, certaines représentations de la distribution de ses effectifs ou de ses fréquences sont possibles, outre bien sûr un tableau récapitulatif de ces effectifs ou de ces fréquences.

## a) Modes de représentation d'une variable qualitative

Une variable qualitative peut être représentée par un diagramme circulaire, en tuyaux d'orgues ou en bandes, en utilisant dans chaque cas un principe de proportionnalité.
Exemple :

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

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

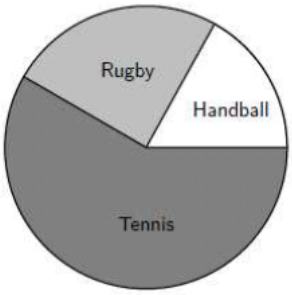



Diagramme circulaire

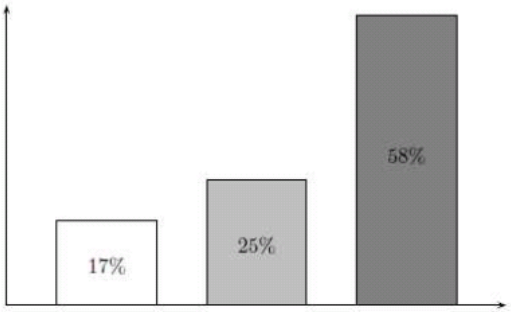



Diagramme en tuyau d'orgue

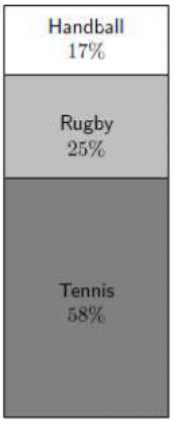



Diagramme en bandes

## b) Modes de représentation d'une variable quantitative

Une variable quantitative est représentable de diverses manières selon sa nature.

### Cas d'une variable discrète

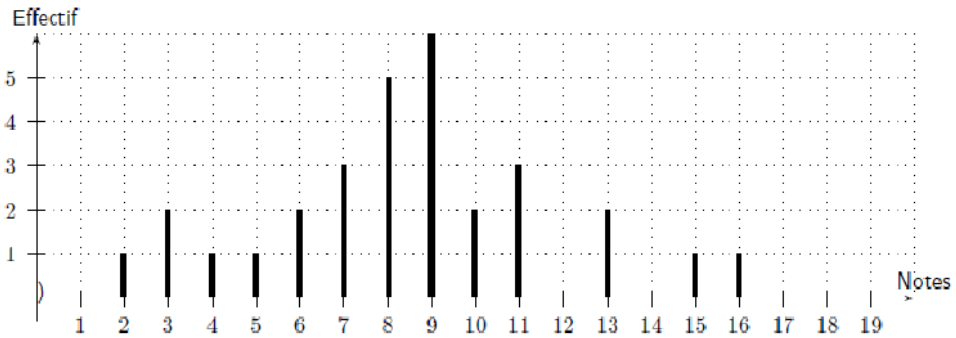
Lorsqu'elle est discrète, elle est représentable par un diagramme en bâtons, qu'il s'agisse des effectifs ou des fréquences (représentés en ordonnées).

Exemple :

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

| 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 |



## Cas d'une variable continue

Lorsque la variable est continue et regroupée par classes, on peut la représenter par un histogramme des effectifs ou des fréquences.

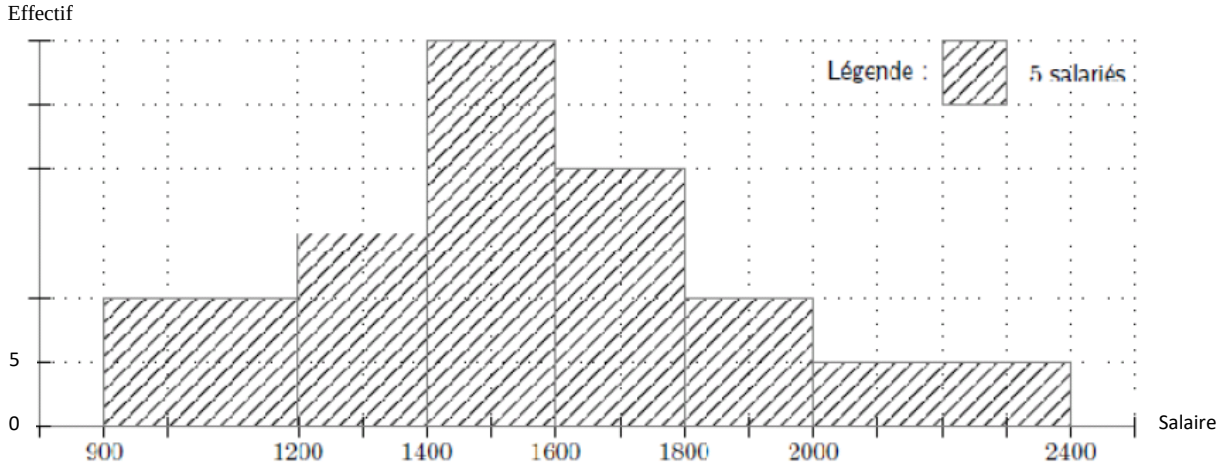
Exemple : Salaires en euros des employés d'une entreprise :

| Salaires | [900; 1200[ | [1200; 1400[ | [1400; 1600[ | [1600; 1800[ | [1800; 2000[ | [2000; 2400[ | TOTAL |
|---|---|---|---|---|---|---|---|
| Effectif | 30 | 30 | 60 | 40 | 20 | 20 | 200 |

On constate que les classes sont d'amplitude inégale, aussi, pour construire l'histogramme de cette série statistique, nous devons définir une unité d'amplitude commune et calculer les effectifs par unité d'amplitude (sous l'hypothèse de régularité de la distribution dans chaque classe).
Si l'on choisit 100€ comme unité d'amplitude commune :

| Salaires | [900; 1200[ | [1200; 1400[ | [1400; 1600[ | [1600; 1800[ | [1800; 2000[ | [2000; 2400[ | TOTAL |
|---|---|---|---|---|---|---|---|
| Effectif | 30 | 30 | 60 | 40 | 20 | 20 | 200 |
| Effectif par unité d'amplitude | 10 | 15 | 30 | 20 | 10 | 5 | |

D'où l'histogramme des effectifs :



# II. Paramètres de position (variables quantitatives)

Lorsqu'une variable est quantitative, il est possible de synthétiser son étude au moyen de différents paramètres. Les premiers d'entre eux sont ceux cherchant à dégager des tendances centrales.

## 1. Moyenne

### a) Définition

Soit $X$ une variable quantitative se déclinant en $n$ modalités $(x_i)_{1\leq i\leq n}$ d'effectifs associés $(n_i)_{1\leq i\leq n}$ (et de fréquences associées $(f_i)_{1\leq i\leq n}$).

On appelle moyenne de $X$ le nombre : $\boldsymbol{\bar{X} = \frac{1}{N}\sum_{i=1}^{i=n} n_i x_i = \sum_{i=1}^{i=n} f_i x_i}$

Dans le cas où la variable est continue et regroupée par classes, les $x_i$ sont choisis comme étant les centres des intervalles correspondants (on fait là aussi l'hypothèse d'une régularité de la distribution dans chaque classe).
Exemple : dans le 2ème exemple fourni au I.3.b) (variable continue), $\bar{X} = 1552{,}50\ €$ (salaire moyen).

### b) Propriété de linéarité

Pour tous réels $a$ et $b$ et toute variable quantitative $X$, $\boldsymbol{\overline{aX+b} = a\bar{X} + b}$

## 2. Mode, classe modale

On appelle mode d'une variable quantitative discrète, la valeur de la série dont l'effectif associé est le plus élevé.
On appelle classe modale d'une variable quantitative continue, la classe de la série dont l'effectif par unité d'amplitude associé est le plus élevé.
Un mode ou une classe modale peut ne pas être unique.

## 3. Médiane

### a) Définition

On appelle médiane d'une variable quantitative $X$ le nombre $M$ qui partage l'effectif total de l'échantillon étudié en deux sous-groupes de même effectif.
Lorsque $X$ est continue, on appelle classe médiane la classe contenant la médiane.

Remarque : si $(x_i)_{1\leq i\leq n}$ sont $n$ réels rangés par ordre croissant,
alors leur médiane est $x_{(n+1)/2}$ si $n$ est impair, et $\frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right)$ si $n$ est pair.
Par exemple, la médiane de {2; 5; 6; 7; 9; 9; 10} est 7 ; celle de {2; 5; 6; 7; 9; 9} est 6,5 ; et celle de {2; 5; 6; 6; 9; 9} est 6.

### b) Procédés de calcul de la médiane

#### Cas d'une variable discrète

Si la variable est discrète, on peut facilement déterminer la médiane en construisant le tableau des effectifs (ou des fréquences) cumulé(e)s croissant(e)s et en utilisant la remarque précédente. Dans le 1er exemple fourni au I.3.b) (variable discrète), ce tableau est le suivant :

| 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 |

L'effectif total étant 30, la médiane $M$ est la moyenne de 8 et 9, soit 8,5.

### Cas d'une variable continue

Si la variable est continue et regroupée en $n$ classes $([x_i; x_{i+1}[)_{1\leq i\leq n}$, la médiane se calcule par interpolation affine avec la ligne polygonale associée à l'histogramme des effectifs (ou des fréquences) cumulé(e)s croissant(e)s, laquelle s'obtient, sous l'hypothèse de régularité de la distribution dans chaque classe, à partir des diagonales des rectangles correspondants.

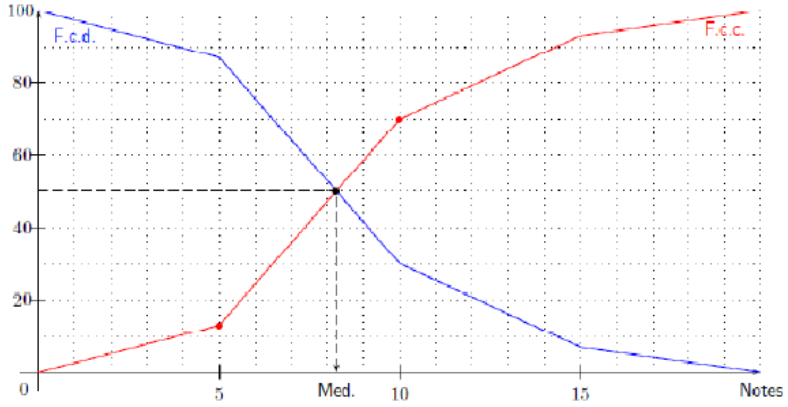
Exemple : considérons la série statistique précédente, regroupée ainsi par classes :

| 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 |

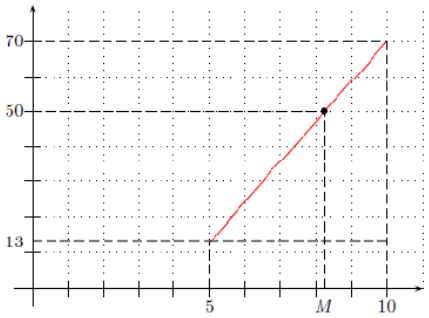
Le tableau des fréquences cumulées (croissantes et décroissantes) est le suivant :

| Notes | [ 0 ; 5 [ | [ 5 ; 10 [ | [ 10 ; 15 [ | [ 15 ; 20 [ |
|---|---|---|---|---|
| Fréq. en % | 13 | 57 | 23 | 7 |
| F.c.c. | 13 | 70 | 93 | 100 |
| F.c.d. | 87 | 30 | 7 | 0 |

La représentation de ces fréquences cumulées fournit deux lignes polygonales et la médiane :




Le calcul de $M$ par interpolation affine correspond à l'application du théorème de Thalès :




$$\frac{M-5}{10-5}=\frac{50-13}{70-13}$$

d'où $M \approx 8{,}25$

## 4. Quantiles, déciles…

On peut définir de manière analogue à la médiane :
Les quartiles $Q_1, Q_2, Q_3$ , qui divisent l'effectif total en quatre parties de même effectif ($Q_2$ étant donc la médiane).
Les déciles $D_1, D_2, \ldots, D_{10}$, qui divisent l'effectif total en dix parties de même effectif ($D_5$ étant donc la médiane).
Les vingtiles, les centiles, les millimes, etc.

# III. Paramètres de dispersion (variables quantitatives)

Les paramètres de position ne suffisent pas à rendre compte d'une série statistique quantitative : deux séries peuvent avoir même moyenne ou même médiane tout en étant très différentes dans la distribution de leurs effectifs respectifs, du point de vue de la dispersion autour des tendances centrales que mesurent la moyenne et la médiane.
On introduit donc d'autres paramètres visant à mesurer cette dispersion.

## 1. Étendue

L'étendue est la différence entre les valeurs maximale et minimale de la série étudiée.

## 2. Interquartile

On appelle interquartile de la série étudiée la différence $I$ entre le troisième et le premier quartiles :

$$\boldsymbol{I = Q_3 - Q_1}$$

## 3. Variance

### a) Définition

La variance $V(X)$ d'une variable $X$ est la moyenne des carrés des écarts entre les valeurs prises par $X$ (centres des classes en cas de continuité) et la moyenne de $X$ :

$$\boldsymbol{V(X) = \frac{1}{N}\sum_{i=1}^{i=n} n_i(x_i - \overline{X})^2 = \sum_{i=1}^{i=n} f_i(x_i - \overline{X})^2}$$

### b) Propriété

La variance d'une variable $X$ est aussi égale à la différence entre la moyenne des carrés des valeurs prises par $X$ (centres des classes en cas de continuité) et le carré de la moyenne de $X$ :

$$\boldsymbol{V(X) = \overline{X^2} - (\overline{X})^2 = \left(\frac{1}{N}\sum_{i=1}^{i=n} n_i {x_i}^2\right) - \left(\frac{1}{N}\sum_{i=1}^{i=n} n_i x_i\right)^2 = \left(\sum_{i=1}^{i=n} f_i {x_i}^2\right) - \left(\sum_{i=1}^{i=n} f_i x_i\right)^2}$$

Exemple : dans le 2ème exemple fourni au I.3.b) (variable continue), $V(X) \approx 106\,619\ €^2$

### c) Propriété de changement d'échelle

Pour toute variable $X$ et tous réels $a$ et $b$ : $\boldsymbol{V(aX + b) = a^2 V(X)}$
Remarque : l'effet de translation ($b$) est nul sur la variance.

## 4. Écart-type

### a) Définition

L'écart-type $\sigma(X)$ d'une variable $X$ est la racine carrée de sa variance :

$$\boldsymbol{\sigma(X) = \sqrt{V(X)}}$$

Exemple : dans le 2ème exemple fourni au I.3.b) (variable continue), $\sigma(X) \approx 327\ €$

### b) Propriété de changement d'échelle

Pour toute variable $X$ et tous réels $a$ et $b$ : $\boldsymbol{\sigma(aX + b) = |a|\sigma(X)}$

## 5. Remarques

La variance et l'écart-type sont des nombres positifs ou nuls.
Ils sont nuls si et seulement si toutes les valeurs de la série sont égales à sa moyenne.
Ils sont d'autant plus grands que la série est dispersée autour de sa moyenne.

# EXERCICES

**Exercice 1**

Que dire d'un pays où le salaire médian est de 30% inférieur au salaire moyen ?

**Exercice 2**

On a fait passer un test de 10 questions concernant la connaissance des consignes de sécurité à 3000 employés d'une grande entreprise. Les nombres de réponses exactes et les effectifs correspondant sont donnés par le tableau suivant :

| Nombre de réponses exactes | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|
| Effectif | 144 | 213 | 786 | 1143 | 501 | 144 | 69 |

1) Représenter cette série statistique.
2) Déterminer sa moyenne, sa médiane, son interquartile, son étendue, sa variance et son écart-type.

**Exercice 3**

Les salaires mensuels en euros des membres du personnel d'une entreprise se répartissent comme suit :

| Sal. | [450; 610[ | [610; 760[ | [760; 1220[ | [1220; 1520[ | [1520; 1820[ | [1820; 2280[ | [2280; 2890[ |
|---|---|---|---|---|---|---|---|
| Fré. | 0,0318 | 0,1227 | 0,2409 | 0,3455 | 0,1955 | 0,05 | 0,0136 |

1) Déterminer sa moyenne, sa médiane, son interquartile, son étendue, sa variance et son écart-type.
2) Déterminer sa moyenne, sa variance et son écart-type si l'on augmente tous les salaires de 2% et qu'on attribue à chaque salarié une prime de fin d'année de 100 €.