

# Chapitre 4. Statistiques descriptives : Paramètres de position

Mathématiques et statistiques appliquées  
Département TC1-IUT de Sceaux

Damien THOMINE

## Objectifs

- Savoir calculer et interpréter la moyenne et la médiane.
- Savoir calculer et interpréter quartiles et déciles.
- Faire le lien avec les représentations de données (histogrammes, fréquences cumulées).

## Plan du cours

- 1 Moyenne
- 2 Médiane
- 3 Fractiles : quartiles, déciles et centiles

# Paramètres statistiques

## Les paramètres statistiques

sont des nombres qui ont pour but de résumer l'essentiel de l'information relative à une variable statistique quantitative.

- Les **paramètres de position** indiquent la valeur “typique” autour de laquelle les observations sont réparties.  
Les deux paramètres de position les plus importants sont la **moyenne** et la **médiane**.
- Les **fractiles**, notamment les **quartiles** et les **déciles**, donnent des informations plus fines sur la série. Ils sont une généralisation de la médiane.
- Les **paramètres de dispersion** mesurent combien les observations s'éloignent de la valeur centrale. Le plus important est l'**écart-type**.

# Section 1

## Moyenne

# Moyenne

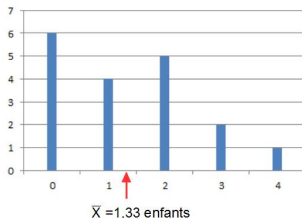
La **moyenne** d'une variable est la somme des valeurs observées  $x_i$  divisée par le nombre total  $n$  d'individus dans la population :

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

**Exemple.** À droite, la série brute d'une enquête sur le nombre d'enfants de 18 familles d'un immeuble.

$$\bar{x} = \frac{0+0+0+0+0+0+1+1+1+1+2+2+2+2+3+3+4}{18} \simeq 1,33$$

Il y a en moyenne 1,33 enfants par famille.



Famille N°	Nb Enfants
1	0
2	0
3	0
4	0
5	0
6	0
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	2
16	3
17	3
18	4

# Calcul de la moyenne à partir des effectifs et fréquences

On peut calculer la moyenne à partir de modalités  $y_1, \dots, y_p$  et des effectifs  $n_i$  ou fréquences  $f_i$  associés.

$$\bar{x} = \frac{y_1 \times n_1 + \dots + y_p \times n_p}{n}$$

**Exemple.** À partir du tableau des effectifs

Nb. d'enfants $y_i$	0	1	2	3	4	Total
Effectif $n_i$	6	4	5	2	1	18

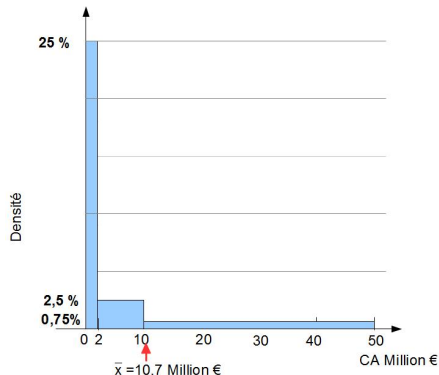
$$\bar{x} = \frac{0 \times 6 + 1 \times 4 + 2 \times 5 + 3 \times 2 + 4 \times 1}{18} \simeq 1,33$$

Famille N°	Nb Enfants
1	0
2	0
3	0
4	0
5	0
6	0
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	2
16	3
17	3
18	4

# Calcul de la moyenne avec des classes

Si on dispose seulement des effectifs par classe, on peut calculer une valeur approchée de la moyenne en assignant à chaque classe  $[e_i, e_{i+1}[$  son **centre**  $c_i = \frac{e_i + e_{i+1}}{2}$ .

On utilise ces valeurs à la place des modalités dans la formule.



**Exemple.** Parmi les PME d'une région, on a

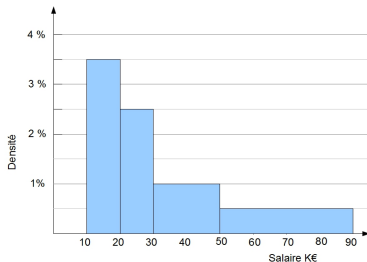
CA annuel en Millions d'euros	Micros Entreprises [0,2[	Petites Entreprises [2,10[	Moyennes Entreprises [10,50[
Fréquence	50%	20%	30%
Centre	$\frac{0+2}{2} = 1$	$\frac{2+10}{2} = 6$	$\frac{10+50}{2} = 30$

$$\bar{x} = 1 \times 0,5 + 6 \times 0,2 + 30 \times 0,3 = 10,7$$

Le CA moyen des entreprises de la région est d'environ 10,7 M€.

# Test

Voici la distribution par classes des salaires annuels des habitants d'une région.



Salaire annuel K€	[10,20[	[20,30[	[30,50[	[50,90[
Fréquence	35%	25%	20%	20%
Centre				

Moyenne =  $\bar{x} \simeq$



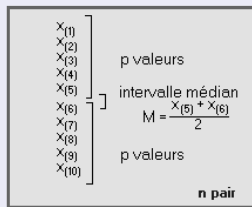
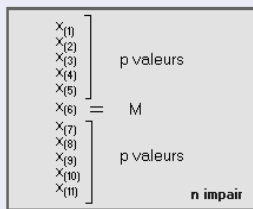
## Section 2

### Médiane

# Médiane

## La médiane

M est la valeur du milieu de la série des données, c'est-à-dire la valeur telle qu'il y ait autant d'observations "au-dessous" qu'"au-dessus".



Si la série brute des valeurs observées est triée par ordre croissant  $x_1 \leq x_2 \leq \dots \leq x_n$  :

- si  $n$  est **impair**, soit  $n = 2p + 1$  alors la médiane est  $M = x_{p+1}$ .
- si  $n$  est **pair**, soit  $n = 2p$  on choisit en général  $M = \frac{x_p + x_{p+1}}{2}$ .

# Exemple et Test

Immeuble A Immeuble B Immeuble C Calculer le nombre médian d'enfants par famille dans les trois immeubles.

Famille N°	Nb Enfants
1	0
2	0
3	0
4	0
5	0
6	0
7	1
8	1
9	1
10	1
11	2
12	2
13	2
14	2
15	2
16	3
17	3
18	4

Famille N°	Nb Enfants
1	0
2	0
3	1
4	1
5	1
6	1
7	1
8	1
9	2
10	2
11	2
12	2
13	3
14	3
15	4
16	4

Famille N°	Nb Enfants
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	1
10	1
11	1
12	1
13	1
14	1
15	2

**Immeuble A :**

$$M_A = \frac{x_9 + x_{10}}{2} = \frac{1+1}{2} = 1 \text{ car } n = 18 = 2 \times 9$$

**Immeuble B :**

$$M_B = \boxed{\phantom{000000}}$$

**Immeuble C :**

$$M_C = \boxed{\phantom{000000}}$$

## Médiane à partir des effectifs et fréquences cumulés

La médiane  $M$  peut s'obtenir à l'aide des fréquences (ou des effectifs) cumulés :

la médiane est la première modalité dont la fréquence cumulée est supérieure à 50% (ou dont l'effectif cumulé est supérieur à la moitié de l'effectif total).

**Exemple.** Pour le nombre d'enfants des 18 familles étudiées,

Nombre d'enfants $x_i$	0	1	2	3	4
Effectifs cumulés $N_i$	6	10	15	17	18
Fréquences cumulées $F_i$	33%	55%	83%	94%	100%

On remarque que 55% des familles ont au plus un enfant et que 33% des familles n'ont pas d'enfant, donc la médiane est  $M = 1$ .

De même "1" est la première valeur dont l'effectif cumulé (10) dépasse la moitié de la moitié effectif total  $\frac{18}{2} = 9$

# Test

Voici le tableau des effectifs et effectifs cumulés concernant la variable "nombre de salariés" sur une population composée des 107 PME d'une ville.

Nombre de salarié $x_i$	0	1	2 <span style="color: red;">2</span>	3	4	5	6	Total
Effectif $n_i$	13	20	35	19	12	5	3	107
Effectifs cumulés $N_i$	13	33	68 <span style="color: green;">68</span>	87	99	104	107	

La médiane est  $M =$

# Médiane à partir des classes

Si on dispose des données par classes, on peut obtenir une valeur approchée de la médiane à l'aide du graphique des fréquences cumulées.

**Exemple.** Le tableau des fréquences cumulées de l'âge des employés d'une entreprise est :

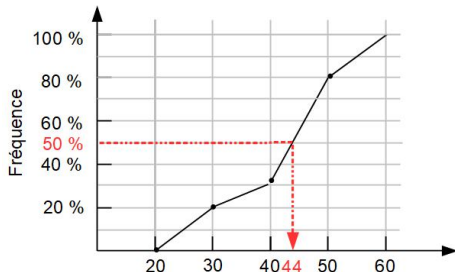
Âge	[20,30[	[30,40[	[40,50[	[50,60[
Fréquence	20%	10%	50%	20%

Ce qui donne les fréquences cumulées

Âge	30	40	50	60
Fréquences cumulées	20%	30%	80%	100%

On lit sur le graphique que environs 50% des employés ont moins de 44 ans.

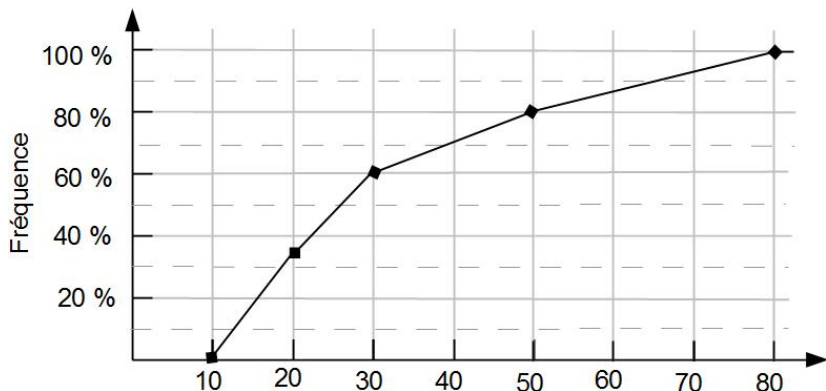
L'âge médian est donc  $M \simeq 44$



# Test

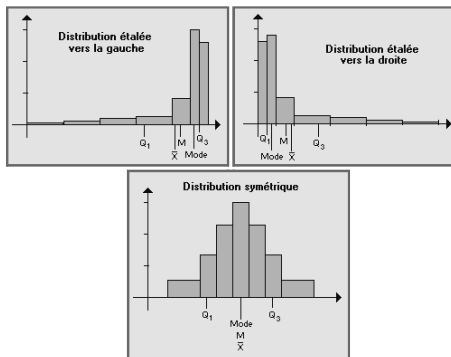
Voici la distribution par classes des salaires annuels des habitants d'une région et ses fréquences cumulées

Salaire annuel K€	[10,20[	[20,30[	[30,50[	[50,80[
Fréquences	35%	25%	20%	20%



# Médiane vs. Moyenne

Dans le cas d'une distribution symétrique, médiane et moyenne coïncident. Cela n'est pas vrai en général.



Distribution étalée vers la gauche :  
 $\Rightarrow \bar{x} = \text{Moyenne} < M = \text{médiane}$

Distribution étalée vers la droite :  
 $\Rightarrow M = \text{Médiane} < \bar{x} = \text{Moyenne}$

**Exemple.** La distribution des salaires est étalée vers la droite.

Selon l'INSEE en France en 2013 :

Salaire mensuel médian = 1 772 € < 2 202 € = Salaire mensuel moyen.



## Section 3

### Fractiles : quartiles, déciles et centiles

# Quartiles

Les **quartiles**  $Q_1$ ,  $Q_2$  et  $Q_3$  sont les trois valeurs qui partagent la série statistique en 4 parties de même effectif.

N°	Note
1	2
2	3
3	4
4	4
5	6 $Q_1$
6	6
7	8
8	9
9	10
10	10 $Q_2$
11	10
12	11
13	11
14	11
15	11 $Q_3$
16	12
17	14
18	14
19	16

$Q_1$  est la donnée de la série qui sépare les 25 % inférieurs des 75% supérieurs

$Q_2$  est la donnée de la série qui sépare en deux parties de même effectifs.  $Q_2$  coïncide avec la médiane.

$Q_3$  est la donnée de la série qui sépare les 75 % inférieurs des 25% supérieurs

**Exemple.** Pour les notes de la classes de 19 étudiants à gauche,  $Q_1 = 6$ ,  $Q_2 = 10$  et  $Q_3 = 11$ .

Le deuxième quartile  $Q_2$  est la médiane.

## Déciles et centiles

Les **déciles**  $D_1, D_2, \dots$  et  $D_9$  sont les valeurs qui partagent la série statistique en 10 parties de même effectif. Particulièrement significatif sont

$D_1$  la donnée qui sépare les 10 % inférieurs des 90% supérieurs

$D_9$  la donnée qui sépare les 90 % inférieurs des 10% supérieurs

**Exemple.** Selon l'INSEE pour le salaire mensuel en France en 2013 on a  $D_1 = 1\,200$  et  $D_9 = 3\,544$ .

C'est-à-dire 10% des français gagnent moins de 1 200 € et 10% plus de 3 544 € par mois.

Les **centiles**  $C_1, \dots$  et  $C_{99}$  sont les valeurs qui partagent la série statistique en 100 parties de même effectif.

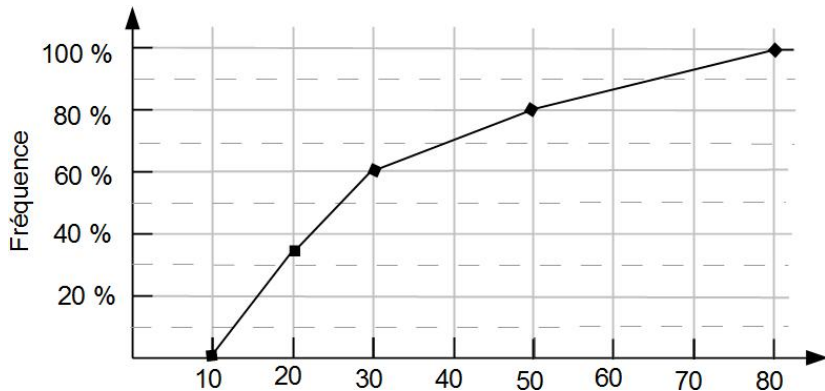
**Exemple.** Pour le salaire mensuel en France on a  $C_{95} = 4\,526$  et  $C_{99} = 8\,061$ .

C'est-à-dire 5% des français gagnent plus de 4 526 € et 1% gagnent plus de 8 061 € par mois.

## Test : à partir du graphique des fréquences cumulés

Voici la distribution par classes des salaires annuel des habitants d'une région et ses fréquences cumulés

Salaire annuel K€	[10,20[	[20,30[	[30,50[	[50,80[
Fréquences	35%	25%	20%	20%



## Test : Fractiles d'une variable discrète

Voici le tableau des effectifs et effectifs cumulés du “Nombre de salariés” sur une population composée des 107 PME d'une ville.

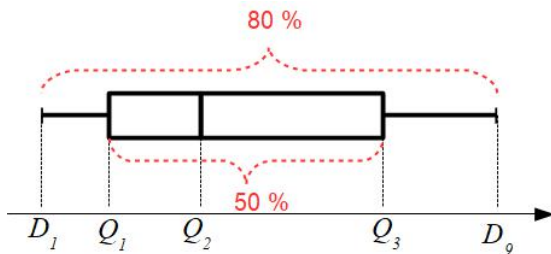
Nb. de salariés $x_i$	0	1	2	3	4	5	6	Total
Effectif $n_i$	13	20	35	19	12	5	3	107
Effectifs cumulés $N_i$	13	33	68	87	99	104	107	

$$Q_1 = \square \text{ et } Q_3 = \square$$

$$D_1 = \square \text{ et } D_9 = \square$$

## Boîte à moustache

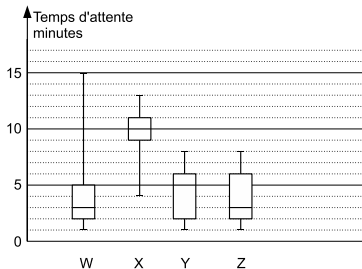
Le **diagramme en boîte** (ou “boîte à moustaches”) permet de représenter les quartiles et déciles d’une distribution.



L'intérêt de ces diagrammes en boîtes est de pouvoir comparer par un moyen visuel plusieurs distributions d'un même caractère sur des populations différentes.

# Test

Une association de consommateurs a testé les services d'assistance téléphonique de quatre assurances W, X, Y et Z. Pour chaque appel au service, on a mesuré le temps d'attente du client.



- Pour le temps d'attente au **service clients de W** on a :

$$D_1 = \boxed{\phantom{00}} \quad D_9 = \boxed{\phantom{00}}$$

$$Q_1 = \boxed{\phantom{00}} \quad Q_2 = \boxed{\phantom{00}} \quad Q_3 = \boxed{\phantom{00}}$$

- 10% des clients de W attendent plus de  et 50% plus de
- Le pire des services clients est .
- Le meilleur est .