

Statistique descriptive

Paramètres de Tendence Centrale et paramètre de dispersion

DR SM, GUENIFID

Les paramètres de tendance centrale

- 1 - Le mode
- 2 - La médiane
- 3 - La moyenne arithmétique

Introduction

Les paramètres de tendance centrale (Paramètres de position)

Synthétisent et
caractérisent
l'ensemble des données par
un nombre unique,
une valeur type



de telle sorte qu'en première approximation la comparaison de
deux séries puisse se ramener à la comparaison
de **deux nombres**.

1- Le Mode (Mo)

- Le mode est la modalité ou la valeur qui a le plus grand effectif, la plus grande fréquence.
- Il est noté *Mo*.

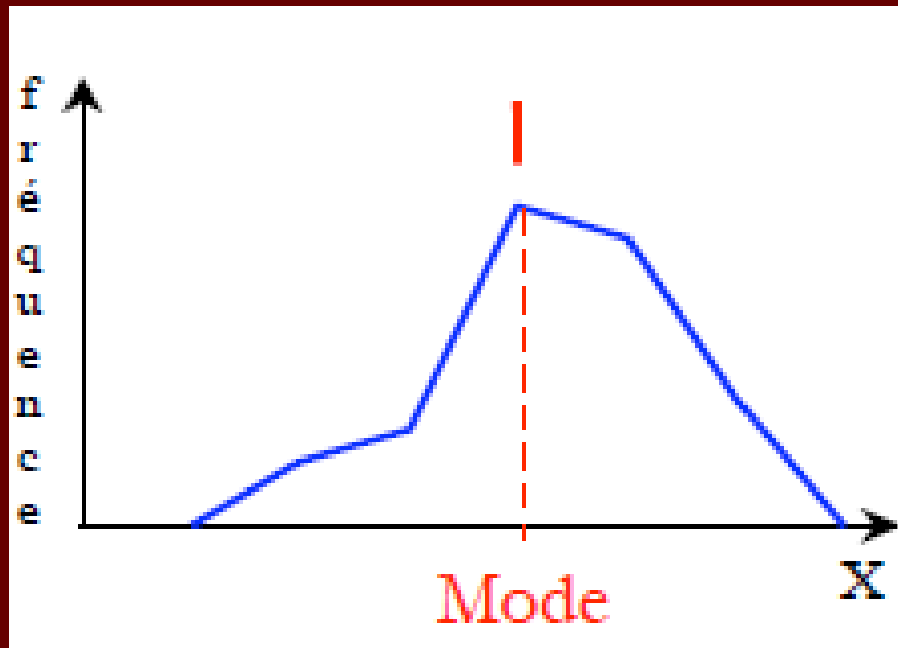
-Une distribution peut avoir un seul mode. Il s'agit alors du sommet de la distribution. On parle de : **distribution uni modale**

-Une distribution peut aussi avoir plusieurs modes :

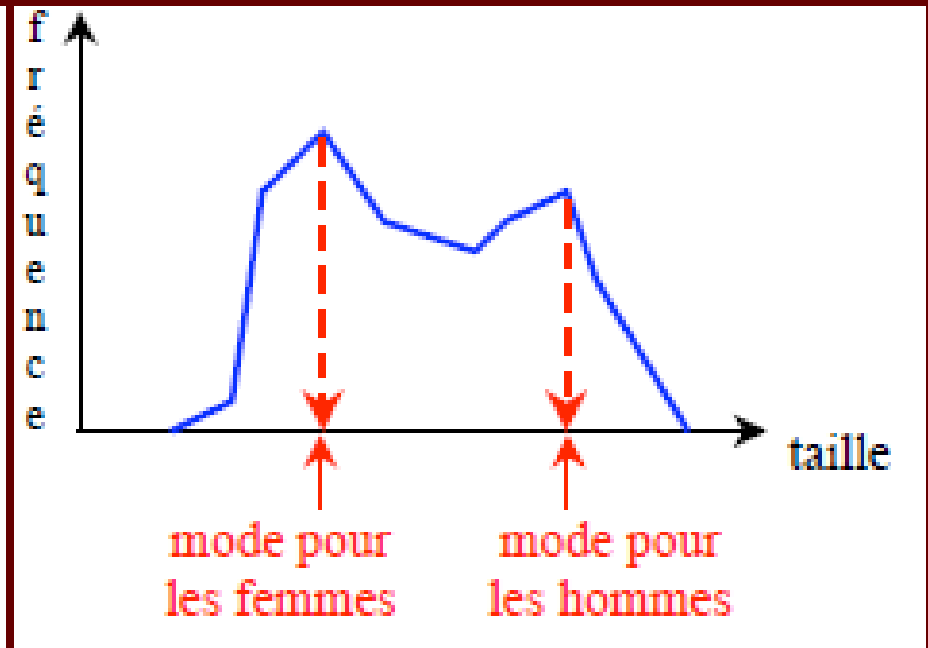
- * **distribution bimodale** (deux modes)

- * **distribution multimodale** (plusieurs modes).

Distribution uni modale



Distribution bi modale



distribution de la taille des individus dans une population adulte

1 - Cas des données non groupées

soit la série statistique :

3, 5, 7, 15, 16, 16, 16, 17, 17, 30

alors

Mo = 16

2- Cas des données groupées :

a- Pour les variables qualitatives :

- Le mode d'une variable qualitative est la modalité la plus fréquemment observée.

Exemple : Perception de l'état de santé

	Effectifs
Très bon	3303
Bon	6337
Moyen	2325
Mauvais	354
Très mauvais	71
Total	12390

**Le mode est la modalité « bon »
L'effectif modal est 6337.**

**Le mode est la seule mesure de tendance centrale applicable
aux variables qualitatives.**

b- Pour les variables quantitatives discrètes :

Le mode est la valeur de la variable statistique qui correspond à l'effectif le plus élevé.

Exemple

Nombre d'enfants par famille

Nombre d'enfants x_i	Nombre de familles n_i
0	4
1	5
2	10
3	16
4	18
5	14
6	7
7	6
TOTAL	80

Mo = 4



Interprétation :

Le nombre d'enfants le plus fréquent dans cet échantillon est égal à 4.

c- Pour les variables quantitatives continues

La classe qui correspond à l'effectif le plus élevé est appelée classe modale

Le mode est le centre de la classe modale

Exemple

Les pesées de 50 nouveau-nés.

Poids (Kg) x_i	Effectifs n_i
2.0 - 2.5	2
2.5 - 3.0	4
3.0 - 3.5	6
3.5 - 4.0	30
4.0 - 4.5	8
TOTAL	50

Classe modale



La classe modale est : 3,5 – 4,0 alors le mode est

$$Mo = (3,5 + 4,0)/2 = 3,75 \quad \text{Donc } \mathbf{Mo = 3,75 \text{ Kg}}$$

Interprétation :

Le poids le plus fréquent dans cet échantillon est égal à 3,75 kg

- Calcul du mode en utilisant la méthode d'interpolation linéaire : si on cherche plus de précision on applique la méthode d'interpolation linéaire en utilisant la formule suivante

$$Mo = b_{inf} + \frac{d1}{d1 + d2} \times k$$

où

b_{inf} : borne inférieure de la classe modale

d1 : différence entre l'effectif de la classe modale et de la classe précédente

d2 : différence entre l'effectif de la classe modale et de la classe suivante

k : amplitude de la classe

Exemple : Moyenne en histoire-géographie dans une classe de terminale S

Classes	Effectifs (ni)
[6; 8[3
[8; 10[6
[10; 12[11
[12; 14]	7
[14; 16[4
Total	31

- La classe modale est **[10;12[**, avec un effectif modal de 11
- $Mo = 10 + [5/(5+4)] * 2 = 11,1$
- La moyenne la plus fréquente dans cet échantillon est égal à 11,1

2- La médiane (Me)

- La médiane est la valeur de la variable qui partage en 2 parties égales ou en 2 sous-ensemble égaux la population.
- Elle divise une série statistique ordonnée en deux ensembles comportant chacun 50% des données. 50% des valeurs sont supérieures à la médiane et 50% inférieures .

1- Cas des données non groupées :

Pour déterminer la médiane d'un échantillon ou d'une population :

- On classe les individus par ordre croissant

- **2 cas :**

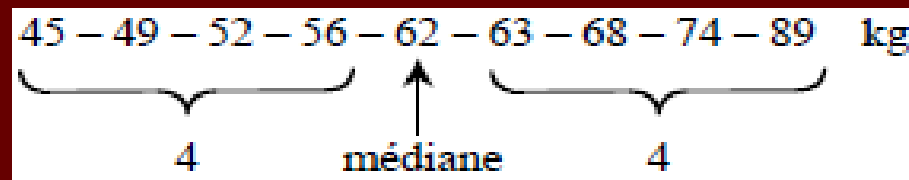
- Si n est impair, la médiane est la valeur d'ordre $\frac{n+1}{2}$
- Si n est pair, la médiane est la moyenne entre la valeur d'ordre $\frac{n}{2}$ et la valeur $\frac{n+2}{2}$

Exemple :

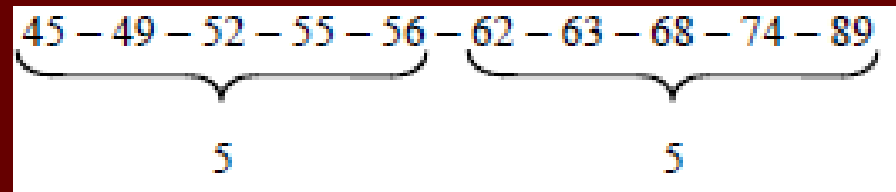
* Soit un échantillon de 9 personnes dont le poids est :

45 – 68 – 89 – 74 – 62 – 56 – 49 – 52 – 63 kg

classés par ordre croissant :



* Soit un échantillon de 10 personnes dont le poids est :



La médiane : $\frac{56 + 62}{2} = 59 \text{ Kg}$

2 - Cas des données groupées :

a- Pour les variables quantitatives discrètes

La médiane est la première valeur de la variable statistique qui correspond à un effectif cumulé sup à $n/2$
(détermination directe)

Exemple

Nombre d'enfants par famille

Me



Nombre d'enfants (Xi)	Nombre de familles (ni)	Effectif cumulé
0	4	4
1	5	9
2	10	19
3	16	37
4	18	51
5	14	65
6	7	72
7	6	80
TOTAL	80	

$n/2 = 80/2 = 40$, la valeur de la variable qui occupe le 40^{ème} rang est égale à 4 donc **Me = 4 enfants**

Interprétation:

Il y a 50 % (soit 40 familles) qui ont moins de 4 enfants et 40 plus de 4 enfants.

b- Pour les variables quantitatives continues :

- Calcul de la médiane par interpolation linéaire

$$\mathbf{Me = binf + \frac{n/2 - S}{N} \times k}$$

où :

binf : borne inférieure de la classe médiane

n : la taille de l'échantillon

S : somme des effectifs de toutes les classes précédant la classe médiane.

nMe : l'effectif de la classe médiane

k: l'amplitude de la classe

Exemple : Le poids de 50 nouveau-nés.

Poids (Kg) x_i	Effectifs n_i
2.0 - 2.5	2
2.5 - 3.0	4
3.0 - 3.5	6
3.5 - 4.0	30
4.0 - 4.5	8
TOTAL	50

$$\text{Me} = 3,5 + \left(\frac{50/2 - 12}{30} \right) 0,5$$

$$\text{Me} = 3,72 \text{ kg}$$

Interprétation :

Il y a 50 % (soit 25) nouveau-nés qui ont un poids inférieur à 3,75 kg et 50 % (25) qui ont un poids supérieur à 3,75 kg.

3- La moyenne arithmétique :

La moyenne est la mesure de tendance Centrale la plus utilisée.

La moyenne est la somme des valeurs divisée par l'effectif total.

1- Cas de données non groupées :

La moyenne arithmétique est la somme des observations divisée par leur nombre.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4 + \dots + X_n}{n}$$

Ce qui donne

$$\overline{X} = \sum x_i / n$$

x_i = valeur de la variable X

n = taille de l'échantillon (effectif total)

2- cas de données groupées :

A- Pour les variables quantitatives discrètes

S'il y a répétition de certaines observations, c'est à dire le nombre x_1 se produit n_1 fois, x_2 se produit n_2 fois,....., la formule précédente devient :

$$\bar{X} = \frac{\sum n_i x_i}{N}$$

x_i = valeur de la variable X

n_i = effectif correspondant à la valeur x_i

N = taille de l'échantillon (effectif total)

Exemple

Nombre d'enfants par famille

xi	ni	ni . xi
0	4	0
1	5	5
2	10	20
3	16	48
4	18	72
5	14	70
6	7	42
7	6	42
	$\Sigma n_i = n = 80$	$\Sigma n_i . x_i = 299$

$$\bar{X} = (1/n) \Sigma n_i x_i = (1/80) \Sigma n_i x_i = \frac{299}{80} = 3,74$$

$\bar{X} = 4$ enfants, il y a en moyenne 4 enfants par famille.

b- Pour les variables quantitatives continues :

$$\bar{X} = \frac{\sum n_i x_{i_{cc}}}{N}$$

$x_{i_{cc}}$ = centre de la classe

n_i = effectif correspondant à la valeur x_i

N = taille de l'échantillon (effectif total)

Exemple

Moyenne en histoire-géographie dans une classe de terminale S

	n_i	C_i	$n_i C_i$
[6; 8[3	7	21
[8; 10[6	9	54
[10; 12[11	11	121
[12; 14]	7	13	91
[14; 16[4	15	60
Total	31		347

$$\overline{X} = \frac{347}{31} = 11,2$$

Paramètres de dispersion

- La moyenne est insuffisante pour caractériser une série statistique elle doit être accompagnée d'un paramètre de dispersion

variance s^2
écart type(S)

Considérons les deux series

Serie1:15,20,25,30,35

Serie 2: 5,15,25,35,45

La moyenne de la serie1=25

La moyenne de la serie 2 = 25

Les deux séries ont la même moyenne
mais visiblement ne se ressemblent pas

Estimation de la **variance** dans un échantillon

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \{x_i - \bar{x}\}^2$$

La **variance** est la moyenne des écarts à la moyenne

Ecart-type de (X) *la racine carrée* de la variance de (X)

La variance et l'écart type renseignent tous deux sur l'étalement de la série autour de la moyenne

La variance et écartype des 2 séries

Variance de la série 1 =62,5

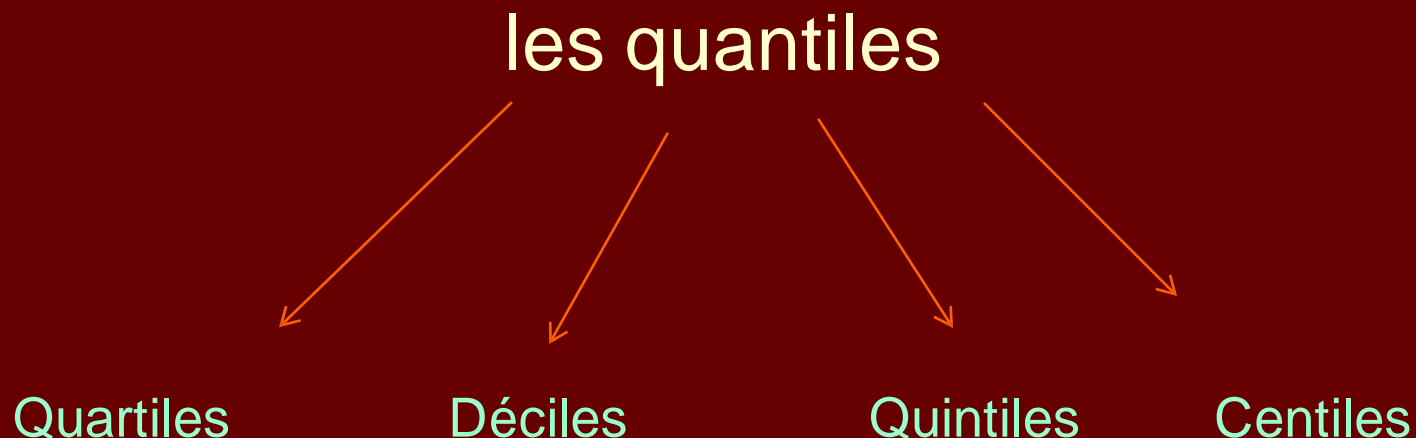
Variance de la série 2 =250

Écartype de la série 1 =7,9

Écartype de la série 2 =15,8

Les quantiles :

- Les quantiles sont des valeurs qui divisent une série statistique ordonnée en plusieurs groupes comprenant la même proportion des données.
- Voici un arbre représentant les quantiles les plus fréquemment utilisés :

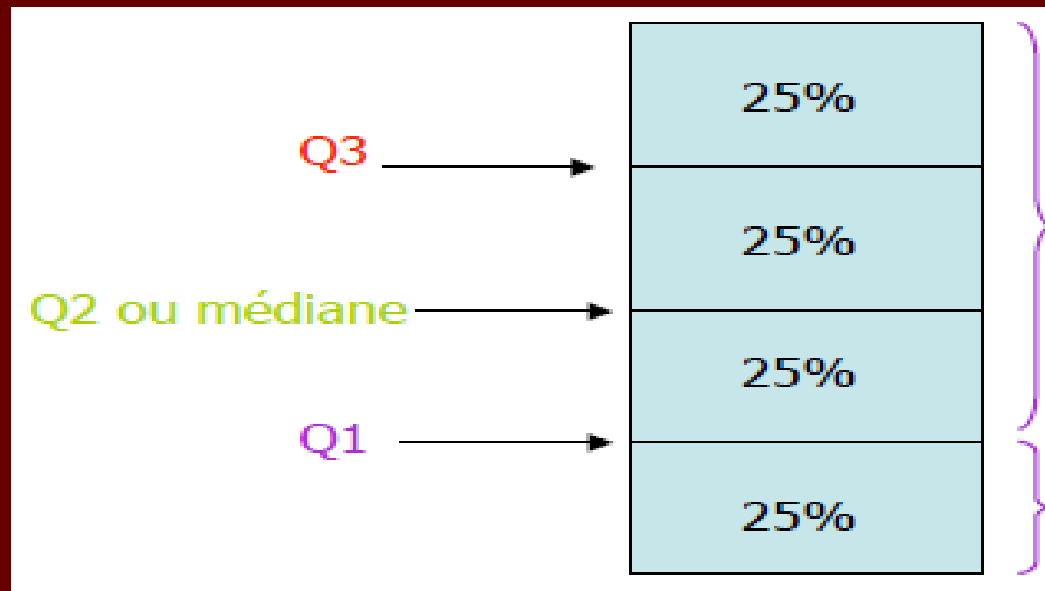


1-Les quartiles :

Les quartiles, notés par Q1, Q2 et Q3, divisent une série statistique ordonnée en **quatre groupes égaux** comprenant chacun 25% des données de la série.

On dit que :

- 25% des données sont inférieures à Q1
- 50% des données sont inférieures à Q2 → correspond à la médiane
- 75% des données sont inférieures à Q3



2-Les déciles :

Les déciles, notés par D1, D2, ..., D8 et D9, divisent une série statistique ordonnée en **dix groupes égaux** comprenant chacun 10% des données de la série.

On dit que :

- 10% des données sont inférieures à D1
- 20% des données sont inférieures à D2
- ...
- 90% des données sont inférieures à D9

3-Les quintiles :

Les quintiles, notés par V_1 , V_2 , V_3 et V_4 , divisent une série statistique ordonnée en **cinq groupes égaux** comprenant chacun 20% des données de la série.

On dit que :

- 20% des données sont inférieures à V_1
- 40% des données sont inférieures à V_2
- 60% des données sont inférieures à V_3
- 80% des données sont inférieures à V_4

4-Les centiles :

Les centiles, notés par C1, C2, ...C98 et C99, divisent une série statistique ordonnée en **100 groupes égaux** comprenant chacun 1% des données de la série.

On dit que :

- 1% des données sont inférieures à C1
- 2% des données sont inférieures à C2
- ...
- 99% des données sont inférieures à C99.

Remarques

- Dans une distribution symétrique, les trois paramètres de tendance centrale (**mode, médiane, moyenne**) sont égaux.
- Mode, médiane et moyenne sont les valeurs centrales les plus **utilisées**

Exemple : l'étude du poids de 50 personnes a donner les résultats suivants (en Kg)

Age	Effectifs
[37-47[2
[46- 55[4
[55- 64[8
[64- 73[14
[73- 82[11
[82- 91[7
[91-100[4
Total	50

1. Quel est le mode de la distribution?
2. Calculer la moyenne
3. Calculer la médiane
4. Calculer le 1er quartile

Classes d'âge	Centre de classe X_{icc}	effectif n_i	Fréquence relative $f_i \%$	Effectif cumulé $n_i \text{ cum}$	$n_i \cdot x_i$
[37-47[41,5	2	4	2	83
[46- 55[50,5	4	8	6	202
[55- 64[59,5	8	16	14	476
[64- 73[68,5	14	28	28	959
[73- 82[77,5	11	22	39	852,5
[82- 91[86,5	7	14	46	605,5
[91-100[95,5	4	8	50	382
Total		50	100 %		3560

CALCUL du mode , médiane moyenne :

La classe modale est : $[64 - 73[$

$$\text{Mo} = \text{binf} + \frac{d1}{d1 + d2} \times k$$

$$\text{Mo} = 64 + \frac{14 - 8}{(14 - 8) + (14 - 11)} \times 9 = 68,5 \text{ kg}$$

Calcul de la moyenne :

$$\bar{X} = \frac{\sum n_i x_{i_{cc}}}{N}$$

$x_{i_{cc}}$: le centre de classe la valeur qui se trouve au milieu de la classes)

n_i : effectif correspondant à la valeur x_i

n : la taille de l'échantillon (effectif total)

$$\bar{X} = \frac{3560}{50} = 71,2 \text{ Kg}$$

Calcul de la médiane :

$$Me = b_{inf} + \frac{(n/2) - s}{n_M} \times k$$

b_{inf} : la borne inférieure de la classe médian

n : la taille de l'échantillon

n_{med} : l'effectif de la classe médiane

k : amplitude de la classe médiane

s : la somme des effectifs précédant la classe médiane

La classe médiane $:: [64 - 73[$

$$Mo = 64 + \frac{50/2 - (8 + 4 + 2)}{14} \times 9 = 71,07 \text{ kg}$$

Calcule du 1 er quartile :

$$Q1 = b_{inf} + \frac{(n/4) - s}{n_{Q1}} \times k$$

b inf : la borne inferieure du Q1

n : la taille de l'échantillon

n Q1 : l'effectif de la classe du Q1

k : amplitude de la classe du Q1

s : la somme des effectifs précédant la classe du premier quartile Q1 a

Q1 appartient à la classe [55- 64[

$$Q1 = 55 + \frac{(50/4) - 6}{14} \times 9 = 59,17$$

Paramètres de dispersion

- La moyenne est insuffisante pour caractériser une série statistique elle doit être accompagnée d'un paramètre de dispersion

variance s^2
écart type(S)

Considérons les deux series

Serie1:15,20,25,30,35

Serie 2: 5,15,25,35,45

La moyenne de la serie1=25

La moyenne de la serie 2 = 25

Les deux séries ont la même moyenne
mais visiblement ne se ressemblent pas

Estimation de la **variance** dans un échantillon

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n \{x_i - \bar{x}\}^2$$

La **variance** est la moyenne des écarts à la moyenne

Ecart-type de (X) *la racine carrée* de la variance de (X)

La variance et l'écart type renseignent tous deux sur l'étalement de la série autour de la moyenne

La variance et écartype des 2 séries

Variance de la série 1 =62,5

Variance de la série 2 =250

Écartype de la série 1 =7,9

Écartype de la série 2 =15,8

Exercice n°2 :

Voici la série, ordonnées dans l'ordre croissant, des 15 notes obtenues en mathématiques par un élève au cours du premier semestre.

4 - 6 - 6 - 9 - 11 - 11 - 12 - 13 -
13 - 13 - 14 - 15 - 17 - 18 - 18 .

1. Quelle est la note moyenne ?
2. Quelle est la note médiane ?
3. Quel est le mode ?