I. Mesures de tendence centrale

La tendance centrale d'une distribution est la valeur autour de laquelle se concentrent en général les données.
Il y a plusieurs façons (indicateurs) de trouver cette valeur.

1. La moyenne arithmétique :

Variables discrètes

C'est la somme des données, divisée par leur nombre.

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{\sum x_i n_i}{\sum n_i}$$

Exemple:
On suppose que la série suivante représente le salaire quotidien en D.A des 10 ouvriers dans un chantier.
1200, 1500, 1000, 800, 2500, 1000, 1200, 1200, 1000, 3000
Calculer la moyenne des salaires de ces ouvrier

1. Pour calculer la moyenne de cette série, il faut d'abord calculer le total S des salaires des ouvriers.
Ce total est calculé comme suit : x1 + x2 + x3 + ... + x10
S = 1200+1500+1000+800+2500+1000+1200+1200+1000+3000 = 14400
La moyenne = S/10 = 14400/10 = 1440

x_i	*Fréquence*	$n_i\ x_i$
800	1	800
1000	3	3000
1200	3	3600
1500	1	1500
2500	1	2500
3000	1	3000

Total = 10

$S = \sum n_i\ x_i = 14400$

La moyenne $= S / 10$
$= 1440$

Variables continues et variables discrète classées

On suppose que les valeurs d'une variable Y sont mis dans un nombre *k* des intervalles ou classes comme suit :

$[y_0 - y_1[,\ [y_1 - y_2[,\ [y_2 - y_3[,\ \ldots\ [y_{k-1} - y_k[$

La moyenne est définie comme la moyenne arithmétique des centres des classes

$$\overline{y} = \sum_{i=1}^{k} \underbrace{f_i}_{\text{Fréquence relative}} \left(\frac{y_{i-1} + y_i}{2}\right) = \frac{1}{N} \sum_{i=1}^{k} \underbrace{n_i}_{\text{Fréquence}} \left(\frac{y_{i-1} + y_i}{2}\right)$$

Exemple:

le tableau suivant représente le salaire mensuel de 105 salariés d'une entreprise économique (en 1000 D.A).

Calculer la moyenne des salaires

Classe x_i	*Fréquence*	*Centre xi*	*n_i x_i*
[17 - 22[	11	19.5	214.5
[22 - 27[	23	24.5	563.5
[27 - 32[	43	29.5	1268.5
[32 - 37[	17	34.5	586.5
[37 - 42[	7	39.5	276.5
[42 - 50[	4	46	184
Total =	*105*		*3093.5*

La moyenne = 3093 / 105
= 29.46190
= 29461.9 DA

Inconvénients :

1: Ne peut pas être calculée pour les données qualitatives
2: Influencé par des valeurs extrêmes
3: On ne peut pas le trouver graphiquement

2. La médiane :

C’est la valeur qui se trouve au centre d'une série statistique, c-à-d permet de la partager en deux sous-ensembles de même nombre d’individus.
(variable qualitative ordinale ou quantitative)

<u>Variables discrètes</u>

- Cas d’une Série

Pour la trouver, la série doit être écrite en ordre croissant ou décroissant et la médiane correspond à la valeur du centre.

Si (N est impair) Alors
| Me ← Valeur du centre
Sinon
| Me ← Moyenne des deux valeurs du centre
Finsi

Série 1 : 11, 5, 23, 14, 58, 51, 47

Série 1 triée : 5, 11, 14, 23, 47, 51, 58

Ordres : 1 2 3 4 5 6 7

La médiane = 23

- Cas d'un tableau de fréquence

La médiane correspond à la valeur qui a une fréquence cumulée de 50 % de l'effectif totale.

Variables continues et variables discrètes classées

L'estimation de la valeur exacte de la médiane peut être obtenue par interpolation linéaire à l'intérieur de la classe médiane à l'aide de la relation suivante :

$$Me = e_{i-1} + a_i \frac{\frac{n}{2} - Fc_{i-1}}{f_i}$$

e_{i-1} : borne inférieure de la classe médiane

a_i : amplitude de la classe médiane

Fc_{i-1} : fréquence cumulée de la classe précédante de la classe médiane

f_i : fréquence de la classe médiane

n : effectif totale

Exemple:

La répartition des enfants scolarisés par âge en 2010-2011 en France

Classe x_i	*Fréquence*	*Fréq cumulé*
[02 - 06[	2.538.643	2.538.643
[06 - 10[	3.220.753	5.759.396
[10 - 14[	3.174.548	8.933.944
[14 - 18[	2.967.358	11.901.302
[18 - 22[	1.925.926	13.827228

Total = 13.827.228

*C M = (1/2) * Effectif total*

= ***6.913.614***

Classe médiane

$$Me = e_{i-1} + a_i \frac{\frac{n}{2} - Fc_{i-1}}{ni}$$

e_{i-1} : 10

a_i : 04

Fc_{i-1}: 5.759.396

f_i : 3.174.548

n : 13.827.228

$$Me = e_{i-1} + \frac{a_i}{fr_i} \times [0.5 - Fr_{i-1}]$$

fr_i : 0.2296

Fcr_{i-1} : 0.4165

La médiane ≈ 11.45

Inconvénients :

1: Ne peut pas être calculée pour les données qualitatives nominales
2: la majorité des valeurs ne sont pas incluses dans son calcul

3. Le mode :

C'est la valeur de la variable ayant la plus grande fréquence.
On parle aussi de classe modale qui a la fréquence la plus élevée. (variable qualitative ou quantitative)

Variables discrètes

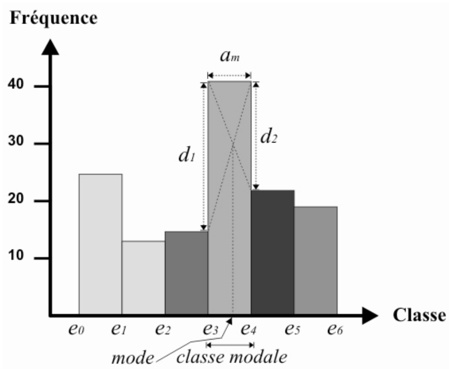
- Si la série contient deux valeurs **(ou plus)** consécutives pour lesquelles la fréquence est la plus élevée, on parle d'une intervalle modale.
- Il peut arriver que la distribution comporte 2 ou plusieurs modes (*non consécutives*), on parle de distribution *bi-modale* ou *pluri-modale*

Variables continues et discrètes classées

En suivant la formule :

$$Mode = e_{i-1} + a_i \frac{d_1}{d_1 + d_2}$$

e_{i-1} : borne inférieure de la classe modale
a_i : amplitude de la classe modale
d_1 : différence de fréquence entre classe modale et sa précédente
d_2 : différence de fréquence entre classe modale et sa suivante



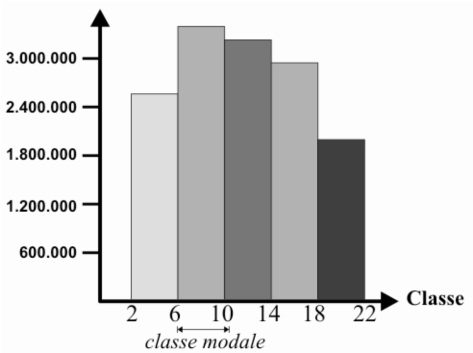
Exemple:

La répartition des enfants scolarisés par âge en 2010-2011 en France

Classe x_i	*Fréquence*
[02 - 06[	2.538.643
[06 - 10[	3.220.753
[10 - 14[	3.174.548
[14 - 18[	2.967.358
[18 - 22[	1.925.926
Total =	*13.827.228*

Classe modale

$$\textbf{Mode} = e_{i-1} + a_i \, \frac{d_1}{d_1 + d_2}$$

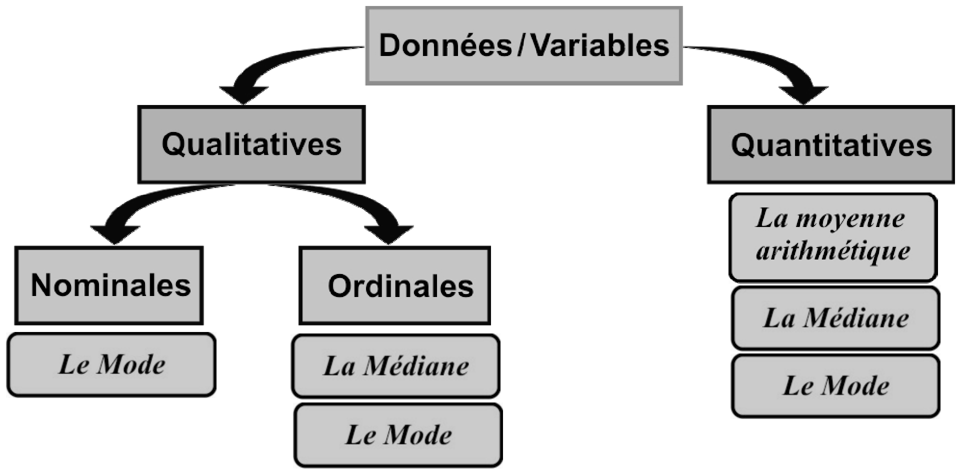


$e_{i-1} = 6$
$a_i = 4$
$d_1 = 682.110$
$d_2 = 46.205$

Le mode $\approx$ *9.75*

Inconvénients :

1: Il peut être inexistant, ou bien il existe plusieurs modes.
2: la majorité des valeurs ne sont pas incluses dans son calcul



II. Mesures de dispersion

L'étude de la variabilité d'une série consiste à analyser ses fluctuations autour d'une valeur centrale.
Voici les mésures les plus utilisées.

1. L'étendue :

Elle est définie comme la différence entre la plus grande et la petite valeur d'une série.

2. La variance :

Elle est définie comme la moyenne arithmétique du carré des écarts des valeurs xi d'une série à leur moyenne arithmétique.

$$V(x) = \frac{1}{N}\sum_{i=1}^{n}(x-\bar{x})^2 = \frac{1}{N}\sum_{i=1}^{k} n_i (x-\bar{x})^2$$

3. L'écart type:

Elle est définie comme la racine carrée de la variance.

$$\delta_x = \sqrt{V(x)} = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x - \bar{x})^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i (x - \bar{x})^2}$$

4. Le coefficient de variation :

Il est désigné par *CV*, se définit par la relation suivante :

$$CV = \frac{\delta_x}{\bar{x}}$$

Le coefficient de variation est une mesure relative de dispersion (puisque l'écart-type est rapporté à la moyenne).
Il s'exprime généralement en %.

Exemple:
le tableau suivant représente le salaire mensuel de 105 salariés d'une entreprise économique (en 1000 D.A).
Calculer la variance et l'écart type des salaires

Classe x_i	*Fréquence*	*Centre xi*	$n_i\ x_i$	$n_i\ x_i^2$
[17 - 22[	11	19.5	214.5	46.010,25
[22 - 27[	23	24.5	563.5	317.532,25
[27 - 32[	43	29.5	1268.5	1.609.092,25
[32 - 37[	17	34.5	586.5	343.982,25
[37 - 42[	7	39.5	276.5	76.452,25
[42 - 50[	4	46	184	33.856
Total =	*105*		*3093.5*	*2.426.925,25*

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

V(x) = 22.245,57

d = 149.15