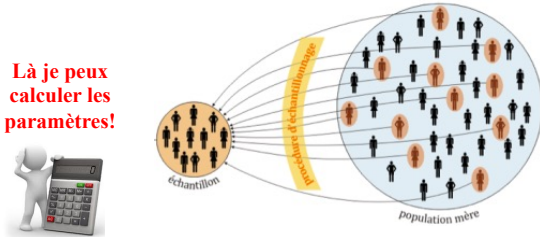


Notions de Statistique descriptive


0

0

Notions	Notions de bases en statistique	
1) Les éléments dont s'occupe la statistique sont appelés :	<input type="checkbox"/> Unité statistique	2) Ces unités sont regroupées dans une : <input type="checkbox"/> Population
3) Lorsque la population est trop importante pour être connue entièrement on prélève un :	<input type="checkbox"/> Échantillon	4) Le nombre calculé sur l'échantillon et qui décrit une des caractéristiques de la population (ex : moyenne, variance...) : <input type="checkbox"/> Paramètre
<p data-bbox="512 1675 627 1749">Là je peux calculer les paramètres!</p> 		

1

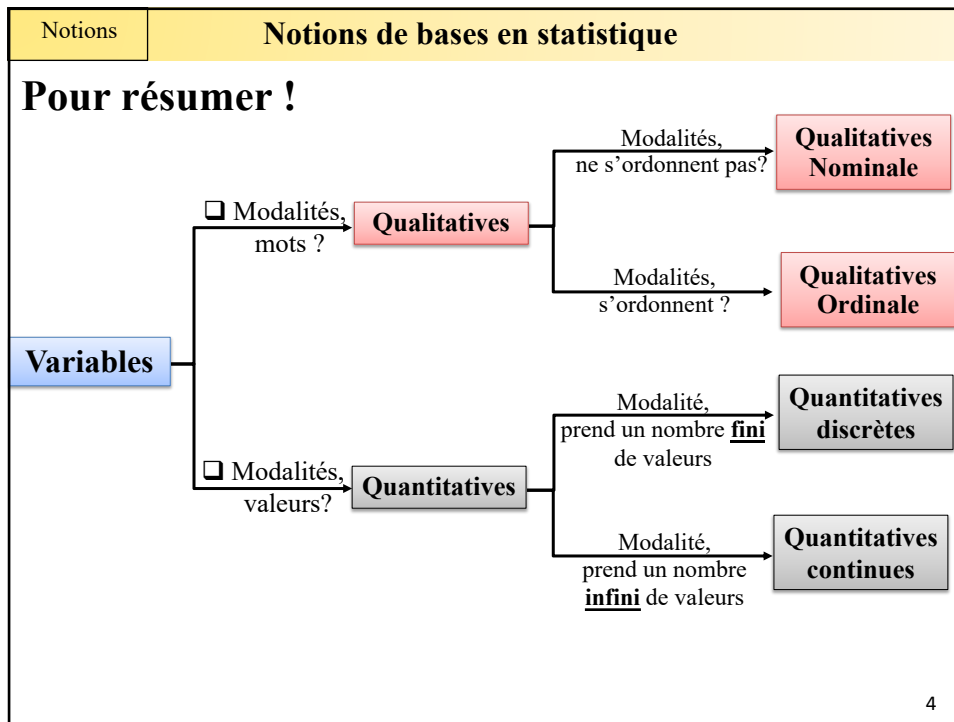
1

Notions	Notions de bases en statistique
	<p>☑ Lorsque les données sont recueillies, la première étape consiste à identifier et décrire les variables statistique ou Caractère.</p> <p style="padding-left: 40px;">5) Est la propriété étudiée sur chaque individu (<u>Unité statistique</u>).</p> <p>📖 Exemple :</p> <p>Une usine fabrique 1 Million de steaks hachés par mois. Pour étudier la contamination de la production en bactérie X. On effectue une analyse sur un lot de 100 barquettes/mois. Il a été estimé une contamination de 3log ± 1.5 UFC/g</p> <div style="text-align: right;">  </div> <ul style="list-style-type: none"> ▪ Variable étudiée : La charge microbienne. ▪ Population statistique : 1million barquettes ▪ L'échantillon : 100 barquettes ▪ Unité statistique : 1Barquette/steak ▪ Paramètres : moyenne (3log) et écart type (1.5log)

2

Notions	Notions de bases en statistique
	<div style="text-align: center;"> <div style="border: 1px solid black; background-color: #a0c4ff; padding: 5px; display: inline-block;">Les variables</div> </div> <div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>☐ Qualitatives</p> <div style="display: flex; justify-content: space-between; width: 100%;"> <div style="text-align: center;"> <p>▪ Ordinales</p> <p>Ex : Degré de sensation: Chaud, très chaud...</p> </div> <div style="text-align: center;"> <p>▪ Nominales</p> <p>Ex : Etat civile : Mariée, célibataire...</p> </div> </div> </div> <div style="text-align: center;"> <p>☐ Quantitatives</p> <div style="display: flex; justify-content: space-around; width: 100%;"> <div style="text-align: center;"> <p>▪ Une variables continues</p> <p>Si elle prend un nombre infini de valeurs (Entre deux valeurs, il peut exister une infinité de valeurs) : Ex : la taille, le poids, la teneur en sucre d'un fruit et toutes les variables mesurables à l'aide d'un instrument.</p> </div> <div style="text-align: center;"> <p>▪ Une variables discrète</p> <p>Si elle prend un nombre fini de valeurs : Ex : nombre d'enfants par foyer, nombre de grappes de raisin par souche, etc.</p> </div> </div> </div> </div> <p style="color: purple; font-size: small;">NB: Une variable statistique ou aléatoire est notée par une lettre majuscule X, Y, et les valeurs qu'elle prend par des lettres minuscules x1, x2..., y1, y2</p>

3




4


Notions	Notions de bases en statistique	
☐ Fréquences absolues, relatives, cumulées		
▪ <i>Fréquence relative :</i>	$f_i = \frac{n_i}{n}$	<p>← n_i est effectif de n à i</p> <p>← n l'effectif total</p>
▪ <i>Effectif cumulé absolue :</i>	$N_i = \sum_{k=1}^i n_k$	
▪ <i>Fréquence relative cumulée :</i>	$F_i = \sum_{k=1}^i f_k$	

5

5

Notions	Notions de bases en statistique
<p> Distribution des fréquences</p> <p>Lorsque n_i est important → Le rangement des data ne suffit pas → <i>Classer</i> → Donner des fréquences à chaque classe</p> <p><input checked="" type="checkbox"/> Distribution des fréquences consiste à regrouper les données dans des classes</p> <p>1. Le nombre de classes par la règle de :</p> <p style="text-align: center;">$YULE : NC = 2.5 \times \sqrt[4]{n}$ ou $STURGE : NC = 1 + (3.3 \times \log_{10} n)$</p> <p>2. L'étendue de la série statistique : $Etendue = Valeur_{max} - Valeur_{min}$</p> <p>3. L'amplitude de classe : $Amplitude = \frac{Etendue}{NC}$</p> <p>4. Construire des classes avec un nombre et une amplitude calculés, incluant les deux valeurs extrêmes (min et max). Ex : Réalisez le classement des notes ?</p> <div style="border: 1px solid #e91e63; padding: 5px; margin-top: 10px; color: #e91e63;"> <p>NB : L'amplitude de classe et le nombre de classe peut être établi d'une façon empirique. Il n'y a cependant pas de règles.</p> </div>	

6

Mesure de tendance centrale																															
<p> Caractéristiques de tendance centrale</p> <p><input type="checkbox"/> Moyenne arithmétique :</p> <div style="display: flex; justify-content: space-around;"> <div style="width: 45%;"> <p>Cas 1 : n données non réparties en classes.</p> $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ </div> <div style="width: 45%;"> <p>Cas 2 : n données réparties en k classes, la classe i étant d'effectif absolu n_i et d'effectif relatif f_i.</p> $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k f_i x_i$ </div> </div> <p><input checked="" type="checkbox"/> Exemples</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>Superficie en ha</th> <th>Nb de fermes</th> <th>Centre de classes</th> <th>Produit</th> </tr> <tr> <th>x_i</th> <th>n_i</th> <th>CC</th> <th>CC * n_i</th> </tr> </thead> <tbody> <tr> <td>[0 ; 10[</td> <td>4</td> <td>5</td> <td>20</td> </tr> <tr> <td>[10 ; 20[</td> <td>10</td> <td>15</td> <td>150</td> </tr> <tr> <td>[20 ; 30[</td> <td>14</td> <td>25</td> <td>350</td> </tr> <tr> <td>[30 ; 40[</td> <td>12</td> <td>35</td> <td>420</td> </tr> <tr> <td>[40 ; 50[</td> <td>10</td> <td>45</td> <td>450</td> </tr> </tbody> </table> <div style="margin-top: 20px; text-align: right;"> <p>La superficie moyenne : $\frac{1390}{50} = 27.8 \text{ ha}$</p> </div>				Superficie en ha	Nb de fermes	Centre de classes	Produit	x_i	n_i	CC	CC * n_i	[0 ; 10[4	5	20	[10 ; 20[10	15	150	[20 ; 30[14	25	350	[30 ; 40[12	35	420	[40 ; 50[10	45	450
Superficie en ha	Nb de fermes	Centre de classes	Produit																												
x_i	n_i	CC	CC * n_i																												
[0 ; 10[4	5	20																												
[10 ; 20[10	15	150																												
[20 ; 30[14	25	350																												
[30 ; 40[12	35	420																												
[40 ; 50[10	45	450																												

7

Mesure de tendance centrale

📖 Médiane

- ❑ La médiane M_e est une moyenne de **position** ; Elle est la valeur qui partage une série en **deux parties** comprenant exactement le même nombre de données de part et d'autre de M_e .

Cas 1 : Effectifs impair : Ordonner (croissant/décroissant) → Trouver le centre
 Ex : {8, 5, 10, 13, 25} → {5, 8, 10, 13, 25} → {5, 8, **10**, 13, 25}

Cas 2 : Effectifs pair : Ordonner → La moyenne des 2 valeurs centrales
 Ex : {8, 5, 10, 13} → {5, 8, 10, 13} → $(8+10)/2$ donc {5, 8, **9**, 10, 13}

📖 Le mode

- ❑ Le **mode** est la valeur la plus fréquente constatée dans les observations.
 Ex : {8, 5, 10, 8, 13, 25, 8} → 8 a la fréquence la + élevée → 8 est la valeur modale

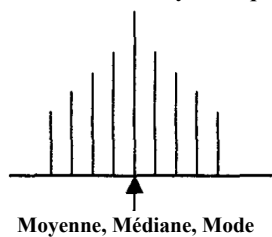
8

8

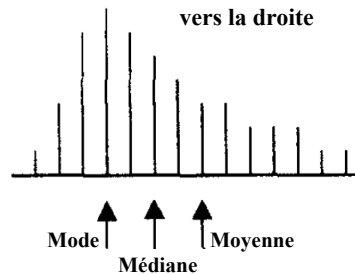
Mesure de tendance centrale

📖 Comparaison des trois indicateurs

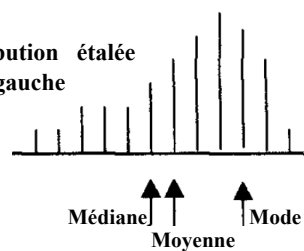
- ☑ Distribution symétrique



- ☑ Distribution étalée vers la droite



- ☑ Distribution étalée vers la gauche



L'étendue, l'intervalle inter quantile, et l'écart type sont les trois paramètres de dispersion absolue les plus courants.

9

9

Mesures de la dispersion

📖 Mesures de la dispersion autour de la médiane

- ❑ **L'étendue (ou l'amplitude)** est la différence entre la plus grande et la plus petite valeur de la distribution : $X = X_{\max} - X_{\min}$

- ❑ **Quantiles (ou fractiles)** sont des caractéristiques de position partageant la série statistique ordonnée en k parties égales ($1/k$). Pour $k = 4$, les quantiles, appelés quartiles, sont trois nombres Q_1 , Q_2 , Q_3 tels que :
 - 25 % des valeurs prises par la série sont inférieures à Q_1 ,
 - 25 % des valeurs prises par la série sont supérieures à Q_3 ,
 - Q_2 est la médiane Me .

- ❑ **L'intervalle interquartile** ($Q_3 - Q_1 = 50\%$) est l'étendue de la distribution sur laquelle se trouvent **concentrée** la moitié des éléments dont les valeurs de X sont les plus **proches de la médiane**. On exclut de la distribution :
 - 25% des valeurs les plus faibles
 - 25% des valeurs les plus fortes

10

10

Mesures de la dispersion

☑ Exemples

{4, 13, 17, 7, 1, 3, 9, 14, 12, 20, 16, 15, 11, 6, 5}

Classer par ordre



{1, 3, 4, 5, 6, 7, 9, 11, 12, 14, 15, 16, 17, 19, 20}

Médiane de chaque série



{1, 3, 4, 5, 6, 7, 9, 11, 12, 14, 15, 16, 17, 19, 20}

{1, 3, 4, 5, 6, 7, 9, 11, 12, 14, 15, 16, 17, 19, 20}

1^{ère} série

Me

2^{ème} série

{1, 3, 4, 5, 6, 7, 9}

{12, 14, 15, 16, 17, 19, 20}

↑
 Q_1

↑
 Q_3

- L'intervalle interquartile : $IQ = Q_3 - Q_1 = 16 - 5 = 11$

11

11

Mesures de la dispersion

Exercice

Ce tableau présente le nombre de mutations par souris traitée par un produit X. A partir de ces données, calculez la médiane et les deux quartiles Q_1 et Q_3 .

Nombre de mutations	0	1	2	3	4	5	6	7
Nombre de souris	7	17	13	14	8	6	0	1

Effectifs cumulés	7	24	37	51	59	65	65	66
--------------------------	----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------


- Effectif total est : $7 + 17 + \dots + 0 + 1 = 66$
- Pour $Me = 66 \times 0.5 = 33$ 66 est paire donc Me est entre 33 et 34^{ème} valeur, donc **Me= 2**
- Pour $Q_1 = 66 \times 0.25 = 16.5$ le est dans la 17^{ème} valeur, donc **$Q_1 = 1$**
- Pour $Q_3 = 66 \times 0.75 = 49.5$ le est dans la 50^{ème} valeur, donc **$Q_3 = 3$**
- Pour IQ : $3-1 = 2$
- Ça veut dire :
 - 25 % des souris ont eu moins d'une mutation
 - 75 % „ „ „ de trois mutation

12


12

Mesures de la dispersion autour de la médiane

Boite à moustache



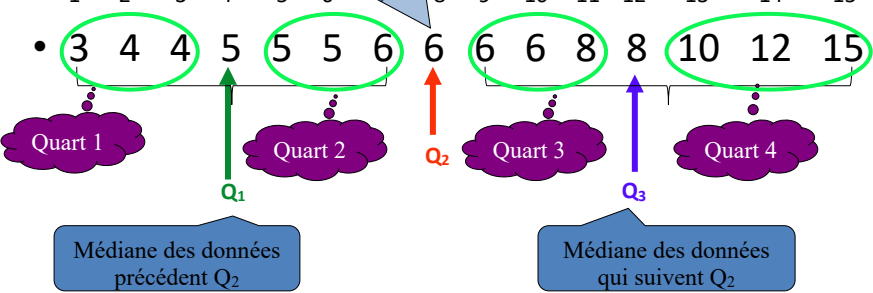
Non!
C'est ça !



Médiane de la distribution 8^{ème} donnée
 $15/2 = 7.5$

• 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

• 3 4 4 5 5 5 6 6 6 6 8 8 10 12 15



Quart 1

Q_1

Médiane des données précédent Q_2

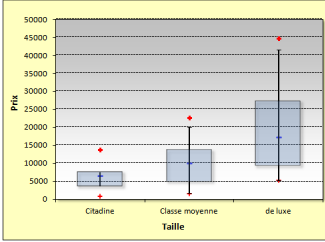
Q_2

Médiane des données qui suivent Q_2

Quart 3

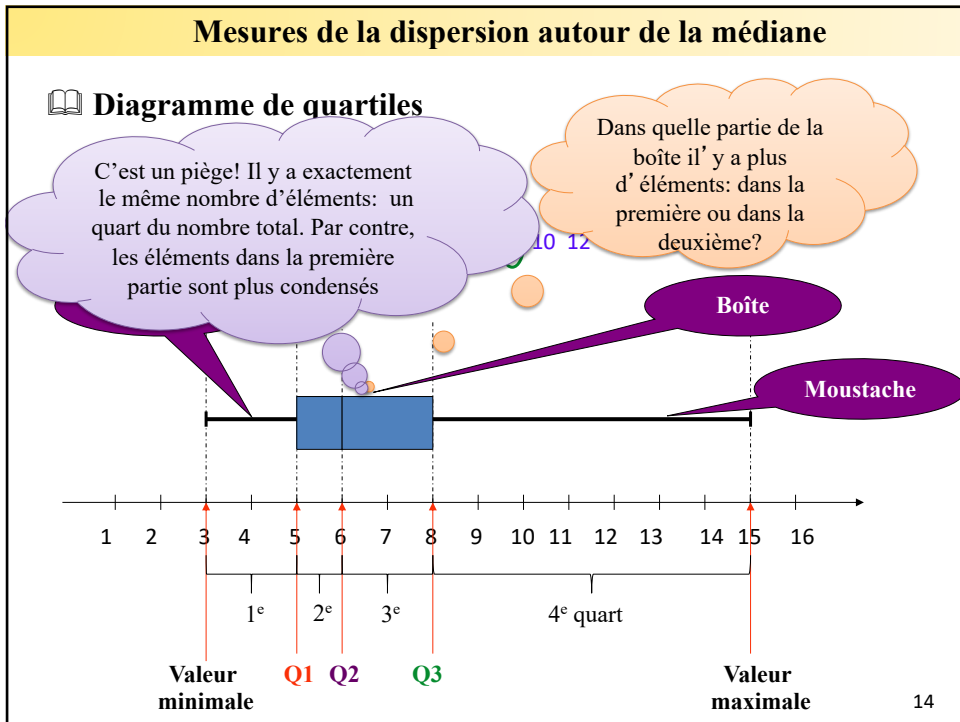
Q_3

Quart 4

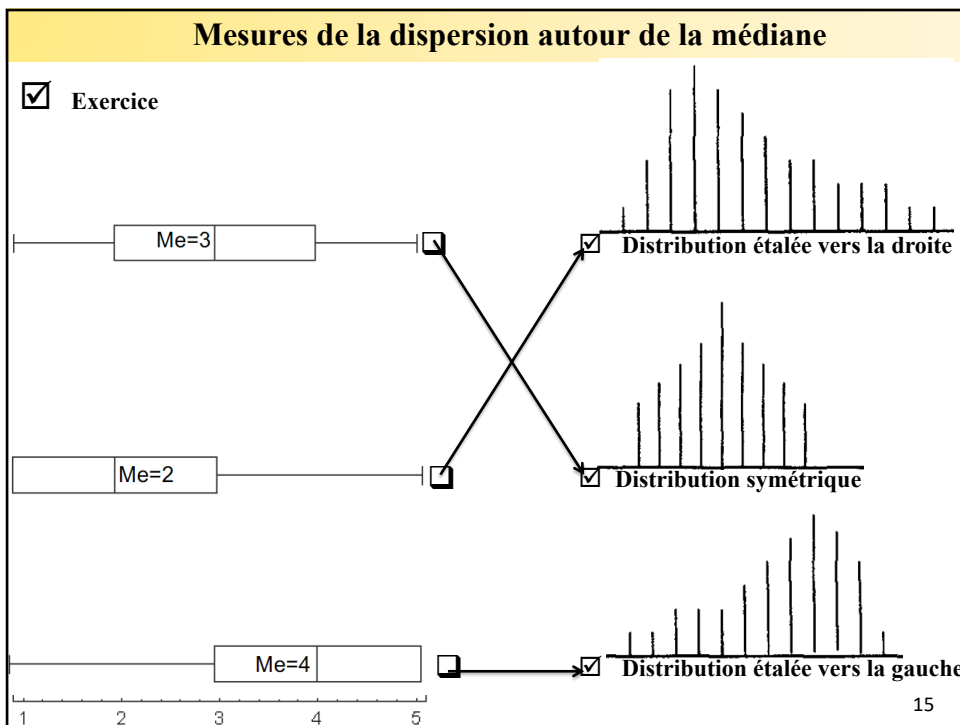


13

13





14




15


Mesures de la dispersion autour de la moyenne


 **Les moyennes**
 Alors, pourquoi la moyenne est si couramment utilisé ? 

- Eh bien, elle est plus facile à calculer !
- Calculer la médiane est embêtant, il fallait trier toutes vos valeurs dans l'ordre !
- Simplement les ajouter tous ensemble, puis diviser par le nombre de valeurs

Arithmétique
 

Moyenne


 Quadratique


 Géométrique

16

16

Mesures de la dispersion autour de la moyenne

La moyenne arithmétique :

- Simple

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

Calculer la moyenne de cette série :
2 4 8 5 2 1 9 7 3

- Pondérée

$$\bar{X} = \frac{1}{n} \sum_{i=1}^h (n_i \cdot x_i)$$

Calculer la moyenne de cette série :
2 4 8 5 2 4 9 4 8 4 1 4

↓

- Élaguée

Calculer la moyenne de cette série :		\bar{X}
1	200 201 201 202 203 204 1990	= 400.3
	200 201 201 202 203 204 1990	= 457.3
1	200 201 201 202 203 204	= 173.1
	200 201 201 202 203 204	= 201.8

17

17

Mesures de la dispersion autour de la moyenne

□ La moyenne géométrique:

▪ Simple

$$G = \left[\prod_{i=1}^n X_i \right]^{\frac{1}{n}}$$

Calculer la moyenne de cette série :

8 5 9 13 25

$$G = [8 \times 5 \times 9 \times 13 \times 25]^{\frac{1}{5}}$$

$$= \sqrt[5]{117000} \cong 10,32$$

$$G = [25^{10} \times 8^{16} \times 4^{25} \times 12^{20}]^{\frac{1}{71}}$$

$$\ln G = \frac{1}{71} [10 \ln 25 + 16 \ln 8 + 25 \ln 4 + 20 \ln 12]$$

$$\ln G = \frac{1}{71} [32,1888 + 32,2711 + 34,6574 + 49,6981]$$

$$\ln G = \frac{149,815}{71} \cong 2,1100704$$

▪ Pondérée

$$G = \left[\prod_{i=1}^h X_i^{n_i} \right]^{\frac{1}{n}}$$

Calculer la moyenne de cette série :

X_i	n_i
25	10
8	16
4	25
12	20

$$G = e^{2,1100704} = 8,2488$$

18

18

Mesures de la dispersion autour de la moyenne

□ La moyenne géométrique:

Est un instrument permettant de calculer des taux moyens, notamment des taux moyens annuels. Son utilisation n'a un sens que si les valeurs ont un caractère multiplicatif.



Soit l'augmentation des prix de l'immobilier est de la façon suivante :

- | Année | Variation % |
|-------|-------------|
| 1 | 9,2 |
| 2 | 12,7 |
| 3 | 8,8 |
| 4 | 7,7 |
| 5 | 3,9 |
| 6 | 1,7 |
| 7 | 0,9 |
| 8 | 2,2 |
| 9 | 4,7 |
| 10 | 3,3 |
- La moyenne arithmétique simple, donne une évolution moyenne de $55,1 / 10 = 5,51 \%$; mais ce résultat est faux compte tenu de la relation entretenue par les taux d'une année sur l'autre.
 - L'utilisation de la moyenne géométrique permet de solutionner ce problème : $G = (1611\ 964,46)^{1/10} = 4,18$
 - Soit une hausse moyenne annuelle de $4,18 \%$.

19

19

Mesures de la dispersion autour de la moyenne

□ *La moyenne harmonique :*

▪ Simple

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Calculer la moyenne H de cette série : 8 5 9 13 25

$$H = \frac{5}{\frac{1}{8} + \frac{1}{5} + \frac{1}{9} + \frac{1}{13} + \frac{1}{25}} = 0,5530342$$

▪ Pondérée


$$H = \frac{n}{\sum_{i=1}^h \frac{n_i}{x_i}}$$

Calculer la moyenne H : Tableau précédant

📖 **Exercice :**

Une machine (a) fabrique 500 barquettes à la vitesse 100 barquettes par heure. Une machine (b) fabrique 300 barquettes à la vitesse 60 barquettes par heure.

Calculer la vitesse moyenne de production dans l'usine.



vitesse moyenne = $\frac{800}{\frac{500}{100} + \frac{300}{60}} = \frac{800}{10} = 80$

= 80 barquette/h

20

20

Mesures de la dispersion autour de la moyenne

□ *La moyenne quadratique :*

▪ Simple

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

Calculer la moyenne de cette série :
-2 4 -7 1 3

▪ Pondérée

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^h (n_i \cdot x_i^2)}$$

Calculer la moyenne de cette série :
-2 5 -4 7 -2 5 -2

Exemple : La formule de la moyenne quadratique sera utilisée dans le calcul de la variance.

21

21

Mesures de la dispersion autour de la moyenne

📖 Variance

☐ Est une mesure **globale de la variation** d'un caractère de part et d'autre de la moyenne arithmétique (quantité d'information). Elle exprime la **dispersion** dans une unité de l'ordre du carré de l'unité de mesure du caractère.

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

↓

Pour faciliter les calculs

➔

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

↓

La formule développée

- Calcul de la moyenne
- Calcul des écarts de la moyenne
- Calcul des carrés des écarts de la moyenne
- Somme des carrés des écarts de la moyenne
- Division par n

- Calcul de la moyenne pondérée au carré
- Calcul des x_i^2
- Calcul des $n_i x_i^2$
- Somme des $n_i x_i^2$
- Division des $n_i x_i^2$ par n
- Soustraction du carré de la moyenne au carré de la moyenne des $n_i x_i^2$

22

22

Mesures de la dispersion autour de la moyenne

☑ Exercice : Calculer la variance des données suivantes :

x_i	2	6	9	11	15
n_i	5	9	4	3	5

La formule simple

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

La formule développée

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

x_i	n_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$n_i (x_i - \bar{x})^2$
2	5	-6	36	180
6	9	-2	4	36
9	4	1	1	4
11	3	3	9	27
15	5	7	49	245
Somme	26			492

x_i	n_i	$n_i x_i$	x_i^2	$n_i x_i^2$
2	5	10	4	20
6	9	54	36	324
9	4	36	81	324
11	3	33	121	363
15	5	75	225	1125
Somme	26	208		2156

$\bar{x} = 208/26 = 8$

$\sigma^2 = 492/26 = 18,92$

$\bar{x} = 208/26 = 8$

$\sigma^2 = (1/26 \times 2156) - (8)^2 = 18,92$

23

23

Mesures de la dispersion autour de la moyenne

Ecart type

Ecart type est égale à la **racine carrée** de la variance



Est l'indicateur le plus souvent utilisé pour mesurer la dispersion d'une série

Coefficient de variation

$$CV = \left(\frac{\sigma}{\bar{x}} \right) \times 100$$

- **L'écart type** est exprimé dans la **même unité** que celle de la variable étudiée.
- **Mais...** on ne peut pas comparer des variables **de nature différente**.
- Avec... le **coefficient de variation** on peut effectuer **des comparaisons** entre variables de nature différente.

24

24

Mesures de la dispersion autour de la moyenne

La covariance

La covariance permet d'étudier les variations simultanées de **deux variables** (X, Y) par rapport à **leur moyenne** respective ; contrairement à la variance qui permet d'étudier les variations d'une variable par rapport à elle-même (une moyenne).

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$



La covariance mesure la relation linéaire entre deux variables.

- **Covariance (+)** : indiquent que les valeurs supérieures à la moyenne d'une variable (X) sont associées aux valeurs **supérieures** à la moyenne de l'autre variable (Y) et que les valeurs inférieures à la moyenne sont associées de **façon similaire**.
- **Covariance (-)** : indiquent que les valeurs supérieures à la moyenne d'une variable sont associées aux valeurs **inférieures** à la moyenne de l'autre variable.

25

25

Association de variables

Il est intéressant d'étudier les variables simultanément (deux variables).

↓

Et on s'interroge souvent sur les liens qui peuvent exister entre elles.

📖 **Représentation des données**

▪ **Tableaux**

Branche	Statut	Salariés	Patrons	Travailleurs familiaux	Total
Agriculture		95	30	125	250
Industrie		160	25	100	285
Commerce		100	40	90	230
Fonction Publique		60	0	0	60
Services		80	15	70	165
Total		495	110	385	990

	A	B	C
1 Individu		Masse en kg	Glycémie
2 A		65	0,9
3 B		92	1,1
4 C		82	1
5 D		78	1
6 E		105	1,3
7 moyenne		84,4	1,06

▪ **Graphique**

Autres

26

26

Association de variables

📖 **Relation entre variables**

On dit qu'il existe une relation entre X et Y (caractères quantitatifs) si l'attribution des modalités de X et de Y ne se fait pas au **hasard**, c'est à dire si les valeurs de X dépendent des valeurs de Y ou vice versa.

↓

Y dépend de X signifie que la connaissance des valeurs de X permet de **prédire** Y. En d'autres termes, si Y dépend de X, on peut trouver une **fonction f** telle que :

$$Y = f(X)$$

↓

La relation entre X et Y peut être établit via un **courbe de régression** ou un **diagramme corrélation** croisant les modalités de X et de Y.

↓

Chaque élément *i* est représenté par le point de coordonnées (X_i, Y_i) . L'ensemble des points forme un **nuage de points** qui permet de caractériser la *forme*, le *sens* et l'*intensité* de cette la relation.

27

27

