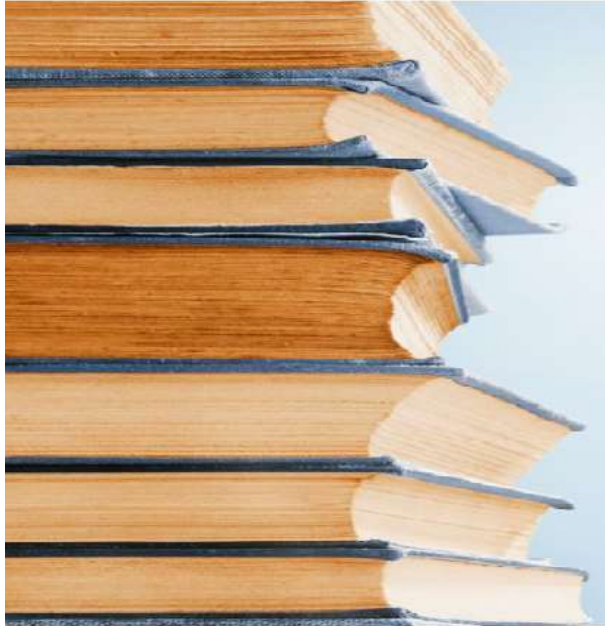


BIOSTATISTIQUE



Master 1. Nutrition et Contrôle de qualité

Dr AIDOUZ Aziouz
Maitre de Conférence
Faculté des Sciences
Université M'Hamed Bougara. Boumerdes

Contenu de la matière : Biostatistique

- 1.- Rappels sur certaines notions statistiques
 - 1.1.- La distribution de fréquence
 - 1.2.- Moyenne arithmétique et moyenne pondérée
 - 1.3.- Mode et médiane
 - 1.4.- Variance, écart type et coefficient de variation
- 2.- La régression linéaire
 - 2.1.- Introduction à la régression linéaire
 - 2.2.- Les modèles de la régression
 - 2.3.- La réalisation de la régression
 - a.- Les différentes étapes de la régression
 - b.- Les tests de signification de la régression
 - c.- Les dangers de la régression
- 3.- L'analyse de variance
 - 3.1.- Introduction à l'analyse de variance
 - 3.2.- Les modèles en analyse de variance

Travail personnel

Réaliser des exercices de statistiques en relation avec la spécialité



Généralités

La statistique est l'étude de la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.

L'analyse des données est utilisée pour décrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

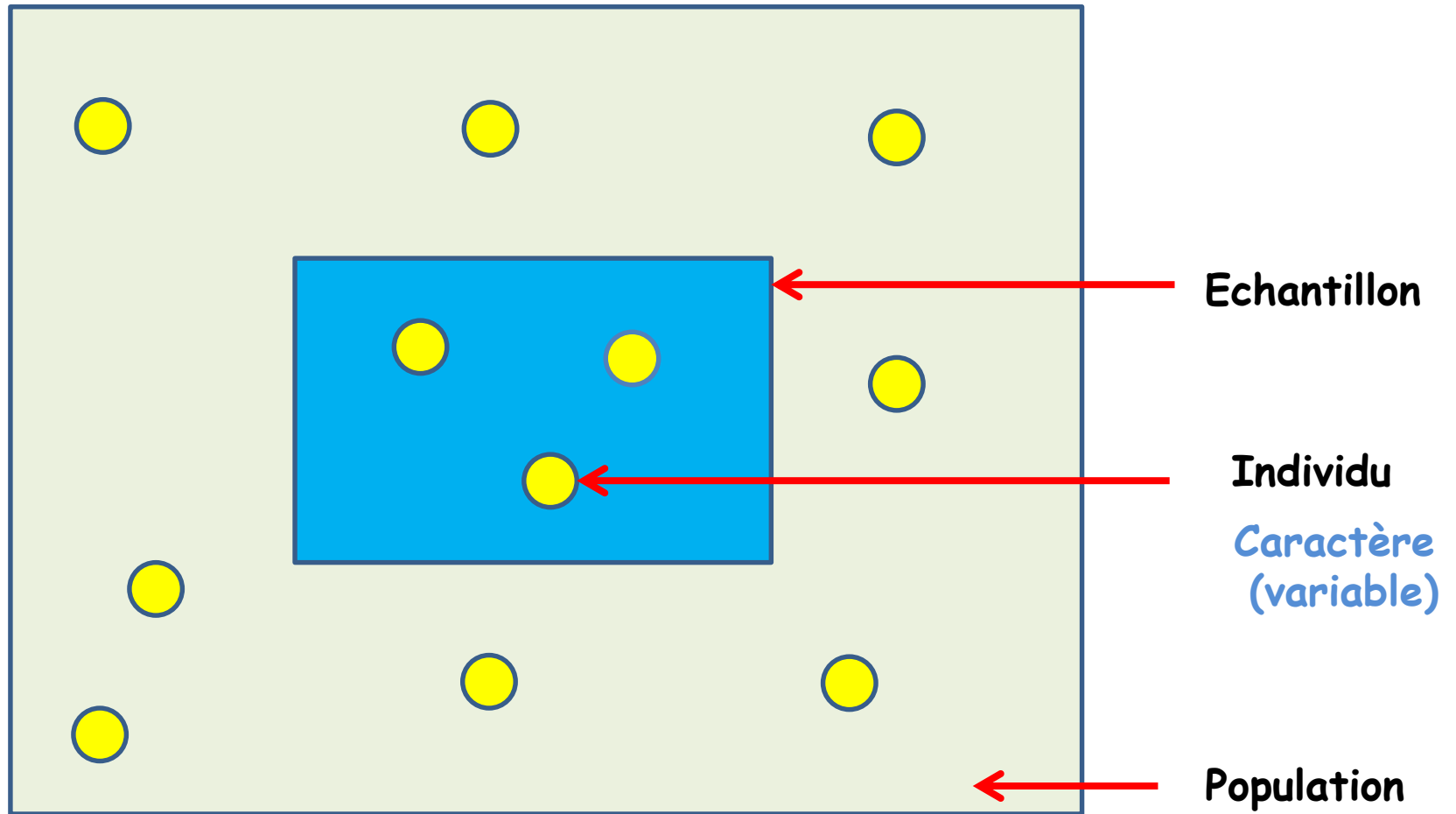
La statistique consiste à :

- Recueillir des données.
- Présenter et résumer ces données.
- Tirer des conclusions sur la population étudiée et d'aider à la prise de décision.
- En présence de données dépendant du temps, nous essayons de faire de la prévision.

Statistique descriptive:

Ensemble des méthodes permettant de décrire, de résumer et de présenter les données observées sous la forme la plus accessible : Tableaux, graphiques, pourcentages et indicateurs ou résumés numériques (moyenne, médiane.....)

Vocabulaire



Population : Ensemble des personnes , objets ou éléments sur lesquels on veut effectuer l'étude statistique.
Cet ensemble est noté Ω .

Ω = ensemble des étudiants.

Ω = ensemble des véhicules.

Echantillon : c'est un sous-ensemble de la population

Individu (unité statistique) : On appelle individu tout élément de la population Ω , il est noté ω (ω dans Ω).

- un individu est tout étudiant de la section

- Si on étudie la production annuelle d'une usine de boîtes de boisson en métal (canettes).

La population est l'**ensemble des boîtes** produites durant l'année et **une boîte** constitue un **individu**.

Variable (caractère) : caractéristique relative à chacun des individus de la population

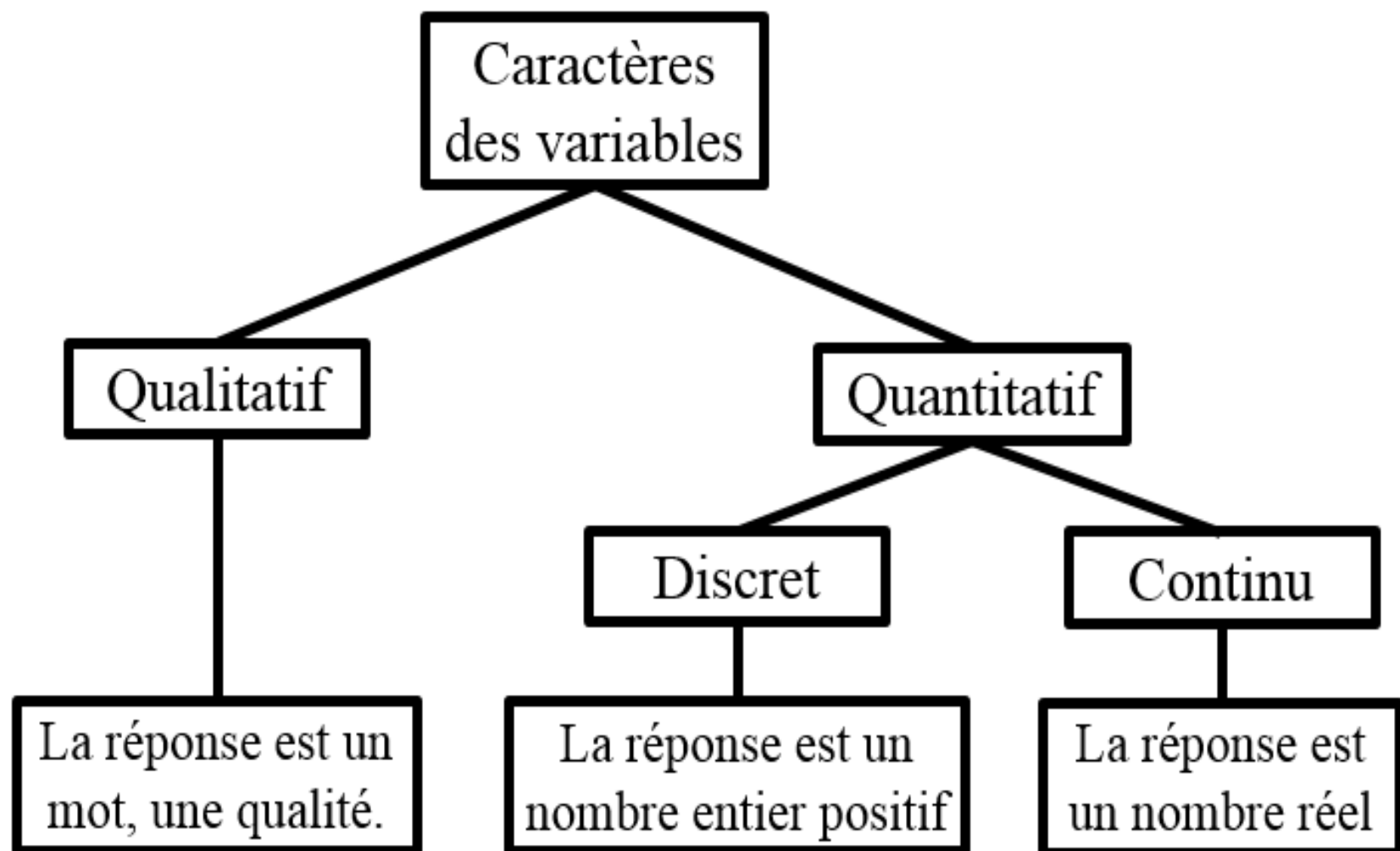
Taille, température, nationalité, couleur des yeux, catégorie socioprofessionnelle ...

Caractères statistiques

Caractère = propriété observable des individus qui prend différents états appelés **modalités**.

(homologue de la **variable aléatoire** en probabilités, les deux termes sont utilisés).

Plusieurs catégories de caractères => méthodes statistiques différentes.



Quelques exemples

❖ Sexe à la naissance d'un individu

Caractère qualitatif à deux modalités (masculin/féminin , peut être codifié en 1/0)

❖ Couleurs de papillons, couleurs de fleurs

Caractère qualitatif à n modalités (n : nombre de couleurs)

❖ Poids, taille, périmètre crânien, diamètre d'un œuf...

Caractères quantitatifs continus

❖ Nombre de sépales , de pétales d'une fleur

Caractères quantitatifs discrets

Statistiques descriptives pour une variable

Dans les crèches de la région parisienne , on a relevé en 1975 un certain nombre de paramètres sur une population de 694 garçons et 622 filles.

Le tableau suivant donne la distribution de périmètres crâniens pour les nouveaux nés considérés

k	Périmètre crânien (cm)	Nombre de garçons	Nombre de filles
1	[26, 27[1	1
2	[27, 28[0	0
3	[28, 29[0	0
4	[29, 30[1	1
5	[30, 31[1	3
6	[31, 32[13	17
7	[32, 33[36	45
8	[33, 34[92	118
9	[34, 35[171	191
10	[35, 36[217	151
11	[36, 37[107	67
12	[37, 38[45	21
13	[38, 39[5	5
14	[39, 40[5	2

Ω : ensemble des nouveaux nés de 1975 des crèches de la région parisienne

Echantillon : se réduit aux crèches étudiées

Si les **sexes de nouveaux nés** étaient étudiés ,
ce serait un **caractère qualitatif**

Le nombre de modalités de ce caractère est 2.

Le périmètre crânien est un caractère quantitatif continu

w : (Individu)

M_G 694

M_F 622

Ranger les données par classes d'amplitudes 1 cm pour le périmètre crânien

Pour ne pas compter deux fois un individu dans deux classes Connexes , il faut convenir d'intervalles semi-ouverts ici
Du type $[, [$.

Pour la 6^{ème} classe , on a observé :

$n_6^G = 13$ garçons

← Indice « sexe »

← Indice « classe »

Dont le périmètre crânien \mathcal{C} [31, 32[

De façon générale : n_i est l'effectif de la i -ème classe (indice supérieur G désigne ici garçon) = le rapport $f_i = n_i/N$ (ici

(ici n_i^G / N_G) de l'effectif de la classe à la population totale est appelée **fréquence relative** :

$$f_i = \sum_{i=1}^{14} \frac{n_i^G}{N_G} = \frac{1}{N_G} \sum_{i=1}^{14} n_i^G$$

Exemple

Sexe	Coca Cola	Age
M	1	30
F	0	34
F	1	28
F	1	31
F	1	33
F	1	28
F	0	33
M	1	30
M	1	40
M	0	30
M	0	30
M	1	31

Coca – Cola : **Oui : 1**
 Non : 2

Variable qualitative : modalités (k)
(1 ou 0)

Variable quantitative

Discrète : $\in \mathbb{N} : (1, 2, 3, 4, \dots)$

Continue : (l'Age) -- [, [

Identifiez le type (et le sous-type) des variables suivantes :

- | | |
|--|-------------------------|
| a) Le nombre d'animaux par laboratoire ;
discret | a) quantitatif |
| b) La niche écologique principale ; | b) qualitatif nominal |
| c) Le modèle de matériel utilisé ;
nominal | c) qualitatif |
| d) La distance en kilomètre entre le prélèvement A et le prélèvement B ; | d) quantitatif continue |
| e) Être végétarien ou non ;
ordinal | e) qualitatif |
| f) Le temps passé à observer le comportement X ;
continue | f) quantitatif |
| g) Avoir ou non une réponse;
ordinal | g) qualitatif |
| h) Le nombre de frères et sœurs.
discret | h) quantitatif |

Caractères statistiques

Caractères qualitatifs

Ne résultent ni d'une mesure par un instrument ni d'un comptage.

* **nominale** : modalités exprimables par des noms et non hiérarchisées. (dichotomique = 2 modalités).

Couleur des yeux, nationalité, présence/absence d'une maladie

* **ordinaire** : traduit le degré d'un état sans que ce degré ne puisse être défini par un nombre. Modalités hiérarchisées.

Stade d'une maladie +, ++, +++

Caractères statistiques

Caractères quantitatifs

= mesurables ; résultent d'une mesure ou d'un comptage.

* **discret** : il peut prendre seulement certaines valeurs = résulte d'un comptage (transformation d'une variable nominale)

Nombre de jeunes par portée, nombre d'individus porteurs d'un caractère

* **continu** : peut prendre n'importe quelle valeur dans un intervalle donné.

Le poids, l'âge, la glycémie

Caractères statistiques

Quantitatif continu

n = taille de l'échantillon

Continues non groupées :

$$X_1, X_2, \dots, X_i, \dots, X_n$$

Continues groupées en k classes ou discrètes:

Classe	$]x_1, x_2]$	$]x_2, x_3]$				
X'	x'_1	x'_2	\dots	x'_i	\dots	x'_k
Effectif	n_1	n_2	\dots	n_i	\dots	n_k

X'_i : médiane de la classe i

Caractères statistiques

Quantitatif discret

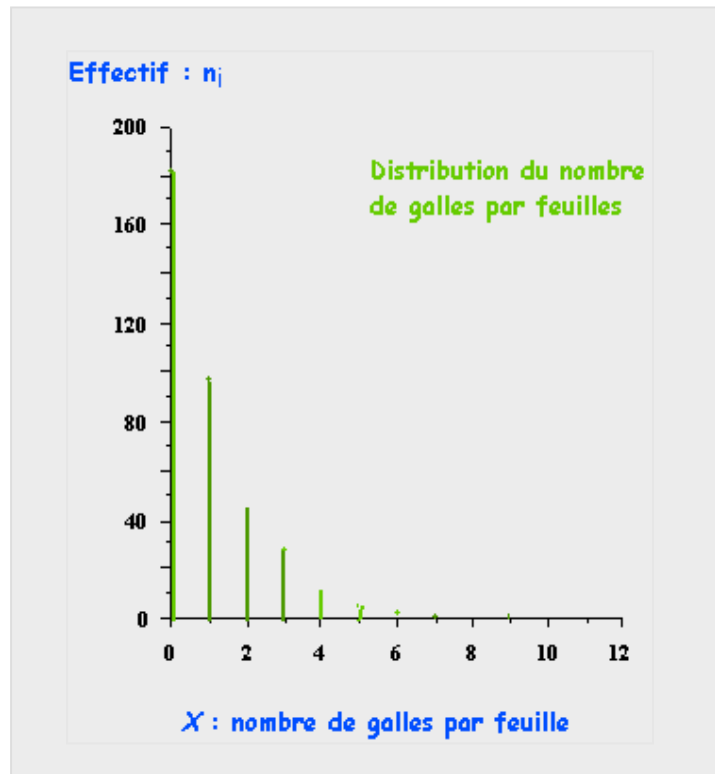
n = taille de l'échantillon

X	x_1	x_2	...	x_i	...	x_k
Effectif	n_1	n_2	...	n_i	...	n_k

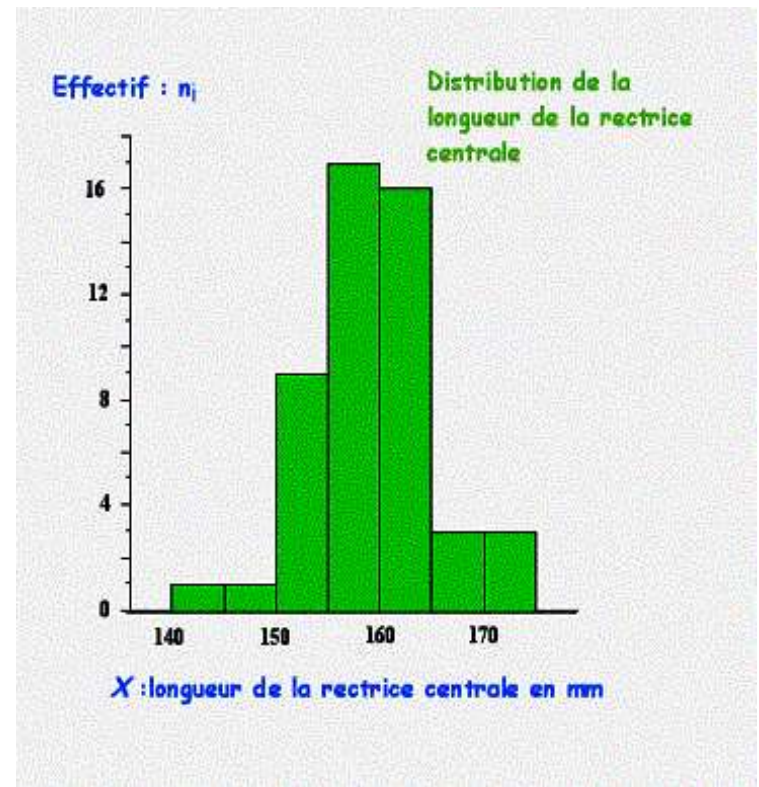
n_i : effectif de la modalité i

Représentation graphique

Caractère discret



Caractère continu



Représentation graphique

Analyse graphique de la distribution

Quantile observé : valeur observée y_i telle que

$$P(Y \leq y_i) = i/n$$

Quantile théorique: valeur x_i telle que

$$P(X \leq x_i) = i/n \text{ / } X \sim N$$

Est-elle unimodale?
Est-elle symétrique?

Peut-on la rattacher à
une distribution
normale?

Indicateurs statistiques

1. Indicateurs de position

▪ La moyenne arithmétique

Soit un échantillon de n valeurs observées $x_1, x_2, \dots, x_i, \dots, x_n$

Données non groupées $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

Données groupées pour caractère discret $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i$

Données groupées pour caractère continu $\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x'_i$

Indicateurs statistiques

1. Indicateurs de position

▪ La moyenne arithmétique

- Facile à calculer
- La somme des écarts à la moyenne est nulle:

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

- Fortement influencée par les valeurs extrêmes
- Représente mal une population hétérogène (polymodale)

Indicateurs statistiques

1. Indicateurs de position

▪ La médiane

La médiane est la valeur de la variable pour laquelle la fréquence cumulée est égale à 0,5 ou 50%. Elle correspond à la valeur pour laquelle 50% des valeurs observées sont supérieures et 50% sont inférieures.

Si n est impair, la médiane est la valeur au rang $(n+1)/2$

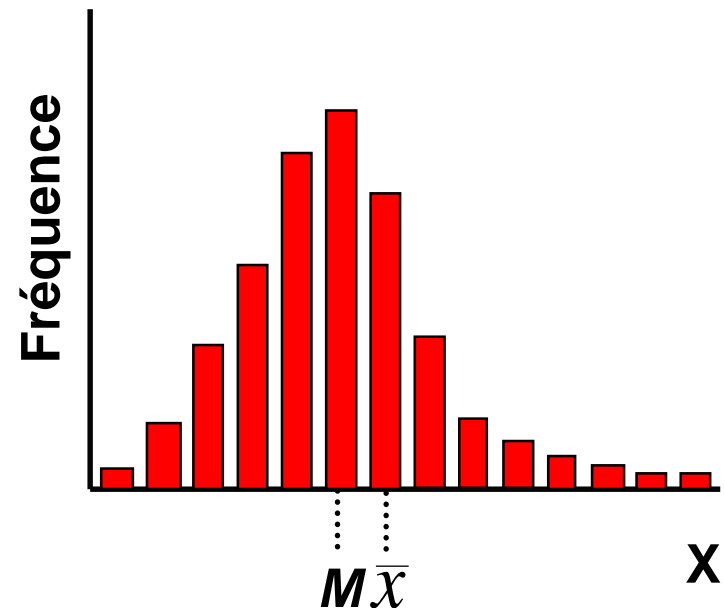
Si n est pair, la médiane est la **moyenne** des valeurs aux rangs $n/2$ et $(n/2) + 1$

Indicateurs statistiques

1. Indicateurs de position

■ La médiane

- Pas influencée par les valeurs extrêmes
- Peu sensible aux variations d'amplitude des classes
- Se prête mal aux calculs statistiques



Indicateurs statistiques

2. Indicateurs de dispersion

▪ La variance

Soit un échantillon de n valeurs observées $x_1, x_2, \dots, x_i, \dots, x_n$

Données non groupées $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$

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

Données groupées continues $s^2 = \frac{1}{n} \sum_{i=1}^k n_i (x'_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x'^2_i - \bar{x}^2$

Indicateurs statistiques

2. Indicateurs de dispersion

■ L'écart-type

C'est la racine carrée de la variance $\sigma = \sqrt{s^2}$

(en anglais « standard deviation », « s.d. »)



Ne pas confondre avec l'erreur standard $\sqrt{\frac{s^2}{n-1}}$
:

(en anglais « standard error », « s.e. »)

Données brutes

Données provenant d'un échantillon de 17 adultes canadiens, janvier 2015.

Identifiant du répondant	Sexe	Âge	Revenu annuel brut	Niveau de scolarité	Nombre d'enfants
1	H	39	63751	universitaire	4
2	F	24	28604	collégial	1
3	F	57	81379	universitaire	3
4	H	54	16641	secondaire	2
5	F	40	72865	collégial	3
6	H	68	69532	collégial	3
7	H	59	17735	secondaire	1
8	H	61	34369	universitaire	2
9	F	21	17206	secondaire	0
10	F	26	15195	collégial	1
11	H	53	54268	universitaire	4
12	F	65	63743	secondaire	3
13	F	70	84963	collégial	0
14	F	45	54040	universitaire	2
15	H	34	56388	collégial	1
16	F	27	74159	collégial	0
17	H	70	24819	collégial	2

Tableau de distribution de fréquences

Tableau permettant de regrouper les données brutes selon leurs modalités ou leurs valeurs et de mentionner la fréquence absolue ou la fréquence relative de chacune d'elles.

Variable qualitative nominale ou à échelle nominale

Données provenant d'un échantillon de 17 adultes canadiens, janvier 2015.

Identifiant du répondant	Sexe	Âge	Revenu annuel brut	Niveau de scolarité	Nombre d'enfants
1	H	39	63751	universitaire	4
2	F	24	28604	collégial	1
3	F	57	81379	universitaire	3
4	H	54	16641	secondaire	2
5	F	40	72865	collégial	3
6	H	68	69532	collégial	3
7	H	59	17735	secondaire	1
8	H	61	34369	universitaire	2
9	F	21	17206	secondaire	0
10	F	26	15195	collégial	1
11	H	53	54268	universitaire	4
12	F	65	63743	secondaire	3
13	F	70	84963	collégial	0
14	F	45	54040	universitaire	2
15	H	34	56388	collégial	1
16	F	27	74159	collégial	0
17	H	70	24819	collégial	2

Mot clé

Qui ?

Où ?

Distribution des 17 adultes canadiens
interrogés selon leur sexe, janvier 2015.

Quoi ?

Quand ?

Distribution des 17 adultes canadiens interrogés selon leur sexe, janvier 2015.

Sexe	Nombre d'adultes canadiens	Pourcentage d'adultes canadiens
Femme	9	52,94%
Homme	8	47,06%
Total	17	100,00%

Distribution des 17 adultes canadiens interrogés selon le plus haut niveau de scolarité qu'ils ont complété, janvier 2015.

Niveau de scolarité	Nombre d'adultes canadiens	Pourcentage d'adultes canadiens	Pourcentage cumulé d'adultes canadiens
Secondaire	4	23,53%	23,53%
Collégial	8	47,06%	70,59%
Universitaire	5	29,41%	100,00%
Total	17	100,00%	

Variable quantitative discrète (ayant un petit nombre de valeurs différentes)

Répartition des 17 adultes canadiens interrogés
selon le nombre d'enfants qu'ils ont, janvier 2015.

Nombre d'enfants	Nombre d'adultes canadiens	Pourcentage d'adultes canadiens	Pourcentage cumulé d'adultes canadiens
0	3	17,65%	17,65%
1	4	23,53%	41,18%
2	4	23,53%	64,71%
3	4	23,53%	88,24%
4	2	11,76%	100,00%
Total	17	100,00%	

Variable quantitative continue (ou quantitative discrète ayant un grand nombre de valeurs différentes)

Distribution des 17 adultes canadiens interrogés selon leur revenu annuel, janvier 2015.

Revenu annuel (\$)	Nombre d'adultes canadiens	Pourcentage d'adultes canadiens	Pourcentage cumulé d'adultes canadiens
[15 000; 30 000[6	35,29%	35,29%
[30 000; 45 000[1	5,88%	41,18%
[45 000; 60 000[3	17,65%	58,82%
[60 000; 75 000[5	29,41%	88,24%
[75 000; 90 000[2	11,76%	100,00%
Total	17	100,00%	

Source : données fictives.

Résumé

Mot clé

Qui?

Où?

Quoi?

Quand?

Distribution (ou répartition) des *unités statistiques étudiées* selon la *variable étudiée, moment de l'étude*.

Variable	Nombre d'unités statistiques	Pourcentage d'unités statistiques	Pourcentage cumulé d'unités statistiques
Modalités ou valeurs (placées en ordre)	Nombres d'unités statistiques pour chaque modalité ou valeur	Pourcentages d'unités statistiques pour chaque modalité ou valeur	Pourcentages cumulés d'unités statistiques pour chaque modalité ou valeur et celles qui précèdent
Total	Nombre total d'unités statistiques	100%	

La moyenne arithmétique

Soit un échantillon de n valeurs observées $x_1, x_2, \dots, x_i, \dots, x_n$

Données non groupées

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

cas de serie statistique

2 3 4 5 7 1 2 3 4 8

$$h = 10$$

$$\sum x_i = (2 + 3 + 4 + 5 + 7 + 1 + 2 + 3 + 4 + 8) = 39$$

$$\bar{x} = \frac{39}{10} = 3,9$$

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

Données groupées pour caractère discret

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

Données groupées pour caractère continu

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

③ Les moyennes

① La moyenne arithmétique \bar{X}

La moyenne arithmétique est égale à la somme des valeurs observées divisée par le nombre d'observations.

* Cas discret

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i X_i = \sum_{i=1}^k f_i X_i$$

Exemple

X_i	n_i	$n_i X_i$
1	40	40
2	80	160
3	60	180
4	20	80
Total	200	460

X_i	n_i	C_i	$n_i C_i$
[0,10[60	5	300
[10,20[120	15	1800
[20,30[140	25	3500
[30,40]	80	35	2800
Total	400		8400

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i C_i = \frac{\sum_{i=1}^k n_i C_i}{n}$$

$$\bar{X} = \frac{8400}{400}$$

$$\boxed{\bar{X} = 21}$$

② La moyenne géométrique

$$G = e^{\log G}$$

$$\log G = \frac{1}{n} \sum_{i=1}^k n_i \log(X_i)$$

xi	ni	nixi
1	40	
2	80	
3	60	
4	20	
Total	200	

+ Cas continue

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i C_i = \sum_{i=1}^k p_i C_i$$

C_i : centre de classe

$$C_i = \frac{\text{born sup} + \text{born inf}}{2}$$

Exemple

On a une classe [10, 20[

$$C_i = \frac{10 + 20}{2} = 15$$

Exemple

X_i	n_i	C_i	$n_i C_i$
[0, 10[60	5	300
[10, 20[120	15	1800
[20, 30[140	25	3500
[30, 40]	80	35	2800
Total	400		8400

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i C_i = \frac{\sum n_i C_i}{n}$$

$$\bar{X} = \frac{8400}{400}$$

$$\boxed{\bar{X} = 21}$$

② La médiane:

La médiane est la valeur qui partage la série statistique en deux parties égales.

* Cas d'un caractère discret:

Pour le cas discret la médiane est déterminée par l'observation.

Exemple

Effectif cumulé croissant

X_i	n_i	ECC
1	40	40
2	80	120
3	60	180
4	20	200
Total	200	

$\Rightarrow 120 > 100$

* Cas d'un caractère continu

Pour le cas continu la médiane est calculée par la relation suivante:

$$Me = \text{born inf} + a_i \cdot \frac{n/2 - N_{i-1}}{n_i}$$

$n/2$: total divisé sur 2

N_{i-1} : ECC du classe avant la classe médiane

n_i : l'effectif du classe médiane

Exemple

X_i	n_i	ECC	a_i
$[0, 5[$	40	40	5
$[5, 10[$	80	120	5
$[10, 15[$	60	180	5
$[15, 20]$	20	200	5
Total	200		

Frequency distribution table

The following figures are the ages of patients admitted to a hospital with poliomyelitis..

8, 24, 18, 5, 6, 12, 14, 3, 23, 9, 18, 16, 1, 2, 3, 5, 11, 13, 15, 9, 11, 11, 7, 10, 6, 9, 5, 16, 20, 4, 3, 3, 3, 10, 3, 2, 1, 6, 9, 3, 7, 14, 8, 1, 4, 6, 4, 15, 22, 2, 1, 4, 6, 4, 15, 22, 2, 1, 4, 7, 1, 12, 3, 23, 4, 19, 6, 2, 2, 4, 14, 2, 2, 21, 3, 2, 1, 7, 19.

Age	Number of patients
0-4	35
5-9	18
10-14	11
15-19	8
20-24	6

Frequency distribution table

- The following figures are the ages of patients admitted to a hospital with poliomyelitis..

8, 24, 18, 5, 6, 12, 14, 3, 23, 9, 18, 16, 1, 2, 3, 5, 11, 13, 15, 9, 11, 11, 7, 10, 6, 9, 5, 16, 20, 4, 3, 3, 3, 10, 3, 2, 1, 6, 9, 3, 7, 14, 8, 1, 4, 6, 4, 15, 22, 2, 1, 4, 6, 4, 15, 22, 2, 1, 4, 7, 1, 12, 3, 23, 4, 19, 6, 2, 2, 4, 14, 2, 2, 21, 3, 2, 1, 7, 19.

Age	Number of patients
0-4	35
5-9	18
10-14	11
15-19	8
20-24	6

- Mean = $\frac{\text{sum of all values}}{\text{total no. of values}}$
- Median = middle value (when the data are arranged in order.)
- Mode = most common value

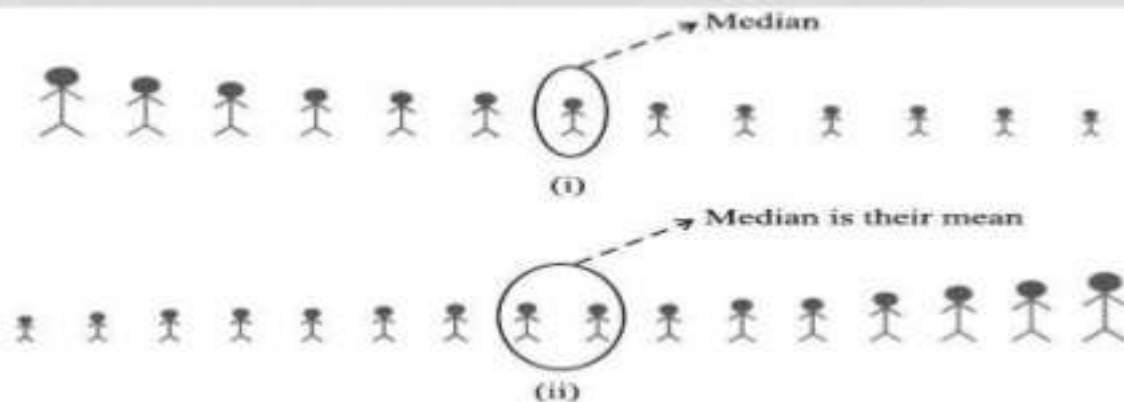


Fig. 9

- For eg..
the income of 7 people per day in rupees are as follows.
5, 5, 5, 7, 10, 20, 102= (total 154)
- Mean = $154/7 = 22$
- Median= 7
- Median, therefore, is a better indicator of central tendency when more of the lowest or the highest observations are wide apart .
- Mode is rarely used as series can have no modes, 1 mode or multiple modes.

- **Standard Deviation :-** most frequently used
“Root Mean Square Deviation”
denoted by greek letter σ or by initials

$$\text{S.D.} = \text{Square root of } \frac{\sum (X - \bar{X})^2}{n}$$

- if sample size is less than 30 in denominator, $(n-1)$
- S.D. gives us idea of the spread of dispersion .
- Larger the standard deviation, greater the dispersion of values about the mean

La médiane

$$n=9$$

$$n/2 = 4,5$$

1 1 1 1 2 2 2 3 3

④

↓

⑤

$$Me = 2$$

$$n = 10$$

$$n/2 = 5$$

1 1 1 1 2 2 2 3 3 4

④

⑤

$$\frac{2+2}{2} = 2 \quad Me = 2$$

La variable statistique, chaque attribut (ou caractère ou caractéristique) a des modalités, ou peut s'exprimer selon une mesure, celles-ci varient d'un individu à l'autre ou d'un groupe d'individus à un autre groupe d'individus. La variable statistique est le nom que l'on donne à ces caractères (attributs, caractéristiques).

Variable quantitatif : c'est un caractère auquel on peut associer un nombre c'est-à-dire, pour simplifier, que l'on peut "mesurer" (grandeur mesurable). Les différentes situations où peuvent se trouver les éléments sont des *mesures*; elles sont ordonnables et la moyenne a une signification. On distingue alors deux types de caractère quantitatif

a - Variable discrète ou discontinue : c'est un caractère quantitatif, un tel caractère ne prend qu'un nombre fini de valeurs (valeur entière dénombrable et sans aucune valeur intermédiaire). Les différentes situations où peuvent se trouver les éléments (observations, mesures, valeurs,...) sont des nombres isolés dont la liste peut être établie a priori. Exemple: (nombre d'enfants, nombre de pétales d'une fleur, nombre de dents,..) : (1 ; 2 ; 3 ; 4 ; 5 ;10 ; 11 ; ...)

b - Variable continue : c'est un caractère quantitatif, un tel caractère peut, théoriquement, prendre toutes les valeurs d'un intervalle de l'ensemble des nombres réels. Toutes les valeurs ne sont pas dénombrables et ne peuvent pas être établies a priori. Ses valeurs sont alors regroupées **en classes** (taille, temps, poids, vitesse, glycémie, altitude, surfaces,...) (1,60 m ; 1,61 m ; 1,62 m ;

Variable qualitative : c'est un caractère qualitatif, dans ce type de variable les modalités ne sont pas quantifiables (pas mesurables) (couleur des yeux, douleur, ...). Ce sont des noms ou ce qui revient au même des sigles ou des codes. Les différentes modalités ne sont pas ordonnables. Attention, même si les modalités sont des codes numériques, les opérations sur les modalités n'ont aucun sens.

Exemple : type de relief avec trois modalités (plaine, montagne, plateau), ou encore taille d'une niche écologique avec quatre modalités (petite, moyenne, grande, très grande).

Les données qualitatives peuvent être assimilées au cas des variables discontinues, en supposant que les différentes variantes du caractère qualitatif sont rangées dans un ordre correspondant par exemple à la suite des nombres entiers positifs (différentes couleurs, différents degrés d'infection...). Les données qualitatives peuvent être réalisées dans deux échelles de mesure : échelle de rangement et l'échelle nominale. Ces données ne sont pas manipulables par l'arithmétique.

a - Dans l'échelle ordinale (de rangement), on parle dans ce cas de **caractère ordinal** (**caractères qui peuvent être** exprimés sur une échelle ordinale) : dans cette échelle chaque *modalité* est explicitement significative du rang pris par chaque individu pour le caractère considéré.

Si E possède N éléments, les modalités seront 1er, 2eme, 3eme, ... neme. Comme on possède juste l'ordre des individus, on ne sait rien de l'intervalle des valeurs. Il existe une certaine relation entre les objets du type plus grand que, supérieur à, plus difficile que, préférée à..... Une transformation ne changeant pas l'ordre des objets est admissible. La statistique la plus appropriée pour décrire la tendance centrale des données est la médiane.

b - Dans l'échelle nominale, les nombres ou symboles identifient les groupes auxquels divers objets appartiennent. C'est le cas des **numéros d'immatriculation** des voitures ou de **sécurité sociale** (chaînes de caractères). Le même nombre peut être donné aux différentes personnes habitant le même département ou de même sexe constituant des sous-classes. Les symboles désignant les différentes sous-classes dans l'échelle nominale peuvent être modifiés sans altérer l'information essentielle de l'échelle. Les seules statistiques descriptives utilisables dans ce cas sont le mode, la fréquence... et les tests applicables seront centrés sur les fréquences des diverses catégories.