

## Introduction aux stat

### I\ Définition

- Les statistiques : dénombrement, décompte.
- La statistique : science, discipline qui permet le raisonnement en utilisant un outil statistique (des tests paramétriques, des tests non-paramétriques...)
- Epidémiologie : c'est la capacité d'observer (étudier) un phénomène dans une population (une maladie par exemple), elle est divisé en :

#### 1 - Epidémiologie descriptive :

- Décrire le phénomène observé
- Donner une identité au phénomène en répondant à 3 questions :
  - ✓ Où ?  $\longrightarrow$  lieu
  - ✓ Quand ?  $\longrightarrow$  temps
  - ✓ Chez qui ?  $\longrightarrow$  personne
- Mesurer / quantifier le phénomène observé par le calcul de :
  - ✓ Moyenne
  - ✓ Mode
  - ✓ Médiane
  - ✓ Variance
  - ✓ Ecart type

#### 2 - Epidémiologie analytique :

- Etudier la relation entre le facteur de risque (FDR) et la maladie observée (facteur viral, physique, génétique, alimentaire, de l'environnement ...)
- Mesurer le phénomène observé par le calcul de :
  - ✓ Risque relatif
  - ✓ Odds ratio (rapport de côtes)

#### • Causalité

- Loi de Pasteur de causalité unidirectionnelle

1 Microbe  $\xrightarrow{\text{donne}}$  1 maladie

- Loi de la multifactorialité

1 Maladie  $\xrightarrow[\text{par}]{\text{est causée}}$  des facteurs de risque + cofacteurs

**II\ La typologie des variables** : il existe 2 types

**1\ Variable quantitative** : c'est une variable qui peut prendre un ensemble de valeur, c-à-d ce sont des variables mesurables ou quantifiables  
Exp : poids, taille, valeur glycémique ...

**a\ Variable quantitative continue** : c'est une variable qui peut prendre un ensemble de valeur qui se situe dans un intervalle  
Exp : poids, taille, glycémie

**b\ Variable quantitative discontinue** (discrète, isolée) : c'est une variable qui ne peut prendre que des valeurs isolées  
Exp : FC (fréquence cardiaque), N<sup>br</sup> d'enfant par famille, N<sup>br</sup> d'élève par classe ...

**2\ Variable qualitative** : c'est une variable qui ne peut pas prendre une valeur numérique  
Exp : couleur des yeux, couleur des cheveux, groupage sanguin, maladie (présent ou absente), stade d'une maladie, CSP (catégorie socioprofessionnelle)

**III\ Notion de modalité**

La modalité se définit comme étant un état ou une catégorie qui qualifie une variable étudiée ou bien une variable à étudier

Exemples :

- Variable quantitative continue : FT (fièvre typhoïde) selon l'âge

Classe d'âge	FT
[ 05 – 10 [	04
[ 10 – 15 [	10
[ 15 – 20 [	15
[ 20 – 25 [	20
Total	49

L'âge (variable) a été étudié selon les modalités : 05 – 10  
10 – 15  
15 – 20  
20 - 25

- Variable quantitative discontinue : N<sup>br</sup> d'enfant par famille

N <sup>br</sup> d'enfant / famille	Fréquence absolue
0	06
1	10
2	20
3	30
Total	66

Le N<sup>br</sup> d'enfant par famille (variable) a été étudié selon les modalités : 0  
1  
2  
3

- Variable qualitative

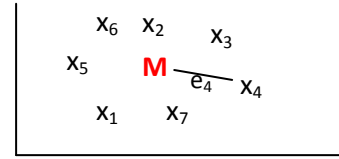
Groupage sanguin	Fréquence absolue
A	20
B	10
AB	50
O	20
Total	100

Le groupage sanguin (variable) des étudiants a été étudié selon les modalités : A  
B  
AB  
O

**C\ Les paramètres de mesure** : 2 types**1\ Paramètre de tendance centrale** (moyenne, mode, médiane)

- **La moyenne « M »** : c'est un paramètre de mesure qui nous permet de centraliser une distribution, une série de valeurs observées (à condition que les valeurs soient ordonnées)

La différence :  $M - x_i = \text{écart } (e_i) = \text{variance}$   
plus la variance est  $\downarrow$  plus la méthode est exacte



$$M = \sum \frac{x_i}{N} \quad x_i : \text{valeur observée} \quad N : \text{taille de l'échantillon}$$

$$M = \sum \frac{n_i x_i}{N} \quad \text{dans le cas de données groupées}$$

**NB :**

**Echantillon** : ensemble d'unité statistique tirée au hasard à partir d'un ensemble, à condition que chaque unité statistique doit avoir une probabilité différente de 0 de faire partir de l'étude, et que l'échantillon soit représentatif

- **Le mode** : valeur de la variable statistique qui a l'effectif le plus élevé

Exp : soit la répartition des familles selon le N<sup>br</sup> d'enfant

N <sup>br</sup> d'enfant / famille	N
0	02
1	06
2	03
3	05

Le mode **1** correspond à l'effectif le plus élevé (06)

NB : dans le cas de données groupées (variable quantitative continue), on parlera de la **classe modale**

exp : soit la répartition de la FT selon les classes d'âge

Classe d'âge	FT
[ 10 – 15 [	02
[ 15 – 20 [	06
[ 20 – 25 [	03
[ 25 – 30 [	05

L'effectif le plus élevé (6) correspond à la classe modale **15 – 20**

- **La médiane** : la valeur de la variable statistique qui divise la distribution de valeurs observées en parties égales, c-à-d, il y a autant de valeurs supérieures que des valeurs inférieures

NB : le calcul de la médiane doit passer par les étapes suivantes :

- déterminer le rang de la médiane
- déterminer la classe médiane proprement dite
- le calcul mathématique de la médiane, donné par la formule :

$$L_2 + \frac{L_1 - L_2}{n_0} \left( \frac{N}{2} - \text{FAC} \right)$$

$L_1, L_2$  : limites de classe                       $N$  : taille de l'échantillon  
 $n_0$  : effectif de la classe médiane  
FAC : fréquence absolue cumulée de la classe précédant la classe médiane

Le calcul de la médiane peut être déterminé par une méthode graphique à partir des fréquences cumulées.

**2\ Paramètre de dispersion** (variance, écart type) : paramètres de mesure qui nous permettent d'étudier l'étalement des mesures par rapport aux valeurs observées

- **La variance « V » :** 
$$V = \sum \frac{n_i (x_i - M)^2}{N - 1}$$
  $x_i$  : valeur observée      N : taille de l'échantillon  
 $n_i$  : effectif      M : moyenne

- **L'écart type «  $\sigma$  » :** 
$$\sigma = \sqrt{V}$$

Exp : gly : ] 0,9 \_ 1,3 [ = ] M-  $\sigma$  \_ M+  $\sigma$  [

### **NB : autres définitions**

- **Effectif** : fréquence absolue : c'est le n<sup>br</sup> d'observation d'une série de valeurs
- **Fréquence relative** : c'est un rapport où le numérateur est composé de valeur observée et le dénominateur est le n<sup>br</sup> d'observation
- **Proportion** : c'est un rapport où le numérateur se trouve dans le dénominateur  
Exp : soit 25 filles et 40 garçons, la proportion des filles =  $\frac{25}{25 + 40} \times 100$
- **Ratio** : c'est un rapport où le numérateur et le dénominateur sont des quantités différentes pour une même valeur  
Exp : sex-ratio :  $\frac{H}{F}$  , H et F sont 2 quantités différentes pour la même variable : sexe  
ratio d'athérogénicité :  $\frac{HDL}{LDL}$       variable : cholestérol
- **Taux** : c'est une probabilité de survenue d'un évènement (maladie, décès, handicap, accident...) par rapport à une population exposée au risque de développer ce phénomène

$$\text{Taux} = \frac{\text{nbr d'accident}}{\text{population exposée}}$$

**NB** : Le risque (pour 1 individu) est le singulier du taux (pour 1 population)

**Exercice 1 :**

Le taux de glucose sanguin déterminé chez 32 patients est donné en g/l

0,85	0,95	1,00	1,06	1,18	} Série statistique ordonnée de la petite valeur « v » vers la grande valeur « V »
0,87	0,97	1,01	1,07	1,14	
0,90	0,97	1,03	1,08	1,14	
0,93	0,98	1,03	1,08	1,15	
0,94	0,98	1,03	1,10	1,17	
0,94	0,99	1,04	1,10	1,19	
			1,11	1,20	

- 1- Quelle est la nature du caractère étudié ?
- 2- Quelle est l'étendue (E) de la série ?
- 3- Donner la répartition par classe de cette série
- 4- Calculer la moyenne, la médiane, le mode, la variance, l'écart type.
- 5- Représenter graphiquement cette distribution.

**Solution**

- 1- nature du caractère étudié : **variable quantitative continue**
- 2- L'étendue de la série :  $E = V - v = 1,20 - 0,85 = 0,35$
- 3- Répartition par classe : il faut respecter les conditions suivantes :

- limites de la classe  $L_1, L_2$  (bornes)
- amplitude de classe (k)
- centre de classe  $\frac{L_1 + L_2}{2}$
- le n<sup>br</sup> de classe  $n = \sqrt{N}$

$$n = \sqrt{N} = \sqrt{32} = 5,65 \approx 6$$

$$E = V - v = k \times n \text{ donc}$$

$$k = \frac{E}{n} = \frac{0,35}{6} = 0,06$$

donc les classes seront : **[0,85 – 0,91[**  
**[0,91 – 0,97[**  
**[0,97 – 1,03[**  
**[1,03 – 1,09[**  
**[1,09 – 1,15[**  
**[1,15 – 1,21[**

$$\text{centre de la classe : } x_1 = \frac{L_1 + L_2}{2} = \frac{0,85 + 0,91}{2} = 0,88$$

- 4- Moyenne :

classes	$n_i$	$x_i$	$n_i x_i$
[0,85 – 0,91[	3	0,88	2,64
[0,91 – 0,97[	4	0,94	3,76
[0,97 – 1,03[	7	1,00	7,00
[1,03 – 1,09[	8	1,06	8,48
[1,09 – 1,15[	6	1,12	6,72
[1,15 – 1,21[	4	1,18	4,72
Total	32		33,32

$$M = \frac{\sum n_i x_i}{N} = \frac{33,32}{32} = 1,04 \text{ g/l}$$