

**Premier chapitre du module de "STATISTIQUE"  
de la première année de médecine**

Dr. Soumia KHARFOUCHI



## Table des matières

INTRODUCTION	5
Chapitre 1. Généralités sur les séries statistiques	7
1. Notions de base et terminologie	7
2. Représentation des données	9
3. Variables statistiques quantitatives	9



## INTRODUCTION

Lorsqu'on veut introduire une certaine discipline, nous devons faire le point sur l'ensemble des termes que l'on doit cerner d'une manière précise. D'abord s'agit-il **des statistiques** ou de **la statistique**? au fait, **les statistiques** sont des ensembles de données ou d'observations que l'on recueille en effectuant des recensements ou des sondages. Ils ont pour origine le besoin des états pour gérer rationnellement leurs ressources. Pour cela, il était nécessaire après collecte d'informations (nécessité de techniques de quantification ; production de données nombreuses, organisées en tableaux) de disposer de méthodes permettant de définir les variations, les évolutions, les ressemblances ou les différences entre régions, entre années, entre catégories. Par contre, à partir de 1843, la statistique désigne l'ensemble de techniques d'interprétation mathématique appliquées à des phénomènes pour lesquels une étude exhaustive de tous les facteurs est impossible, à cause de leur grand nombre ou de leur complexité. C'est la science qui utilise les méthodes mathématiques pour étudier et analyser des statistiques en vue

- d'en accroître les connaissances scientifiques ;
- de planifier des stratégies ;
- d'aider à la prise de décision.

Elle vise à décrire, à résumer et à interpréter des phénomènes dont le caractère essentiel est la variabilité. Elle fournit de la manière la plus rigoureuse possible des éléments d'appréciation utiles à l'explication ou à la prévision de ces phénomènes, mais elle n'explique ni ne prévoit aucun d'entre eux (Vigueron 1997). C'est ainsi que la statistique a fini par envahir tous les domaines de la vie scientifique, passant des affaires publiques à l'agriculture, la biologie, la chimie, la physique, les communications, l'économie, l'électronique, les sciences politiques, la sociologie et bien sûr la médecine.

La statistique descriptive est une branche de la statistique qui se résume à l'ensemble des méthodes à partir desquelles on recueille, ordonne, réduit et condense les données. A cette fin, la statistique descriptive utilise des paramètres, des graphiques et des méthodes dites d'analyse des données. Le développement des performances des ordinateurs a simplifié considérablement les tâches du statisticien et lui a permis d'exploiter de très grandes masses de données.



## Généralités sur les séries statistiques

### 1. Notions de base et terminologie

**1.1. Population / Echantillon / Individu.** - **La population** correspond à l'ensemble des individus sur lequel porte l'étude ou la prévision, (il est généralement difficile de l'étudier dans sa totalité), et **l'échantillon** représente la fraction de cette population qui est réellement observée ou étudiée.

- La notion **d'individu** est très large : les éléments d'un échantillon ou d'une population sont appelés généralement des individus, cependant cette notion peut être remplacé par plusieurs dénominations : unité statistique, sujet, objet, élément, observation, mesure, doses, . . . toutefois, dès que la dénomination est choisie aucune ambiguïté ne doit persister.

### 1.2. Caractère / Modalité / Variable.

1.2.1. **Le caractère.** les éléments d'un ensemble sont décrits par un caractère. Cela revient à établir une correspondance entre chaque élément  $i$  de l'ensemble  $E$  et l'ensemble  $X$  des modalités ou des valeurs du caractère : chaque élément de  $E$  a une modalité (caractère qualitatif) ou une valeur (caractère quantitatif) et une seule dans  $X$ . Ainsi le caractère peut être défini comme une des caractéristiques ou des attributs d'un individu,

1.2.2. **Modalité.** la modalité (respectivement la mesure) est l'une des formes particulière d'un caractère. Les différentes situations où les éléments de  $E$  peuvent se trouver à l'égard d'un caractère qualitatif considéré, sont les différentes modalités du caractère qualitatif  $X$ . Dans le cas où le caractère  $X$  est quantitatif, les différentes situations où les éléments de  $E$  peuvent se trouver sont des mesures. Ces modalités ou ces mesures doivent être à la fois incompatibles (un élément de  $E$  ne peut prendre qu'une seule modalité) et exhaustive (à chaque élément de  $E$  doit pouvoir correspondre une modalité de  $X$ ) de sorte que chaque élément de  $E$  ait une modalité et une seule dans  $X$ .

1.2.3. **Variable.** Dans chaque étude statistique il est très important de considérer la nature des données (observations, caractères, attributs) que l'on va tester. D'elle dépend la nature des opérations possibles et donc des statistiques utilisables dans chaque situation. Il est donc primordial de préciser la nature de chaque variable, ou caractère. Il existe deux types de variables (ou observations), celles-ci peuvent être soit quantitatives soit qualitatives. Ces variables peuvent être mesurées d'où l'importance du choix des échelles de mesures, c'est-à-dire, des règles permettant d'affecter une valeur à chaque individu de la population ou de l'échantillon.

**Variable quantitative.** 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 :

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 ; ...)

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 établi a priori. Ses valeurs sont alors regroupées en classes (taille, temps, poids, vitesse, glycémie, altitude, surfaces,...) (1,60 m ; 1,61 m ; 1,62m ; ...)

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...).

**1.3. Fréquences absolues, relatives et cumulées.** Désignée par « n », « f » ou « F » la notion de fréquence peut être exprimée de plusieurs manières :

- \* Fréquence absolue (effectif)
- \* Fréquence relative (ou fréquence)
- \* Fréquences cumulées

1.3.1. *Fréquences absolues = Effectifs.* Le terme de fréquence absolue désigne les effectifs : a chaque modalité  $x_i$  du caractère X, peut correspondre un ou plusieurs individus dans l'échantillon de taille  $N$ . On appelle effectif (ou fréquence absolue) de la modalité  $x_i$ , le nombre  $n_i$  où  $n_i$  est le nombre d'individus de chacune des modalités

1.3.2. *Fréquences relatives = Fréquences.* On appelle fréquence de la modalité  $x_i$ , le nombre  $f_i$  tel que

$$f_i = \frac{n_i}{N},$$

Remarques :

Rq1 : Le pourcentage est une fréquence exprimée en pour cent. Il est égal à 100  $f_i$ .

Rq2 : L'emploi des fréquences ou fréquences relatives s'avère utile pour comparer deux distributions de fréquences établies à partir d'échantillons de tailles différentes.



1.3.3. *Effectifs et Fréquences cumulés.* On appelle effectifs cumulés (resp. fréquences cumulées) en  $x_i$ , le nombre

$$N_i = \sum_{p=1}^i n_p \text{ (resp. } F_i = \sum_{p=1}^i f_p \text{)}.$$

Remarques :  $k$  étant le nombre de modalités du caractère X.

Rq1 : la taille de l'échantillon est  $N = \sum_{i=1}^k n_i$ .

Rq2 :  $\sum_{i=1}^k f_i = 1$ .

## 2. Représentation des données

Il existe plusieurs niveaux de description statistique : la présentation brute des données, des présentations par tableaux numériques, des représentations graphiques et des résumés numériques fournis par un petit nombre de paramètres caractéristiques.

**2.1. Tableaux statistiques.** En général une série statistique à caractère discret se présente sous la forme :

Valeurs	$X_1$	$X_2$	...	$X_k$
Effectifs	$n_1$	$n_2$	...	$n_k$
Fréquences	$f_1$	$f_2$	...	$f_k$

Plutôt que réécrire ce tableau on écrira souvent : la série  $(x_i, n_i)_{1 \leq i \leq k}$ . (On n'indique pas le nombre de valeurs lorsqu'il n'y a pas d'ambiguïté). Souvent on notera  $N$  l'effectif total de cette série donc  $N = n_1 + n_2 + \dots + n_k$ .

**2.2. Représentations graphiques et statistiques descriptives.** Les représentations graphiques sont très importantes en statistique descriptive. Elles ont l'avantage de renseigner immédiatement sur l'allure générale de la distribution. Elles facilitent l'interprétation des données recueillies. La représentation graphique des données montre la forme générale de la distribution et donne une image de la grandeur des nombres qui constituent les données. D'autres statistiques simples sont utilisées pour représenter le centre de la distribution et les mesures liées à la dispersion des observations autour de cette tendance centrale.

## 3. Variables statistiques quantitatives

L'étude descriptive d'une variable statistique quantitative (aussi bien discrète que continue) se fait en quatre étapes :

- Description préliminaire
- Caractéristiques de position centrale
- Caractéristiques de dispersion
- Caractéristiques de symétrie et d'aplatissement

Une variable quantitative peut être :

Discrète : si la variable ne prend qu'un nombre fini de valeurs (ces valeurs sont appelées modalités et notées  $x_i$ ) .

Continue : si la variable prend ses valeurs dans un intervalle (classe )

### 3.1. représentation graphique des séries statistiques.

3.1.1. *Tableau statistique.* Quand la variable statistique  $X$  est discrète, on compte, pour chaque valeur de  $X$ , le nombre d'individus prenant cette valeur ; c'est l'effectif de la valeur. On aboutit à un tableau du type :

Valeurs de la variable $x_i$	Effectifs $n_i$
$x_1$	$n_1$
$x_2$	$n_2$
$\vdots$	$\vdots$
$x_p$	$n_p$

avec  $X_1 < X_2 < \dots < X_p$ .

Quand la variable statistique  $X$  est continue, on regroupe les valeurs en classes. Les classes sont des intervalles semi-ouverts  $[e_i, e_{i+1}[$ . Leur amplitude est le nombre :  $e_{i+1} - e_i$  et leur centre, le nombre :  $\frac{x_{i+1} + x_i}{2}$ .

Pour chaque classe, on compte le nombre d'individus qui prennent une valeur supérieure ou égale à  $e_i$  et inférieure à  $e_{i+1}$  : c'est l'effectif de la classe. On aboutit à un tableau de type :

Classes	Effectifs $n_i$
$[e_0; e_1[$	$n_1$
$[e_1; e_2[$	$n_2$
$\vdots$	$\vdots$
$[e_{p-1}; e_p[$	$n_p$

Remarques

- Quand le nombre de valeurs prises par la variable statistique est trop grand, on traite la variable discrète comme une variable continue.
- Quand on regroupe les valeurs par classes, on essaye d'avoir des classes de même amplitude et pas trop nombreuses. Mais, souvent, les valeurs extrêmes posent problème, c'est pourquoi les premières ou dernières classes sont soit ouvertes, soit d'amplitude différente des autres classes.

**Exemple 1 (cas discret)**

Afin d'étudier la production en fruits d'une certaine variété de fraises, 150 arbrisseaux ont été sélectionnés dans un champ. On a dénombré le nombre de fruits qu'ils portent. Le tableau suivant a été obtenu :

Nombre de fruits	8	10	16	20	24	32	42
Nombre d'arbrisseaux	12	23	41	24	22	16	12

**Exemple 2 : (Cas continu)**

Dans le but de déterminer le temps de réaction des gens par rapport au son, l'expérience suivante a été entreprise :

50 personnes ont été réunies et pour chacune d'elles, on a enregistré le temps mis pour réagir après avoir entendu un signal sonore. Les résultats de l'expérience ont été reportés dans le tableau suivant :

T. de réaction	$[0, 45; 0, 51[$	$[0, 51; 0, 57[$	$[0, 57; 0, 63[$	$[0, 63; 0, 69[$	$[0, 69; 0, 75[$	$[0, 75; 0, 81[$
Nbre d'individus	2	8	18	16	4	2

3.1.2. *Comment représenter une série statistique ?* - Pour représenter une variable statistique discrète, on utilise un diagramme en bâtons (chaque bâton a une hauteur proportionnelle à l'effectif et/ou à la fréquence).

- Pour représenter une variable statistique continue, on trace un histogramme. L'histogramme est constitué de rectangles juxtaposés dont la surface est proportionnelle à l'effectif de la classe correspondante.

Si les classes ont des amplitudes égales, la hauteur des rectangles est proportionnelle à l'effectif. Si les classes ont des amplitudes inégales, on représente la classe ayant la plus petite amplitude ; puis on compense une amplitude  $k$  fois plus grande par une hauteur  $k$  fois plus petite.