

# CHAPITRE I - DEFINITIONS ET TERMINOLOGIE

Il serait peut-être préférable, dès le départ, de recourir à une série d'exemples. Ils serviront, en premier lieu, comme illustrations pour dégager graduellement le sens donné aux termes techniques. Ils nous seront surtout utiles pour cerner de manière précise les notions introduites. Ils permettront enfin, si besoin est, de dissiper d'éventuelles ambiguïtés qui pourraient s'installer.

Les tout premiers termes de la statistique descriptive sont :

- Population
- Individu
- Caractère
- Modalité
- Caractère qualitatif
- Caractère quantitatif
- Série statistique
- Tableau statistique
- Fréquence absolue
- Fréquence relative
- Variable statistique discrète
- Variable statistique continue

Ce sont ces mots là que nous devons comprendre. Les exemples qui suivent vont nous aider pour cela.

## Exemple 1 :

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 :

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 :

Temps 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]
Nombre d'individus	2	8	18	16	4	2

**Exemple 3 :**

Un laboratoire pharmaceutique a mis au point un produit à effet anesthésiant. L'expérimentation de ce produit a été menée sur un groupe de 70 cobayes. Dans la mesure où il est attendu que la réponse au produit peut différer d'un cobaye à l'autre, il est prévu différents dosages. La mesure étalon (dose) est une seringue dont la contenance est  $1,5\text{mm}^3$

Lors des essais, on injecte à chaque animal un nombre suffisant de doses pour aboutir à une anesthésie complète.

Quantité de produit (en $\text{mm}^3$ )	1,5	3	4,5	6	7,5
Nombre de cobayes	21	35	10	3	1

**Exemple 4 :**

Un lac naturel a été exploité pour l'élevage de poissons.

Deux ans après y avoir introduit les premiers spécimens on a procédé à un prélèvement aléatoire d'un échantillon de 200 poissons.

Chaque poisson capturé est pesé, marqué et ensuite relâché dans le lac. De cette manière un poisson ne peut être pris et pesé plus d'une fois.

Les pesées ont été rapportées dans le tableau suivant

Poids (en Kg)	[0.1 - 0.3[	[0.3 - 0.5[	[0.5 - 0.7[	[0.7 - 0.9[	[0.9 - 1.5[
Nombre De poissons	8	82	24	80	6

**Exemple 5 :**

Activités et mouvement hospitalier au département de médecine du CHU de Constantine durant l'année 1994

Services	Consultations
Cardiologie	7.729
Dermatologie	2.155
Endocrinologie	1.307
Hématologie	4.003
Infectieux	5.961
Médecine interne	13.415
Médecine nucléaire	2.906
Neurologie	7.361
Pédiatrie	30.949
Pneumologie	1.918
Psychiatrie	10.909
Radiothérapie	8.434
Rhumatologie	5.095
<b>Total</b>	<b>102.142</b>

### Exemple 6 :

Dans le tableau suivant ont été répartis les cas hospitalisés et notifiés relevant des maladies à déclaration obligatoire.

(C.H.U. de Constantine, année 1993)

Maladies	Nombre
Tuberculose	229
Méningites	224
Kyste hydatique	198
Fièvre typhoïde	152
Amibiase	46
Brucellose	19
Hépatites virales	18
Tétanos	6
Rage	3
Paludisme	1
Autres	22
<b>Total</b>	<b>918</b>

#### a) Population :

L'ensemble sur lequel porte l'étude statistique s'appelle **population**. On dit aussi (mais rarement) **univers statistique**.

Les éléments de cet ensemble s'appellent **individus** ou **unités statistiques**.

L'utilisation des mots population et individu au lieu des mots ensemble et élément remonte aux origines de la statistique qui s'est d'abord investie dans l'étude des groupes humains (démographie). L'extension à des domaines aussi vastes que divers s'est faite en conservant cette terminologie de base.

*La population peut être donc un groupe de personnes, d'objets, de choses concrètes ou même abstraites.*

- Dans l'exemple 1, la population est constituée de l'ensemble des 150 arbrisseaux. Chacun de ces arbrisseaux est un individu de la population.
- Dans l'exemple 2, les 50 personnes soumises au test constituent la population.
- Le groupe des 70 cobayes est la population du 3<sup>ème</sup> exemple.
- Dans l'exemple 4, les individus sont les poissons. L'ensemble des 200 poissons forme la population (bien que l'on ait utilisé le mot échantillon).
- Nous profiterons de l'exemple 5 pour apporter quelques remarques et mises au point.

Dans cet exemple il faut d'abord préciser que ce sont les consultations qui représentent la population statistique. L'individu statistique est la consultation faite et non la personne qui s'est présentée. Il se pourrait qu'une même personne se soit présentée à plusieurs services pour des consultations différentes (cardiologie et pneumologie par exemple).

#### Remarque 1 :

Avant toute étude, la population statistique doit être cernée avec exactitude par sa spécificité et par le nombre de ses éléments.

- Nous devons énoncer avec clarté la caractéristique identificatrice des individus de cette population. De cette manière, pour chaque élément nous pouvons affirmer, sans hésitation, qu'il appartient ou non à la population.  
Dans notre exemple, l'individu est la consultation, mais si nous avons considéré la population comme étant l'ensemble des malades nous aurions obtenu un tableau différent (duquel seraient exclues les répétitions des mêmes malades sur des services différents ou sur le même service plusieurs fois dans l'année).  
Quelles consultations doit-on considérer comme individus de la population ? celles effectuées sur des personnes venant de l'extérieur ou doit-on également leurs ajouter celles faites lors des visites sur les malades hospitalisés ?
- Le nombre définitif d'individus constituant la population doit être, également, toujours indiqué. On l'appelle l'effectif total et on le représente souvent par la lettre N. Nous avons, pour les exemples cités :

Exemple 1 :  $N = 150$

Exemple 2 :  $N = 50$

Exemple 3 :  $N = 70$

Exemple 4 :  $N = 200$

Exemple 5 :  $N = 102.142$

Exemple 6 :  $N = 918$

### Remarque 2 :

Tout sous-ensemble d'une population donnée est lui-même une population. Ce sous-ensemble s'appelle **échantillon** ou **sous-population**.

Dans l'exemple 4, la population des 200 poissons a été désignée par le mot échantillon. La raison est que, dans ce cas-ci, ce groupe est lui-même sous-ensemble de la population de tous les poissons du lac et qui peut être, elle-même, objet d'une étude.

### **b) Caractère**

Quand une population est proposée à une étude statistique c'est en fonction d'une propriété particulière que possèdent ses éléments.

Un groupe d'étudiants, par exemple, peut faire l'objet d'une étude concernant la taille ou le poids de ses éléments ; ou bien leurs notes de mathématiques ou de biologie ou encore leur âge ou l'éloignement de leur domicile par rapport à l'université ... etc.

***Le trait (ou propriété) choisi pour l'étude statistique s'appelle caractère.***

Les caractères évoqués dans les exemples précédents sont, respectivement :

- Exemple 1 : Nombre de fruits portés par l'arbrisseau.
- Exemple 2 : Temps de réaction des personnes par rapport au son.
- Exemple 3 : Quantité de produit nécessaire pour aboutir à l'anesthésie des cobayes.
- Exemple 4 : Poids des poissons.
- Exemple 5 : Nature des consultations médicales au département de médecine.
- Exemple 6 : Maladies faisant l'objet de déclaration obligatoire.

### c) Modalités

Chaque individu de la population se situe dans une position particulière vis à vis du caractère étudié.

Il existe, pour chaque caractère, plusieurs positions sur lesquelles se répartissent tous les individus de la population.

#### *Les différentes positions que peut prendre un caractère s'appellent modalités.*

Ainsi, dans l'exemple 1, nous avons sept modalités. Ce sont les valeurs 8 - 10 - 16 - 20 - 24 - 32 et 42.

Chaque arbrisseau est associé à une, et une seulement, de ces modalités ; c'est le nombre de fruits qu'il porte. Ce nombre figure parmi les 7 cités plus haut.

Un arbrisseau ne peut, évidemment pas, être associé à plusieurs modalités.

Dans tous les cas, le découpage du caractère en modalités se fait de manière à ce qu'un individu ne peut avoir plus d'une modalité.

Ainsi, dans l'exemple 2, vous remarquez que les intervalles sont semi-ouverts (exemple :  $[0.45 - 0.51[$  puis  $[0.51 - 0.57[$  ). Ceci pour écarter la possibilité qu'un individu ait deux modalités. La personne dont le temps de réaction est exactement 0.51 seconde sera, de cette manière, classée uniquement dans le 2<sup>ème</sup> intervalle. Elle aurait été classée dans les deux intervalles s'ils avaient été fermés.

Notation : D'usage, on numérote les modalités de 1 à k. La modalité numéro  $i$  est notée  $C_i$

### d) Effectifs

Lorsque la population est répartie sur les différentes modalités, nous obtenons pour chacune d'elles un nombre. C'est le nombre des individus ayant cette modalité. On l'appelle l'effectif de la modalité ou aussi sa **fréquence absolue**.

On note, habituellement,  $n_i$  l'effectif correspondant à la modalité  $C_i$  (numéro  $i$ )

Ainsi, pour les 7 modalités de l'exemple 1, nous avons, dans l'ordre, les effectifs suivants : 12 - 23 - 41 - 24 - 22 - 16 et 12.

Bien entendu, puisqu'il n'y a pas répétition des mêmes individus sur les différentes modalités, la somme de tous les effectifs est égale à l'effectif total.

$$\sum_{i=1}^k n_i = N$$

### e) Série statistique

Nous avons vu qu'à chaque modalité ( $C_i$ ) est associé son effectif ( $n_i$ ).

Les couples ( $C_i ; n_i$ ) forment une suite qu'on appelle **série statistique**.

Ainsi, pour les exemples que nous avons, la série statistique est :

- Pour l'exemple 1 : (8 ; 12), (10 ; 23), (16 ; 41), (20 ; 24), (24 ; 22),  
(32 ; 16), (42 ; 12)

- Pour l'exemple 2 : ( $[0.45 - 0.51[$  ; 2), ( $[0.51 - 0.57[$  ; 8), ( $[0.57 - 0.63[$  ; 18),  
( $[0.63 - 0.69[$  ; 16), ( $[0.69 - 0.75[$  ; 4), ( $[0.75 - 0.81[$  ; 2)

- Pour l'exemple 6 : ( Tuberculose ; 229), ( Méningites ; 224),  
 ( Kyste hydatique; 198), ( Fièvre typhoïde ; 152),  
 ( Amibiase ; 46), ( Brucellose ; 19), ( Hépatites virales ; 18),  
 ( Tétanos ; 6), ( Rage ; 3), ( Paludisme ; 1), ( Autres; 22).

### f) Tableau statistique

Un tableau statistique est un tableau à deux colonnes. La colonne de gauche est toujours réservée aux modalités du caractère à étudier tandis que celle de droite est utilisée pour indiquer les fréquences absolues (effectifs) de ces différentes modalités.

Nous avons ainsi un tableau de la forme :

Modalités du caractère	Effectifs correspondants aux modalités
$C_1$	$n_1$
$C_2$	$n_2$
...	...
$C_i$	$n_i$
...	...
$C_k$	$n_k$
<b>Total</b>	<b>N</b>

### Remarques :

1. Un travail statistique commence par la collecte de l'information. Cette collecte se fait par des mesures, des observations ou des enquêtes.  
 Il s'agit d'observer sur chaque individu de la population la modalité du caractère auquel il se rattache. les données ainsi recueillies s'appellent des données brutes et constituent toute l'information en notre possession.  
 Après cette phase commence le traitement statistique dont la première opération est l'établissement d'un tableau statistique.  
 Le tableau statistique renferme donc toute l'information qu'on a rassemblée.
2. Le tableau statistique (qui, finalement, sert à représenter la série statistique) peut aussi être donné sous forme de lignes (c'est le cas des exemples 1 ; 2 ; 3 et 4).
3. Aux deux colonnes du tableau viennent s'ajouter d'autres colonnes au fur et à mesure que le traitement statistique se développe. Le tableau obtenu à chaque étape continu de s'appeler tableau statistique.

### g) Fréquences relatives

La fréquence relative qui correspond à chaque modalité  $C_i$ , et qu'on note  $f_i$ , est le rapport de la fréquence absolue  $n_i$  à l'effectif total N.

C'est-à-dire :

$$f_i = n_i / N$$

$$f_i = n_i / N$$

Pour l'exemple 1 nous avons :

Modalités du caractère $C_i$ (Nombre de fruits)	Fréquences absolues $n_i$ (Nombre d'arbrisseaux)	Fréquences relatives $f_i$
8	$n_1 = 12$	$f_1 = n_1 / N = 0.080$
10	$n_2 = 23$	$f_2 = n_2 / N = 0.153$
16	$n_3 = 41$	$f_3 = n_3 / N = 0.273$
20	$n_4 = 24$	$f_4 = n_4 / N = 0.160$
24	$n_5 = 22$	$f_5 = n_5 / N = 0.147$
32	$n_6 = 16$	$f_6 = n_6 / N = 0.107$
42	$n_7 = 12$	$f_7 = n_7 / N = 0.080$
<b>Total</b>	<b>N = 150</b>	<b>1</b>

### Remarques :

- La première colonne qu'on ajoute habituellement au tableau statistique est celles des fréquences relatives et ceci se justifie par plusieurs raisons :
  - Les études descriptives se font généralement pour des besoins de représentation et de visualisation. Avec les fréquences relatives la comparaison devient immédiate entre des populations d'effectifs différents lorsqu'elles sont décrites suivant le même caractère. Ce qui n'est pas le cas quand on se restreint aux fréquences absolues.
  - Très souvent, les études descriptives préparent à des études inductives qui font appel à la statistique inférentielle. C'est une théorie qui se base sur le calcul des probabilités. Ce qui permet le passage entre le concret et la théorie est le fait que lorsque le nombre d'observations (ou d'expériences) augmente indéfiniment la fréquence relative converge vers la probabilité (théorème central limite, démontré par J. Bernoulli).
- Naturellement, la somme des fréquences relatives est égale à l'unité,

$$\sum_{i=1}^k f_i = \sum_{i=1}^k (n_i / N) = (1/N) \sum_{i=1}^k n_i = N/N = 1$$

Mais dans la pratique, parfois, les erreurs d'arrondi font que l'on s'écarte légèrement de 1.

### h) Caractère qualitatif - Caractère quantitatif

Revenons aux 6 exemples que nous avons considérés précédemment et observons-les attentivement. Nous allons constater que les caractères étudiés à travers ces exemples se différencient les uns des autres par leurs natures même.

En effet, nous avons :

- Exemple 1 : Nombre de fruits portés par l'arbrisseau.
- Exemple 2 : Temps de réaction des personnes par rapport au son.
- Exemple 3 : Quantité de produit nécessaire pour aboutir à l'anesthésie des cobayes.
- Exemple 4 : Poids des poissons.
- Exemple 5 : Nature des consultations médicales au département de médecine.

Remarquez que, dans les quatre premiers exemples, les modalités du caractère se mesurent par des nombres qui indiquent l'intensité de ce caractère chez les individus.

On peut parler en terme de **quantité** et dire que le caractère est plus fort chez un individu que chez un autre.

Tandis que dans les exemples 5 et 6 les modalités ne peuvent être désignées que par leur **qualité**. Dans l'exemple 5 nous pouvons seulement dire que tel ou tel individu a la tuberculose ou la rage ; mais nous ne pouvons pas associer aux individus des nombres qui mesureraient leurs positions vis à vis du caractère.

nous allons, ainsi, adopter les définitions suivantes :

*Un caractère est dit **quantitatif** quand ses différentes modalités sont mesurables par des nombres qui en indiquent l'intensité.*

*Un caractère est dit **qualitatif** quand ses différentes modalités ne peuvent être désignées que par leurs qualités.*

### **i) Variable statistique discrète – Variable statistique continue**

Lorsque le caractère est quantitatif on l'appelle **variable statistique**.

Revenons à nos exemples et limitons nous à ceux où le caractère est quantitatif ; c'est-à-dire les exemples de 1 à 4.

Ici aussi, nous avons à discerner entre deux choses. Les variables statistiques des exemples 1 et 3 se distinguent des deux autres par le fait qu'elles ne peuvent prendre que des valeurs **isolées**. On ne peut pas imaginer qu'un arbrisseau porte 0,1345 fruits par exemple !

De même, Les doses de produit étant spécifiées dans l'exemple 3, la variable statistique ne peut prendre que des valeurs particulières et isolées.

Définition :

*Une variable statistique est dite **discrète** lorsque ses modalités ne peuvent être que des nombres isolés.*

Remarquons que le mot isolé ne veut pas dire, systématiquement, entier. Il est vrai que dans l'exemple 1 toutes les valeurs sont entières, mais ce n'est pas le cas dans l'exemple 3.

Dans les exemples 2 et 4 nous pouvons nous attendre à ce que les variables prennent n'importe quelle valeur dans les intervalles considérés. Les variables dans ces deux exemples peuvent parcourir continûment l'étendue où elle sont étudiée.

nous dirons donc,

*Une variable statistique est dite **continue** quand elle peut prendre n'importe quelle valeur dans un intervalle donné.*