Chapitre 1 : Notions de base

1. Introduction :

La statistique est l'ensemble des méthodes mathématiques qui ont pour objectif la collecte, le traitement et l'interprétation de données d'observation relatives à un groupe d'individus ou d'échantillons.

La statistique est présente dans beaucoup de domaines scientifiques tel que :

La sociologie : les méthodes statistiques sont utilisées comme techniques de traitement des données obtenues par des sondages ou des recensements.

La physique : l'étude de la mécanique et de la thermodynamique statistiques permet de déduire du comportement de particules individuelles un comportement global (passage du microscopique au macroscopique).

La finance : calcul des risques et de prévention….etc

2. Exemples d'études statistiques :

Exemple 1 : Considérons les données suivantes concernant la répartition suivant le sexe des habitants d'une ville:

Hommes	Femmes
29722	31444

Exemple 2 : une enquête a été réalisée auprès de 100 étudiants pour connaitre leur âge :

Age (années)	18	19	20	21	22	23	24
Nombre d'étudiants	15	13	8	21	18	16	9

Exemple 3 : On procède à un prélèvement de 200 poissons pris dans un lac naturel suivant leur poids :

Poids (en g)	[100; 300[	[300; 500[	[500; 700[	[700; 900[	[900; 1500[
Nombre de poissons	8	82	24	80	6

3. Termes statistiques :

3.1. Population :

La population statistique est l'ensemble des éléments sur lesquels porte l'étude. Les éléments de la population sont appelés "individus statistiques" ou "unités statistiques".

Si la population comporte N individus, on note l'ensemble des individus $\Omega = \{w_1, w_2, \ldots, w_N\}$ tel que w_i représente le $i^{ème}$ individu avec $i \in \{1, 2, 3, \ldots, N\}$.

1- La population étudiée dans l'exemple 1 est l'ensemble des hommes et de femmes dans une ville avec N = 31444 + 29722 = 61166

2- La population étudiée dans l'exemple 2 est l'ensemble des étudiants avec N = 100.

3- La population étudiée dans l'exemple 3 est l'ensemble des poissons avec N = 200

3.2. Echantillon :

On appelle échantillon un sous ensemble de la population considérée. Le nombre d'individus dans l'échantillon représente la taille de l'échantillon

Exemple : Dans une population formée par les étudiants de $2^{ème}$ année Génie Electrique (unité statistique = étudiant), l'échantillon représente un groupe d'étudiants.

Remarque 1 : La notion d'échantillon est fondamentale car en règle générale la population entière n'est pas disponible ou observable (comme le cas de l'exemple 3).

3.3. Variable (caractère) statistique :

C'est la propriété ou l'aspect que l'on se propose d'observer dans la population ou l'échantillon

La variable étudiée dans l'exemple 1 est : le sexe.

La variable étudiée dans l'exemple 2 est : l'âge des étudiants.

La variable étudiée dans l'exemple 3 est : .le poids des poissons.

3.4. Modalité:

Les modalités sont les différents aspects ou valeurs qu'une variable peut prendre, on appelle M l'ensemble des modalités.

Si le nombre de modalités est noté r; l'ensemble des modalités de la variable statistique X sera noté $M = \{x_1, x_2, \ldots, x_r\}$

Dans l'exemple 1, la variable « sexe » peut prendre deux modalités possibles : « hommes » et « « femmes » donc M= {Hommes, Femmes}.

Dans l'exemple 2, la variable « l'âge des étudiants» peut prendre les modalités suivantes : M= {18, 19, 20, 21, 22, 23, 24}.

Dans l'exemple 3, la variable « le poids des poissons » peut prendre les modalités suivantes : M= { [100, 300[, [300, 500[, [500, 700[,[700, 900[, [900, 1500[, }.

4. Types de variables statistiques :

On distingue deux types de variables :

<u>a- Variables qualitatives :</u> Une variable statistique est dite qualitative lorsque ses modalités ne sont pas mesurables. Elle peut être de type :

- Nominal : les modalités ne sont pas ordonnées.

Exemple : la variable statistique : « couleur des yeux » possède par exemple les modalités : noire, marron, verte,…etc, mais elles ne peuvent pas être ordonnées.

- Ordinal : les modalités sont ordonnées.

Exemple : la variable statistique : « mention au BAC » possède les modalités : passable, assez-bien, bien, très bien, execellent qui peuvent pas être ordonnées.

b-<u>Variables quantitatives :</u> Une variable statistique est dite quantitative lorsque ses modalités sont mesurables. Elle peut être de type :

- Discret : les modalités sont dénombrables (nombre fini).

Exemple : la variable statistique : « nombre d'enfants par famille » possède des modalités finies : 0, 1, 2, 3,….

- Continu : les modalités sont définies sur un intervalle continu (nombre infini).

Exemple : la variable statistique : « la taille » possède des valeurs dans un certain intervalle, on peut par exemple classer les étudiants suivant leur taille dans des intervalles : [60kg, 70kg[, [70kg, 80kg[…

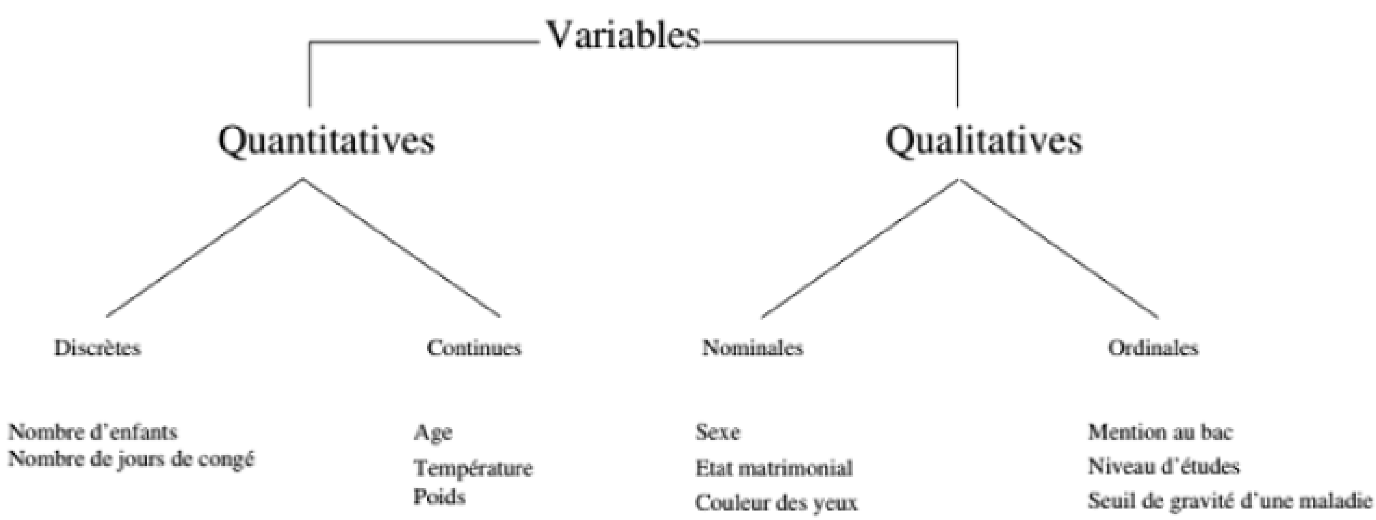

Figure 1- Types de variables statistiques

La variable «sexe ».de l'exemple 1 est qualitative nominale (pas d'ordre logique ou pas ordonnée)
La variable de l'exemple 2 « Age des étudiants» est quantitative discrète.
La variable de l'exemple 3 « le poids des poissons » est quantitative continue.

5. **Effectif (fréquence absolue):**

L'effectif de la modalité x_i est noté n_i et désigne le nombre d'individus de la population présentant la modalité x_i. L'effectif total de la population N est donné par :

$$N = \sum_{i=1}^{r} n_i = n_1 + n_2 + \cdots + n_r$$

6. **Fréquence (fréquence relative):**

La fréquence de la modalité x_i est notée f_i définie par $f_i = \frac{n_i}{N}$ et elle désigne la proportion d'individus de la population présentant la modalité x_i. Elle peut s'exprimer sous la forme d'un nombre décimal ou sous la forme d'un pourcentage.

Remarque 2 :

Soit X une variable à r modalités, on a les propriétés suivantes :

$0 \leq f_i \leq 1, \qquad \forall i \in \{1, 2, 3, \ldots, r\}$

$\sum_{i=1}^{r} f_i = 1 \left(ou\ en\ pourcentage\ \sum_{i=1}^{r} f_i = 100 \right)$

1-Dans l'exemple 1 :

Modalité x_i (sexe)	Hommes	Femmes
Effectif n_i	29722	31444
Fréquence $f_i = \frac{n_i}{N}$	$\frac{29722}{61166} = 0.4859$	$\frac{31444}{61166} = 0.5141$

2-Dans l'exemple 2 :

Modalité x_i: Age (années)	18	19	20	21	22	23	24
Effectif n_i : nombre d'étudiants	15	13	8	21	18	16	9
Fréquence $f_i = \frac{n_i}{N}$	0.15	0.13	0.08	0.21	0.18	0.16	0.09

3- Dans l'exemple 3 :

Modalité x_i : poids des poissons	[100, 300[	[300, 500[	[500, 700[	[700, 900[	[900, 1500[
Effectif n_i : nombre de poissons	8	82	24	80	6
Fréquence $f_i = \frac{n_i}{N}$	0.04	0.41	0.12	0.4	0.03

7. Amplitude d'une classe:

On appelle amplitude.de la classe $[a_i, b_i[$ le réel A_i représentant la longueur de l'intervalle $[a_i; b_i[$ tel que $A_i = b_i - a_i$

8. Centre d'une classe:

On appelle centre.de la classe $[a_i, b_i[$ le réel c_i représentant le milieu de l'intervalle $[a_i; b_i[$ tel que $c_i = \frac{a_i+b_i}{2}$

Pour l'exemple 3, on aura :

classe $[a_i, b_i[$	[100; 300[	[300; 500[	[500; 700[	[700; 900[	[900; 1500[
Amplitude	200	200	200	200	600
Centre	200	400	600	800	1200

9. Effectifs cumulés croissants:

On appelle effectif cumulé croissant, noté n_icc associé à la valeur x_i la somme des effectifs passés tel que :

$$n_i cc = n_{i-1} cc + n_i = \sum_{j=1}^{i} n_j \qquad i = 2, 3, \dots, r$$

Avec $n_1cc = n_1$

10. Effectifs cumulés décroissants:

On appelle effectif cumulé décroissant, noté n_icd associé à la valeur x_i la différence des effectifs passés tel que :

$$n_i cd = n_{i-1} cd - n_{i-1} = N - \sum_{j=1}^{i-1} n_j \qquad i = 2, 3, \dots, r$$

Avec $n_1cd = N$ le nombre total des individus dans la population

Pour l'exemple 2 :

x_i: Age (années)	n_i : nombre d'étudiants	n_icc	n_icd
18	15	15	100
19	13	28	85
20	8	36	72
21	21	57	64
22	18	75	43

23	16	91	25
24	9	100	9

Pour l'exemple 3 :

Poids des poissons x_i	Nombre de poissons n_i	$n_i cc$	$n_i cd$
[100, 300[	8	8	200
[300, 500[	82	90	192
[500, 700[	24	114	110
[700, 900[	80	194	86
[900, 1500[	6	200	6

11. Fréquences cumulées croissantes:

On appelle fréquence cumulée croissante, noté $f_i cc$ associé à la valeur x_i la somme des fréquences passés tel que :

$$f_i cc = f_{i-1} cc + f_i = \sum_{j=1}^{i} f_j \qquad i = 2, 3, \dots, r$$

Avec $f_1 cc = f_1$

12. Fréquences cumulées décroissantes:

On appelle fréquence cumulée décroissante, noté $f_i cd$ associé à la valeur x_i la différence des fréquences passés tel que :

$$f_i cd = f_{i-1} cd - f_{i-1} = 1 - \sum_{j=1}^{i-1} f_j \qquad i = 2, 3, \dots, r$$

Avec $f_1 cd = 1$

Pour l'exemple 2 :

x_i: Age (années)	n_i : nombre d'étudiants	f_i	$f_i cc$	$f_i cd$
18	15	0.15	0.15	1
19	13	0.13	0.28	0.85
20	8	0.08	0.36	0.72

21	21	0.21	0.57	0.64
22	18	0.18	0.75	0.43
23	16	0.16	0.91	0.25
24	9	0.09	1	0.09

Pour l'exemple 3 :

Poids des poissons x_i	Nombre de poissons n_i	f_i	$f_i$cc	$f_i$cd
[100, 300[	8	0.04	0.04	1
[300, 500[	82	0.41	0.45	0.96
[500, 700[	24	0.12	0.57	0.55
[700, 900[	80	0.4	0.97	0.43
[900, 1500[	6	0.03	1	0.03

Remarque 3 :

Arrondir à « n » chiffres après la virgule consiste à supprimer tous les autres chiffres qui viennent à droite du $n^{ème}$ chiffre en tenant compte du $(n+1)^{ème}$ chiffre tel que :

Si le $(n+1)^{ème}$ chiffre $\in \{0, 1, 2, 3, 4\}$ ($n < 5$) alors pas de changement dans le $n^{ème}$ chiffre.

Si le $(n+1)^{ème}$ chiffre $\in \{5, 6, 7, 8, 9\}$ ($n \geq 5$) alors ajouter « 1 » au $n^{ème}$ chiffre.

Exemple :

Arrondir la valeur de X à 4 chiffres après la virgule :

$X = 3.26783998$, le $5^{ème}$ chiffre < 5 donc $X = 3.2678$

$X = 1.78236439$, le $5^{ème}$ chiffre ≥ 5 donc $X = 1.7824$