

Chapitre I : Statistique descriptive

1- Définition de la statistique

La statistique est la science qui englobe un ensemble de méthodes mathématiques (numériques et graphiques) permettant de résumer l'information relative à une population étudiée selon une ou plusieurs variables. Le but d'une étude statistique est généralement de déterminer certaines caractéristiques moyennes d'**une population** qu'on appelle aussi **un univers**. Les éléments de cette population peuvent être des individus, des objets réels, ou des éléments abstraits.

Les éléments de la population sont généralement appelés des **unités d'observation**. Les caractéristiques auxquelles on s'intéresse (Age, Consommation, Résultat) sont appelées des variables car leur valeur varie en fonction de l'unité observée.

D'habitude, la taille de la population est trop élevée pour que l'on puisse examiner ou observer tous ces individus. On doit alors se limiter à **un échantillon**, c'est-à-dire, un sous ensemble de la population. La population est l'échantillon exhaustif. On désigne par N (parfois $N=\infty$) la taille de la population et par n la taille d'un échantillon.

On notera une variable par son initiale majuscule (A, C, R) ou généralement par X, Y, Z.....etc. Les valeurs possibles numériques ou non numériques (voir exemples ci-dessous) ou **modalités** d'une variable seront indiquées par la même lettre minuscule affectées d'indice : $x_1, X_2, \dots, y_1, Y_2, \dots$ etc. Les modalités d'une variable X sont toutes différentes.

La statistique est composée de deux disciplines : La statistique descriptive et la statistique inférentielle.

A- La statistique descriptive

C'est un ensemble de méthodes numériques et graphiques. Elle permet grâce à des indices mathématiques de caractériser des populations ayant au moins un caractère en commun. L'objectif de la statistique descriptive est donc d'obtenir des éléments de caractérisation de la population étudié.

B- La statistique inférentielle

C'est la discipline qui permet de juger le caractère étudié dans la population, soit en considérant que les données obtenues expliquent bien la population considérée, soit que ces données sont insuffisantes ou permettent de rejeter une hypothèse préalablement fixée. Elle est connue en général comme la discipline des tests d'hypothèse.

2- Eléments de La statistique descriptive

2-1- Population ou univers : un ensemble d'individus ou d'éléments ayant en commun une caractéristique ou une propriété.

2-2- Echantillon : un sous-ensemble ou une partie de la population. L'étude de l'échantillon a pour objectif d'extrapoler l'information obtenue de cet échantillon à la population entière (on explique le tout par la partie); surtout quand la population est formée d'un nombre infini d'éléments.

2-3- Unité statistique : un élément de la population qui peut faire l'objet d'une mesure ou d'une observation.

2-4- Variable ou caractère : Une forme physique observable sur un individu (couleur, longueur,....etc.). En général, on fait le lien entre le caractère et la variable du fait que le caractère lui-même peut varier d'un individu à un autre. On distingue plusieurs types de variables :

- Variable quantitative : lorsque les modalités sont des nombres qui expriment des quantités (revenu de 50000 dinars, taille de 185 cm,.....etc.) ;

- Variable quantitative continue : si l'ensemble de modalités est un intervalle de nombres réels (poids entre 0 et 300 kg, taille entre 20 et 50 cm,.....etc.) ;

- variable quantitative discrète : si l'ensemble des valeurs possibles est fini ou infini mais dénombrable (nombres de frères, nombre d'accidents d'un assuré) ;

- variable qualitative ou catégorielle : lorsque les modalités représentent des qualités (sexe : masculin, féminin) ;

- variable en catégories ordonnées : lorsque les modalités ne sont pas des quantités numériques mais peuvent être ordonnées (état de patient : il va mal, il est stable, il va mieux).

2-5- Effectif ou fréquence absolue : associée à une valeur d'un caractère, est le nombre de fois où cette valeur du caractère a été observée.

2-6- Fréquence relative : associée à une valeur d'un caractère, est le rapport de la fréquence absolue correspondante à cette valeur de caractère au nombre d'individus de l'échantillon.

2-7- Série statistique ou distribution statistique : associée à un caractère est, l'ensemble des valeurs du caractère, avec un regard, les fréquences absolues ou relatives correspondantes.

La méthode statistique comporte essentiellement trois phases :

- Une phase matérielle où il s'agit de rassembler des données, de les regrouper et de les présenter sous forme de tableaux ou graphes.
- Une phase analytique qui consiste à réduire les données à un nombre limité de paramètres caractéristiques (moyenne, variance, mode.....etc.) susceptibles de décrire la série statistique.

L'ensemble de ces deux phases constitue l'objet essentiel de la statistique descriptive (ou déductive) dont les résultats restent limités aux échantillons étudiés.

- Une phase interprétative qui est la base de la statistique inductive, est qui permet de déduire des résultats obtenus sur un échantillon des conclusions relatives à l'ensemble de la population d'où est extrait cet échantillon. Ces conclusions doivent tenir compte de la marge d'erreur due au fait que les données sont seulement partielles. Les méthodes utilisées n'ont de sens que si elles sont justifiées par des résultats ultérieurs.

Une série statistique peut être présentée sous forme de tableaux ou de graphes.

1- Représentation en tableaux

1-1- Tableau de fréquence à un caractère : il établit la correspondance entre deux séries de nombres, l'une constitue par les valeurs du caractère étudié, l'autre par les effectifs correspondants (ou les fréquences relatives correspondantes).

- Exemple d'une série quantitative discrète : l'échantillon est un immeuble de 64 familles, le caractère étudié étant le nombre d'enfants par famille.

Tableau 1

Nombre d'enfants	0	1	2	3	4	5	Total
Nombres de familles	16	18	14	11	3	2	64
Fréquence relative	0.25	0.281	0.218	0.172	0.047	0.031	1

- Exemple d'une série qualitative : l'échantillon est l'immeuble de 64 familles de l'exemple précédant, le caractère est la profession du chef de famille.

Tableau 2

Profession	effectif	Fréquence relative
Ouvriers et employés	24	0.375
Cadres moyens et supérieurs	9	0.140
Commerçants	10	0.156
Fonctionnaires	15	0.234
Professions libérales	6	0.093
	64	1

1-2- Tableau de fréquences cumulées : La série statistique peut être présentée sous une forme dite cumulée, de deux manières suivantes :

- Cumul par valeurs inférieures ou effectifs cumulés croissants

Tableau 3

Nombres d'enfants	Effectifs cumulés croissants
Moins de 0 enfant	0 familles
Moins de 1 enfant	16 familles
Moins de 2 enfants	$16 + 18 = 34$ familles
Moins de 3 enfants	$34 + 14 = 48$ familles
Moins de 4 enfants	$48 + 11 = 59$ familles
Moins de 5 enfants	$59 + 3 = 62$ familles
Moins de 6 enfants	$62 + 2 = 64$ familles

- Cumul par valeurs supérieures ou effectifs cumulés décroissants

Tableau 4

Nombres d'enfants	Effectifs cumulés décroissants
0 enfant ou plus	64 familles
1 enfant ou plus	$64 - 16 = 48$ familles
2 enfants ou plus	$48 - 18 = 30$ familles
3 enfants ou plus	$30 - 14 = 16$ familles
4 enfants ou plus	$16 - 11 = 5$ familles
5 enfants ou plus	$5 - 3 = 2$ familles
6 enfants ou plus	$2 - 2 = 0$ familles

NB : On peut remarquer que pour chaque valeur du caractère, la somme des effectifs cumulés croissants et des effectifs cumulés décroissants est égale à l'effectif total.

1-3- Tableau de fréquences à deux caractères : Si l'on s'intéresse à deux caractères différents dans un même échantillon, il est possible de représenter l'ensemble des renseignements dans un même tableau (tableau à double entrée).

Tableau 5

X \ Y	2	3	4	5	6	Total nombre d'appartements
2	8	5	2	0	0	15
3	5	7	4	2	0	18
4	3	6	8	9	5	31
Total nombre de familles	16	18	14	11	5	64

X: Nombre de personnes vivant dans un appartement ; Y : nombre de pièces par appartement.

- L'intersection d'une ligne et d'une colonne du tableau est le nombre de fois où l'on a observé X personnes vivants dans un appartement de y pièces. Exp : on a observé 7 fois 3 personnes vivantes dans un appartement de 3 pièces.

1-4- Cas d'une série quantitative continue : Afin de rendre la série statistique plus commode à étudier, il est nécessaire de regrouper les valeurs du caractère en intervalles successifs et contigus, tels que dans chaque intervalle ou classe, on ne distingue pas les valeurs du caractère qui y sont comprises. Les nombres entre lesquels sont comprises ces valeurs constituent **les limites de classe**. Dans chaque classe, on remplace les valeurs du caractère observées par une valeur unique. Celle du milieu de l'intervalle ou **centre de classe**.

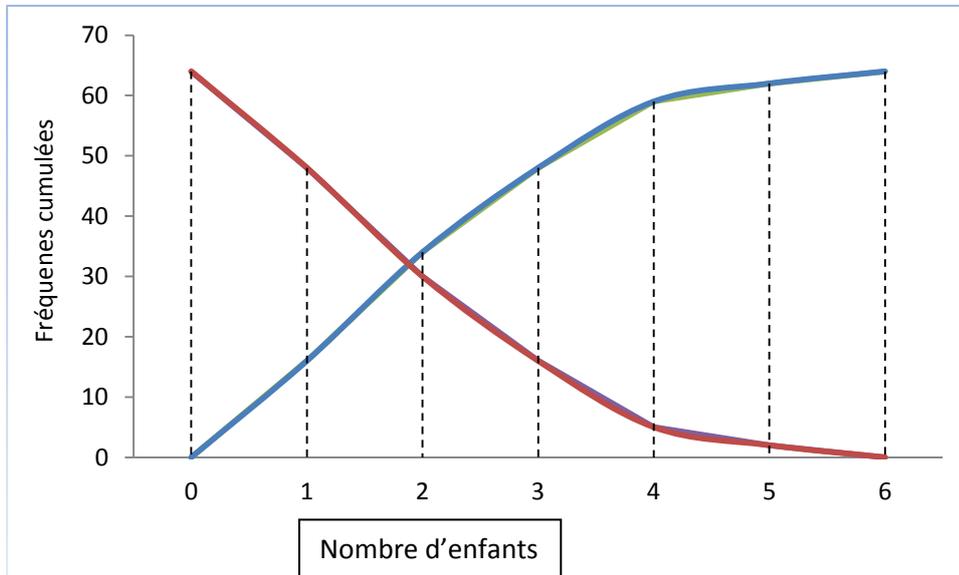
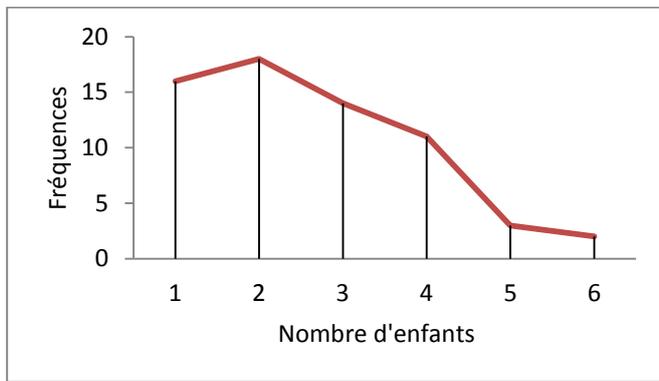
Tableau 6

Classes	Centre de classe	Effectif	Effectif cumulé croissant	Effectif cumulé décroissant
1.55-1.59	1.57	3	3	$80 - 3 = 77$
1.60-1.64	1.62	12	15	$77 - 12 = 65$
1.65-1.69	1.67	18	33	$65 - 18 = 47$
1.70-1.74	1.72	25	58	$47 - 25 = 22$
1.75-1.79	1.77	15	73	$22 - 15 = 7$
1.80-1.84	1.82	5	78	$7 - 5 = 2$
1.85-1.89	1.87	2	80	$2 - 2 = 0$
		80		

2- représentations graphiques

Les représentations graphiques ont l'avantage d'offrir une meilleure vue d'ensemble de la série statistique que les tableaux. Elles permettent par simple lecture, de voir les caractéristiques essentielles de la série, et aussi de comparer des séries différentes.

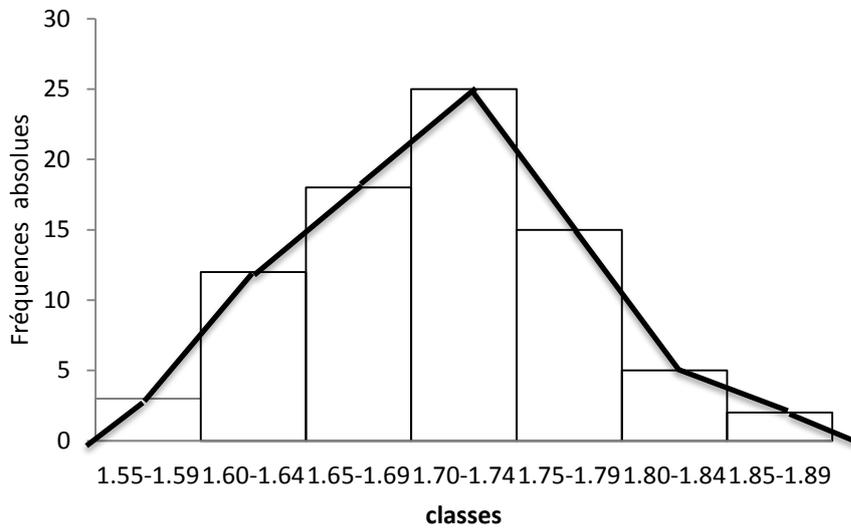
2-1- Caractère discontinu-diagramme en bâtons : Lorsque le caractère est discontinu, on utilise le diagramme en bâtons : les valeurs du caractère sont portées en abscisses, les fréquences correspondantes sont représentées par des traits pleins, en ordonnées. Si l'on joint les sommets des bâtons, on obtient **le polygone des fréquences**. Exp : diagramme en bâtons des tableaux 1, 3, 4 respectivement.



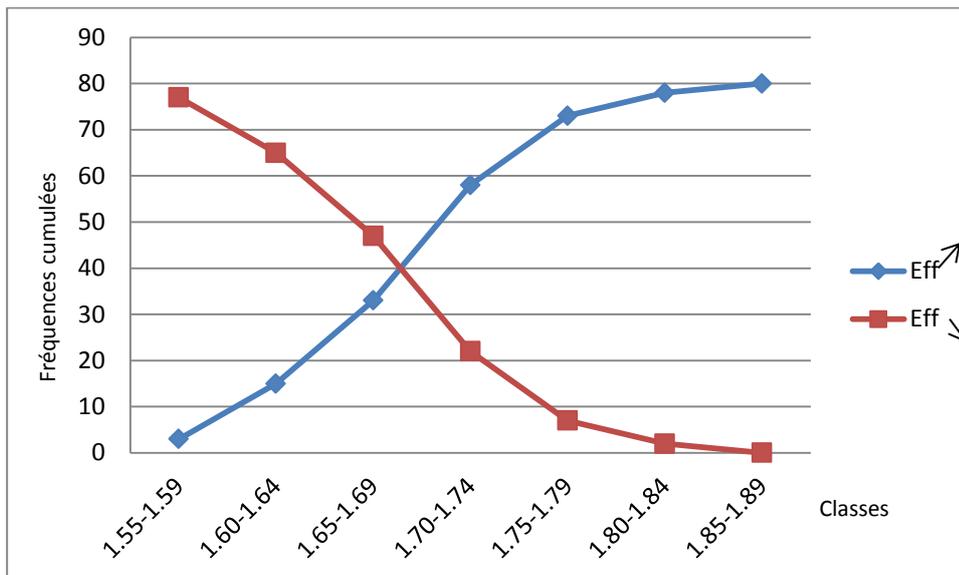
2-2- Caractère continu-Histogramme : Dans le cas d'un caractère continu, on utilise l'histogramme, qui constitue une généralisation du diagramme en bâtons à la notion de classe.

a) Séries à classes égales : chaque classe est représentée par un rectangle dont la base est égale à l'intervalle de la classe et dont la hauteur est égale à l'effectif correspondant. L'histogramme est constitué en fait par le contour polygonal enveloppant l'ensemble de ces rectangles.

- **Le polygone des fréquences absolues ou des fréquences relatives** : il s'obtient en joignant les points dont les abscisses sont les milieux des différentes classes et dont les ordonnées sont les effectifs ou les fréquences relatives correspondantes. Exemple relatif au tableau 6.



- **Le polygone des effectifs cumulés** : on observant que les effectifs cumulés croissants correspondent aux frontières supérieures des différentes classes, et les effectifs cumulés décroissants aux frontières inférieures des classes. On peut, à l'aide du tableau 6, construire les polygones des effectifs cumulés suivants.



b) Séries à classes inégales : une classe dont l'étendue est égale à n fois l'intervalle de classe fondamentale est représentée avec une ordonnée égale à l'effectif de cette classe divisé par n, de telle sorte que l'aire relative à cette classe soit bien proportionnelle à son effectif.

Exemple : une enquête portant sur les salaires mensuels perçus dans une certaine entreprise a fourni les renseignements suivants :

Salaire mensuel en Euro	Effectif		Salaire mensuel en Euro	Effectif
Entre 2000 et 3000	34		Entre 7000 et 8000	1.5
Entre 3000 et 4000	52		Entre 8000 et 9000	1.5
Entre 4000 et 5000	60		Entre 9000 et 10000	1.5
Entre 5000 et 6000	20		Entre 10000 et 11000	1.5
Entre 6000 et 7000	8			
Entre 7000 et 11000	6			

La dernière classe vaut 4 intervalles de classe ; il faut donc réaménager sa présentation ainsi :

$$\text{nb sous classe} = \frac{\text{Étendue de la classe inégale}}{\text{Intervalle de la classe fondamentale}}$$

$$\text{ni sous classe} = \frac{\text{ni de la classe inégale}}{\text{Nb de sous classe}}$$

