

Chapitre 1 : Statistique descriptive (Partie 1)

1)- Définitions:

- ✚ **Statistique** : la statistique est ensemble des méthodes qui servent à **organiser** les épreuves fournissant des observations, à **analyser** celles-ci et à **interpréter** les résultats.
- ✚ **Statistique descriptive** : a pour but de décrire c-à-d de résumer ou représenter les données.

Questions typiques

- Représentation graphique
- Paramètres de position
- Paramètre de dispersion
- Paramètre de forme.

2)- Notions de bases :

- **POPULATION** La collection d'objets ou de personnes étudiées (élèves, habitants, voitures...).
- **INDIVIDU** élément de la population étudiée. (un élève, un habitant, une voiture,...).
- **ECHANTILLON** partie de la population étudiée. Nombre d'individus dans un échantillon noté n est appelé taille de l'échantillon
- **VARIABLE (CARACTERE)** propriété commune aux individus de la population, que l'on veut étudier.

Exemple 1: Nous résumons les différents concepts dans cet exemple :

Population : l'ensemble des tous les employés d'une usine.

Individu : chaque employé de l'usine.

Caractère : le salaire, l'état matrimonial, le nombre d'enfants,... etc.

Les modalités du caractère : marié, célibataire, divorcé et veuf sont les modalités de l'état matrimonial, par exemple.

3) Type de caractères

Caractère quantitatif : Modalités mesurables (tout ce qu'on peut mesurer par un instrument)

Discret	Continu
<ul style="list-style-type: none"> Les modalités prennent des valeurs isolées (ne peut prendre qu'un nombre fini ou dénombrable de valeurs) <p>Exemple: Nombre d'enfants dans une famille</p> <p>0,1,2,.....etc</p>	<ul style="list-style-type: none"> Les modalités prennent des valeurs très rapprochées (peut prendre toutes les valeurs d'un intervalle de l'ensemble des nombres réels) <p>Exemple: Taux du glucose dans le sang 0.88, 0.89, 1.01, 1.03, 1.27...etc</p>

Caractère qualitatif : modalités non mesurables (on ne peut associer ni valeur numérique ni un ordre naturel (type de voiture, couleur des cheveux,...)).

Ordinale	Nominale
<p>Les modalités sont des qualités peuvent être ordonnées</p> <p>Exemples:</p> <p>Taille de vêtements (S, M,L, XL...)</p>	<p>Pas d'ordre</p> <p>Exemples:</p> <ul style="list-style-type: none"> le sexe La couleur des yeux (Bleu, vert, Marron...)

4) Représentation des données :

Après la réalisation d'une étude statistique, les données recueillies peuvent être résumées et représentées sous forme de

- Tableaux statistiques
- Graphes
- Paramètres statistiques

- **Notations**

Effectif ou fréquence absolue : (noté n_i) nombre d'apparitions de la valeur associé à un caractère dans un échantillon.

Fréquence relative : (noté f_i)

$$f_i = \frac{n_i}{n}$$

Série statistique : l'ensemble des valeurs du caractère avec en regard, les fréquences absolue ou relative correspondantes.

Remarque: On a toujours

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

$$0 \leq f_i \leq 1, \quad \sum_{i=1}^k f_i = 1$$

Présentation sous forme des tableaux statistiques

1) Série statistique dans le cas d'une variable quantitative discrète

Exemple 1 :

Nombre d'enfant	0	1	2	3	4	5	total
Nombre de famille ou effectif : n_i	16	18	14	11	3	2	64
Fréquence relative : f_i	0,25 0	0,28 1	0,21 8	0,17 2	0,04 7	0,03 1	1

Population étudié : les familles

L'échantillon sur lequel porte l'étude : familles d'un immeuble ; $n=64$.

Le caractère étudié est le nombre d'enfants par famille. C'est un caractère quantitatif discret.

2) Série statistique dans le cas d'une variable quantitative continue

Exemple: poids de 161 nouveau-nés

Classe	Limites de la classe (kg)	Centre de classe x_i	Effectif n_i	Fréquence relative f_i	n_i^{\uparrow}
--------	---------------------------	------------------------	----------------	--------------------------	------------------

1	2.2-2.5	2.35	5	0.031	5
2	2.5-2.8	2.65	11	0.068	16
3	2.8-3.1	2.95	24	0.148	40
4	3.1-3.4	3.25	40	0.248	80
5	3.4-3.7	3.55	42	0.259	122
6	3.7-4.0	3.85	20	0.124	142
7	4.0-4.3	4.15	13	0.08	155
8	4.3-4.6	4.45	6	0.037	161

3) Série statistique dans le cas d'une variable qualitative

Exemple: L'analyse du sang de 100 individus à donné les résultats suivants

Groupe SANGUIN	Effectif n_i	Fréquence relative f_i
O	40	0.40
A	43	0.43
B	12	0.12
AB	5	0.05

Remarque : On peut découper une série de valeurs en classe c.-à-d. passer de caractère discret au caractère continue en appliquant l'une des formules suivantes:

1) Si $n < 50$

$$\text{Nombre de classes} = k \cong \sqrt{n}$$

2) Formule de Sturge

$$k = 1 + 3.332 \log_{10}(n)$$

3) Formule de Yule

$$k = 2.5 \sqrt[4]{n}$$

Amplitude des classes:

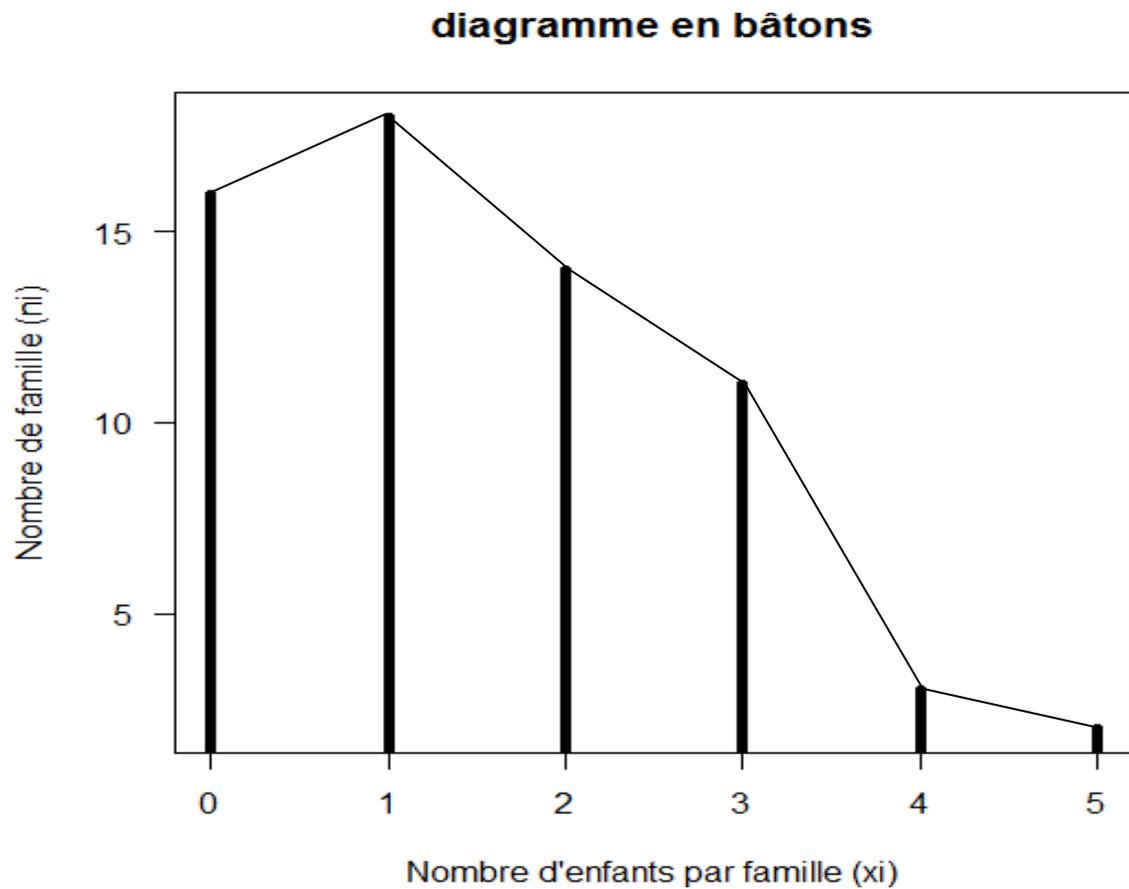
Il faut mieux d'utiliser des classes de même amplitude a donnée par

$$a = \frac{X_{\max} - X_{\min}}{k}$$

Représentation graphique d'une série statistique

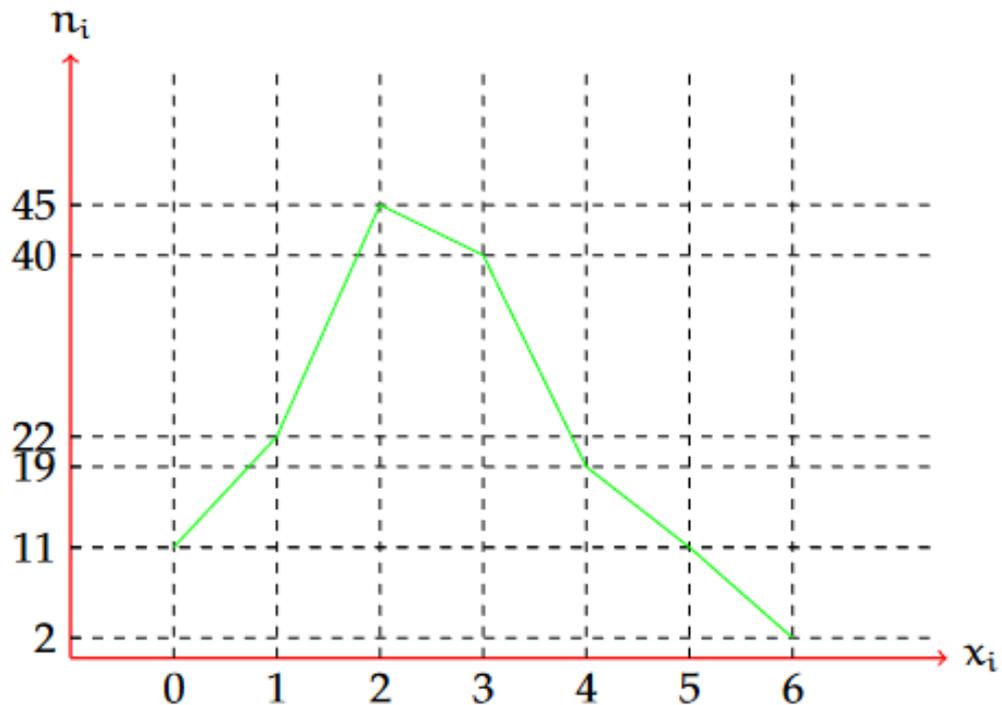
1) Caractère discret (discontinue)

- **Diagramme en Bâtons:** c'est un ensemble de bâtons ayant pour abscisses les valeurs X_1, X_2, \dots, X_n du caractère et en chacun des points d'abscisses x_i une ordonnée proportionnelle à l'effectif n_i de X_i



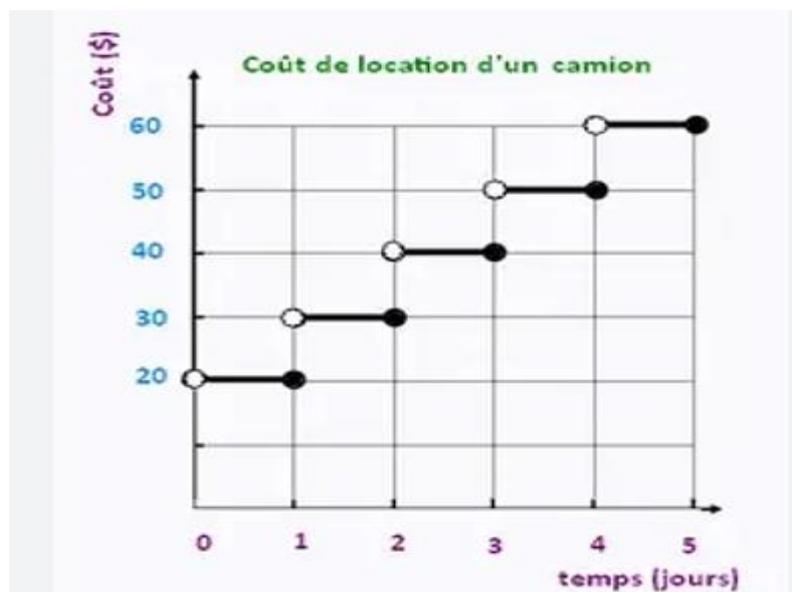
- **Polygone des Fréquences**

On l'obtient en joignant par des segments de droite les extrémités des bâtons



- **Diagramme cumulatif (diagramme en escaliers)**

En abscisse figurent, encore une fois, les observations de la variable considérée, tandis qu'en ordonnée figurent maintenant les effectifs cumulés, les fréquences cumulées ou les pourcentages cumulés. Dans le diagramme cumulatif les bâtons ont des longueurs proportionnelles aux effectifs cumulés (ou aux fréquences cumulées)



Exemple 1:

Tracer le diagramme en escaliers de l'exemple : Répartition de 150 grenouilles

Exemple 2

Tracer le diagramme en escaliers de tableau suivant

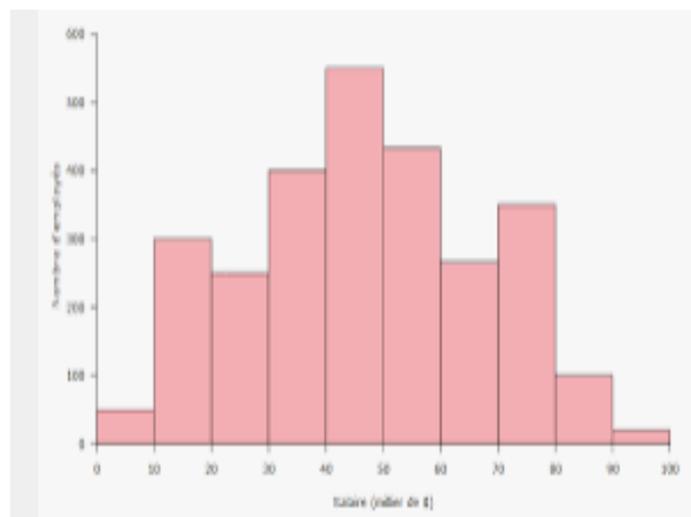
Valeurs	effectif	fréquence
0	10	0.2
1	5	0.1
3	10	0.2
4	15	0.3
5	10	0.2

2) Caractère continue

- **Histogramme:** c'est un ensemble de rectangles ayant pour largeur l'amplitude de la classe et pour hauteur l'effectif de la classe

Exemple:

tracer l'histogramme de l'exemple « poids des nouveaux nés »



- **Polygone des effectifs ou polygone des fréquences**

Est la ligne brisée joignant les milieux des bases supérieures des différents rectangles adjacents.

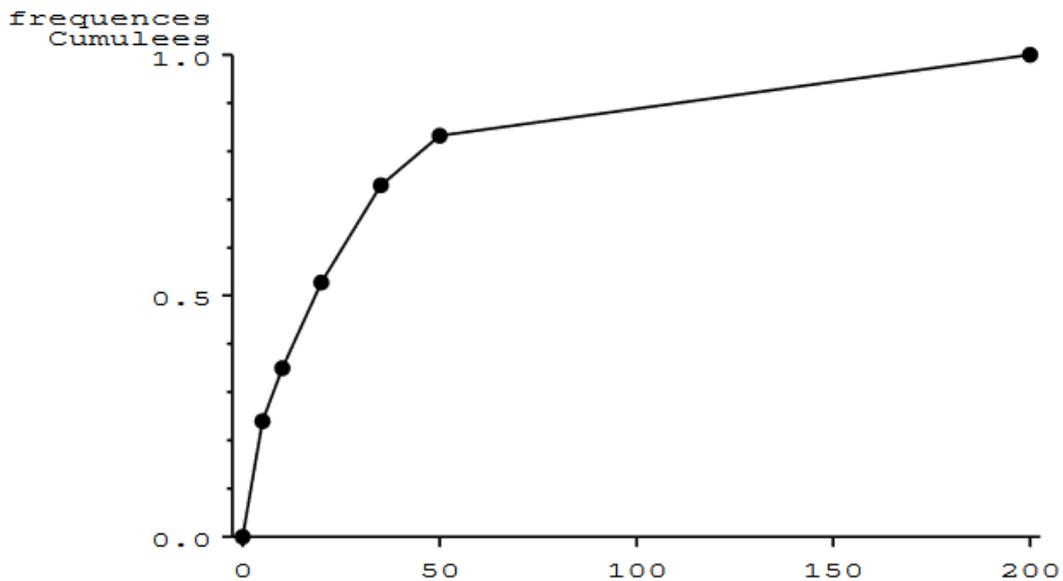
Exemple:

tracer la polygone des effectifs pour le même exemple

La courbe cumulative croissante et décroissante (poids des nouveaux nés)

Pour la variable quantitative continue étudiée, chaque classe considérée doit d'abord être représentée par un point unique dont l'abscisse est la borne

supérieure/inférieure de la classe et l'ordonnée est l'effectif (ou la fréquence, ou le pourcentage) cumulé de cette classe. La courbe cumulative est alors la courbe joignant les points en question



3) Dans le cas d'une variable qualitatives

Les représentations graphiques que l'on rencontre avec les variables qualitatives sont assez nombreuses. Les deux plus courantes, qui sont aussi les plus appropriées, sont :

- le diagramme en colonnes (en rectangles),
- le diagramme en secteurs (diagramme en camembert)

