

Socle Commun

INTRODUCTION AUX STATISTIQUE DESCRIPTIVE ET PROBABILITÉS

Avec exercices corrigés

1^{ÈRE} ANNÉE MATHÉMATIQUES ET DE
L'INFORMATIQUE

Mr : ABDELKEBIR Saad

Chapitre 1.

Notions de base et vocabulaire statistique

ABDELKEBIR SAAD

INTRODUCTION :

La statistique a pour objet de recueillir des observations portant sur des sujets présentant une certaine propriété et de traduire ces observations par des nombres qui permettent d'avoir des renseignements sur cette propriété.

Le but de la statistique descriptive est de structurer et de représenter l'information contenue dans les données.

1- Notions de base et vocabulaire statistique :

1-1 La population :

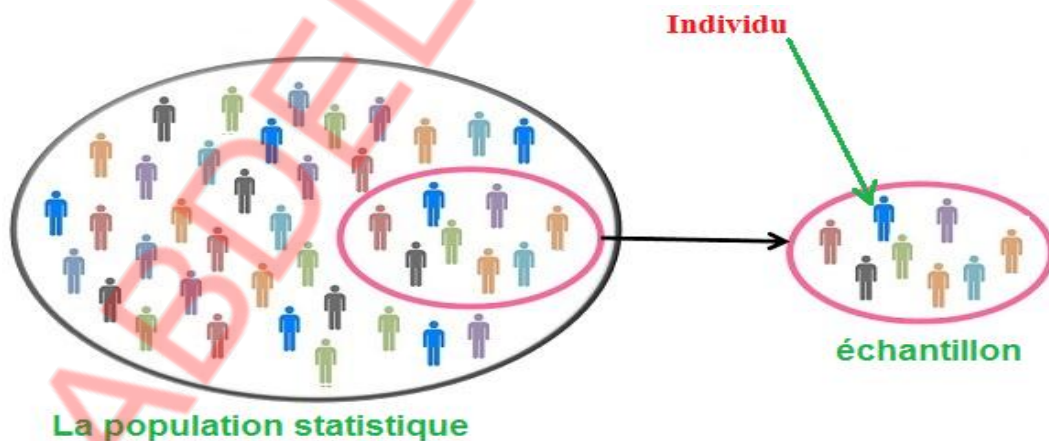
Définition :

La population statistique est l'ensemble sur lequel porte l'étude statistique. C'est l'univers de référence, noté : $\Omega = \{x_1, x_2, \dots, x_N\}$.

1-2 Individus :

Définition :

Les individus ou unités statistiques sont les éléments qui constituent la population. Ils possèdent tous le caractère étudié, et sont notés x_i avec i variant de l'ensemble de nombre naturelle (i.e : 1 à N) pour une population de taille N .



1-3 Les caractères statistiques :

Le caractère désigne une grandeur ou un attribut, observable sur un individu et susceptible de varier prenant ainsi différents états appelés modalités.

Les caractères statistiques sont de deux types : **quantitatifs** et **qualitatifs**.

1-3-1- Les caractères qualitatifs :

Ce type de variable **concerne des valeurs qui correspondent aux qualités non quantifiables des individus**. On ne peut pas dire qu'une catégorie est supérieure ou inférieure à une autre.

Exemple :

- i- **Le sexe** est un exemple de ce type de variable.

On dit qu'elles sont qualitatives parce que les différences entre les catégories sont précisément qualitatives.

- ii- **Couleur des cartables** des élèves d'une classe. On dit qu'elles sont qualitatives parce que les différences entre les catégories sont précisément qualitatives.

1-3-2- Les caractères quantitatifs :

La variable quantitative **concerne des valeurs dans un ensemble standard de valeurs numériques**. Cela signifie que l'on peut la mesurer. La variable quantitative comprend deux types de variables :

- | |
|--|
| • <u>Discrète</u> . L'ensemble est fini ou dénombrable.
Exemple : Le nombre d'enfants dans une famille. Le nombre de pièces dans chaque appartement. |
| • <u>Continue</u> . L'ensemble est infini et ne peut être numéroté. Cela signifie qu'il contient un certain intervalle.
Exemple : pourrait être le temps. |

2- LES TABLEAUX STATISTIQUES

Les traitements des données s'effectuent à base d'un tableau en fonction de caractère, il existe des tableaux correspondants.

2.1. **Qu'est-ce qu'un effectif ?**

- **L' effectif total** « N »: c'est le nombre de valeurs dans la série statistique.
- **L' effectif d'une valeur donnée** « n_i »: c'est le nombre de fois ou la valeur apparaît pour cette série.

Exemple 2.1:

- Prenons la série 2, 2, 4, 34, 11, 4, 2
- L'effectif total pour cette série est de 7 puisqu'il y a 7 valeurs (des nombres ici).
- L'effectif de la valeur 2 est de 3 puis que la valeur 2 apparaît 3 fois dans la liste.
- Pour la valeur 4 c'est 2, puisque le nombre 4 apparaît à 2 reprises

Exemple 2.2 :

- Prenons la série : bleu, bleu, vert, noir, rouge, vert, bleu, noir, noir
- L'effectif total pour cette série est de 9 puisqu'il y a 9 valeurs (Ici, des couleurs).
- L'effectif de la valeur bleu est de 3 puisque le bleu apparaît 3 fois dans la liste.
- Pour la valeur vert c'est 2, puisque le vert apparaît à 2 reprises

2.2. Qu'est-ce qu'une fréquence ?

La fréquence d'une valeur donnée, c'est le quotient (la division) de l'effectif de la valeur par l'effectif total. Intuitivement, elle indique la proportion de la présence de la valeur dans la liste.

Exemple 2.3 :

- Prenons la série : 2, 2, 4, 34, 11, 4, 2, 1, 9, 9
- L'effectif de la valeur 2 est de 3 puisque le nombre 2 apparaît 3 fois dans la liste.
- L'effectif pour cette série est de 10 puisqu'il y a 10 valeurs.
- La fréquence = (effectif de la valeur « 3 » / effectif total N=10) = 3 / 10

2.3. L'effectif cumulé croissant « resp décroissant »

Définition :

- L'effectif cumulé croissant, ECC, noté « $N_i (C)$ » d'une valeur est la somme des effectifs de cette valeur avec la précédente.

$$N_0 = 0, N_1 = n_1, N_2 = n_1 + n_2, \dots, N_n = n_1 + n_2 + \dots + n_n = N$$

- L'effectif cumulé décroissant, ECD, noté « $N_i (D)$ » d'une valeur est la somme des effectifs de cette valeur avec la suivante.

$$N_0 = N, N_1 = N - n_1, N_2 = N - n_1 - n_2, \dots, N_n = N - n_1 - n_2 - \dots - n_n = 0$$

2.4. Fréquence cumulée croissante « resp décroissante »

Définition :

Somme des fréquences de la classe étudiée et des fréquences des classes qui lui sont inférieures.

On distingue :

- **Les fréquences cumulées croissantes** : la fréquence cumulée croissante noté

« $F_i(C)$ » de la dernière classe vaut 1. Tels que :

$$F_0 = 0, F_1 = f_1, F_2 = f_1 + f_2, \dots, F_n = f_1 + f_2 + \dots + f_n = 1$$

- **Les fréquences cumulées décroissantes** : la fréquence cumulée décroissante noté « $F_i(D)$ » de la première classe vaut 1. Tels que :

$$F_0 = 1, F_1 = 1 - f_1, F_2 = 1 - f_1 - f_2, \dots, F_n = 1 - f_1 - f_2 - \dots - f_n = 0$$

Ce tableau se traduit de la manière suivante :

Individus (Caractère) X_i	Effectifs n_i	f_i	$N_i(C)$	$N_i(D)$	$F_i(C)$	$F_i(D)$
X_1	n_1	f_1	0	N	0	1
X_2	n_2	f_2	n_1	$N - n_1$	f_1	$1 - f_1$
X_3	n_3	f_3	$n_1 + n_2$	$N - n_1 - n_2$	$f_1 + f_2$	$1 - f_1 - f_2$
N	$\sum_{i=1}^n n_i$	1	$n_1 + n_2 + n_3 = N$	$N - n_1 - n_2 - n_3 = 0$	$f_1 + f_2 + f_3 = 1$	$1 - f_1 - f_2 - f_3 = 0$

Remarque :

Soit la série statistique X_i Avec l'effectifs correspondante n_i . tq : $i \in \{1, 2, \dots, k\}$
 Nous avons :

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

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

$$0 \leq f_i \leq 1. \forall i \in \{1, 2, \dots, k\}$$

Exemple 2.1

Au cours d'un recensement effectué par rapport à l'âge des élèves sur 35 élèves ; on a obtenu les résultats suivants :

10 ; 11 ; 10 ; 12 ; 10 ; 13 ; 12 ; 11 ; 14 ; 13 ; 15 ; 16 ; 15 ; 13 ; 14 ; 11 ;
 12 ; 11 ; 13 ; 14 ; 14 ; 15 ; 16 ; 13 ; 16 ; 15 ; 14 ; 14 ; 13 ; 12 ; 11 ; 15 ;
 12 ; 14 et 15.

Donc,

Effectif total est N=35.

x_i	n_i	f_i	$N_i(C)$	$N_i(D)$	$F_i(C)$	$F_i(D)$
10	03	3/35	0	35	0	1
11	05	5/35	3	32	3/35	32/35
12	05	5/35	8	27	8/35	27/35
13	06	6/35	13	22	13/35	22/35
14	07	7/35	19	16	19/35	16/35
15	06	6/35	26	9	26/35	9/35
16	03	3/35	32	3	32/35	3/35
16	03	3/35	35	0	35/35=1	0
N	$\sum_{i=1}^7 n_i$ = 35					

3- Représentation graphique des données :

La représentation graphique est une synthèse de l'information qui fait apparaître la forme globale de la distribution des données.

Le choix du type de graphe dépend de la nature des variables.

3-1 - Cas d'une variable quantitative et qualitative discontinue « discrète »

i- 1^{ère} cas le variable qualitative :

1- Diagramme en camembert ou en cercle :

Le principe du graphe consiste à diviser un cercle en secteurs proportionnels aux fréquences des classes en partant de la position « Midi » pour disposer dans le sens horaire les secteurs représentant les classes.

i.e :

$$N \rightarrow 360^\circ \quad \text{Or} \quad 1 \rightarrow 360^\circ$$

$$n_i \rightarrow (\alpha_i)^\circ \quad \text{Or} \quad f_i \rightarrow (\alpha_i)^\circ$$

$$(\alpha_i)^\circ = \frac{n_i \times 360^\circ}{N} \quad \text{Or} \quad (\alpha_i)^\circ = f_i \times 360^\circ$$

Alors :

Nous calculons la mesure d'angle de $(\alpha_i)^\circ$, pour chaque n_i .

Exemple 3.1 :

Voir le tableau suivant :

Couleur	Effectif	Fréquence	Angle
Roux	2	8.0%	29°
Bruns	6	24.0%	86°
Blonds	7	28.0%	101°
Noirs	4	16.0%	58°
Châtains	6	24.0%	86°

Répartition de 25 Sujets en fonction de la couleur des cheveux.

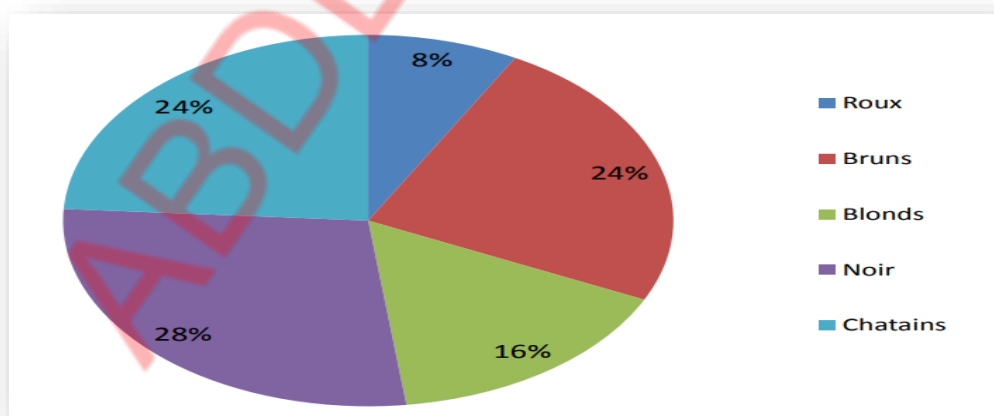


Diagramme en camembert ou en cercle ou bien diagramme circulaire

2^{ème} cas le variable quantitative :

Ce type de variable est associée généralement à un diagramme en bâtons où l'axe horizontal des abscisses porte les valeurs prises par la variable (x_i) tandis que l'axe vertical des ordonnées porte l'effectif absolu (n_i) observé.

Exemple 3.2

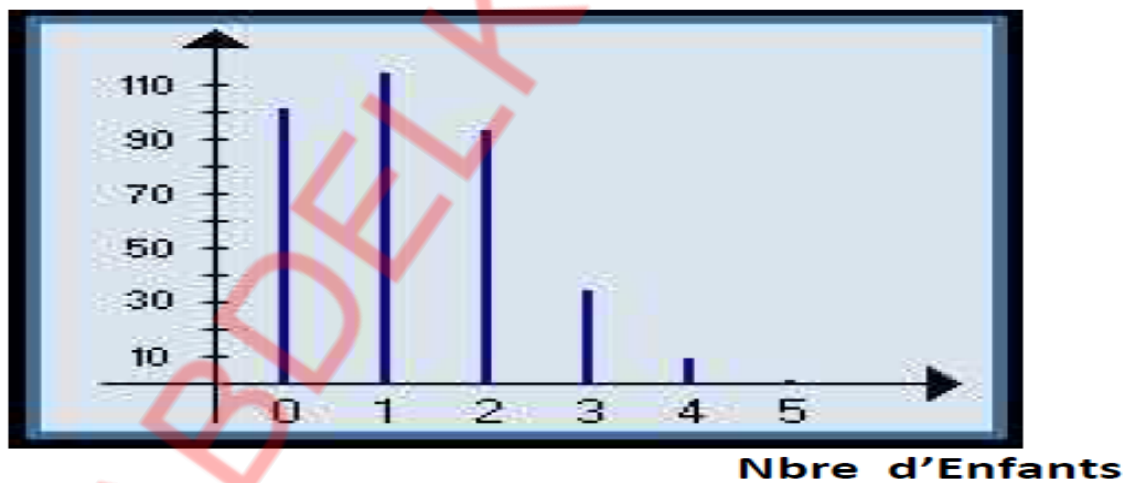
Soit le nombre d'enfant par famille est représenté sur le graphe ci-dessous.

Pour les données non groupées, on peut représenter les fréquences absolues, relatives ou cumulées par un diagramme en bâtons.

Nombre d'Enfants x_i	1	2	3	4	5	6	N
Nombre de Famille ayant ce nombre d'Enfant n_i	103	115	95	35	10	2	360

Donc le diagramme correspondant du tableau statistique précédent :

Effectifs



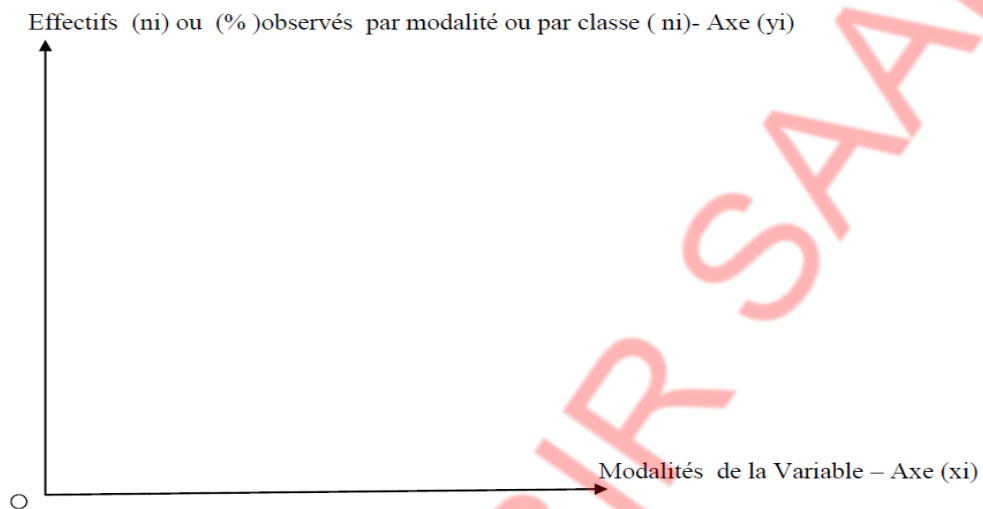
Histogramme en Bâtons

3-2 - Cas d'une variable quantitative et qualitative continue

il doit être identique a celui du tableau dont il est issu.

Des coordonnées :

- 1- les modalités de la variable sur l'axe (x).
- 2- les effectifs (n_i) ou les fréquence (f_i) sur l'axe (y).



3-2-1. Histogrammes des effectifs ou bien fréquences :

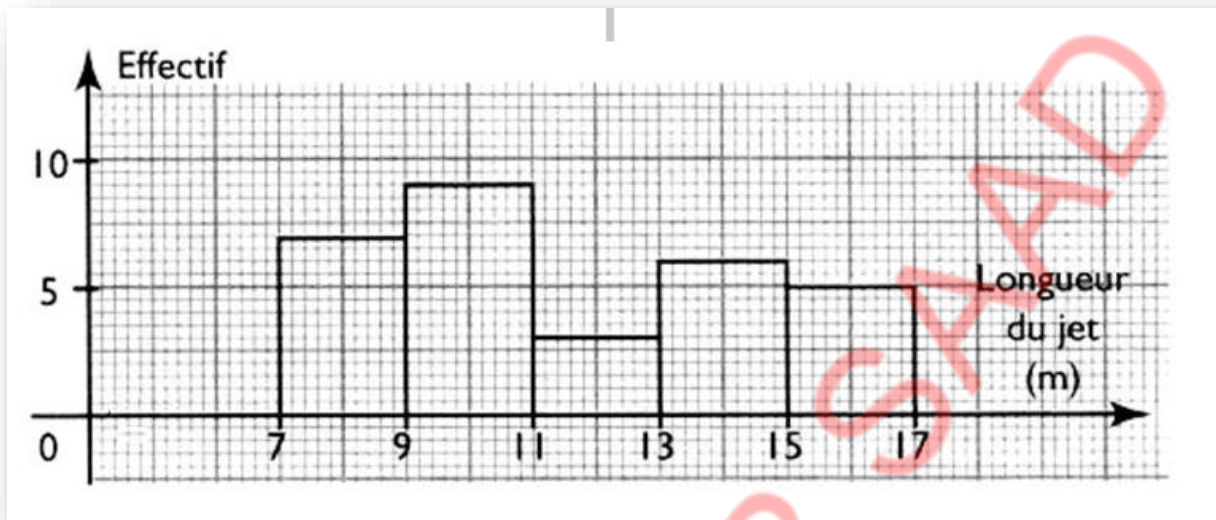
Est un graphique servant à représenter les distributions des fréquences. Il est constitué d'un ensemble de rectangles adjacents, dont chacune des bases coïncide avec un intervalle de classe et chacune des surfaces mesure la fréquence de la classe correspondante.

Exemple 3.3

On a relevé les moyennes des notes des 30 élèves d'une classe d'un établissement scolaire. Les résultats sont regroupés dans le tableau ci-dessous :

Moyenne	Effectif
[7 ; 9 [7
[9 ; 11 [9
[11 ; 13 [3
[13 ; 15 [6
[15 ; 17 [5

Représentation graphique :

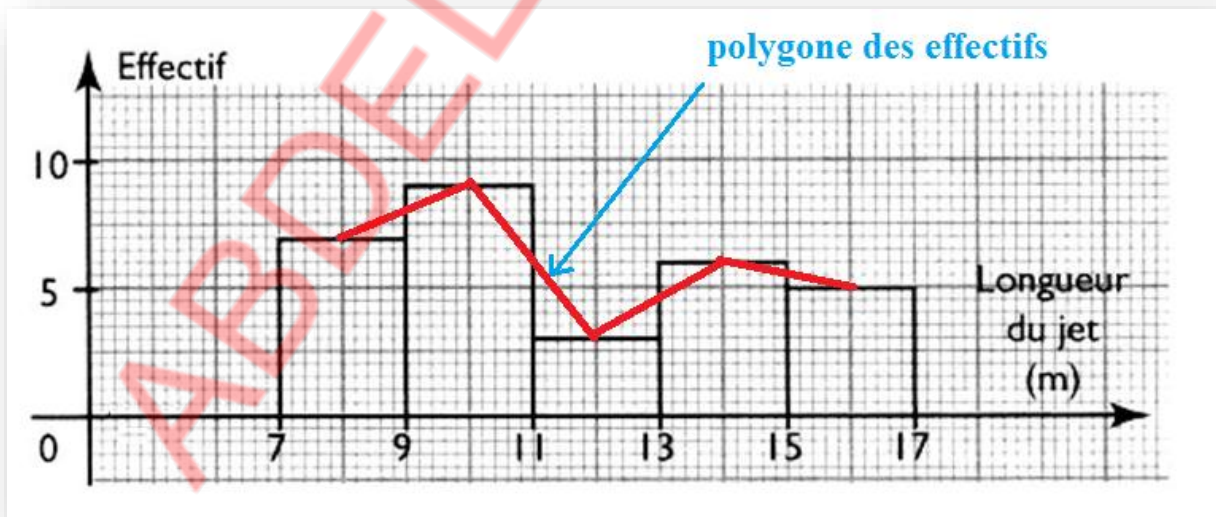


Histogramme des effectifs

3-2-2. Le polygone des effectifs ou bien des fréquences :

Le polygone est obtenu à partir de l'histogramme en rejoignant le point milieu du sommet de chaque rectangle au milieu du sommet du rectangle adjacent.

Donc d'après l'exemple 3.3. Nous dessinons le polygone sur l'histogramme pour l'exemple précédent.



Le polygone des effectifs