République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
ECOLE SUPÉRIEURE EN INFORMATIQUE
8 Mai 1945 - Sidi-Bel-Abbès

الجمهوريـــــة الجزائريـــــة الديمقراطيـــــة الشعبيـــــة
وزارة التعليـــــم العالـــــي والبحـــــث العلمـــــي
المدرســـة العليـــا للإعـــلام الآلـــي
8 مـاي 1945 - سيـدي بلعبـاس

Statistique Descriptive

Présenté par Pr. Nabil KESKES.

Année 2021-2022

.

PLAN

- Introduction
- Definition
- Concepts de Base
- Variable discrète
- Variable continue
- Conclusion

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Ecole Supérieur d'Informatique
De Sidi Bel Abbés

مدرسة العليا للإعلام الالي
بسيدي بلعباس

1. Introduction

L'analyse des données est utilisée pour décrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

La statistique consiste à :

- ✓Recueillir des données.
- ✓Présenter et résumer ces données.
- ✓Tirer des conclusions sur la population étudiée et d'aider à la prise de décision.
- ✓En présence de données dépendant du temps, nous essayons de faire de la prévision.

2. Concepts

2.1 population

On appelle **population** l'ensemble sur lequel porte notre étude statistique. Cet ensemble est noté Ω .

On considère l'ensemble des étudiants de la section A. On s'intéresse aux nombre de frères et sœurs de chaque étudiant. Dans ce cas

Ω = ensemble des étudiants.

Si l'on s'intéresse maintenant à la circulation automobile dans une ville, la population est alors constituée de l'ensemble des véhicules susceptibles de circuler dans cette ville à une date donnée. Dans ce cas

Ω = ensemble des véhicules.

2.2 Individu (unité statistique)

On appelle individu tout élément de la population Ω , il est noté ω

Dans l'exemple précédent , un individu **est tout étudiant de la section.**

2.3 Caractère (variable statistique)

On appelle caractère (ou variable statistique, dénotée V.S) toute application

$$X : \Omega \to C.$$

L'ensemble C est dit : ensemble des valeurs du caractère X (c'est ce qui est mesuré ou observé sur les individus)

Exemple

Taille, température, nationalité, couleur des yeux, catégorie socioprofessionnelle,...

2.4 Modalités

Les modalités d’une variable statistique sont les différentes valeurs que peut prendre celle-ci.

Exemple

Variable est " situation familiale "
Modalités sont " célibataire, marié, divorcé "

Variable est" statut d’interrupteur "
Modalités sont " 0 et 1 ".

Variable est " catégories socioprofessionnelles "
Modalités sont " Employés, ouvriers, retraités,... "

3. variable statistique discrète

nous considérons la situation suivante :

$$X : \Omega \to \{x_1, x_2, ..., x_n\},$$

avec $\mathrm{Card}(\Omega) := N$ est le nombre d'individus dans notre étude.

Le caractère statistique (ou variable statistique, dénotée V.S) peut prendre un nombre fini raisonnable de valeurs (note, nombre d'enfants, nombre de pièces, ...). Dans ce cas, le caractère statistique étudié est alors appelé un caractère discret.

Exemple

Une enquête réalisée dans un village porte sur le nombre d'enfants à charge par famille. On note X le nombre d'enfants, les résultats sont données par ce tableau :

x_i	0	1	2	3	4	5	6
n_i (Effectif)	18	32	66	41	32	9	2

Nous avons

- Ω *ensemble des familles.*
- ω *une famille.*
- X *nombre d'enfants par famille*

3.1 Effectif partiel (fréquence absolue)

Pour chaque valeur x_i, on pose par définition

$$n_i = Card\{\omega \in \Omega : \quad X(\omega) = x_i\}.$$

n_i : le nombre d'individus qui ont le même x_i, ça s'appelle effectif partiel de x_i.

Exemple

on a 66 est le nombre de familles qui ont 2 enfants.

x_i	...	2	...
n_i (Effectif)	...	66	...

3.2 Effectif cumulé

Pour chaque valeur xi, on pose par définition

$$N_i = n_1 + n_2 + ... + n_i.$$

L'effectif cumulé N_i d'une valeur est la somme de l'effectif de cette valeur et de tous les effectifs des valeurs qui précèdent.

on a 50 est le nombre de familles qui ont un nombre d'enfants inférieur à 1. Nous le regardons dans le tableau suivant :

x_i	0	1	2	3	4	5	6
N_i	18	50	116	157	189	198	200

Interprétation : N_i est le nombre d'individus dont la valeur du caractère est inférieur ou égale à x_i. De ce fait, l'effectif total est donné par

$$N = \text{card}(\Omega) = \sum_{i=1}^{n} n_i.$$

Dans notre exemple précédent, nous avons $N = 200$.

3.3 fréquence partielle

Pour chaque valeur *xi,* on pose par définition

$$f_i := \frac{n_i}{N}.$$

fi s'appelle la fréquence partielle de **xi**. La fréquence d'une valeur est le rapport de l'effectif de cette valeur par l'effectif total.

Exemple

x_i	$\cdots$	2	$\cdots$
n_i (*Effectif*)	$\cdots$	66	$\cdots$
f_i (*fréquence*)	$\cdots$	$\frac{66}{200} = 0.33$	$\cdots$

3.4 Fréquence cumulée

Pour chaque valeur xi, on pose par définition

$$Fi = f1 + f2 + ... + fi.$$

La quantité F_i s'appelle la fréquence cumulée de x_i.

Dans l'exemple précédent, 0.785 représente 78.5% de familles dont le nombre d'enfants est inférieur ou égale à 3.

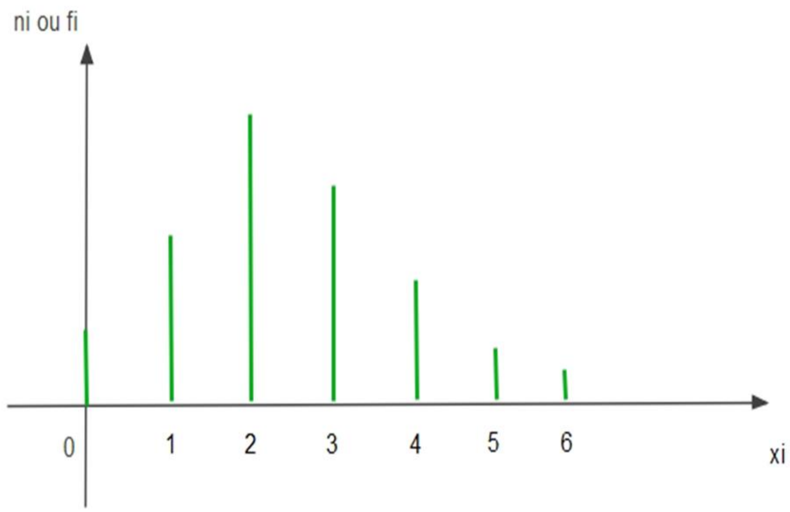
3.5 Représentation graphique des séries statistiques

3.5.1 Distribution à caractère quantitatif discret

A partir de l'observation d'une variable quantitative discrète, deux diagrammes permettent de représenter cette variable : **le diagramme en bâtons** et **le diagramme cumulatif**

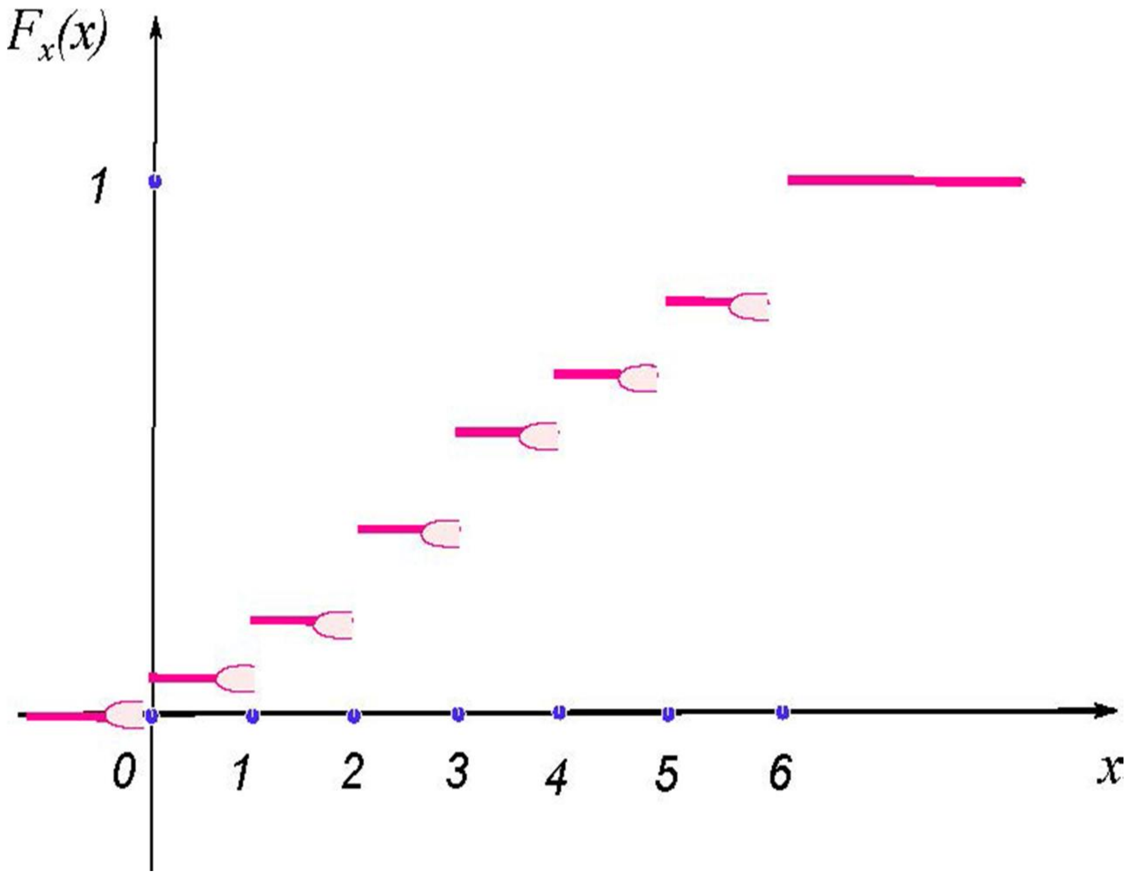
Exemple

Pour l'illustration, nous prenons l'exemple précédent de départ (nombre d'enfants par famille). Nous rappelons le tableau statistique associe.



x_i	0	1	2	3	4	5	6
n_i	18	32	66	41	32	9	2

Exemple



– Si $x < 0$, alors $F_x(x) = 0$.

– Si $x \in [0, 1[$, alors $F_x(x) = 0.09$.

...

– Si $x \geq 6$, alors $F_x(x) = 1$.

Distribution à caractère qualitatif

A partir de l'observation d'une variable qualitative, deux diagrammes permettent de représenter cette variable : le diagramme en bandes (dit tuyaux d'orgue) et le diagramme à secteurs angulaires (dit camembert).

Tuyaux d'orgues

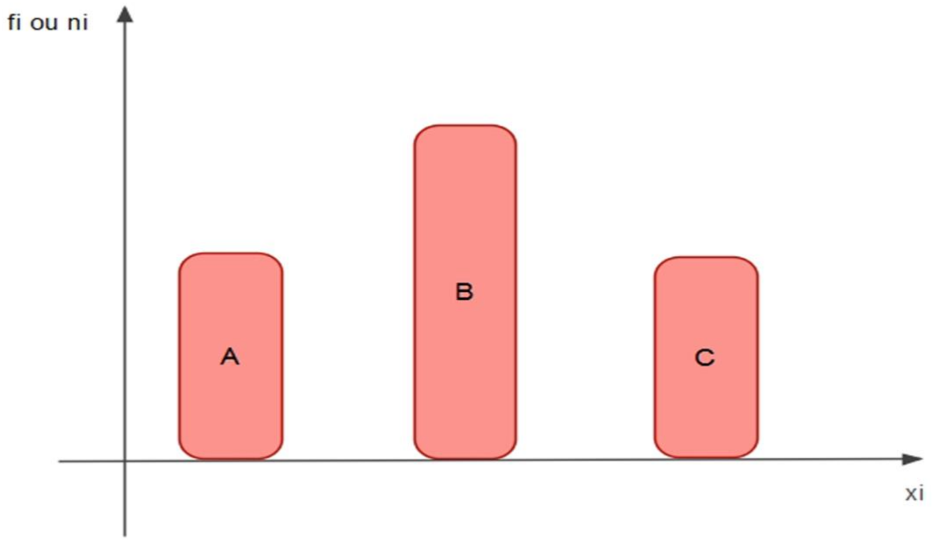
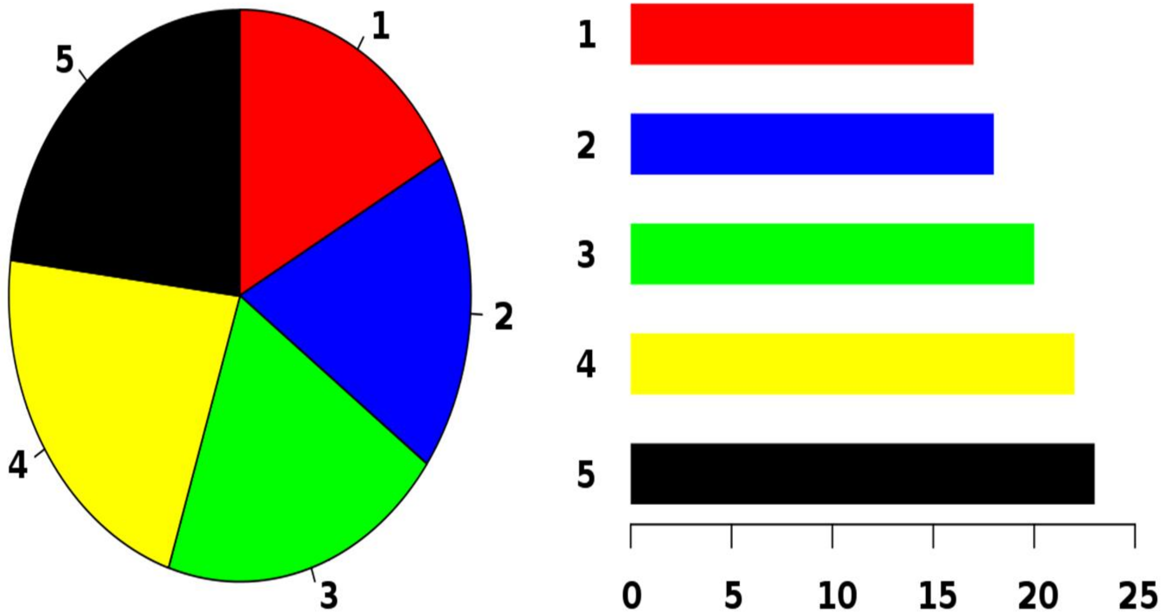


Diagramme par secteur (diagramme circulaire)

Les diagrammes circulaires, ou semi-circulaires, consistent à partager un disque ou un demi-disque, en tranches, ou secteurs, correspondant aux modalités observées et dont la surface est proportionnelle à l'effectif, ou à la fréquence, de la modalité



Le degré d'un secteur est déterminé à l'aide de la règle de trois de la manière suivante :
N – 360
ni – di (degré de la modalité i).

di = ni × 360 / N

.

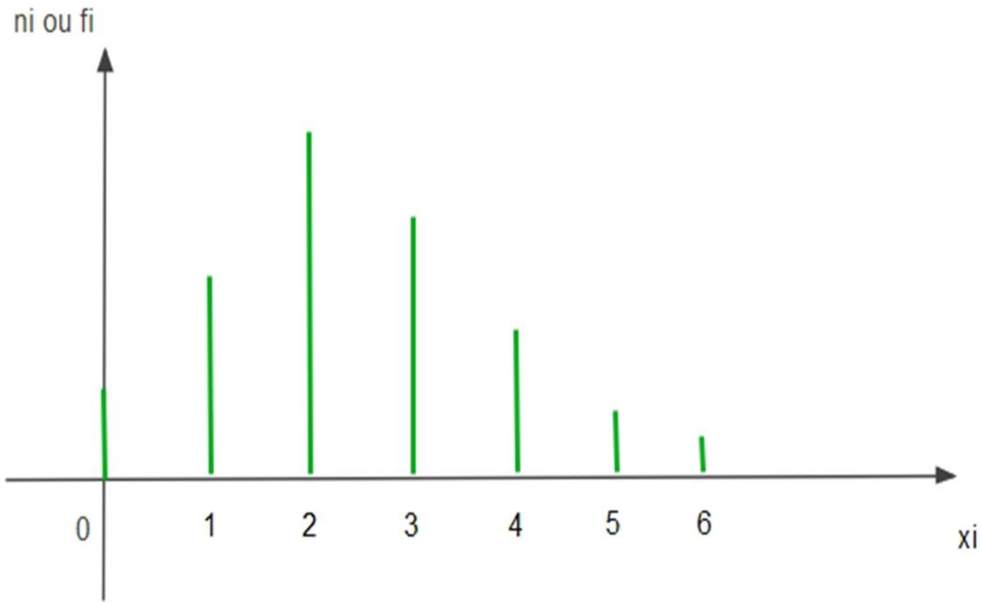
Paramètres de position (caractéristique de tendance centrale)

Le mode

Le mode d'une V.S est la valeur qui a le plus grand effectif partiel (ou la plus grande fréquence partielle) et il est dénoté par *Mo.*

Exemple

le mode est égal à 2 qui correspondant au plus grand effectif.



La médiane

On appelle médiane la valeur Me de la V.S X qui vérifie la relation suivante :
$Fx(Me-) < 0.5 < Fx(Me+) = Fx(Me)$.

La médiane partage la série statistique en deux groupes de même effectif.

Exemple

$$F_x(0) = 0 < 0.5 \leq F_x(0^+) = 0.09$$

n'est pas satisfaite. Donc, la médiane est différente de 0. Par contre, nous avons

$$F_x(2^-) = 0.25 < 0.5 \leq F_x(2^+) = F(2) = 0.58.$$

Donc, $Me = 2$.

La moyenne

On appelle moyenne de *X, la quantité*

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{n} n_i x_i = \sum_{i=1}^{n} f_i x_i,$$

avec *N = Card(). On peut donc exprimer et calculer la moyenne dite "arithmétique" avec* des effectifs ou avec des fréquences.

Exemple

Si $\overline{x} = 2.46$, alors nous avons au moyenne une famille de quartier a 2.46 d'enfants.

La valeur de la moyenne est abstraite. Comme dans l'exemple précédent, *x = 2.46 est* un chiffre qui ne correspond pas à un fait concret.

Paramètres de dispersion (variabilité)

L’étendue

La différence entre la plus grande valeur et la plus petite valeur du caractère, donnée par la quantité

$$e = xmax - xmin,$$

Il donne une première idée de la dispersion des observations.

Paramètres de dispersion (variabilité)

La variance

On appelle variance de cette série statistique *X, le nombre*

$$Var(X) = \sum_{i=1}^{n} f_i(\overline{x} - x_i)^2$$

L'écart type

La quantité

$$\sigma_X = \sqrt{Var(x)}$$

Paramètres de dispersion (variabilité)

Il sert à mesurer la dispersion d'une série statistique autour de sa moyenne.

- *Plus il est petit, plus les caractères sont concentrés autour de la moyenne (on dit que la série est homogène).*
- *Plus il est grand, plus les caractères sont dispersés autour de la moyenne (on dit que la série est hétérogène).*

Exercices

- Le gérant d'un magasin vendant des articles de consommation courante a relevé pour un article particulier qui semble connaître une très forte popularité, le nombre d'articles vendus par jour. Son relevé a porté sur les ventes des mois de Mars et Avril, ce qui correspond à 52 jours de vente. Le relevé des observations se présente comme suit :

7 13 8 10 9 12 10 8 9 10 6 14 7 15 9 11 12 11 12 5 14 11 8 10 14 12 8

5 7 13 12 16 11 9 11 11 12 12 15 14 5 14 9 9 14 13 11 10 11 12 9 15.

1. Quel type est la variable statistique étudiée.
2. Déterminer le tableau statistique en fonction des effectifs, des fréquences, des effectifs cumulés et des fréquences cumulés.
3. Tracer le diagramme des bâtonnés associé à la variable X.
4. Soit F_x la fonction de répartition. Déterminer F_x.
5. Calculer le mode Mo et la moyenne arithmétique $\overline{x}$.

6. *Déterminer à partir du tableau puis à partir du graphe, la valeur de la médiane Me.*
7. *Calculer la variance et l'écart-type.*

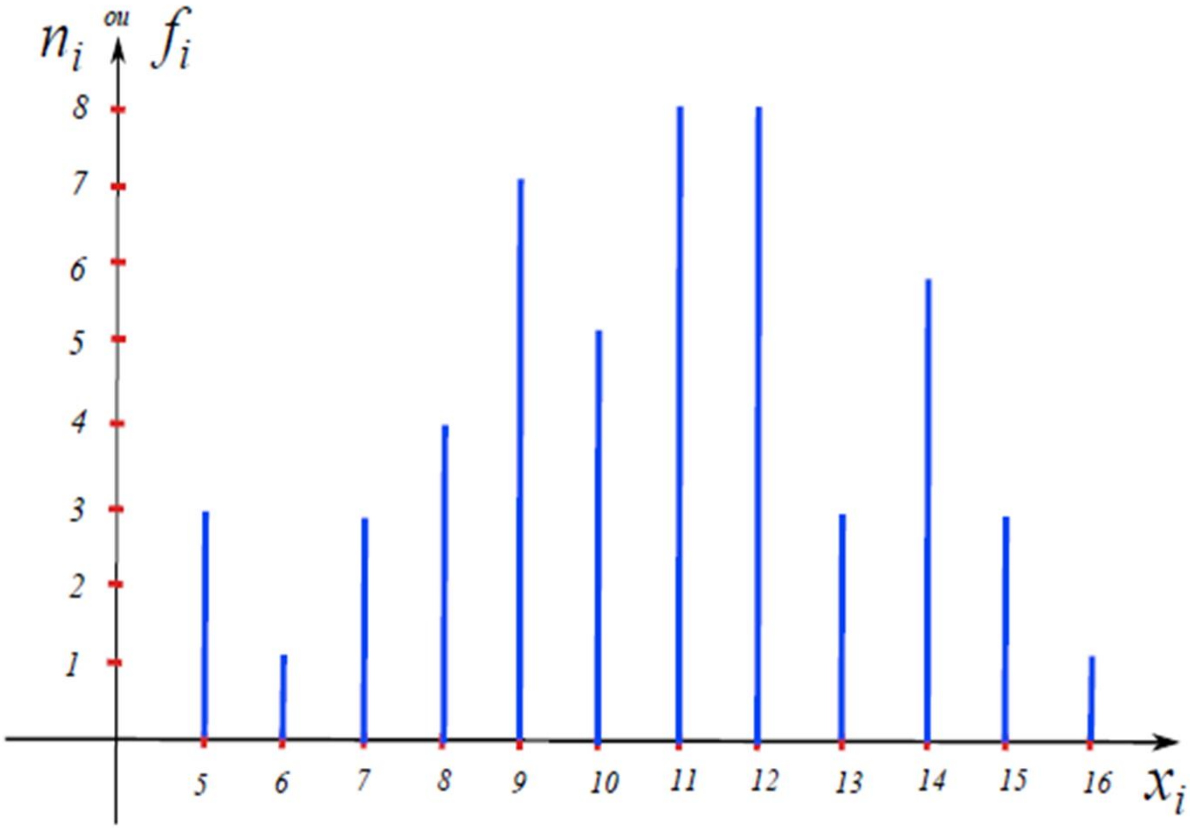
Solution 1 -

La population est les 52 jours et la variable statistique étudiée est le nombre d'articles vendus par jour. Son type est bien évidement quantitatif discret (nombre).

2 - Le tableau statistique est donné par

x_i	5	6	7	8	9	10	11	12	13	14	15	16	$\sum$
n_i	3	1	3	4	7	5	8	8	3	6	3	1	$N = 52$
f_i	3/52	1/52	3/52	4/52	7/52	5 /52	8 /52	8/52	3 /52	6 /52	3/52	1/52	1
N_i	3	4	7	11	18	23	31	39	42	48	51	52	$\varnothing$
F_i	3/52	4/52	7/52	11/52	18/52	23/52	31/52	39/52	42/52	48/52	51/52	1	$\varnothing$

3 - L'élaboration du diagramme des bâtonnets de X,



4 - La fonction de répartition est donnée par

$$F_x(x) = \begin{cases} 0, & si \quad x < 5, \\ 3/52, & si \quad 5 \leq x < 6, \\ 4/52, & si \quad 6 \leq x < 7, \\ 7/52, & si \quad 7 \leq x < 8, \\ \ddots & \quad \ddots \\ 1, & si \quad x \geq 16. \end{cases}$$

5 - Le mode est la valeur de la variable qui a le plus grand effectif, c'est à dire, $n_i = 8$.
Donc,

$$M_o = 11 \quad \text{et} \quad M_o = 12.$$

La moyenne arithmétique est donnée par ;

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{12} n_i x_i = \sum_{i=1}^{12} f_i x_i.$$

Par conséquent,

$$\overline{x} = \frac{1}{52}(3 \times 5 + 1 \times 6 + 3 \times 7 + ... + 1 \times 16) = \frac{555}{52} = 10.67.$$

6 - La médiane est la valeur de la variable qui divise la population de la série statistique en deux parties égales. Nous avons,

$$F_x(11^-) = \frac{23}{52} < 0.5 \leq F_x(11^+) = F(Me) = \frac{31}{52}.$$

Donc, $Me = 11$.

7 - Nous commençons par la variance,

$$Var(X) = \frac{1}{N}\sum_{i=1}^{n} n_i x_i^2 - \overline{x}^2.$$

Après calcule, on trouve

$$Var(X) = 7.64.$$

Par conséquent, l'écart type est calculé à partir de

$$\sigma_X = \sqrt{Var(x)} = 2.76.$$

6. Conclusion

La statistique descriptive a pour but d'étudier un phénomène à partir de données. Cette description se fait à travers la présentation des données (la plus synthétique possible), leur représentation graphique et le calcul de résumés numériques.