

Statistiques



Dr ZAIZ Khaoula

Université d' El Oued

Faculté de sciences exactes.

Département de mathématiques

Email : khaoula.za@gmail.com

Table des matières



Objectifs	4
Introduction	5
I - Pré-requis	6
II - Exercice : Pré-test	7
III - Exercice : Pré-test	8
IV - Chapitre I : Généralités sur la statistique	9
1. Terminologie	9
2. Exercice : Connaissance	10
3. Différents types des variables statistiques	11
4. Exercice : Compréhension	12
5. Exercice : Analyse	12
6. Série de TD	12
V - Chapitre 2 : Série statistiques à une variable	13
1. Étude d'une variable statistique discrète	13
2. Effectif, fréquence	14
2.1. Effectif	14
2.2. Fréquence	14
3. Exercice : Connaissance	15
4. Exercice : Application	15
5. Représentation graphique des séries statistiques	16
5.1. Généralités	16
6. Distribution à caractère qualitatif	17
6.1. Tuyaux d'orgues	17
6.2. Diagramme par secteur (diagramme circulaire)	17
6.3. Distribution à caractère quantitatif discret	18
7. Exercice : Analyse	18
8. Paramètres de position (caractéristique de tendance centrale)	20

8.1. <i>Le mode</i>	20
8.2. <i>La médiane</i>	20
8.3. <i>La moyenne</i>	21
9. Paramètres de dispersion (variabilité)	21
9.1. <i>L'étendue</i>	21
9.2. <i>La variance</i>	21
9.3. <i>L'écart type</i>	22
10. Série de TD	22
Bibliographie	23
Webographie	24

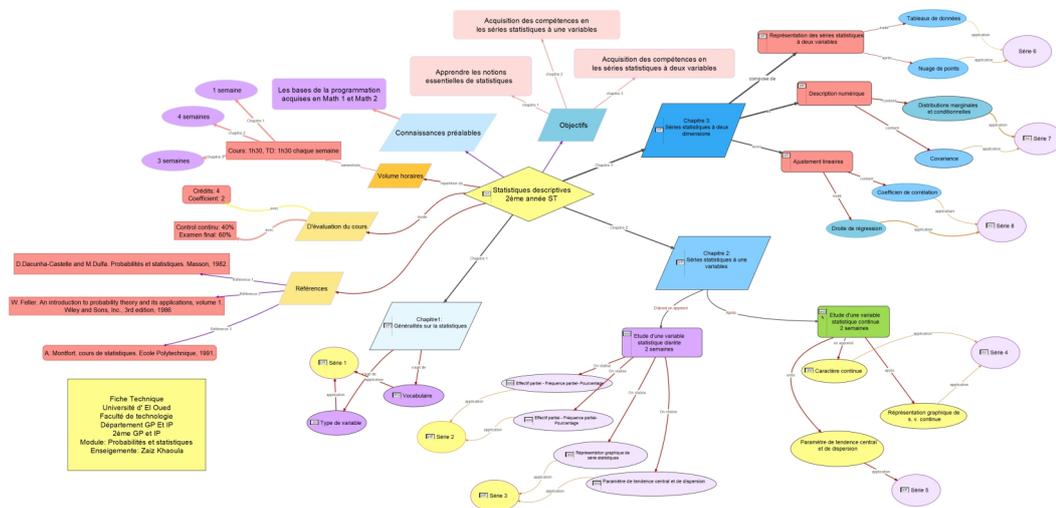
Objectifs

Le cours vise principalement à introduire et faire méditer les concepts fondamentaux et méthodes élémentaires de la statistique pour permettre un apprentissage autonome ultérieur de méthodes complémentaires.

- **Connaissance**
Connaître les définitions de base de statistique, aussi les différents types de variables statistiques et ses caractères.
- **Compréhension**
Préciser les types de variables statistiques, si elle est qualitative, quantitative, discrète ou continue.
- **Application**
Calculer les effectifs, fréquences, les paramètres de position et les paramètres de dispersion.
Représenter graphiquement de variables statistiques.
- **Analyse**
Classer les variables statistiques d'après ses caractères et ses graphes.
- **Jugement évaluation**
Analyser les propriétés d'une série statiques d'après ces caractéristique de tendance centrale et variabilité.
Évaluation finale de contenu étudié.

Introduction

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.



Pré-requis



Les bases de la programmation acquises en Math 1 et Math 2.

Exercice : Pré-test

II

La fonction $f(x) = \ln|x-2|$ est continue dans \mathbb{R}

- Oui
- Non

Exercice : Pré-test



La limite de $f(x) = \frac{3x^2 - 1}{x^2 + 1}$ quand x tend vers 1 est

Chapitre I : Généralités sur la statistique

IV

La statistique est l'étude de la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.

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

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie en passant par la psychologie et bien sûr les sciences de l'ingénieur.

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.

1. Terminologie

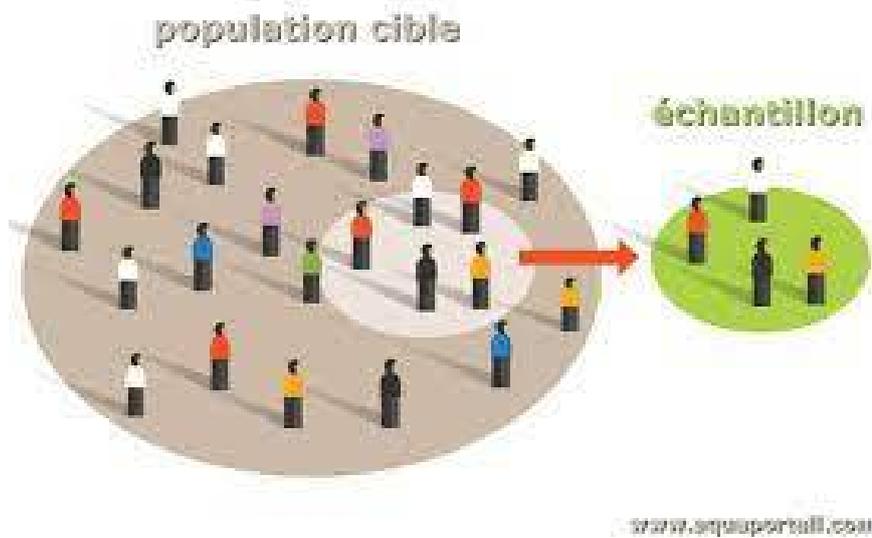
◆ *Rappel*

Les **statistiques descriptives** visent à étudier les caractéristiques d'un ensemble d'observations comme les mesures obtenues lors d'une expérience. L'expérience est l'étape préliminaire à toute étude **statistique**. Il s'agit de prendre "contact" avec les observations. De manière générale, la **méthode statistique** est basée sur le concept suivant.

🔑 *Définition*

-
- **Épreuve statistique**: on appelle **épreuve statistique** l'expérience que l'on provoque.
 - **Population**: On appelle **population** un ensemble d'éléments caractérisés par un critère permettant de les identifier sans ambiguïté. Cet ensemble est noté Ω .
 - **Échantillon**: On appelle **échantillon** un sous ensemble de la **population** considérée Ω .

- **Individu**: Les éléments de la population sont appelés **individu** (ou unité statistique), il est noté ω (ω dans Ω).



Exemple

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

Ω = ensemble des étudiants.

- Un individu est tout étudiant de la section A.

Définition : Caractère

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

$$X : \Omega \rightarrow C$$

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

Remarque

Soit Ω un ensemble. On appelle et on note $\text{Card}(\Omega)$, le nombre d'éléments de Ω .

$\text{Card}(\Omega) :=$ nombre d'éléments de $\Omega = N$.

2. Exercice : Connaissance

On appelle variable, une caractéristique que l'on étudie.

- Vrai
- Faux

3. Différents types des variables statistiques

Rappel

Il existe deux grandes catégories de caractères: les caractères **qualitatifs** et les caractères **quantitatifs**.

Définition : caractère qualitatif

- Les **caractères qualitatifs** sont ceux dont les modalités ne peuvent pas être ordonnées c'est-à-dire que si l'on considère deux caractères pris au hasard, on ne peut pas dire de l'un des caractères qu'il est inférieur ou égal à l'autre.
- La catégorie socioprofessionnelle des individus d'une population donnée (artisan, ouvrier, etc.) est un **caractère qualitatif**, la situation matrimoniale (célibataire, veuf, etc.) aussi. On appelle également **caractères nominaux**, les caractères qualitatifs.

Exemple

Les modalités de la variable sexe sont masculin (codé M) et féminin (codé F). Le domaine de la variable est {M, F}.

Définition : caractère quantitatif

- Les **caractères quantitatifs** sont des caractères dont les modalités peuvent être ordonnées. Ainsi, l'âge, l'espérance de vie ou le salaire d'un individu sont des caractères quantitatifs.
- Au sein des **caractères quantitatifs**, on peut distinguer les caractères mesurables des caractères dont on peut ordonner les modalités sans que celles-ci ne soient **mesurables**.
- On peut faire des opérations algébriques (addition, division, etc.) sur les caractères **mesurables**, alors que l'on ne peut pas les faire sur les caractères **non mesurables**. On appelle également variables **ordinales**, les **variables quantitatives** que l'on peut seulement ordonner.

Exemple

L'évaluation de la qualité d'un produit avec cinq modalités: très bon, bon, passable, mauvais, très mauvais.

Selon la forme des valeurs de la variable, on distingue deux types de caractères quantitatifs, ceux qui sont discrets et ceux qui sont continus.

Définition : Caractère discret

Les **caractères discrets** sont ceux dont le nombre de **modalités** est fini ou dénombrable. Leurs valeurs peuvent être ou non des nombres entiers.

Exemple

Le nombre de pages d'un livre, le nombre de personnes dans une famille.

Définition

On appelle **V.S continue** (ou **caractère continu**) toute application de population et à valeurs réelles et qui prend un nombre "important" de valeurs (Les caractères continus sont ceux qui ont une infinité de modalités).

Remarque

- Parmi ces critères, certains sont **quantitatifs**, comme l'âge, le poids, la taille. On peut en effet effectuer des calculs numériques sur ces critères: poids moyen, taille maximale, taille minimale, etc. D'autres critères ne sont pas quantifiables, car on ne peut pas effectuer de calculs dessus. Ils sont qualitatifs. C'est le cas du sexe par exemple. On peut connaître l'effectif masculin et l'effectif féminin d'une population, mais la notion de « sexe moyen » n'a pas de sens et ne peut d'ailleurs pas être calculée.
 - Afin de différencier les deux type de critères, les critères **qualitatifs** sont appelés des caractères et les critères **quantitatifs** des variables. On désigne par modalités les différentes catégories d'un caractère **qualitatif** et on qualifie de valeurs les différents chiffres d'une variable.

Cf. "Video1"

4. Exercice : Compréhension

La variable statistique "revenu brut" est-elle :

- qualitative
- quantitative

5. Exercice : Analyse

La variable statistique "couleur de maisons d'un quartier" est-elle

- qualitative
- quantitative
- discrète
- continue

6. Série de TD

sserrr.pdf (cf. sserrr)

Chapitre 2 : Série statistiques à une variable

V

Une série statistique est simplement une liste de mesures (qui provient généralement d'une étude). Chaque mesure s'appelle une valeur. Une valeur peut être un nombre, mais ce n'est pas obligatoirement le cas.

1. Étude d'une variable statistique discrète

◆ *Rappel*

Le **caractère statistique*** 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.

Dans toute la suite du chapitre, nous considérons la situation suivante :

$$X: \Omega \rightarrow \{x_1, \dots, x_n\}$$

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

2. Effectif, fréquence

2.1. Effectif

 *Définition : Effectif partiel*

L'**effectif partiel** : est le nombre de fois ou cette valeur apparaît.

L'**effectif total** N : est le nombre total d'individus de la population étudiée.

 *Exemple*

Marie a lancé 20 fois un dé à six faces. Voici les nombres qui sont sortis :

2 - 6 - 5 - 2 - 4 - 1 - 6 - 3 - 1 - 4 - 2 - 1 - 6 - 5 - 6 - 3 - 2 - 1 - 2 - 4.

1. Ce relevé s'appelle une série statistique.
2. On réunit ces résultats dans un tableau d'effectifs en regroupant les valeurs de dés :

nombre x_i	1	2	3	4	5	6	total
effectifs n_i	4	5	2	3	2	4	20

Par exemple, l'**effectif** de la valeur « 4 » est de 3. Ceci veut dire que le nombre 4 est sorti 3 fois.

L'**effectif total** est de 20 lancers.

 *Définition : Effectif cumulé*

Pour chaque valeur x_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.

Autrement dit

$$N_i = n_1 + n_2 + \dots + n_i$$

 *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
N_i (Effectif cumulé)	18	50	116	157	189	198	200

2.2. Fréquence

 *Définition*

On appelle **fréquence**, souvent notée f_i , d'une valeur du caractère le quotient de l'**effectif** de la valeur sur l'**effectif total**.

$$f_i = \frac{\text{effectif de la valeur}}{\text{effectif total}} = \frac{n_i}{N}$$

 **Exemple**

Voici les notes relevées lors d'une interrogation dans une classe.

Note	8	10	12	15	16	20
Effectif n_i	4	8	10	5	2	1

1. L'**effectif total** N est: $4+8+10+5+2+1=30$.
2. La **fréquence** de la note 8 est $\frac{4}{30} = \frac{2}{15}$

Donc, On obtient ainsi le tableau suivant:

Note	8	10	12	15	16	20
Effectif n_i	4	8	10	5	2	1
Fréquence f_i	$\frac{2}{15}$	$\frac{4}{15}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{15}$	$\frac{1}{30}$

 **Remarque**

Soit défini comme précédemment. Alors,

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

 **Définition**

Pour chaque valeur x_i , on pose par définition

$$F_i = f_1 + f_2 + \dots + f_i$$

Le **fréquence cumulé** F_i d'une valeur est la somme de fréquence de cette valeur et de tous les **fréquences** des valeurs qui précèdent.

3. Exercice : Connaissance

l'effectif cumulé de x_n est

Le fréquence cumulé est la division de ; et

4. Exercice : Application

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

Nous rappelons le tableau statistique associe

Calculer le fréquence de nombre 4

5. Représentation graphique des séries statistiques

5.1. Généralités

◆ Rappel

1. Le graphique est un mode d'expression qui permet "visuellement" de saisir et de mémoriser un certain nombre d'informations. C'est pourquoi, lors de la présentation des résultats statistiques et complémentairement aux tableaux, on utilise souvent une représentation graphique.

Cette **représentation graphique*** peut répondre à deux types d'objectifs:

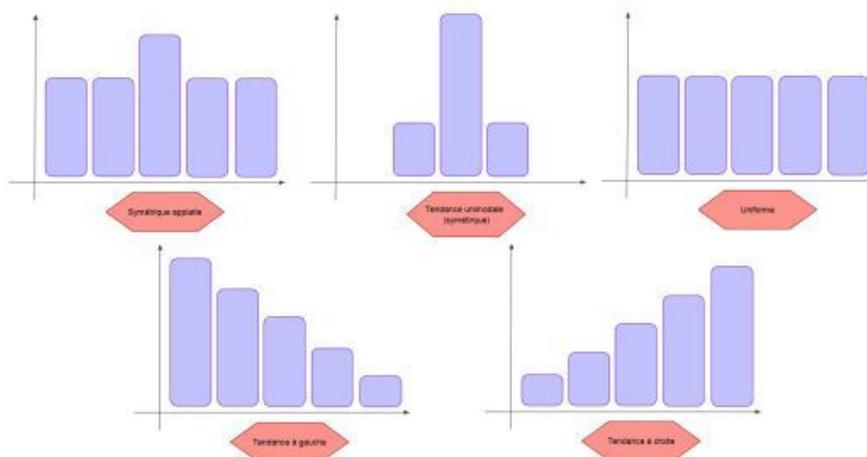
- Être un moyen de communication et permettre de véhiculer une information. C'est ainsi que certains graphiques figurent dans des articles de magazines et journaux, dans des brochures de présentation de résultats commerciaux ou (et) comptables et même dans certaines publicités . Ceci prouve qu'un bon graphique est souvent plus explicite qu'un long discours.
- Être un instrument de travail et permettre une vue d'ensemble synthétique du phénomène étudié, ce qui et facilite l'analys

1. Typologie des graphiques.

En matière statistique, tout graphique est caractérisé par :

- La nature de la série qu'il représente:
 - graphiques chronologiques.
 - graphiques spatiaux, quantitatifs ou qualitatifs
- Le système de coordonnées qu'il utilise :
 - coordonnées cartésiennes.
 - coordonnées polaires.
 - coordonnées pluri linéaires.

La synthèse: visualiser d'un seul coup d'œil les principales caractéristiques (mais on perd une quantité d'informations).



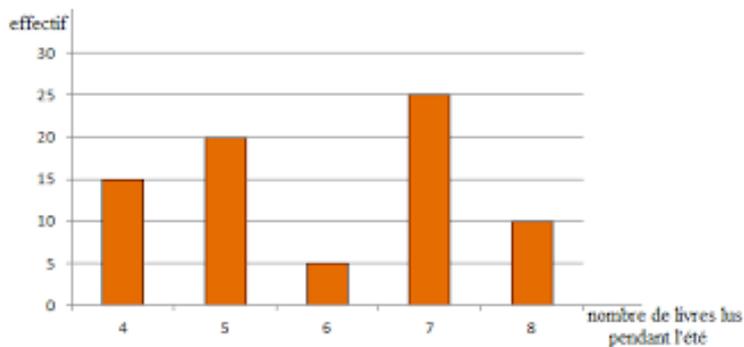
Quelques caractéristiques du graphique

6. 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).

6.1. Tuyaux d'orgues

Définition



Tuyaux d'orgues

Nous portons en abscisses les modalités, de façon arbitraire. Nous portons en ordonnées des rectangles dont la longueur est proportionnelle aux effectifs, ou aux fréquences, de chaque modalité

6.2. Diagramme par secteur (diagramme circulaire)

Définition

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

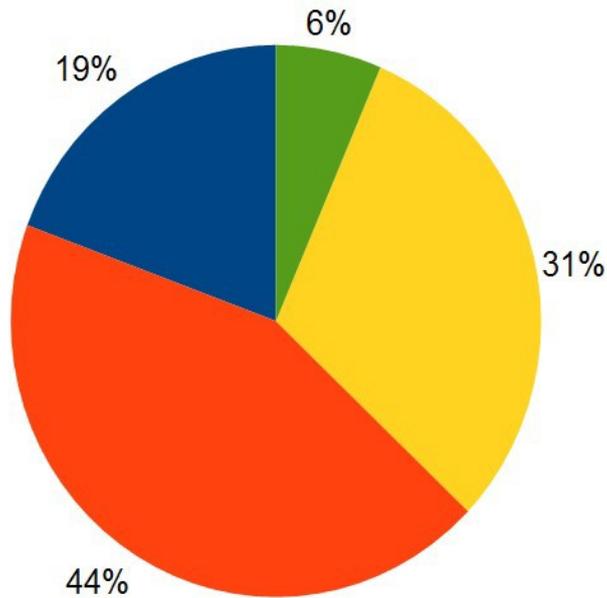


Diagramme par secteur

6.3. 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.

6.3.1. Diagramme en bâtons

Définition

On veut représenter cette répartition sous la forme d'un diagramme en bâtons. À chaque marque correspond un bâton. Les hauteurs des bâtons sont proportionnelles aux effectifs représentés

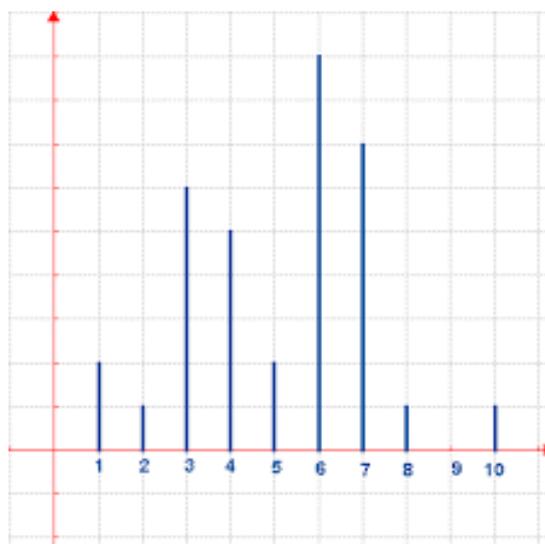
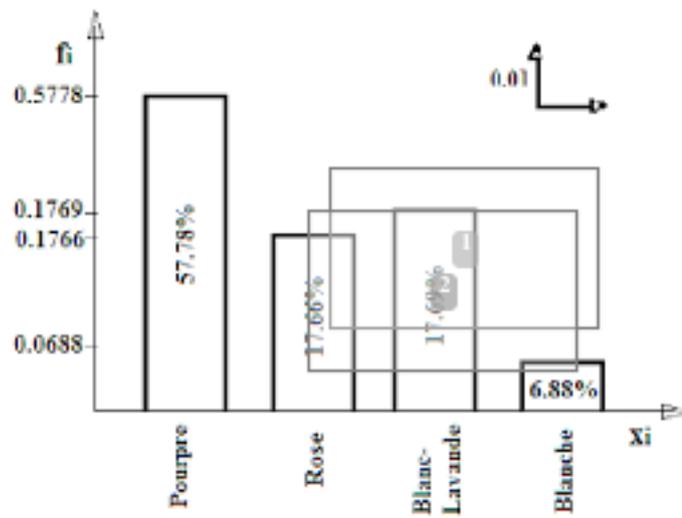


Diagramme à bâtons

7. Exercice : Analyse

Le figure suivant représente une variable statistique qualitative



- 1. vrai
- 2. faux

8. Paramètres de position (caractéristique de tendance centrale)

8.1. Le mode

Définition

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

Exemple

Considérons les couleurs de dix feutres numérotés d'une trousse d'écolier:

noir, jaune, rouge, bleu, vert, bleu, rouge, rouge, noir et rouge.

Le tableau statistique considéré de variable statistique précédent est:

x_i	noir	jaune	bleu	rouge	vert	total
n_i	2	1	2	4	1	10
f_i	0.2	0.1	0.2	0.4	0.1	1

Donc, Le mode est le rouge.

Remarque

On peut avoir plus d'un mode ou rien.

8.2. La médiane

Définition

On appelle médiane la valeur **Me** de la variable statistique qui vérifie la relation suivante :

$$F_x(Me) < 0.5 < F_x'(Me)$$

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

Remarque

La série doit toujours être ordonnée par ordre croissant ou décroissant.

N : Taille de l'échantillon.

Si, $N = 2p$ (paire) alors :

$$Me = \frac{X_p + X_{p+1}}{2}$$

Si, $N = 2p + 1$ (impaire) alors :

$$Me = X_{p+1}$$

Exemple

- $E = \{1; 5; 5; 6; 89; 22\}$

$$N = 6 = 2 * 3 = 2 * p$$

$$\text{alors, } Me = \frac{X_3 + X_4}{2} = \frac{5 + 6}{2} = 5.5$$

donc, $5 < Me < 6$

$$- E = \{1; 5; 6; 7; 8; 10; 24\}$$

$$N = 7 = 2 * 3 + 1 = 2 * p + 1$$

$$\text{alors, } Me = X_{p+1} = 7$$

8.3. La moyenne

Définition

On appelle moyenne de \bar{x} , la quantité

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

ou N est l'effectif total de la série statistique.

9. Paramètres de dispersion (variabilité)

Les indicateurs statistiques de dispersion usuels sont **l'étendue**, **la variance** et **l'écart type**.

9.1. L'étendue

Définition

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

par la quantité

$$e = x_{\max} - x_{\min},$$

s'appelle **l'étendue** de la V.S X.

Le calcul de l'étendue est très simple. Il donne une première idée de la dispersion des observations. C'est un indicateur très rudimentaire et il existe des indicateurs de dispersion plus élaborés (voir ci-dessous)

9.2. La variance

Définition

On appelle variance de cette série statistique X, le nombre

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

Bibliographie



D.Dacunha-Castelle and M.Duflo. Probabilités et statistiques : problèmes à temps fixe. Masson, 1982.



