Statistique Déscriptive

Chapitre 1: Distribution Statistique

Faculté des Sciences. Rabat

Département de Mathématiques

2016-17

Introduction

En présence d'un ensemble de données chiffrées, on a un désir spontané de simplification.
Selon des critères, la statistique cherche d'une part à représenter, ordonner et à classer des données; d'autre part, à résumer la multiplicité et la complexité des notions par des caractéristiques synthétiques.
Le statisticien est ainsi conduit à collecter des données, construire des graphiques, à déterminer des caractéristiques centrales et à calculer des caractéristiques de dispersion.
L'organisation, la description et la présentation des données sous forme de tableaux ou de graphiques font l'objet de la " *statistique descriptive*".
L'interprétation et les conclusions que l'on peut tirer d'un ensemble de données font l'objet de la "*statistique Inférentielle*"

Population

- Toute étude statistique concerne un ensemble Ω appelé **population** dont les éléments sont appelés des **individus**.

Definition

- Une population c'est l'ensemble d'individus ou d'objets qui possèdent un ou plusieurs critères spécifiques en commun.
- Une population statistique est dite finie si l'on peut déterminer avec précision le nombre d'individus qui la composent sinon elle est dite infinie.

Example

i) Dans une étude sur le sport, la population peut être l'ensemble des personnes qui pratique un sport.
ii) Dans une étude sur les revenus mensuels dans une entreprise, la population peut être l'ensemble des personnes qui travaille dans cette entreprise.

Variables Statistique

- L'étude statistique consiste en l'analyse d'une **variable** X appelé parfois caractère qui sert à décrire l'aspect d'une population objet de l'étude. On distigue deux types de variables: **qualitatives et quantitatives**.

Definition

- Une variable X est dite qualitative si les valeurs prises sont des mots ou des lettres.
- Une variable X est dite quantitative si les valeurs prises sont des nombres réels.

Examples

On considère la population marocaine Ω.

i) La couleur des cheveux,état du temps constaté à rabat, mode de transport pour ce rendre à la fac définissent des variables qualitatives.

ii) La taille , le poids, le salaire et l'âge sont des variables quantitatives.

Variables Statistique

- On distingue deux types de variables quantitatives, **discrète** et **continue**

Definition

- Une variable quantitative X est dite discrète si les valeurs qu'elle peut prendre sont isolées les unes des autres.

- Une variable quantitative X est dite continue si elle peut prendre toutes les valeurs d'un intervalle de $\mathbb{R}$ ou une réunion d'intervalles de $\mathbb{R}$ ou l'ensemble des réel $\mathbb{R}$.

Examples

i) Les performances en saut en hateurs de 100 athlètes est une variable quantitative discrète.
ii) La consommation en carburant aux 100 km d'un nouveau modèle d'une voiture est une variable quantitative continue.

Echantillon

- Pour obtenir un renseignement exact concernant une variable X, il faut étudier tous les individus de la population. Quand cela n'est pas possible, on restreint l'étude à une partie de la population appelée **échantillon**.

Definition

Un échantillon est une partie finie représentative de la population c'est donc un sous ensemble E de Ω.

Effectifs - Fréquences - Fréquences cumulées

- L'étude concrète d'une variable X donne N valeurs qui constituent la distribution statistique de X (aussi appelé série statistique).

Effectifs - Fréquences - Fréquences cumulées

- L'étude concrète d'une variable X donne N valeurs qui constituent la distribution statistique de X (aussi appelé série statistique).

- Cette distribution est, en générale, présentée d'une façon groupée:

Effectifs - Fréquences - Fréquences cumulées

- L'étude concrète d'une variable X donne N valeurs qui constituent la distribution statistique de X (aussi appelé série statistique).

- Cette distribution est, en générale, présentée d'une façon groupée:

- Sous la forme $\{(x_i,\ n_i)\ /\ 1 \leq i \leq p\}$ dans le cas d'une variable qualitative ou quantitative discrète (avec $x_1 < x_2 < \cdots < x_p$ dans le cas d'une variable quantitative discrète).

Effectifs - Fréquences - Fréquences cumulées

- L'étude concrète d'une variable X donne N valeurs qui constituent la distribution statistique de X (aussi appelé série statistique).

- Cette distribution est, en générale, présentée d'une façon groupée:

- Sous la forme $\{(x_i,\ n_i)\ /\ 1 \leq i \leq p\}$ dans le cas d'une variable qualitative ou quantitative discrète (avec $x_1 < x_2 < \cdots < x_p$ dans le cas d'une variable quantitative discrète).

- Sous la forme d'intervalles ou classes $\{(]x_i,\ x_{i+1}],\ n_i)\ /\ 1 \leq i \leq p\}$ dans le cas d'une variable quantitative continue .

Effectifs - Fréquences - Fréquences cumulées

- L'étude concrète d'une variable X donne N valeurs qui constituent la distribution statistique de X (aussi appelé série statistique).

- Cette distribution est, en générale, présentée d'une façon groupée:

- Sous la forme $\{(x_i,\ n_i)\ /\ 1 \le i \le p\}$ dans le cas d'une variable qualitative ou quantitative discrète (avec $x_1 < x_2 < \cdots < x_p$ dans le cas d'une variable quantitative discrète).

- Sous la forme d'intervalles ou classes $\{(]x_i,\ x_{i+1}],\ n_i)\ /\ 1 \le i \le p\}$ dans le cas d'une variable quantitative continue .

- n_i est **l'effectif** de la valeur x_i ou de la classe $]x_i,\ x_{i+1}]$. On a évidemment $N = \sum_{i=1}^{p} n_i$.

Effectifs - Fréquences - Fréquences cumulées

Definition

l'effectif n_i est le nombre d'individus de la population ou de l'échantillon pour lesquels X prend la valeur x_i (dans le cas d'une variable qualitative ou quantitative discrète) ou une valeur de l'intervalle $]x_i, x_{i+1}]$ (dans le cas d'une variable quantitative continue).
La somme des effectifs est appelée la taille de la population ou de l'échantillon et est notée N. $N = n_1 + n_2 + \cdots + n_p$

Definition

On appelle **fréquence** de la valeur x_i ou de la classe $]x_i, x_{i+1}]$ le nombre réel

$$f_i = \frac{n_i}{N} \text{ On a évidement } \sum_{i=1}^{p} f_i = 1$$

C'est la proportion de l'effectif d'une valeur de la variable par rapport à N la taille totale de la population ou de l'échantillon.

Effectifs - Fréquences - Fréquences cumulées

Definition

On appelle **fréquence cumulée** de la valeur x_i ou de la classe $]x_i, x_{i+1}]$ le nombre réel

$$F(x) = \sum_{\{i / x_i \leq x\}} f_i$$

C'est la proportion des unités statistiques de la population ou de l'échantillon qui possèdent une valeur inférieure ou égale à une valeur x donnée d'une variable **quantitative.**

Tableau de distribution

Definition

Une distribution statistique est une représentation des données collectées dans un tableau où figurent les valeurs que prenne la variable, les effectifs, les fréquences et les fréquences cumulées relatives à chaque valeur ou ensemble de valeurs prises par la variable.

Example

i)Variable qualitative: La répartition des adultes d'une résidence selon le niveau d'instruction.

Niveau d'instruction	effectifs n_i	fréquences f_i
Primaire	36	0.11
Secondaire	81	0.25
Universitaire	208	0.64
Total	$N = 325$	1

Tableau de distribution

Example

ii) Variable quantitative discrète: Les performances au saut en hauteur (en cm) de 10 athlètes sont: 191, 194, 197, 191, 200, 203, 200, 197, 203, 203.

Hauteur en cm	effectifs n_i	fréquences f_i	fréquences cumulées $F(x)$
191	2	0.2	0.2
194	1	0.1	0.3
197	2	0.2	0.5
200	2	0.2	0.7
203	3	0.3	1
Total	$N = 10$	1	

Tableau de distribution

Example

iii) Variable quantitative continue: Etude de la consommation aux 100 km de 20 voitures d'un nouveau modèle: 5.56, 5.35, 5.98, 5.77, 5.18, 5.66, 5.28, 5.11, 5.58, 5.49, 5.59, 5.33, 5.55, 5.45, 5.76, 5.23, 5.57, 5.52, 5.8, 6.0.

Consommation en litre	effectifs n_i	fréquences f_i	fréquences cumulées $F(x)$
$[5, 5.2]$	2	0.1	0.1
$]5.2, 5.4]$	4	0.2	0.3
$]5.4, 5.6]$	8	0.4	0.7
$]5.6, 5.8]$	4	0.2	0.9
$]5.8, 6]$	2	0.1	1
Total	$N = 20$	1	

Représentations graphiques d'une distribution de variables qualitatives

- **Les tuyaux d'orgues**

Représentations graphiques d'une distribution de variables qualitatives

- **Les tuyaux d'orgues**
- Les tuyaux d'orgues des effectifs (respectivement des fréquences) de la distribution statistique $\{(x_i,\ n_i)\ /\ 1 \leq i \leq p\}$ (respectivement $\{(x_i,\ f_i)\ /\ 1 \leq i \leq p\}$) s'obtient en traçant sur un repére orthonormé, pour tout $i = 1, \cdots, p$, un rectangle de base de centre x_i et de hauteur égale à l'effectif ou la fréquence de la valeur x_i.

Représentations graphiques d'une distribution de variables qualitatives

- **Les tuyaux d'orgues**
- Les tuyaux d'orgues des effectifs (respectivement des fréquences) de la distribution statistique $\{(x_i,\, n_i)\ /\ 1 \leq i \leq p\}$ (respectivement $\{(x_i,\, f_i)\ /\ 1 \leq i \leq p\}$) s'obtient en traçant sur un repére orthonormé, pour tout $i = 1, \cdots, p$, un rectangle de base de centre x_i et de hauteur égale à l'effectif ou la fréquence de la valeur x_i.
- Sur l'axe des abscisses on représente les valeurs ou modalités de la variable, alors que sur l'axe des ordonnées on représente les effectifs ou les fréquences selon que l'on désire tracer un diagramme des effectifs ou des fréquences.

Représentations graphiques d'une distribution de variables qualitatives

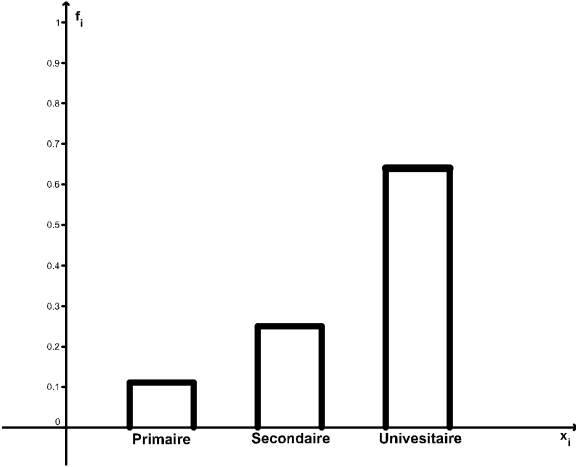


Diagramme en tuyau d'orgue des fréquences

Représentations graphiques d'une distribution de variables qualitatives

- **Représentation circulaire**

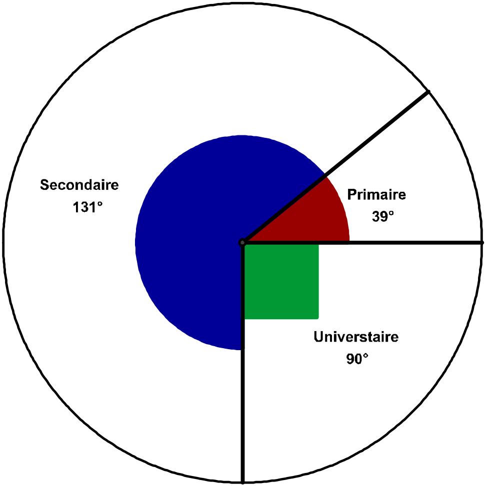
Représentations graphiques d'une distribution de variables qualitatives

- **Représentation circulaire**
- C'est une représentation où chaque modalité ou valeur est représentée par une portion du disque. Si S est l'aire du disque, l'aire d'une portion est égale à $f \times S$, où f est la fréquence de la valeur ou la modalité correspondante.

Représentations graphiques d'une distribution de variables qualitatives

- **Représentation circulaire**
- C'est une représentation où chaque modalité ou valeur est représentée par une portion du disque. Si S est l'aire du disque, l'aire d'une portion est égale à $f \times S$, où f est la fréquence de la valeur ou la modalité correspondante.
- L'angle α de chaque portion s'obtient en multipliant la fréquence par 360° l'angle du disque ($\alpha = f \times 360$)

Représentations graphiques d'une distribution de variables qualitatives



Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Diagramme en bâton**

Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Diagramme en bâton**
- Le diagramme en bâton des effectifs (respectivement des fréquences) de la distribution statistique $\{(x_i,\ n_i)\ /\ 1 \leq i \leq p\}$ (respectivement $\{(x_i,\ f_i)\ /\ 1 \leq i \leq p\}$) s'obtient en traçant sur un repére orthonormé les "bâtons" A_iB_i, c' est à dire les segments joignant les point $A_i(0,\ x_i)$ et $B_i(x_i,\ n_i)$ (respectivement $B_i(x_i,\ f_i)$) pour $1 \leq i \leq p$.

Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Diagramme en bâton**
- Le diagramme en bâton des effectifs (respectivement des fréquences) de la distribution statistique $\{(x_i,\ n_i)\ /\ 1 \leq i \leq p\}$ (respectivement $\{(x_i,\ f_i)\ /\ 1 \leq i \leq p\}$) s'obtient en traçant sur un repére orthonormé les "bâtons" A_iB_i, c' est à dire les segments joignant les point $A_i(0,\ x_i)$ et $B_i(x_i,\ n_i)$ (respectivement $B_i(x_i,\ f_i)$) pour $1 \leq i \leq p$.
- Sur l'axe des abscisses on représente les valeurs ou modalités de la variable, alors que sur l'axe des ordonnées on représente les effectifs ou les fréquences selon que l'on désire tracer un diagramme des effectifs ou des fréquences.

Représentations graphiques d'une distribution de variables quantitatives discrètes

Example

La distribution des performances au saut en hauteur de 100 athlètes sont représentées dans le tableau suivant:

Hauteur en cm	effectifs n_i	fréquences f_i	fréquences cumulées $F(x)$
191	6	0.06	0.06
194	17	0.17	0.23
197	41	0.41	0.64
200	27	0.27	0.91
203	9	0.09	1
Total	100	1	

Représentations graphiques d'une distribution de variables quantitatives discrètes

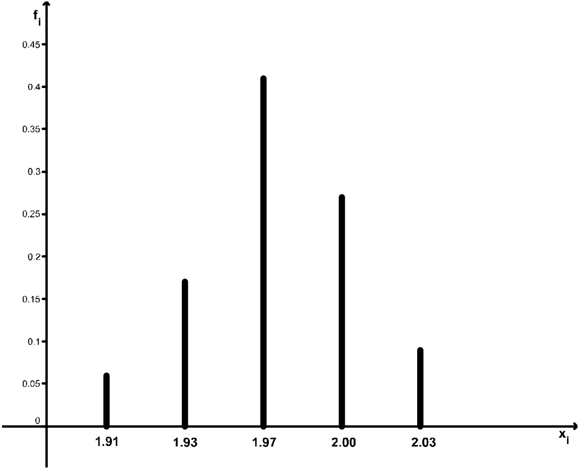


Diagramme en baton

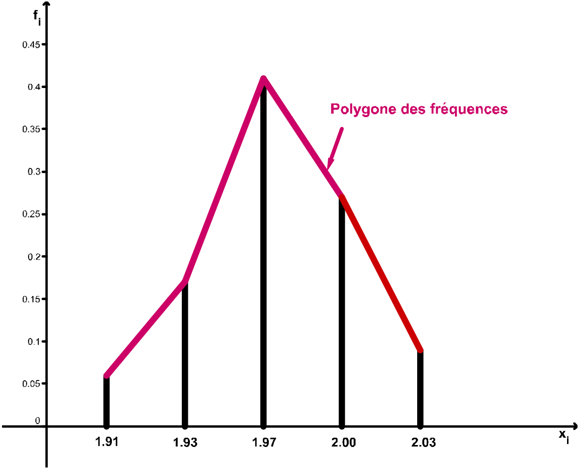
Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Polygône des fréquences**

Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Polygône des fréquences**
- C'est une ligne brisée joignant les points de coordonnées (x_i, f_i). C'est aussi la ligne qui joint les sommets des bâtonnés du diagramme en bâton.

Représentations graphiques d'une distribution de variables quantitatives discrètes



Polygône des fréquences

Représentations graphiques d'une distribution de variables quantitatives discrètes

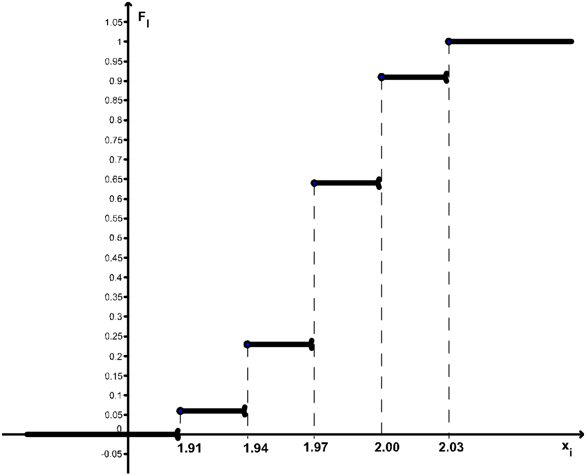
- **Courbe des fréquences cumulées**

Représentations graphiques d'une distribution de variables quantitatives discrètes

- **Courbe des fréquences cumulées**
- C'est une courbe en escaliers qui représente la fonction:

$$F(x) = 0 \text{ si } x < x_1 \text{ et } F(x) = \sum_{j : x_j \leq x} f_j \text{ sinon}$$

Représentations graphiques d'une distribution de variables quantitatives discrètes



Courbe des fréquences cumulées

Représentations graphiques d'une distribution de variables quantitatives continues

- Considérons une variable continue X dont les valeurs se situent dans un intervalle I.

Représentations graphiques d'une distribution de variables quantitatives continues

- Considérons une variable continue X dont les valeurs se situent dans un intervalle I.
- On divise cet intervalle en k classes disjointes $]a_i, a_{i+1}]$, $i = 1, ..., p$. On prendra toujours des classes de même amplitude $(a_{i+1} - a_i = \text{constante})$.

Représentations graphiques d'une distribution de variables quantitatives continues

- Considérons une variable continue X dont les valeurs se situent dans un intervalle I.
- On divise cet intervalle en k classes disjointes $]a_i, a_{i+1}]$, $i = 1, ..., p$. On prendra toujours des classes de même amplitude $(a_{i+1} - a_i = \text{constante})$.

- Considérons une variable continue X dont les valeurs se situent dans un intervalle I. Pour tout i, on note n_i le nombre de valeurs de X dans la classe $]a_i, a_{i+1}]$ qu'on appelle effectif de cette classe.

Représentations graphiques d'une distribution de variables quantitatives continues

- Considérons une variable continue X dont les valeurs se situent dans un intervalle I.
- On divise cet intervalle en k classes disjointes $]a_i, a_{i+1}]$, $i = 1, ..., p$. On prendra toujours des classes de même amplitude $(a_{i+1} - a_i = \text{constante})$.
- Considérons une variable continue X dont les valeurs se situent dans un intervalle I. Pour tout i, on note n_i le nombre de valeurs de X dans la classe $]a_i, a_{i+1}]$ qu'on appelle effectif de cette classe.
- Le choix du nombre de classes est laissé au soin de l'utilisateur. Plus le nombre d'observations est grand plus le nombre de classes est élévé. On admet cependant, pour aider à la compréhention, que ce nombre devrait être entre 5 et 15.

Représentations graphiques d'une distribution de variables quantitatives continues

- Pour dresser le tableau de distribution des effectifs et des fréquences on pourra suivre les étapes suivantes:

Représentations graphiques d'une distribution de variables quantitatives continues

- Pour dresser le tableau de distribution des effectifs et des fréquences on pourra suivre les étapes suivantes:
- **Etape 1:** Déterminer p le nombre de classes à considérer dans l'étude. Pour N l'effectif de la population ou de l'échantillon, on peut calculer le nombre de classes selon l'une des deux formules suivantes:

  i) Règle de Sturge: $P = 1 + 3.3 \times \log_{10}(N)$

  ii) Règle de Yule: $P = 2.5 \times \sqrt[4]{N}$

  Avec $p =$ l'entier le plus proche de P.

Représentations graphiques d'une distribution de variables quantitatives continues

- Pour dresser le tableau de distribution des effectifs et des fréquences on pourra suivre les étapes suivantes:
- **Etape 1:** Déterminer p le nombre de classes à considérer dans l'étude. Pour N l'effectif de la population ou de l'échantillon, on peut calculer le nombre de classes selon l'une des deux formules suivantes:
  i) Règle de Sturge: $P = 1 + 3.3 \times \log_{10}(N)$
  ii) Règle de Yule: $P = 2.5 \times \sqrt[4]{N}$
  Avec $p =$ l'entier le plus proche de P.
- **Etape 2:** Calculer l'étendue $e = x_{max} - x_{min}$. Où x_{min} (resp. x_{max}) est la valeur minimale (resp. la valeur maximale) de la variable X .

Représentations graphiques d'une distribution de variables quantitatives continues

- **Etape 3:** Diviser l'étendue e par p le nombre de classes, pour avoir la valeur de l'amplitude des classes noté a. $a = \dfrac{e}{p}$

Représentations graphiques d'une distribution de variables quantitatives continues

- **Etape 3:** Diviser l'étendue e par p le nombre de classes, pour avoir la valeur de l'amplitude des classes noté a. $a = \dfrac{e}{p}$
- **Etape 4:** On construit alors les classes $[x_{min}, x_{min} + a], \;]x_{min} + a, x_{min} + 2a], \; \cdots \; , \;]x_{min}(p-1)\, a, x_{min} + p\, a]$

Représentations graphiques d'une distribution de variables quantitatives continues

- **Etape 3:** Diviser l'étendue e par p le nombre de classes, pour avoir la valeur de l'amplitude des classes noté a. $a = \dfrac{e}{p}$
- **Etape 4:** On construit alors les classes $[x_{min}, x_{min} + a], \;]x_{min} + a, x_{min} + 2a], \; \cdots \; , \;]x_{min}(p-1)\, a, x_{min} + p\, a]$
- **Etape 5:** S'assurer que chaque observation appartient à une et une seule classe.

Représentations graphiques d'une distribution de variables quantitatives continues

Example

- Etude de la consommation aux 100 km de 20 voitures d'un nouveau modèle: 6.11, 6.05, 5.98, 5.77, 5.18, 5.66, 5.28, 5.11, 5.58, 5.49, 5.62, 5.33, 5.55, 5.45, 5.76, 5.23, 5.57, 5.52, 5.8, 6.0.

- Pour la méthode de Sturge $P = 1 + 3.3 \times \log_{10}(20) = 5.293$.

- Pour la méthode de Yule $P = 2.5 \times \sqrt[4]{20} = 5.287$,
  D'où le nombre de classe est $p = 5$.

- Nous avons $x_{min} = 5.11$ et $x_{max} = 6.11$.
- D'ou $e = 6.11 - 5.11 = 1$ et $a = \dfrac{e}{p} = \dfrac{1}{5} = 0.2$.

Example

Consommation en litre	effectifs n_i	fréquences f_i	fréquences cumulées $F(x)$
$[5.11, 5.31]$	4	0.2	0.2
$]5.31, 5.51]$	3	0.15	0.35
$]5.51, 5.71]$	6	0.3	0.65
$]5.71, 5.91]$	3	0.15	0.8
$]5.91, 6.11]$	4	0.2	1
Total	20	1	

Représentations graphiques d'une distribution de variables quantitatives continues

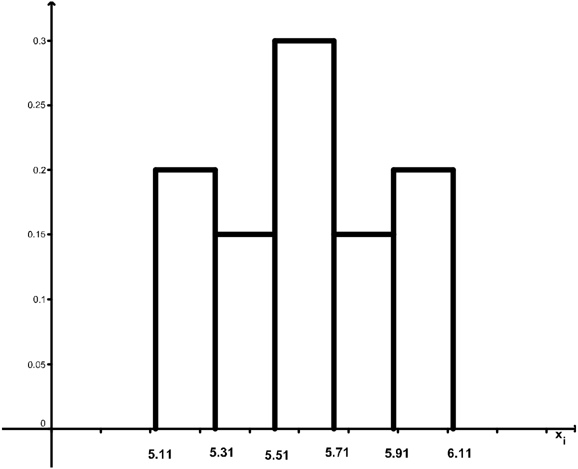
- **Histogramme**

Représentations graphiques d'une distribution de variables quantitatives continues

- **Histogramme**
- L'histogramme des effectifs (respectivement des fréquences) de la distribution statistique $\{(]a_i, a_{i+1}], n_i) \,/\, 1 \leq i \leq p\}$ (respectivement $\{(]a_i, a_{i+1}], f_i) \,/\, 1 \leq i \leq p\}$) s'obtient en traçant sur un repère orthonormé, pour tout $i = 1, \cdots, p$, un rectangle de base la longueur du segment $]a_i, a_{i+1}]$ et de hauteur égale à l'effectif ou la fréquence de cette classe.

- **Histogramme**
- L'histogramme des effectifs (respectivement des fréquences) de la distribution statistique $\{(]a_i, a_{i+1}], n_i) \,/\, 1 \leq i \leq p\}$ (respectivement $\{(]a_i, a_{i+1}], f_i) \,/\, 1 \leq i \leq p\}$) s'obtient en traçant sur un repère orthonormé, pour tout $i = 1, \cdots, p$, un rectangle de base la longueur du segment $]a_i, a_{i+1}]$ et de hauteur égale à l'effectif ou la fréquence de cette classe.
- Sur l'axe des abscisses on représente les bornes des classes $]a_i, a_{i+1}]$ de la variable c'est à dire les points $a_1, a_2, \cdots, a_p, a_{p+1}$, alors que sur l'axe des ordonnées on représente les effectifs ou les fréquences selon que l'on désire tracer un histogramme des effectifs ou des fréquences.

Représentations graphiques d'une distribution de variables quantitatives continues



Histogramme

Représentations graphiques d'une distribution de variables quantitatives continues

- **Polygône des fréquences**

Représentations graphiques d'une distribution de variables quantitatives continues

- **Polygône des fréquences**
- Le polygône des fréquences de la distribution $\{(]a_i,\ a_{i+1}],\ f_i)\ /\ 1 \leq i \leq p\}$ est la ligne brisée joignant les points de coordonnées (c_i, f_i) où $c_i = \dfrac{a_i + a_{i+1}}{2}$, $i = 1, \cdots, p$.

Représentations graphiques d'une distribution de variables quantitatives continues

- **Polygône des fréquences**
- Le polygône des fréquences de la distribution $\{(]a_i, a_{i+1}], f_i) \,/\, 1 \leq i \leq p\}$ est la ligne brisée joignant les points de coordonnées (c_i, f_i) où $c_i = \dfrac{a_i + a_{i+1}}{2}$, $i = 1, \cdots, p$.
- Lorsque la borne inférieure de la première classe est observée c'est à dire l'intervalle est fermé en a_1 (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_0, 0)$ et (c_1, f_1) où $c_0 = a_1 - \dfrac{a}{2}$.

Représentations graphiques d'une distribution de variables quantitatives continues

- **Polygône des fréquences**
- Le polygône des fréquences de la distribution $\{(]a_i,\, a_{i+1}],\, f_i)\,/\,1 \leq i \leq p\}$ est la ligne brisée joignant les points de coordonnées (c_i, f_i) où $c_i = \dfrac{a_i + a_{i+1}}{2}$, $i = 1, \cdots, p$.
- Lorsque la borne inférieure de la première classe est observée c'est à dire l'intervalle est fermé en a_1 (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_0,\, 0)$ et (c_1, f_1) où $c_0 = a_1 - \dfrac{a}{2}$.
- Lorsque la borne inférieure de la première classe n'est pas observée c'est à dire l'intervalle est ouvert en a_1, on complète la courbe en joignant les points $(a_1,\, 0)$ et (c_1, f_1).

Représentations graphiques d'une distribution de variables quantitatives continues

- Lorsque la borne supérieure de la dernière classe est observée c'est à dire l'intervalle est fermé en a_{p+1} (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_p,\ f_p)$ et $(c_{p+1}, 0)$ où $c_{p+1} = a_{p+1} + \dfrac{a}{2}$.

Représentations graphiques d'une distribution de variables quantitatives continues

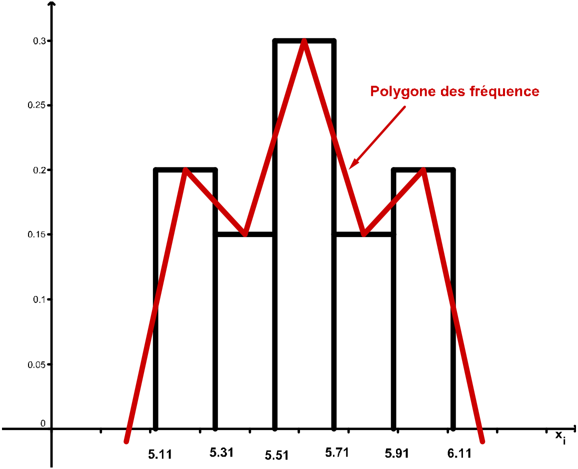
- Lorsque la borne supérieure de la dernière classe est observée c'est à dire l'intervalle est fermé en a_{p+1} (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_p,\ f_p)$ et $(c_{p+1}, 0)$ où $c_{p+1} = a_{p+1} + \dfrac{a}{2}$.

- Lorsque la borne supérieure de la dernière classe n'est pas observée c'est à dire l'intervalle est ouvert en a_{p+1}, on complète la courbe en joignant les points $(c_p,\ f_p)$ et $(a_{p+1}, 0)$.

- Lorsque la borne supérieure de la dernière classe est observée c'est à dire l'intervalle est fermé en a_{p+1} (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_p,\ f_p)$ et $(c_{p+1}, 0)$ où $c_{p+1} = a_{p+1} + \frac{a}{2}$.

- Lorsque la borne supérieure de la dernière classe n'est pas observée c'est à dire l'intervalle est ouvert en a_{p+1}, on complète la courbe en joignant les points $(c_p,\ f_p)$ et $(a_{p+1}, 0)$.
- Représentation du polygône des fréquences de la distribution..

Représentations graphiques d'une distribution de variables quantitatives continues



Polygone des fréquences

Représentations graphiques d'une distribution de variables quantitatives continues

- **Courbe des fréquences cumulées**

Représentations graphiques d'une distribution de variables quantitatives continues

- **Courbe des fréquences cumulées**
- La courbe des fréquences cumulées de la distribution $\{(]a_i,\ a_{i+1}],\ f_i)\ /\ 1 \leq i \leq p\}$ s'obtient en joignant les points de coordonnées (a_{i+1}, F_i) où $F_i = f_1 + \cdots + f_i$, $i = 1, \cdots, p$ et $(x, 1)$ pour $x \geq a_{p+1}$.

Représentations graphiques d'une distribution de variables quantitatives continues

- **Courbe des fréquences cumulées**
- La courbe des fréquences cumulées de la distribution $\{(]a_i,\ a_{i+1}],\ f_i)\ /\ 1 \leq i \leq p\}$ s'obtient en joignant les points de coordonnées (a_{i+1}, F_i) où $F_i = f_1 + \cdots + f_i$, $i = 1, \cdots, p$ et $(x, 1)$ pour $x \geq a_{p+1}$.

- Lorsque la borne inférieure de la première classe est observée c'est à dire l'intervalle est fermé en a_1, $F(a_1) \neq 0$, (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_0,\ 0)$ et (a_2, F_1) où $c_0 = a_1 - \dfrac{a}{2}$ et $F_1 = f_1$.

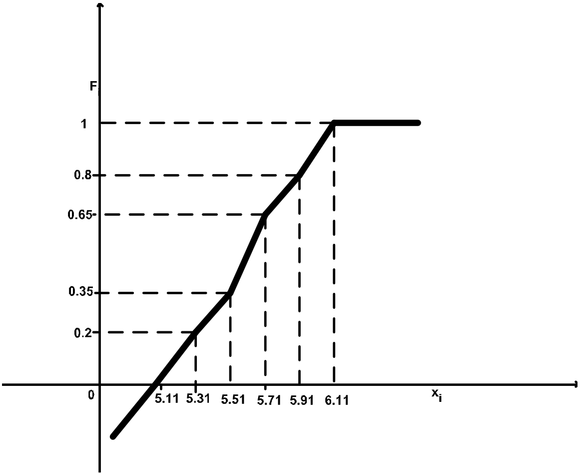
Représentations graphiques d'une distribution de variables quantitatives continues

- **Courbe des fréquences cumulées**
- La courbe des fréquences cumulées de la distribution $\{(]a_i,\, a_{i+1}],\, f_i) \,/\, 1 \le i \le p\}$ s'obtient en joignant les points de coordonnées (a_{i+1}, F_i) où $F_i = f_1 + \cdots + f_i$, $i = 1, \cdots, p$ et $(x, 1)$ pour $x \ge a_{p+1}$.

- Lorsque la borne inférieure de la première classe est observée c'est à dire l'intervalle est fermé en a_1, $F(a_1) \neq 0$, (comme c'est le cas dans l'exemple), on complète la courbe en joignant les points $(c_0,\, 0)$ et (a_2, F_1) où $c_0 = a_1 - \dfrac{a}{2}$ et $F_1 = f_1$.

- Lorsque la borne inférieure de la première classe n'est pas observée c'est à dire l'intervalle est ouvert en a_1, $F(a_1) = 0$, on complète la courbe en joignant les points $(a_1,\, 0)$ et (a_2, F_1).

Représentations graphiques d'une distribution de variables quantitatives continues



Courbe des fréquences cumulées

Statistique Déscriptive

Chapitre 2: Les mesures de tendance centrale et de dispersion

Faculté des Sciences Rabat

Département de Mathématiques

2016-17

Les mesures de tendance centrale

La tendance centrale se propose de synthétiser l'ensemble d'une série statistique en faisant ressortir une position centrale de la valeur du caractère étudié. Il existe plusieurs mesures de tendance centrale.

Le mode , la médiane et la moyenne

Le mode: Variable qualitative ou quantitative discrète

Definition

Le mode est une valeur de la variable pour laquelle l'effectif ou la fréquence est maximal(e). Le mode est noté m_d.
Une distribution peut être unimodale, bimodale ou plurimodale.

Examples

Considérons la distribution des couleurs des voitures dans un parking

x_i	Rouge	Blanche	Verte	Jaune	Noire	Grise
n_i	2	7	5	7	5	7

l'effectif maximal est 7

La variable est qualitative. Ici on a trois modes: Blanche, Jaune et Grise. Cette distribution est plurimodale.

Le mode: Variable qualitative ou quantitative discrète

Examples

Considérons la distribution des notes d'un groupe d'étudiants.

x_i	8/20	9/20	10/20	11/20	12/20	13/20	14/20
n_i	2	7	12	17	11	6	3

l'effectif maximal est 17

La variable est quantitative discrète. On a $m_d = 11/20$. Cette distribution est unimodale.

Le mode:Variable quantitative continue

Dans le cas d'une variable quantitative continue, les données sont regroupées en classes. Si les classes sont toutes de même amplitude, une classe modale est celle dont l'effectif ou la fréquence est le plus élevé(e).

Example

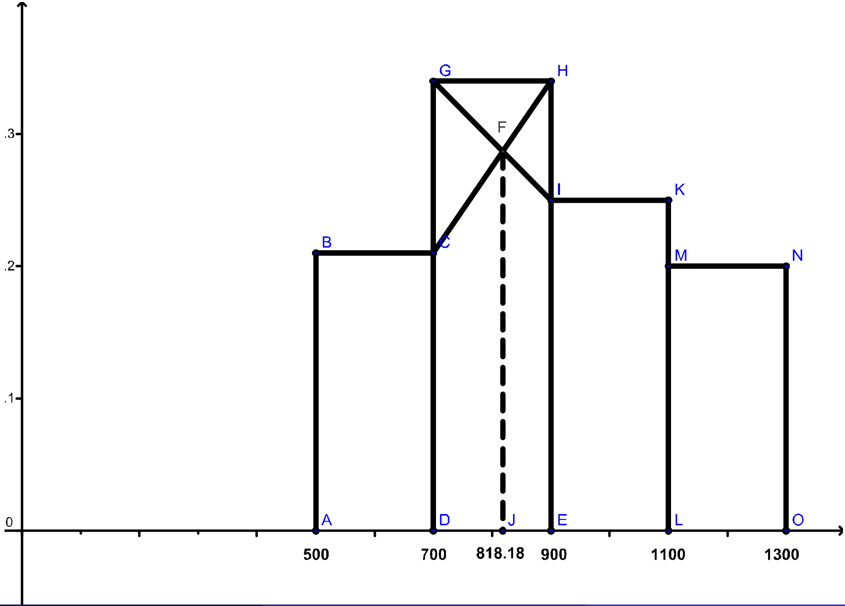
Soit la distribution suivante

$[x_i, x_{i+1}[$	$[500, 700[$	$[700, 900[$	$[900, 1100[$	$[1100, 1300]$
f_i	0.21	0.34	0.25	0.2

la fréquence maximale est 0.34, donc la classe modale est $[700, 900[$.

le mode m_d (qui appartient à la classe modale) est déterminé par interpolation linéaire selon la formule suivante:

Le mode:Variable quantitative continue



Le mode:Variable quantitative continue

le mode m_d (qui appartient à la classe modale) est déterminé par interpolation linéaire selon la formule suivante:

$$m_d = x_{i+1} - a \times \frac{(f_{i+1} - f_{i+2})}{(f_{i+1} - f_{i+2}) + (f_{i+1} - f_i)}$$

Où x_{i+1} est la borne supérieure de la classe modale, a l'amplitude commune à toutes les classes, f_{i+1} la fréquence de la classe modale, f_i la fréquence de la classe qui précède la classe modale et f_{i+2} la fréquence de la classe qui suit la classe modale.

Example

Application: La classe modale est $[700, 900[$, $x_{i+1} = 900$, $a = 200$, $f_{i+1} = 0.34$, $f_i = 0.21$, $f_{i+2} = 0.25$ et

$$m_d = 900 - 200 \times \frac{(0.34 - 0.25)}{(0.34 - 0.25) + (0.34 - 0.21)} = 818.18$$

Le mode:Variable quantitative continue

Remarque: Si les classes ne sont pas de même amplitude, on doit obligatoirement corriger les effectifs et les fréquences (c'est à dire rendre les classes de même amplitude) avant de:

$$\begin{cases} \textbf{Construire l'histogramme} \\ \textbf{Construire le polygône des fréquences} \\ \textbf{déterminer la classes modale} \end{cases}$$

considérons l'exemple suivant:

Le mode:Variable quantitative continue

Les salaires mensuels (en milliers de dirhams) du personnel d'une entreprise se répartissent comme suit:

Classe	Effectif n_i	fréquence f_i	fréquence cumulée $F(x_{i+1})$
$]2,3]$	15	0, 19	0, 19
$]3,4]$	20	0, 25	0, 44
$]4,6]$	20	0, 25	0, 69
$]6,10]$	24	0, 31	1
Total	79	1	

Les classes ne sont pas de même amplitude, il faut donc corriger les données. la plus petite amlitude est $a = 1$.

Le mode:Variable quantitative continue

Classe	Effectif corrigé	fréquence
$[2, 3]$	15	$0, 19$
$]3, 4]$	20	$0, 25$
$]4, 5]$	10	$0, 125$
$]5, 6]$	10	$0, 125$
$]6, 7]$	6	$0, 0775$
$]7, 8]$	6	$0, 0775$
$]8, 9]$	6	$0, 0775$
$]9, 10]$	6	$0, 0775$
Total	79	1

Il est clair que $]3, 4]$ est la classe modale.

Application: $m_d = 4 - 1 \times \dfrac{(0.25 - 0.125)}{(0.25 - 0.125) + (0.25 - 0.19)} = 3.324$

La médiane

Definition

La médiane est la valeur m de la variable qui partage les éléménts de la série statistique, préalablement classés par ordre croissant, en deux groupes d'effectifs égaux: 50% des individus présentent une valeur inférieure ou égale à la médiane et 50% des individus présentent une valeur supérieure ou égale à la médiane

La médiane Variable quantitative discrète

Soient $x_1, x_2, \cdots, x_N$ les valeurs prises par la variable. On les ordonne de la plus petite à la plus grande et on note $x_{(1)}$ la plus petite valeur $x_{(2)}$ la deuxième valeur, $\cdots$, $x_{(i)}$ la $i^{ème}$ valeur, $\cdots$ $x_{(N)}$ la plus grande valeur. Alors on a

$$m = \begin{cases} x_{(\frac{N+1}{2})} & \text{si } N \text{ est impair} \\ \dfrac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2} & \text{si } N \text{ est pair} \end{cases}$$

Examples

i) Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	3	8	4	9	3	3
effectifs cumulés	3	11	15	24	27	30

On a $N = 30$

donc N est pair d'où $\dfrac{N}{2} = 15$ et

$$m = \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2} = \frac{x_{(15)} + x_{(16)}}{2} = \frac{30+40}{2} = 35.$$

$x_{(16)} = 40$ car le premier effectif cumulé supérieur ou égal à 16 est 24 et $x_{(24)} = 40$.

Examples

ii) Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	4	9	4	8	3	3
effectifs cumulés	4	13	17	25	28	31

On a $N = 31$

donc N est impair d'où $\dfrac{N+1}{2} = 16$ et $m = x_{(16)} = 30$ car le premier effectif cumulé supérieur ou égal à 16 est 17 et $x_{(17)} = 30$.

La médiane Variable quantitative continue

La médiane est la solution de l'équation $F(x) = 0,5$.
Pour la déterminer, on commence par déterminer la classe médiane $]x_i, x_{i+1}]$ qui vérifie

$$F(x_i) < 0,5 \text{ et } F(x_{i+1}) \geq 0,5$$

La médine m (qui appartient à la classe médiane) est

$$m = x_i + (x_{i+1} - x_i)\frac{0.5 - F(x_i)}{F(x_{i+1}) - F(x_i)}$$

Où x_{i+1} est la borne supérieure de la classe médiane, x_i la borne inférieure de la classe médiane, $F(x_{i+1})$ la fréquence de la classe médiane et $F(x_i)$ la fréquence de la classe qui précède la classe médiane.

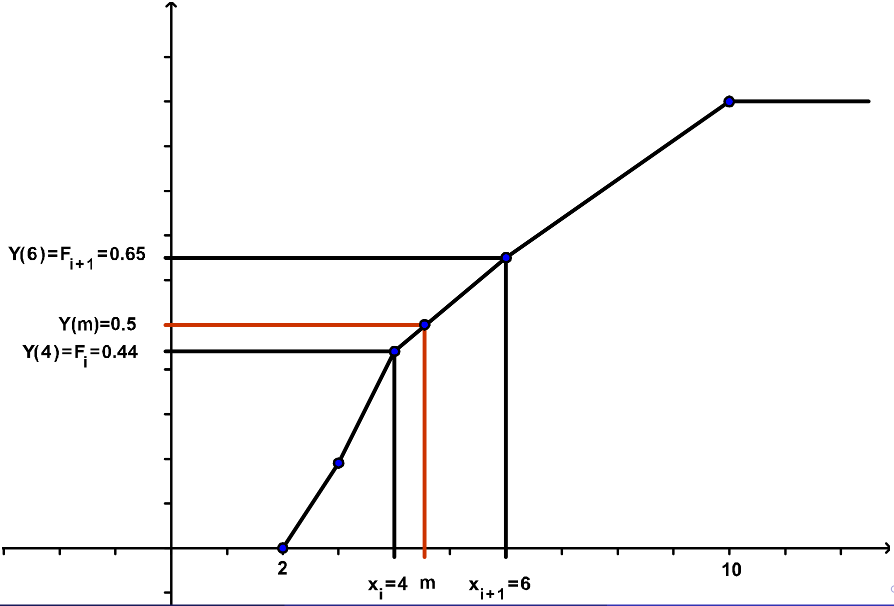
Example

Reprenons l'exemple de la distribution des salaires mensuels (en milliers de dirhams) du personnel d'une entreprise:

Classe	Effectif n_i	fréquence f_i	fréquence cumulée $F(x_{i+1})$
$]2,3]$	15	$0,19$	$0,19$
$]3,4]$	20	$0,25$	$0,44$
$]4,6]$	20	$0,25$	$0,69$
$]6,10]$	24	$0,31$	1
Total	79	1	

il est clair que la classe médiane est $]4,6]$.

La médiane Variable quantitative continue



La médiane Variable quantitative continue

On a $F(4) = 0,44 < 0.5$ et $F(6) = 0.64 > 0.5$, la classe médiane est donc $]4,6]$.
$x_i = 4$, $x_{i+1} = 6$, $F(x_i) = 0.44$, $F(x_{i+1}) = 0.69$ et

$$m = 4 + (6-4)\frac{0,5-0,44}{0,69-0,44} = 4,48$$

La Moyenne arithmétique Variable quantitative discrète

La moyenne arithmétique notée $\overline{x}$, est égale à la somme des valeurs distinctes de la variable multipliées par leurs effectifs respectifs divisée par la somme des effectifs.

$$\overline{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{\sum_i n_i x_i}{N}$$

et comme $f_i = \dfrac{n_i}{N}$ on a aussi $\overline{x} = \sum_i f_i x_i$.

La Moyenne arithmétique Variable quantitative discrète

Example

Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	4	8	4	9	3	3

On a $N = 31$

$$\overline{x} = \frac{10 \times 4 + 20 \times 8 + 30 \times 4 + 40 \times 9 + 50 \times 3 + 60 \times 3}{4 + 8 + 4 + 9 + 3 + 3} = \frac{1010}{31} = 32.58$$

La Moyenne arithmétique Variable quantitative continue

La moyenne arithmétique notée toujours $\overline{x}$, est égale à la somme des centres des classes de la variable multipliées par leurs effectifs respectifs divisée par la somme des effectifs.

$$\overline{x} = \frac{\displaystyle\sum_i n_i c_i}{\displaystyle\sum_i n_i} = \frac{\displaystyle\sum_i n_i c_i}{N}$$

où c_i est le centre de de la classe associée à l'effectif n_i.
et comme $f_i = \dfrac{n_i}{N}$ on a aussi $\overline{x} = \displaystyle\sum_i f_i c_i$

La Moyenne arithmétique Variable quantitative continue

Example

Reprenons l'exemple de la distribution des salaires mensuels (en milliers de dirhams) du personnel d'une entreprise:

Classe	Effectif n_i	fréquence f_i	fréquence cumulée $F(x_{i+1})$
$]2,3]$	15	$0,19$	$0,19$
$]3,4]$	20	$0,25$	$0,44$
$]4,6]$	20	$0,25$	$0,69$
$]6,10]$	24	$0,31$	1
Total	79	1	

$$\overline{x} = \frac{15 \times 2,5 + 20 \times 3,5 + 20 \times 5 + 24 \times 8}{15 + 20 + 20 + 24} = \frac{399,5}{79} = 5,05$$

Les mesures de dispersion

Les indicateurs de dispersion sont nombreux, les plus courants sont:

L'étendue, l'écart interquartile,
la variance, l'ecart-type et le cœfficient de variation.

L'étendue Variable quantitative discrète

Definition

L'étendue mesure l'écart entre la plus petite valeur de la variable et la plus grande:

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

où x_{min} (resp. x_{max}) est la valeur minimale (resp. maximale) prises par la variable.

Example

Soient les 4 series statistiques suivantes

a) 10, 10, 10, 10, 20, 30, 30, 30, 30

$$\overline{x} = \frac{4 \times 10 + 1 \times 20 + 4 \times 30}{9} = \frac{180}{9} = 20$$

b) 20, 22, 21, 20, 20, 19, 18, 20, 20

$$\overline{x} = \frac{18 + 19 + 5 \times 20 + 21 + 22}{9} = \frac{180}{9} = 20$$

c) 1, 4, 6, 8, 20, 32, 34, 36, 39

$$\overline{x} = \frac{1 + 4 + 6 + 8 + 20 + 32 + 34 + 36 + 39}{9} = \frac{180}{9} = 20$$

d) 10, 12, 14, 16, 20, 24, 26, 28, 30

$$\overline{x} = \frac{10 + 12 + 14 + 16 + 20 + 24 + 26 + 28 + 30}{9} = \frac{180}{9} = 20$$

Ces quatre séries ont la même moyenne $\overline{x} = 20$ et la même médiane $m = 20$.

Example

Pourtant ces séries sont trés différentes. Cette différence provient de leur dispersion, en effet:
$Etendue(a) = 30 - 10 = 20$; $Etendue(b) = 22 - 18 = 4$;
$Etendue(c) = 39 - 1 = 38$; $Etendue(d) = 30 - 10 = 20$.
Quoique les séries a) et d) ont la même étendue, les valeurs de la série d) contrairement à celles de la série a) sont uniformément espacées.

L'étendue Variable quantitative discrète

Definition

L'étendue est la différence entre la borne supérieure de la dernière classe et la borne inférieure de la première classe.

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

où x_{min} (resp. x_{max}) est la borne inférieure (resp. supérieure) de la première (resp. dernière) classe.

Les quartiles

Nous savons que la médiane divise la distribution en deux parties égales. Il existe d'autres indicateurs utiles:

a) Les quartiles qui divise la distribution en quatre (4) parties égales

b) Les déciles qui divise la distribution en dix (10) parties égales

c) Les ceniles qui divise la distribution en cent (100) parties égales

Les quartiles sont notés Q_1, Q_2 et Q_3 et on a $F(Q_1) = 0.25$, $F(Q_2) = 0.5$ et $F(Q_3) = 0.75$.

La médiane est le 2^{eme} quartile, le 5^{eme} décile et le 50^{eme} centile. Des techniques similaires à celles utilisées pour déterminer la médiane permettent de déterminer ces indicateurs.

On cosidère une série statistique dont les valeurs du caractère étudié, ont été rangés dans un ordre croissant:

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_{n-1} \leq x_n$$

La médiane m_e sépare la série en deux séries de même effectif. La série inférieure dont les valeurs du caractère sont inférieures ou égale à la médiane m_e, et la série supérieure dont les valeurs du caractère sont supérieures ou égale à la médiane m_e.
On appelle premier (resp. troisième) quartile, la médiane de la série inférieure (resp. supérieure) on le note Q_1 (resp. Q_3).

Example

i) Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	3	8	4	9	3	3
effectifs cumulés	3	11	15	24	27	30

On a $N = 30$ et $m = 35$

x_i	10	20	30
n_i	3	8	4
effectifs cumulés	3	11	15

série inférieure avec $N_1 = 15$

x_i	40	50	60
n_i	9	3	3
effectifs cumulés	9	12	15

série supérieure avec $N_1 = 15$

donc N_1 est impair d'où $\dfrac{N_1 + 1}{2} = 8 \Longrightarrow Q_1 = x_{(\frac{N_1+1}{2})} = x_{(8)} = 20$ et $Q_3 = x_{(\frac{N_1+1}{2})} = x_{(8)} = 40$.

Example

ii) Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	4	9	5	8	3	4
effectifs cumulés	4	13	18	26	29	33

On a $N = 33$ et $m = 30$.

x_i	10	20	30
n_i	4	9	3
effectifs cumulés	4	13	16

série inférieure avec $N_1 = 16$

x_i	30	40	50	60
n_i	1	8	3	4
effectifs cumulés	1	9	12	16

série supérieure avec $N_1 = 16$

donc N_1 est pair d'où $\dfrac{N_1}{2} = 8 \Longrightarrow$

$$Q_1 = \frac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \frac{x_{(8)} + x_{(9)}}{2} = 20 \text{ et}$$

$$Q_3 = \frac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \frac{x_{(8)} + x_{(9)}}{2} = \frac{40+40}{2} = 40.$$

les quartiles d'une variable quantitative continue

le premier quartile

$$\left.\begin{array}{c} x_i \leq Q_1 \leq x_{i+1} \\ F(x_i) \leq 0,25 \leq F(x_{i+1}) \end{array}\right\} \text{ et } Q_1 = x_i + (x_{i+1} - x_i)\frac{0,25 - F(x_i)}{F(x_{i+1}) - F(x_i)}$$

Pour le troisième quartile

$$\left.\begin{array}{c} x_i \leq Q_3 \leq x_{i+1} \\ F(x_i) \leq 0,75 \leq F(x_{i+1}) \end{array}\right\} \text{ et } Q_3 = x_i + (x_{i+1} - x_i)\frac{0,75 - F(x_i)}{F(x_{i+1}) - F(x_i)}$$

les quartiles d'une variable quantitative continue

Example

Reprenons la distribution des salaires mensuels.

Classe	Effectif n_i	fréquence f_i	fréquence cumulée $F(x_{i+1})$
$]2,3]$	15	$0,19$	$0,19$
$]3,4]$	20	$0,25$	$0,44$
$]4,6]$	20	$0,25$	$0,69$
$]6,10]$	24	$0,31$	1
Total	79	1	

$0.19 \leq F(Q_1) = 0.25 \leq 0.44 \Longrightarrow 3 \leq Q_1 \leq 4$, d'où

$$Q_1 = 3 + (4-3) \times \frac{0,25 - 0,19}{0,44 - 0,19} = 3,24$$

$0.69 \leq F(Q_3) = 0.75 \leq 1 \Longrightarrow 6 \leq Q_3 \leq 10$, d'où

$$Q_3 = 6 + (10-6) \times \frac{0,75 - 0,69}{1 - 0,69} = 6,19.$$

l'écart interquartile

Q_1 étant le premier quartile et Q_3 le troisième quartile, l'écart interquartile est la différence entre le troisième et le premier quartile, il est noté $\mathbf{R}(Q) = Q_3 - Q_1$.

$[Q_1, Q_3]$ est appelé intervalle interquartile. Il contient 50% des observations, le reste se réparti avec 25% à gauche de Q_1 et 25% à droite de Q_3.

L'écart interquartile $\mathbf{R}(Q)$ est la largeur de l'intervalle interquartile. C'est une mesure de longueur de cet intervalle et donc une mesure de dispersion des données autours de la médiane.

- Plus il est grand, plus les données sont dispersées autours de la médiane.
- Plus il est petit, plus les données sont proches de la médiane.

Example

Reprenons l'exemple de la distribution des salaires mensuels.
L'intervalle $[3,24; 6,19]$ est l'intervalle interquartile sa largeur est l'écart interquartile $\mathbf{R}(Q_3 - Q_1) = 6,19 - 3,24 = 2,85$.

Diagramme en boîte

Ce diagramme est aussi appelé boîte à moustaches. Il utilise la valeur du 1^{er} quartile Q_1 (qui correspond à 25% des effectifs), la valeur du $2^{ème}$ quartile $Q_2 = m$ (la médiane qui correspond à 50% des effectifs), la valeur du $3^{ème}$ quartile Q_3 (qui correspond à 75% des effectifs), l'intervalle interquartile $\mathbf{R}(Q)$ et les valeurs minimum et maximum de la série.

Diagramme en boîte

On représente sur un axe gradué (horizontal ou vertical) les différentes valeurs de la série Q_1, Q_2, Q_3, x_{min}, x_{max} ainsi que $Q_1 - 1.5 \times \mathbf{R}(Q)$ et $Q_3 + 1.5 \times \mathbf{R}(Q)$.

Le diagramme est formé d'un rectangle ayant pour extrémité inférieure le 1^{er} quartile et pour extrémité supérieure le $3^{ème}$ quartile. A l'interieur de ce rectangle, on trace un segment représentant la médiane.

A gauche et à droite de ce rectangle, on trace deux segments appelé "moustaches" inférieure et supérieure qui ont pour extrémité respectivement $Q_1 - 1.5 \times \mathbf{R}(Q)$ et $Q_3 + 1.5 \times \mathbf{R}(Q)$.

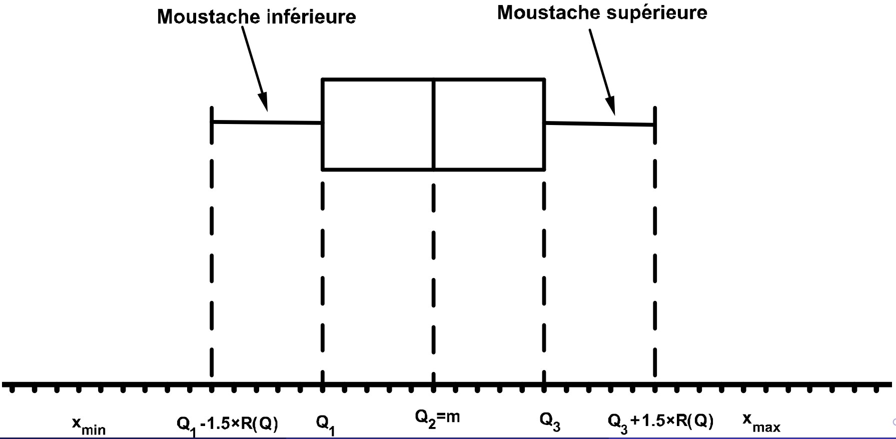

Moustache inférieure
Moustache supérieure
x_min
Q_1-1.5×R(Q)
Q_1
Q_2=m
Q_3
Q_3+1.5×R(Q)
x_max

Diagramme en boîte

La boîte a pour largeur la distance interquartile $(Q_3 - Q_1)$, et les moustaches sont basées généralement sur 1.5 fois la largeur de la boîte.

Dans ce cas, une valeur est atypique si elle dépasse de 1.5 fois l'écart interquartile à gauche du 1^{er} quartile ou à droite du $3^{ème}$ quartile.

La boîte à moustaches permet de répondre à certaines questions:

Diagramme en boîte

- Existe-t-il des observations atypiques? en les repérant et les identifiant

Diagramme en boîte

- Existe-t-il des observations atypiques? en les repérant et les identifiant
- La distribution est-elle symétrique? en repérant la position de la médiane dans la boîte.

Diagramme en boîte

- Existe-t-il des observations atypiques? en les repérant et les identifiant
- La distribution est-elle symétrique? en repérant la position de la médiane dans la boîte.

- La partie centrale (50% des effectifs) est-elle plus ou moins concentrée ou étalée par rapport au reste de la distribution?

Diagramme en boîte

- Existe-t-il des observations atypiques? en les repérant et les identifiant
- La distribution est-elle symétrique? en repérant la position de la médiane dans la boîte.
- La partie centrale (50% des effectifs) est-elle plus ou moins concentrée ou étalée par rapport au reste de la distribution?
- Comparaisons de distributions selon des groupes? Pour comparer les distributions d'une même variable selon les groupes, on juxtapose sur le même graphique les boîtes à moustaches définies respectivement pour les groupes en utilisant la même échelle.

Example

Deux groupes de S3 Statistique comparent leurs résultats du Contrôle final et déclarent: "nos classes ont le même profil puisque dans les deux cas la médiane et le mode des résultats est 10". Qu'en pensez-vous ?

notes	5	6	7	8	9	10	11	12	13	14	15	16	17
groupe 1	4	4	3	3	3	4	3	2	2	3	2	2	1
groupe 2	1	3	4	4	5	7	4	3	1	2	1	0	2

Vérifier que les deux médianes valent 10 et déterminer les quartiles de chaque série. Tracer côte à côte les diagrammes en boites de ces deux séries.

Diagramme en boîte

Les effectifs cumulés des deux groupes est:

notes	5	6	7	8	9	10	11	12	13	14	15	16	17
groupe 1	4	8	11	14	17	21	24	26	28	31	33	35	36
groupe 2	1	4	8	12	17	24	28	31	32	34	35	35	37

$N_1 = 36$ est pair d'où $\dfrac{N_1}{2} = 18 \Longrightarrow$

$$m_1 = \frac{x_{(\frac{N_1}{2})} + x_{(\frac{N_1}{2}+1)}}{2} = \frac{x_{(18)} + x_{(19)}}{2} = \frac{10+10}{2} = 10.$$

$N_2 = 37$ est impair d'où $\dfrac{N_2+1}{2} = 19 \Longrightarrow m_2 = x_{(\frac{N_2+1}{2})} = x_{(19)} = 10.$

Diagramme en boîte

Les séries inférieures et supérieurs du groupe 1 sont:

notes	5	6	7	8	9	10
groupe 1	4	4	3	3	3	1
groupe 2	1	3	4	4	5	1

série inférieure avec $N_{i1} = N_{i2} = 18$

notes	10	11	12	13	14	15	16	17
groupe 1	3	3	2	2	3	2	2	1
groupe 2	5	4	3	1	2	1	0	2

série supérieure avec $N_{s1} = N_{s2} = 18$

Diagramme en boîte

Les effectifs des séries inférieures et supérieurs du groupe 1 et 2 sont:

notes	5	6	7	8	9	10
groupe 1	4	8	11	14	17	18
groupe 2	1	4	8	12	17	18

série inférieure avec $N_{i1} = N_{i2} = 18$

notes	10	11	12	13	14	15	16	17
groupe 1	3	6	8	10	13	15	17	18
groupe 2	5	9	12	13	15	16	16	18

série supérieure avec $N_{s1} = N_{s2} = 18$

Diagramme en boîte

On a $N_{i1} = N_{i2} = 18$ est pair d'où:

$$Q_{11} = \frac{x_{(\frac{N_{i1}}{2})} + x_{(\frac{N_{i1}}{2}+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2} = 7 \text{ et}$$

$$Q_{12} = \frac{x_{(\frac{N_{i2}}{2})} + x_{(\frac{N_{i2}}{2}+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2} = 8.$$

On a $N_{s1} = N_{s2} = 18$ est pair d'où:

$$Q_{31} = \frac{x_{(\frac{N_{s1}}{2})} + x_{(\frac{N_{s1}}{2}+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2} = 13 \text{ et}$$

$$Q_{32} = \frac{x_{(\frac{N_{s2}}{2})} + x_{(\frac{N_{s2}}{2}+1)}}{2} = \frac{x_{(9)} + x_{(10)}}{2} = 11.5.$$

L'écart interquartile des deux groupes est: $\mathbf{R}(Q1) = 13 - 7 = 6$ et $\mathbf{R}(Q2) = 11.5 - 8 = 3.5$.

$$\Longrightarrow \begin{cases} Q_{11} - 1.5 \times \mathbf{R}(Q1) = -2 & Q_{31} + 1.5 \times \mathbf{R}(Q1) = 22 \\ Q_{12} - 1.5 \times \mathbf{R}(Q2) = 2.75 & Q_{32} + 1.5 \times \mathbf{R}(Q2) = 16.75 \end{cases}$$

Diagramme en boîte

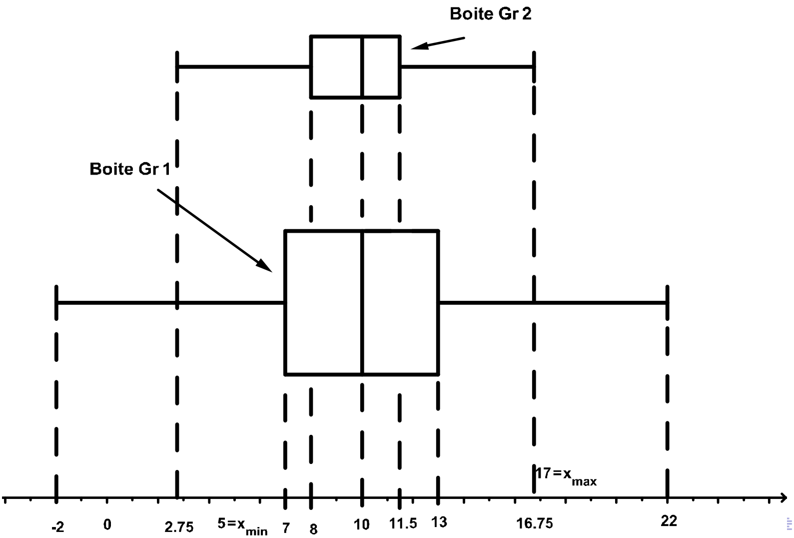


Diagramme en boîte

Le graphique ci-dessus met bien en évidence que l'écart interquartile est plus resserré pour le groupe 2 que le groupe 1 donc les élèves du groupe 2 ont globalement un niveau plus homogène que ceux de du groupe 1. On peut remarquer que 17 est une valeur atypique pour le groupe 2 tandis que le groupe 1 n'a pas de valeur atypique. La distribution du groupe 1 est symetrique car la boîte est symetrique par rapport au segment de la médiane tandis que celle du groupe 2 est asymetrique à gauche.

Diagramme tige et feuille

Un diagramme “ tige et feuille ” est une autre façon de résumer et représenter un ensemble de données de la distribution d'une variable quantitative. C'est un diagramme plus instructif pour les bases de données relativement petites (moins de 100 unités). Il se situe à mi chemin entre le tableau de distribution et le graphique. Comment construire un diagramme “ tige et feuille ”?

• Séparer chaque nombre en une tige qui contient tous les chiffres sauf le dernier et une feuille, soit le dernier chiffre. Les tiges ont autant de chiffres que nécessaire, alors que la feuille n'a qu'un seul chiffre.

• On place les tiges sur une colonne verticale avec la plus petite tige en haut.

• On écrit chaque feuille à droite de sa tige en ordre croissant.

Notons qu'une valeur est répétée autant de fois qu'elle apparaît.

Diagramme tige et feuille

Les avantages d'une telle présentation sont multiples:

- Toutes les valeurs y sont nommées et ordonnées
- Ce tracé ressemble quand on le tourne à un diagramme en bâtons.
- On peut y ajouter l'effectif de chaque tige.
- On peut y lire facilement le nombre de données, la valeur la plus grande, la plus petite, la plus fréquente ainsi que les éventuelles valeurs aberrantes.
- On peut repérer facilement la médiane, les quartiles, les déciles.
- On peut remarquer la symétrie ou l'asymétrie (lorsque sa forme générale est désaxée).

Diagramme tige et feuille

On considère une série de taux d'hémoglobine dans le sang (en g.l^{-1}) mesuré chez des adultes présumés en bonne santé. La série ordonnée est: 105 110 112 112 118 119 120 120 125 125 126 127 128 130 132 133 134 135 138 138 138 138 141 142 144 145 146 148 148 148 149 150 150 150 151 151 153 153 153 154 154 154 155 156 156 158 160 160 160 163 164 164 165 166 166 168 168 170 172 172 176 179. Un tracé en tiges et feuilles donne :

Diagramme tige et feuille

Tige	Feuille															Effecti
10	5															1
11	0	2	2	8	9											5
12	0	0	5	5	6	7	8									7
13	0	2	3	4	5	8	8	8	8							9
14	1	2	4	5	6	8	8	8	9							9
15	0	0	0	1	1	3	3	3	4	4	4	5	6	6	8	15
16	0	0	0	3	4	4	5	6	6	8	8					11
17	0	2	2	6	9											5

On peut lire ainsi que la valeur 105 est la plus petite valeur qui semble être une valeur aberrante, que 179 la plus grande valeur, que 120 figure 2 fois dans la série, 138 figure 4 fois.

Diagramme tige et feuille

Pour calculer la médiane , on a $N = 62$ pair et $\frac{N}{2} = 31 \Longrightarrow m = \frac{x_{(31)} + x_{(32)}}{2} = \frac{149+150}{2} = 149.5$, pour calculer le 1^{er} quartile, on a $\frac{N}{2} = 31$ impair et $\frac{\frac{N}{2}+1}{2} = 16 \Longrightarrow Q_1 = x_{(16)} = 133$ et pour calculer le $3^{ème}$ quartile, on a $\frac{N}{2} = 31$ impair et $\frac{\frac{3N}{2}+1}{2} = 47 \Longrightarrow Q_3 = x_{(47)} = 160$.

Un diagramme dos à dos de tige et feuille peut être employé pour comparer deux bases de données. Ci-dessous, nous représentons les notes sur 100 de deux groupes du cours de statistiques d'un examen en utilisant le diagramme dos à dos de tige et feuille:

La variance et l'écart-type Variable quantitative discrète

Effectifs	Feuille	Tlge	Feuille	Effectiffs
0		0	5	1
2	3 1	2	4 5 7	3
4	4 4 3 3	3	1 2 2 8 8 9	6
5	9 9 6 6 4	4	3 3 3 4 7 7 7	7
10	7 5 5 4 4 4 4 2 2 1	5	4 4 4 6 6 8 8 8 9	9
12	9 9 8 7 7 7 3 3 2 1 1 1	6	1 2 4 4 5 5 9	7
6	9 8 7 5 5 2	7	3 3 4 6 6 6	6
6	6 6 6 3 1 1	8	2 5 9	3
3	4 3 2	9	1	1

La variance et l'écart-type Variable quantitative discrète

Definition

La variance $V(x)$ est la moyenne arithmétique des carrés des écarts des valeurs de la variable à la moyenne arithmétique

$$V(x) = \frac{1}{N}\sum_i n_i(x_i - \overline{x})^2 = \sum_i f_i(x_i - \overline{x})^2 \quad \text{où } N = \sum_i n_i$$

La racine carrée de la variance est appelée l'écart-type

$$\sigma(x) = \sqrt{\frac{1}{N}\sum_i n_i(x_i - \overline{x})^2} = \sqrt{\sum_i f_i(x_i - \overline{x})^2}$$

Relation de König

$$\sum_i n_i(x_i - \overline{x})^2 = \sum_i n_i x_i^2 - N\overline{x}^2 \Longrightarrow V(x) = \frac{1}{N}\left(\sum_i n_i x_i^2\right) - \overline{x}^2$$

Example

Considérons la distribution suivante

x_i	10	20	30	40	50	60
n_i	4	8	4	9	3	3

on a $N = 31$ et $\overline{x} = 32.58$

$$\begin{aligned}
V(x) &= \frac{4(10-32.58)^2 + 8(20-32.58)^2 + 4(30-32.58)^2}{31} \\
&\quad + \frac{9(40-32.58)^2 + 3(50-32.58)^2 + 3(60-32.58)^2}{31} \\
&= \frac{6993.5484}{31} = 225.598 \\
\sigma(x) &= \sqrt{225.598} = 15.02
\end{aligned}$$

La variance et l'écart-type Variable quantitative continue

Definition

La variance $V(x)$ est la moyenne arithmétique des carrés des écarts des centres des classes à la moyenne arithmétique

$$V(x) = \frac{1}{N}\sum_i n_i(c_i - \overline{x})^2 = \sum_i f_i(c_i - \overline{x})^2$$

où c_i est le centre de la classe associée à n_i.
La racine carrée de la variance est appelée l'écart-type

$$\sigma(x) = \sqrt{\frac{1}{N}\sum_i n_i(c_i - \overline{x})^2} = \sqrt{\sum_i f_i(c_i - \overline{x})^2}$$

Relation de König

$$\sum_i n_i(c_i - \overline{x})^2 = \sum_i n_i c_i^2 - N\overline{x}^2 \Longrightarrow V(x) = \frac{1}{N}\left(\sum_i n_i c_i^2\right) - \overline{x}^2$$

Variance

Example

Reprenons la distribution des salaires mensuels.

Classe	Effectif n_i	fréquence f_i	fréquence cumulée $F(x_{i+1})$
$]2,3]$	15	0, 19	0, 19
$]3,4]$	20	0, 25	0, 44
$]4,6]$	20	0, 25	0, 69
$]6,10]$	24	0, 31	1
Total	79	1	

on a $\overline{x} = 5.05$

$$\begin{aligned} V(x) &= \frac{15(2.5-5.05)^2 + 20(3.5-5.05)^2 + 20(5-5.05)^2}{79} \\ &= +\frac{24(8-5.05)^2}{79} = \frac{354.497}{79} = 4.487 \\ \sigma &= \sqrt{4.487} = 2.118 \end{aligned}$$

Le cœfficient de variation

tous les indicateurs de dispersion que nous avons vu jusqu'à présent dépendent des unités de mesure de la variable. Ils ne permettent pas de comparer la dispersion de distributions statistiques hétérogènes.
Le cœfficient de variation, qui est un nombre sans dimension, permet cette comparaison lorsque les valeurs de la variable sont positives. Il s'écrit

$$CV = \frac{\sigma(x)}{\overline{x}}$$

Si $CV < 0,5$ alors la dispersion n'est pas importante. Si $CV > 0,5$ alors la dispersion est importante.

Le cœfficient de variation

Example

Dans une maternité on a relevé le poids (en kg) à la naissance de 47 nouveau-nés. Les données collectées sont résumées dans le tableau suivant:

classe	n_i	c_i	$n_i c_i$	$(c_i - \overline{x})$	$(c_i - \overline{x})^2$	$n_i(c_i - \overline{x})^2$
$]2,5;3,0]$	8	2,75	22,00	−0,73	0,5329	4,2632
$]3,0;3,5]$	15	3,25	48,75	−0,23	0,0529	0,7935
$]3,5;4,0]$	20	3,75	75,00	0,27	0,0729	1,4580
$]4,0;4,5]$	4	4,50	18,00	0,52	0,2704	1,0816
Total	47		163,75			7,5963

$\overline{x} = \dfrac{163,75}{47} = 3,48$, $\sigma(x) = \sqrt{\dfrac{7,5963}{47}} = \sqrt{0,1616} = 0,4019$ et

$CV = \dfrac{0,4019}{3,48} = 0,1154$

Le cœfficient de variation étant faible, le poids à la naissance est concentré autour de la moyenne.

Moments

Definition

Le moment d'ordre r d'une variable statistique est la quantité

$$m_r = \frac{1}{N}\sum_i n_i\, x_i^r \quad \text{ou} \quad m_r = \frac{1}{N}\sum_i n_i\, c_i^r \ \text{où} \ N = \sum_i n_i$$

Pour $r = 0$, $m_0 = 1$.
Pour $r = 1$, $m_1 = \overline{x}$ la moyenne arithmétique.
Le moment centré d'ordre r d'une variable est la quantité

$$\mu_r = \frac{1}{N}\sum_i n_i\,(x_i - \overline{x})^r \quad \text{ou} \quad \mu_r = \frac{1}{N}\sum_i n_i\,(c_i - \overline{x})^r \ \text{où} \ N = \sum_i n_i$$

Pour $r = 0$, $\mu_0 = 1$.
Pour $r = 1$, $\mu_1 = 0$
Pour $r = 2$, $\mu_2 = V(x)$ la variance.

Changement d'origine et d'unité

Definitions

•On appelle changement d'origine l'opération consistant à ajouter (ou soustraire) la même quantité $b \in \mathbb{R}$ à toutes les observations:
$y_i = x_i + b,\ \ i = 1, \cdots, n.$

•On appelle changement d'unité l'opération consistant à multiplier (ou diviser) par la même quantité $a \in \mathbb{R}$ toutes les observations:
$y_i = a\ x_i,\ \ i = 1, \cdots, n.$

•On appelle changement d'origine et d'unité l'opération consistant à multiplier toutes les observations par la même quantité $a \in \mathbb{R}$ puis à ajouter la même quantité $b \in \mathbb{R}$ à toutes les observations:
$y_i = a\ x_i + b,\ \ i = 1, \cdots, n.$

Changement d'origine et d'unité

Theorem

Si on effectue un changement d'origine et d'unité sur une variable X, alors

- *Sa moyenne est affectée du même changement d'origine et d'unité,* $\overline{y} = a\,\overline{x} + b$
- *Sa variance et son écart-type sont affectés par le carré du changement d'unité et pas par le changement d'origine,* $V_y = a^2\,V_x$ *et* $\sigma_y \sqrt{V_y} = |a|\,\sigma_x$.

Preuve: Si $y_i = a\ x_i + b$, alors

$$
\begin{aligned}
\overline{y} &= \frac{1}{n}\sum_{i=1}^{n}(a\ x_i = b) = a\ (\frac{1}{n}\sum_{i=1}^{n} x_i) + b = a\ \overline{x} + b \\
V_y &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(a\ \ x_i + b - a\ \ \overline{x} - b)^2 \\
&= a^2\ \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = a^2\ V_x \\
\sigma_y &= \sqrt{V_y} = \sqrt{a^2\ V_x} = |a|\ \sigma_x
\end{aligned}
$$

Changement d'origine et d'unité

Remarque:

• Les paramètres de position (mode, médiane et moyenne) sont tous affectés par un changement d'origine et d'unité.

• Les paramètres de dispersion sont tous affectés par un changement d'unité mais pas par un changement d'origine (sauf le coefficient de variation).

Centrer et reduire une variable

Centrer et reduire une variable statistique quantitative X consiste la remplacer par la variable: $\dfrac{X-\overline{X}}{\sigma(X)}$.

$X-\overline{X}$ pour la centrer (moyenne 0).

La variable $\dfrac{X-\overline{X}}{\sigma(X)}$ a pour moyenne arithmétique 0 elle est centré.

Diviser par l'écart-type $\sigma(X)$ pour la reduire (écart-type $=1$).

La variable $\dfrac{X-\overline{X}}{\sigma(X)}$ a pour variance et écart-type 1 elle est reduite.

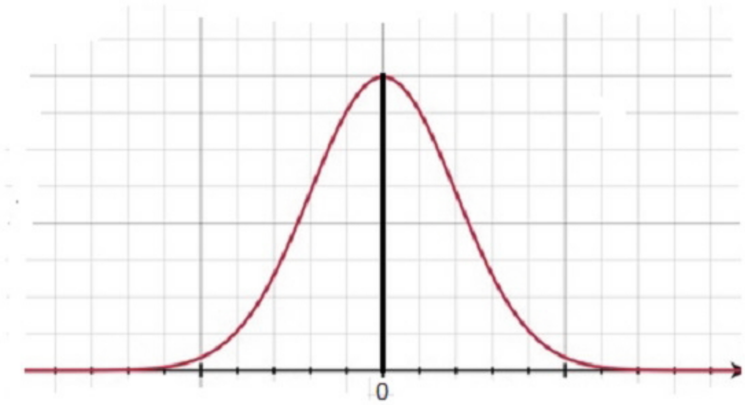
Symétrie et asymétrie

Definition

Une distribution est dite symétrique si le mode, la médiane et la moyenne sont confondus.
Une distribution qui n'est pas symétrique est dite asymétrique

Remarque: Une variable statistique est symétrique si ses valeurs sont réparties de manière symétrique autour de la moyenne c'est à dire si le polygône des fréquences a la forme d'une clôche comme dans la figure ci-après.

forme d'une cloche



Symetrie et asymetrie

A la différence de la médiane et du mode, la moyenne arithmétique est fortement influencée par les valeurs extrêmes. Lorsque les valeurs sont distribuées de manière symétrique, la moyenne arithmétique coïncide avec la médiane. Lorsque la distribution est asymétrique, la moyenne arithmétique dépasse la médiane si les valeurs extrêmes sont élevées et se situe en dessous de la médiane si les valeurs extrêmes sont basses.

Definition

Une distribution est dite asymétrique à droite, si la courbe du polygône des fréquences est étalée à droite, on a généralement:

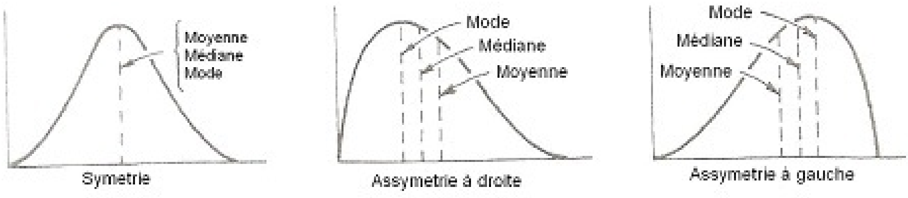
$$\text{mode} < \text{médiane} < \text{moyenne}.$$

Definition

Une distribution est dite asymétrique à gauche, si la courbe du polygône des fréquences est étalée à gauche, on a généralement:

$$\text{moyenne} < \text{médiane} < \text{mode}.$$

Symétrie et asymétrie



le coefficient d'asymétrie a pour rôle de fournir une mesure de dissymétrie d'une distribution

Coefficient de d'asymétrie

Coefficient de d'asymétrie de Pearson

Le premier coefficient d'asymétrie de Pearson est basé sur une comparaison de la moyenne et de la médiane, et est normalisé par l'écart-type. Il est calculé à partir de la formule suivante:

$$A_{P1} = 3\frac{\overline{X} - m}{\sigma} \quad \text{où } \overline{X} \text{ est la moyenne, } m \text{ la médiane et } \sigma \text{ l'écart-type.}$$

Lorsque la distribution statistique est unimodale, on utilise le second coefficient de Pearson basé sur une comparaison de la moyenne et du mode, et est normalisé par l'écart-type. Il est définie par

$$A_{P2} = \frac{\overline{X} - m_d}{\sigma} \quad \text{où } \overline{X} \text{ est la moyenne, } m_d \text{ le mode et } \sigma \text{ l'écart-type.}$$

Coefficient de d'asymétrie

Coefficient d'asymétrie de Yule

Le coefficient d'asymétrie de Yule est basé sur les positions des trois quartile et est normalisé par l'ecart interquartile. Il est calculée à partir de la formule suivante:

$$A_Y = \frac{Q_1 + Q_3 - 2 \times Q_2}{\mathbf{R}(Q)}$$

où Q_1, Q_2, Q_3 les 3 quartiles et $\mathbf{R}(Q)$ l'écart interquartile.

Coefficient de d'asymétrie de Fisher

Le coefficient d'asymétrie de Fisher est basé sur le moment d'ordre 3 et est normalisé par l'écart-type. Il est calculée à partir de la formule suivante:

$$A_F = \frac{\mu_3}{\sigma^3}$$ où μ_3 est le moment d'ordre 3 et σ l'écart-type.

Coefficient de d'asymétrie

Tous les coefficients d'asymetrie ont les mêmes propriétés.

- Si la distrbution est symétrique, le coefficient est nul.
- Si la distribution est asymétrique à droite c'est à dire la courbe est étalée à droite, alors et le coeficient est positif.
- Si la distribution est asymétrique à gauche c'est à dire la courbe est étalée à gauche, alors et le coeficient est négatif.

Les paramètres d'asymétrie ne sont pas affectés par un changement d'unité ou d'origine.

Coefficient de d'asymétrie (Exemple)

On considère la série statistique suivante (masse en grammes des oeufs de poule d'un élevage).

masse: x_i	40	45	50	55	60	65	70	75	80	85	90
Effectif: n_i	16	20	75	141	270	210	165	63	21	12	7

	$\overline{x}$	V	σ	μ_3	$m = Q_2$	m_d	Q_1	Q_3	$\mathbf{R}(Q)$
Groupe 1	62.5	73.8	8.59	91.125	60	60	55	70	15

A_{P1}	A_{P2}	A_Y	A_F
0.87	0.29	0.33	0.14

La distrbution des masses est asymetrie à droite car les coefficients d'asymetrie sont positifs.

Coefficient d'aplatissement

Le coefficient d'aplatissement mesure le degré d'aplatissement d'une distribution. On l'obtient à partir du moment centré de quatrième ordre μ^4.

$\mu^4 = \dfrac{1}{N}\displaystyle\sum_{k=1}^{n} n_i(x_i - \overline{x})^4$ pour une variable quantitative discrète.

$\mu^4 = \dfrac{1}{N}\displaystyle\sum_{k=1}^{n} n_i(c_i - \overline{x})^4$ pour une variable quantitative continue.

- Coefficient de d'aplatissement de Pearson

$$\beta_2 = \frac{\mu^4}{V(X)^2} \quad \text{où } V(X) \text{ est la variance}$$

- Coefficient d'aplatissement de Ficher

$F_2 = \beta_2 - 3$ où β_2 est le coefficient de d'aplatissement de Pearson

Coefficient d'aplatissement

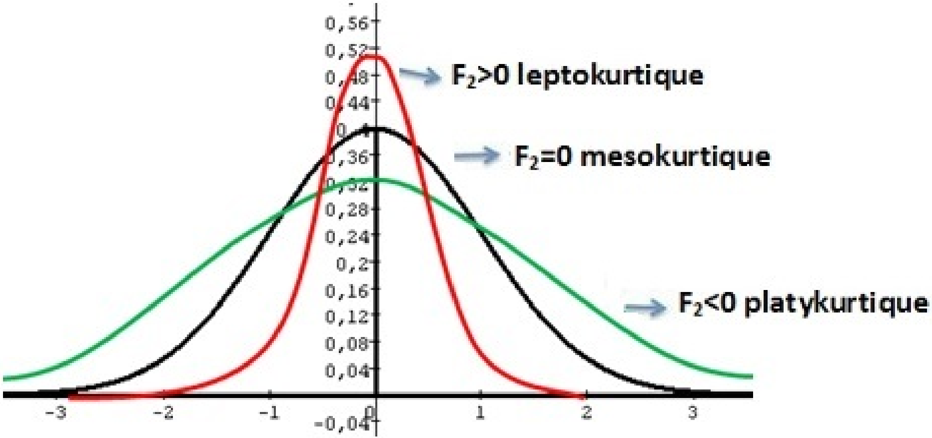
3 est le degré d'aplatissement d'une loi gaussienne centrée réduite.

Si $F_2 = 0$, le polygone statistique de la variable centrée réduite $\frac{X - \overline{X}}{\sigma}$ à le même aplatissement qu'une courbe en cloche, on dit que la variable est mesokurtique.

Si $F_2 > 0$, le polygone statistique de la variable centrée réduite est moins aplati qu'une courbe en cloche, la concentration des valeurs de la série autour de la moyenne est forte, on dit que la variable est leptokurtique.

Si $F_2 < 0$, le polygone statistique de la variable centrée réduite est plus aplati qu'une courbe en cloche, la concentration des valeurs autour de la moyenne est faible, on dit que la variable est platykurtique.

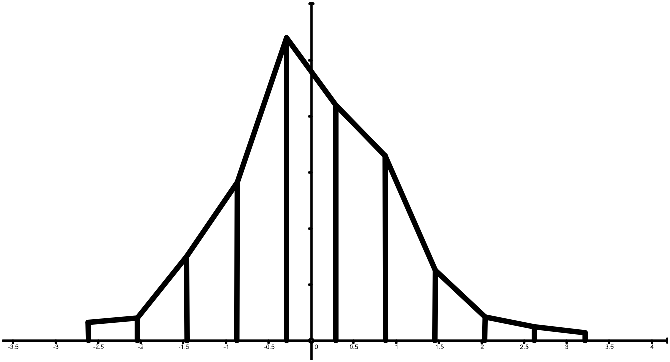
Coefficient d'aplatissement



Coefficient d'aplatissement

Reprenons la distribution des masse des oeufs final de l'exemple 2.3.1. , $\mu_4 = 17523.91$, $V(X) = 73.8$, $\beta_2 = 3.22$ et $F_2 = 0.22 > 0 \Longrightarrow$ la variable est leptokurtique et le polygone statistique de la variable centrée réduite est moins aplati qu'une courbe en cloche, la concentration des valeurs de la série autour de la moyenne est forte.

Coefficient d'aplatissement



Applications: Le théorème de Tchebychebv

Nous avons vu qu'il existe plusieurs mesures de positions et de dispersions. La moyenne est sans doute la mesure de position la plus répandue alors que la variance et l'écart-type sont les mesures de dispersion les plus utilisées. Nous allons voir comment en n'utilisant que la moyenne et l'écart-type, il est possible d'explorer des données.
Le théorème de Tchebychebv permet d'évaluer le pourcentage des données qui se trouvent à k écart-types de la moyenne, pour un entier k donné.

Theorem

Pour un entier $k \geq 2$, au moins $100 \times (1 - \dfrac{1}{k^2})\%$ des observations, d'une série de données, se trouvent à k écart-type de la moyenne de cette série.

Applications: Le théorème de Tchebychebv

Example

Les notes de 100 étudiants d'un contrôle de statistique ont une moyenne $\overline{x} = 14$ avec un écart-type $\sigma(x) = 1$. combien d'étudiants ont une note entre 12 et 16?

Remarquons que $12 = \overline{x} - 2\sigma(x)$ et que $16 = \overline{x} + 2\sigma(x)$. Ainsi, d'après le théorème de Tchebychev, le pourcentage d'étudiants ayant obtenue une note entre 12 et 16 est supérieur ou égal à $100 \times (1 - \dfrac{1}{2^2})\% = 75\%$.

Le pourcentage garanti par le théorème de Tchebychev peut être améliorer sous certaines conditions.

Applications: Le théorème de Tchebychebv

Règle Empirique

Si les observations sont réparties de manière symétrique autour de la moyenne alors

- Approximativement 68% des valeurs sont à un écart-type de la moyenne.

Règle Empirique
Si les observations sont réparties de manière symétrique autour de la moyenne alors

- Approximativement 68% des valeurs sont à un écart-type de la moyenne.
- Approximativement 95% des valeurs sont à deux écart-type de la moyenne.

Applications: Le théorème de Tchebychebv

Règle Empirique
Si les observations sont réparties de manière symétrique autour de la moyenne alors

- Approximativement 68% des valeurs sont à un écart-type de la moyenne.
- Approximativement 95% des valeurs sont à deux écart-type de la moyenne.
- Approximativement toutes les valeurs sont à trois écart-type de la moyenne.