

Chapitre 2

Statistique Descriptive à Une dimension :

Etude d'Une Série Statistique Simple

2.1 Objectif de la Statistique Descriptive

L'objectif majeure de la S-D est de pouvoir résumer l'information contenue dans une s.s.s en remplaçant toute la série par quelques paramètres. Ces paramètres doivent être interprétables pour pouvoir comprendre les interactions qui régissent toute la série.

L'idée est de remplacer le vrac des données par une seule qui se prête à l'interprétation.

Pour résumer l'information, nous devons au préalable réaliser deux tâches :

- Ordonner les données
- Visualiser les données à l'aide de graphiques.

2.2 Mise en ordre d'une série statistique simple

Définitions

- On appelle série statistique simple (s.s.s) une suite finie de nombres réels (x_1, x_2, \dots, x_n) qui sont les valeurs mesurées ou observées d'un caractère X sur les individus d'une population. L'ordre des nombres étant l'ordre des observations.
- Le nombre d'éléments d'une série statistique simple est appelé l'effectif de la série (ici n).

Remarque

L'étude sera faite sur un caractère quantitatif. Le cas qualitatif peut être ramené par codage au cas quantitatif discret.

Cette opération dépend pour un caractère quantitatif de sa nature discret ou continu.

Nous allons utiliser l'une des deux présentations condensées d'une série statistique simple.

1er Cas : Caractère Discret : série statistique à valeurs isolées.

- On ordonne les valeurs x_1, x_2, \dots, x_n par ordre croissant stricte; on obtient la suite y_1, y_2, \dots, y_r avec $y_1 < y_2 < \dots < y_r$ et $r < n$.
(y_1, y_2, \dots, y_r) est appelée l'échantillon ordonnée.

Avec $y_1 = \inf(x_1, x_2, \dots, x_n)$ et $y_r = \sup(x_1, x_2, \dots, x_n)$

- On appelle effectif de la valeur y_i et on note n_i le nombre d'éléments de la s.s.s **qui sont** égaux à y_i .
- Nous avons la relation :

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

- La suite des couples $((y_i, n_i))_{i=1,2,\dots,r}$ est appelée s.s à valeurs isolées.
- On appelle effectif cumulé de la valeur y_i et on note N_i le nombre d'éléments de la s.s.s qui sont inférieurs ou égaux à y_i .

Nous avons les relations suivantes :

$$\begin{cases} N_1 = n_1 \\ N_i = N_{i-1} + n_i \\ N_r = n \end{cases}$$

- On appelle fréquence de la valeur y_i et on note f_i proportion d'éléments de la s.s.s qui sont égaux à y_i . Nous avons la relation :

$$\sum_{i=1}^r f_i = 1$$

On appelle fréquence cumulée de la valeur y_i et on note F_i la proportion d'éléments de la s.s.s qui sont inférieurs ou égaux à y_i .

Nous avons les relations suivantes :

$$\begin{cases} F_1 = f_1 \\ F_i = F_{i-1} + f_i \\ F_r = 1 \end{cases}$$

2ème Cas : Caractère Continu : série statistique à valeurs classées.

- Soit s un entier positif strictement et soient z_1, z_2, \dots, z_s ($s+1$) nombres réels tels que :

$$\circ \quad z_1 < z_2 < \dots < z_s$$

$$\circ \quad \forall i = 1, \dots, n ; z_0 \leq x_i < z_s$$

Les intervalles semi-ouverts $[z_{i-1}, z_i[= \delta_i$ ($i=1,2,\dots,s$) sont appelées des classes.

- Le nombre d'éléments de la série qui appartiennent à la classe δ_i est appelé l'effectif de la classe δ_i . on le note n_i . On a :

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

- La suite des couples $(([z_{i-1}, z_i [))_{i=1,2,\dots,s}$ est appelée s.s à valeurs classées.

Elle représente sous forme simplifiée la série statistique simple (x_1, x_2, \dots, x_n) .

- On appelle effectif cumulé de la classe δ_i et on note N_i le nombre d'éléments de la s.s.s qui appartiennent à la classe δ_i . Nous avons les relations suivantes :

$$\begin{cases} N_1 = n_1 \\ N_i = N_{i-1} + n_i \\ N_s = n \end{cases}$$

- On appelle fréquence de la classe δ_i et on note f_i proportion d'éléments de la s.s.s qui sont appartiennent à la classe δ_i .

Nous avons la relation :

$$\sum_{i=1}^r f_i = 1$$

- On appelle fréquence cumulée de la classe δ_i et on note F_i la proportion d'éléments de la s.s.s qui appartiennent à la classe δ_i .

Nous avons les relations suivantes :

$$\begin{cases} F_1 = f_1 \\ F_i = F_{i-1} + f_i \\ F_s = 1 \end{cases}$$

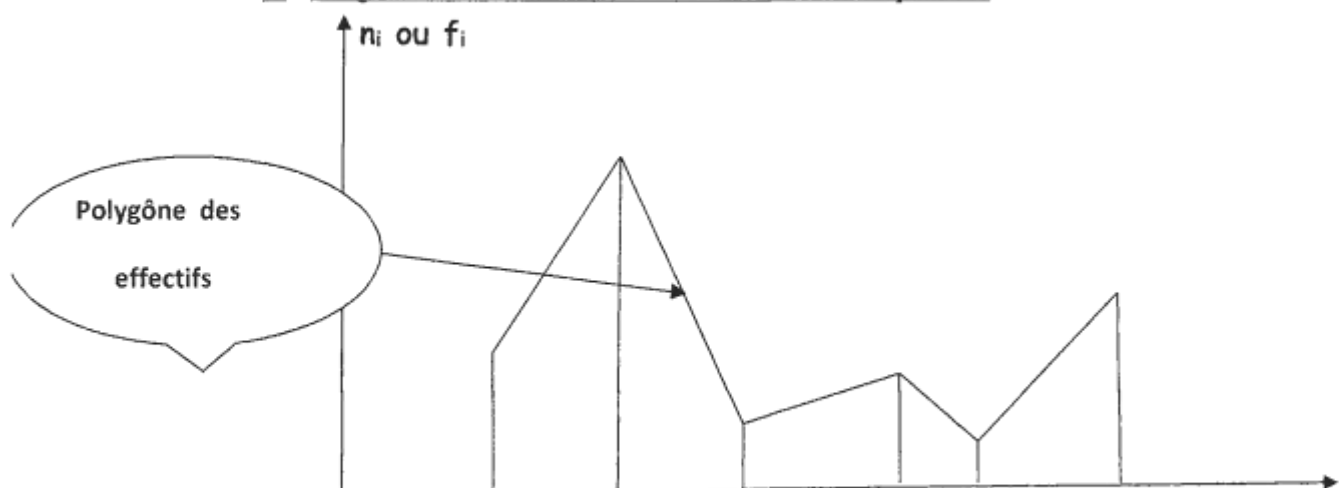
2.3 Représentation d'une série statistique simple

Il y a au moins deux représentations possibles d'une série statistique simple :

- Cas d'une s.s.v.i.**

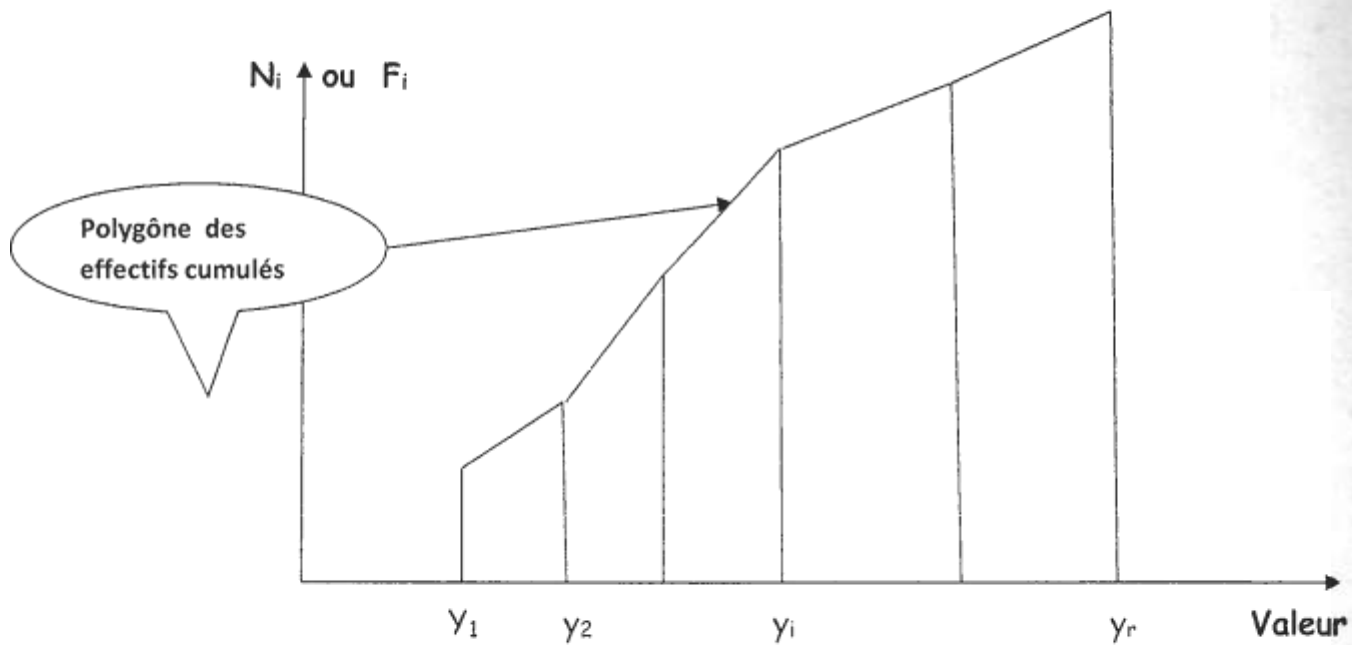
- Table des vecteurs (y_i, n_i, N_i, f_i, F_i).
- Diagramme en bâtons des effectifs ou des fréquences.
- Diagramme en bâtons des effectifs cumulés ou des fréquences cumulées.

- a- **Diagramme en bâtons des effectifs ou des fréquences**



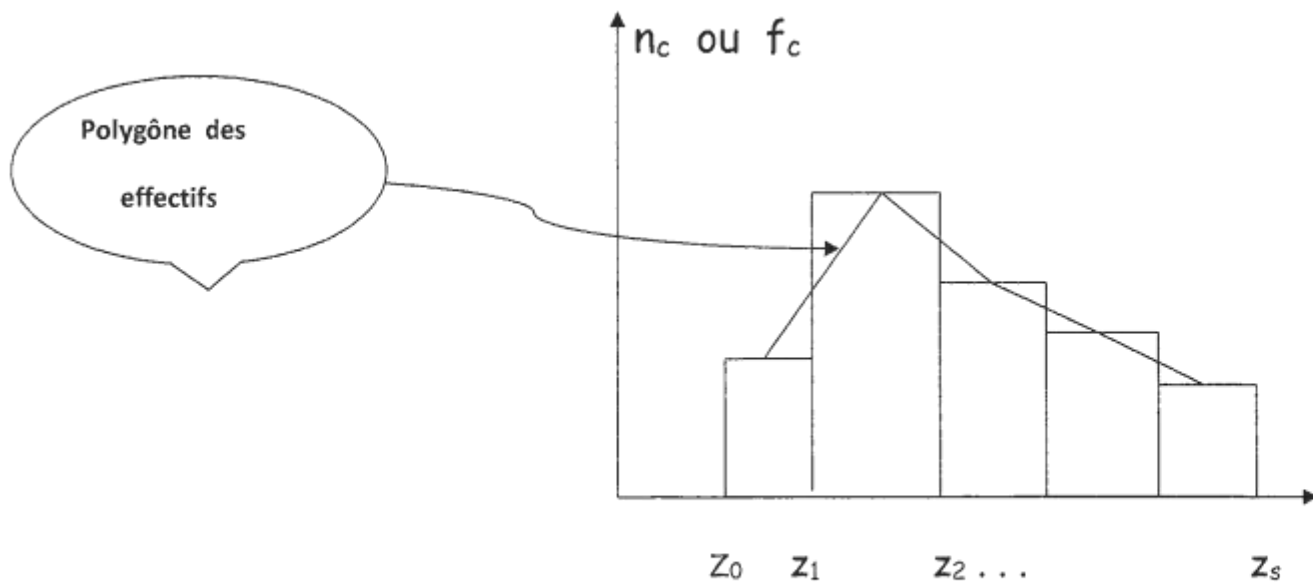
y_1 y_2 y_i y_r Valeur

b- Diagramme en bâtons des effectifs cumulés ou des fréquences cumulées



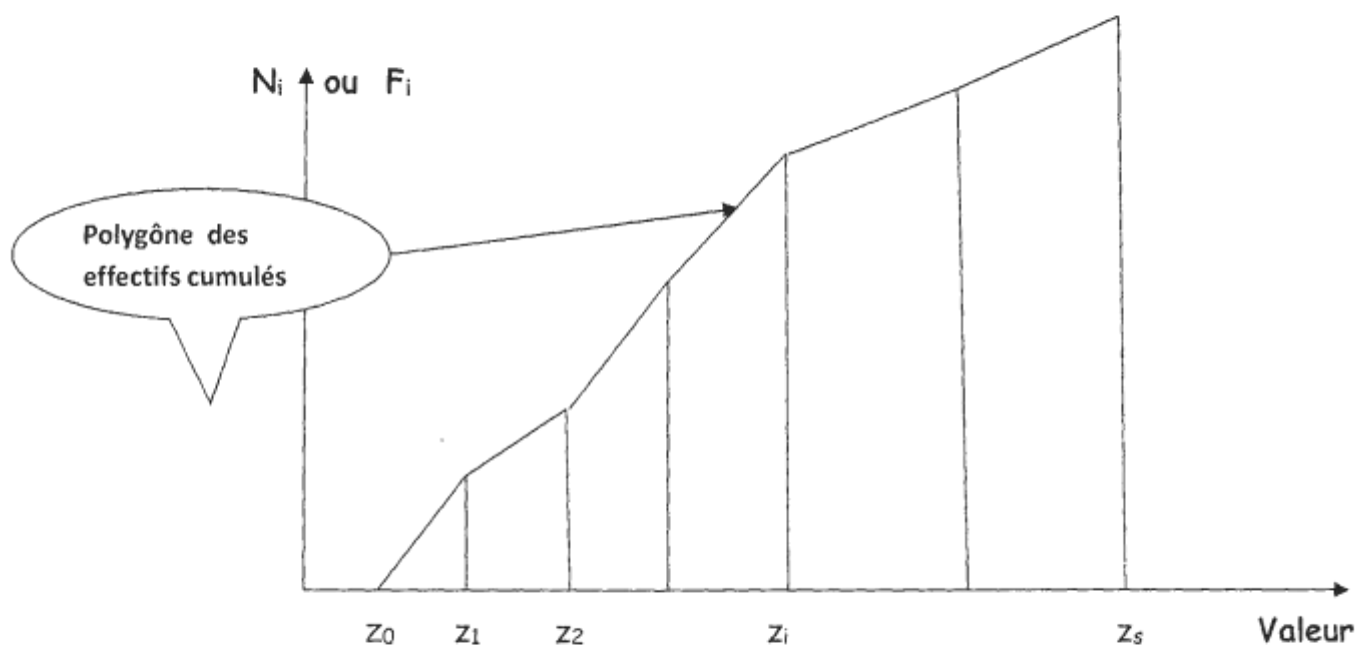
c- Cas d'une s.s.v.c.

- Table des vecteurs ($\delta_i, n_i, N_i, f_i, F_i$).
- Histogramme des effectifs ou des fréquences.
- Diagramme en bâtons des effectifs cumulés ou des fréquences cumulées.
- Histogramme des effectifs ou des fréquences



Indications pour le tracé de l'histogramme

- d- La longueur du côté parallèle à l'axe des ordonnées du rectangle représentatif d'une classe n'est pas en général proportionnelle à l'effectif ou à la fréquence de cette classe ; elle est proportionnelle à une quantité n_{ic} appelée effectif corrigé de la classe, ou à une quantité f_{ic} appelée fréquence corrigée de la classe, que l'on détermine de la manière suivante : si la longueur de la classe est égale à k_i fois une unité de longueur que l'on se fixe à priori, alors l'effectif corrigé est $(1/k_i)$ fois l'effectif, et la fréquence corrigée est $(1/k_i)$ fois la fréquence.
- e- Dans le cas où toutes les classes ont la même longueur, la longueur du côté parallèle à l'axe des ordonnées du rectangle est proportionnelle à l'effectif ou à la fréquence.
- c- Diagramme en bâtons des effectifs cumulés ou des fréquences cumulées



Remarques

- La courbe des fréquences cumulées est une courbe croissante.
- L'ordonnée du point d'abscisse z_s est proportionnel à 1 (ou à n).

2.4 Etude analytique d'une série statistique simple

Soit (x_1, x_2, \dots, x_n) la série statistique simple étudiée. Nous retenons trois objectifs principaux pour son étude analytique.

Objectif₁ - Déterminer des coefficients qui **remplacent** toutes les valeurs de la série et qui **résumant** l'information apportée par la série : Ce sont les **caractéristiques de valeur centrale**.

Objectif₂ - Déterminer des coefficients qui mesurent la **dispersion** des éléments de la série autour d'une valeur centrale donnée : Ce sont les **caractéristiques de valeur centrale**.

Objectif₃ - Déterminer des coefficients qui caractérisent la **forme** de la courbe des fréquences : Ce sont les **caractéristiques de forme**.

Il est souhaitable que les caractéristiques des séries statistiques satisfassent à certaines conditions, qui ont été énoncées par Yule :

- 1) La caractéristique est définie de façon objective.
- 2) La valeur de la caractéristique dépend de toutes les observations, et non d'une partie seulement.
- 3) La caractéristique peut être interprétée d'une façon simple.
- 4) La caractéristique est facile à déterminer.
- 5) La caractéristique est peu sensible aux fluctuations d'échantillonnage.
- 6) La caractéristique se prête aisément aux calculs algébriques ultérieurs.

A.- CARACTERISTIQUES DE VALEUR CENTRALE.

A.1) Moyenne arithmétique

On considère la s.s.s. (x_1, x_2, \dots, x_n) .

Définition

On appelle moyenne arithmétique de la série statistique simple (x_1, x_2, \dots, x_n) , et on note \bar{x} , la quantité réelle $(x_1 + x_2 + \dots + x_n)/n$.

Calcul de la moyenne arithmétique

- 1) Calcul direct, à partir de la définition.
- 2) Si la série est présentée sous la forme d'une série à valeurs isolées, $((y_1, n_1), ((y_2, n_2), \dots, ((y_r, n_r),)$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r n_i y_i = \sum_{i=1}^r f_i y_i$$

- 3) Si la série est présentée sous la forme d'une série à valeurs classées $(([z_{i-1}, z_i[, n_i))_{i=1, \dots, s}$, on peut effectuer l'approximation qui consiste à prendre tous les éléments de la série appartenant à une classe déterminée égaux au milieu de cette classe.

Soit θ_i le milieu de la classe $[z_{i-1}, z_i[$,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s n_i \theta_i = \sum_{i=1}^s f_i \theta_i$$

A.2) La moyenne arithmétique pondérée

Lorsque l'on calcule une moyenne arithmétique, on accorde implicitement aux n individus sur lesquels on fait les mesures x_1, x_2, \dots, x_n la même importance représentée par le poids $1/n$ attribué à chacun,

Mais si l'on attribue aux n individus des poids respectifs p_1, p_2, \dots, p_n différents, tels que $\forall i, 0 \leq p_i \leq 1$, et $p_1 + p_2 + \dots + p_n = 1$, on définit alors la moyenne arithmétique pondérée par la formule :

$$\bar{x} = \sum_{i=1}^n p_i x_i$$

Avantages et inconvénients de la moyenne arithmétique

La moyenne arithmétique se prête facilement au calcul algébrique. Mais elle est sensible aux valeurs extrêmes de la série qui peuvent être aberrantes.

B) La médiane

On considère la s.s.s. (x_1, x_2, \dots, x_n) .

Objectif

On souhaiterait déterminer un élément de la série tel que 50 % des éléments de la série soient inférieurs à cet élément, et 50 % supérieurs. Lorsque cet élément existe, on l'appelle médiane de la série.

Détermination de la médiane

Deux cas sont à envisager.

Premier Cas : Si la série est présentée sous la forme d'une série à valeurs isolées.

On range les éléments de la série statistique simple (x_1, x_2, \dots, x_n) par ordre croissant, sans grouper les valeurs égales ; on obtient la suite de nombres $(x'_1, x'_2, \dots, x'_n)$ telle que $x'_1 \leq x'_2 \leq \dots \leq x'_n$.

1^{er} cas :

Si $n=2p+1$, on appelle médiane de la série statistique simple (x_1, x_2, \dots, x_n) , et on note $me(x)$, le nombre réel x'_{p+1} .

2^{ème} cas :

Si $n=2p$, on appelle intervalle médian de la série statistique simple (x_1, x_2, \dots, x_n) l'intervalle (x'_p, x'_{p+1}) .

Détermination Graphique

On cherche une classe d'effectifs cumulés $] N_{i-1}, N_i]$ qui contient $p+1$; elle correspond à la classe $] y_{i-1}, y_i]$ et alors $me(x) = y_i$.

Remarque

L'objectif n'est pas toujours réalisé, même approximativement. On applique la définition.

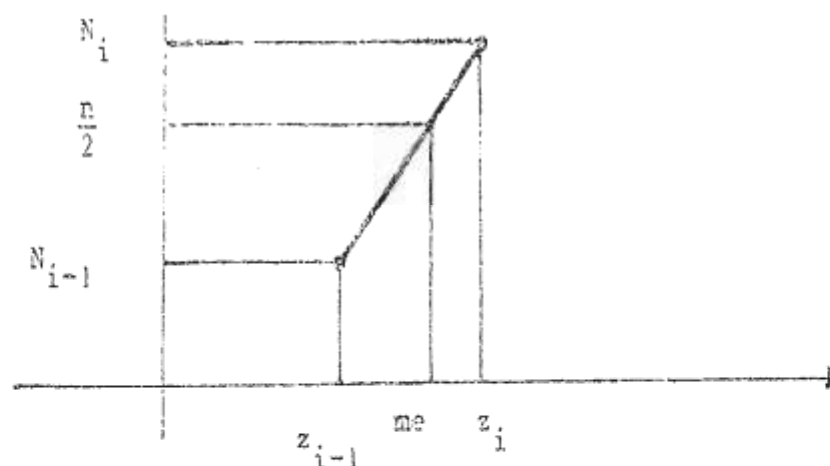
Deuxième Cas : Si la série est présentée sous la forme d'une série à valeurs classées.

Détermination Graphique

on peut déterminer une valeur approchée de la médiane de la façon suivante :

on cherche à quelle classe d'effectifs (respectivement ; fréquences) cumulés appartient la valeur $n/2$ (resp. $1/2$) ; si c'est à l'intervalle $[N_{i-1}, N_i[$

(resp. $[F_{i-1}, F_i[$), alors on détermine une valeur approchée de la médiane par interpolation linéaire à l'intérieur de la classe $[z_{i-1}, z_i[$:



$$me(x) = z_{i-1} + (z_i - z_{i-1}) \frac{\frac{n}{2} - N_{i-1}}{n_i}$$
$$me(x) = z_{i-1} + (z_i - z_{i-1}) \frac{\frac{1}{2} - F_{i-1}}{f_i}$$

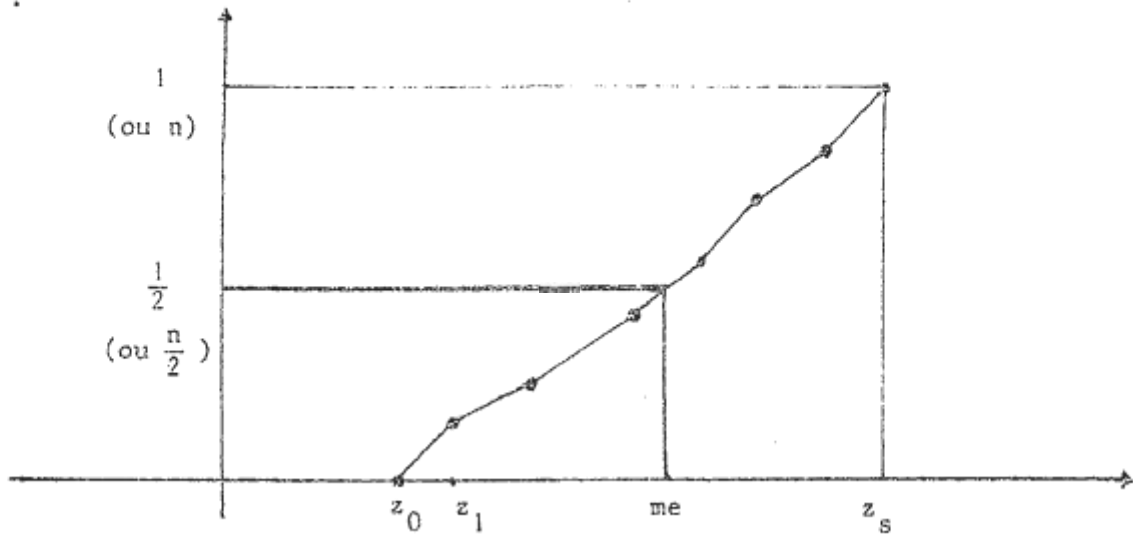
Remarque

Faire une interpolation linéaire revient à supposer que la variation de l'effectif cumulé peut être représentée par un segment de droite, c'est-à-dire que les éléments de la classe à laquelle appartient la médiane sont répartis de façon uniforme dans cette classe.

Recherche graphique

On construit la courbe des fréquences cumulées (ou des effectifs cumulés) :

és) :



C) Les quantiles

On considère la s.s.s. (x_1, x_2, \dots, x_n) .

Objectif

Soit : $0 < p < 1$, on souhaiterait déterminer un élément de la série tel que la proportion p des éléments de la série soient inférieurs à cet élément, et la proportion $(1-p)$ soient supérieurs.

Détermination du quantile d'ordre p

Deux cas sont à envisager.

Premier Cas : Si la série est présentée sous la forme d'une série à valeurs isolées.

On range les éléments de la série statistique simple (x_1, x_2, \dots, x_n) par ordre croissant, sans grouper les valeurs égales ; on obtient la suite de nombres $(x'_1, x'_2, \dots, x'_n)$, telle que $x'_1 \leq x'_2 \leq \dots \leq x'_n$.

1^{er} cas :

Si $n \cdot p \notin \mathbb{N}$, on appelle quantile d'ordre p de la série statistique simple (x_1, x_2, \dots, x_n) , et on note $Q_p(x)$, le nombre réel x'_{np+1} .

2^{ème} cas :

Si $n \cdot p \in \mathbb{N}$, on appelle intervalle quantile d'ordre p de la série statistique simple (x_1, x_2, \dots, x_n) l'intervalle (x'_{np}, x'_{np+1}) .

Détermination Graphique (1^{er} cas)

On cherche une classe d'effectifs cumulés $] N_{i-1}, N_i]$ qui contient $np+1$; elle correspond à la classe $] y_{i-1}, y_i]$ et alors $Q_p(x) = y_i$.

Remarque

L'objectif n'est pas toujours réalisé, même approximativement.

Deuxième Cas : Si la série est présentée sous la forme d'une série à valeurs classées.

Détermination Graphique

On peut déterminer une valeur approchée du quantile d'ordre p de la façon suivante : on cherche à quelle classe d'effectifs (respectivement fréquences) cumulés appartient la valeur $n.p$ (resp. p) ; si c'est à l'intervalle $[N_{i-1}, N_i[$ (resp. $[F_{i-1}, F_i[$) , alors on détermine une valeur approchée de la médiane par interpolation linéaire à l'intérieur de la classe $[z_{i-1}, z_i[$:

$$Q_p(x) = z_{i-1} + (z_i - z_{i-1}) \frac{n.p - N_{i-1}}{n_i}$$

$$Q_p(x) = z_{i-1} + (z_i - z_{i-1}) \frac{p - F_{i-1}}{f_i}$$

D) Mode

On considère la s.s.s. (x_1, x_2, \dots, x_n) .

Objectif

On souhaiterait déterminer un élément de la série qui se répète le plus souvent.

Détermination du mode

Deux cas sont à envisager.

Premier Cas : Si la série est présentée sous la forme d'une série à valeurs isolées.

Définition

On appelle valeur modale ou mode de la série statistique à valeurs isolées (y_1, y_2, \dots, y_r) toute valeur y_i dont l'effectif (resp la fréquence) est supérieure à celui des deux valeurs voisines y_{i-1} et y_{i+1} .

Deuxième Cas : Si la série est présentée sous la forme d'une série à valeurs classées.

Définition

On appelle classe modale de la série statistique à valeurs classées $([z_1, z_2[, [z_2, z_3[, \dots, [z_{s-1}, z_s])$ toute classe $[z_{i-1}, z_i[$ dont l'effectif (resp la fréquence) corrigé(e) est supérieure à celui des deux classes voisines $[z_{i-2}, z_{i-1}[$ et $[z_i, z_{i+1}[$.

Remarques

- 1) La série peut admettre plusieurs modes relatifs.
- 2) La détermination se fait de manière graphique

E) Médiade

On considère la s.s.s. (x_1, x_2, \dots, x_n) . On suppose que le caractère X est cumulatif. Cela veut dire que $(x_1 + x_2 + \dots + x_n)$ a un sens et peut-être interprétée.

Objectif

On cherche un élément de la série qui partage celle-ci en deux sous-séries de même masse totale moitié.

- On définit la masse m_i de la classe $c_i = [z_{i-1}, z_i[$ par la somme des x_j qui sont dans la classe c_i qui peut-être approchée par $n_i * \theta_i$.
- On définit la masse cumulée M_i de la classe c_i par $(m_1 + m_2 + \dots + m_i)$.
- On définit la fréquence de masse r_i de la classe c_i par le rapport m_i / M_t
- On définit la fréquence de masse cumulée R_i de la classe c_i par le rapport $(m_1 + m_2 + \dots + m_i) / M_t$

Détermination Graphique

- On peut déterminer une valeur approchée de la médiane de la façon suivante :
On cherche à quelle classe de masses cumulées (respectivement fréquences de masses cumulées) cumulée appartient la valeur $M_t/2$ (resp. $1/2$) ; si c'est à l'intervalle $[M_{i-1}, M_i[$ (resp. $[R_{i-1}, R_i[$) , alors on détermine une valeur approchée de la médiane par interpolation linéaire à l'intérieur de la classe $[z_{i-1}, z_i[$:

$$Mle(x) = z_{i-1} + (z_i - z_{i-1}) * \frac{\frac{M_t}{2} - M_{i-1}}{m_i}$$

$$Mle(x) = z_{i-1} + (z_i - z_{i-1}) * \frac{\frac{1}{2} - R_{i-1}}{r_i}$$

B.- CARACTERISTIQUES DE DISPERSION.

i) Objectifs

Trois objectifs :

- O_1 : Dispersion des éléments de la série par rapport à une valeur centrale v_c .
- O_2 : Dispersion des éléments de la série dans un intervalle inter-quantile.
- O_3 : Etude de la concentration de la série statistique.

O_1 - Dispersion par rapport à une valeur centrale v_c

La valeur centrale peut être la moyenne \bar{x} , la médiane, le quantile d'ordre α , la médiane, etc...

Pour cela, on définit l'indice de dispersion par rapport à une valeur centrale v_c comme la quantité s_{v_c} définie par la moyenne arithmétique des carrés des écarts des éléments de la série par rapport à v_c

$$s_{vc}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - v_c)^2$$

Sa racine carrée s_{vc} est appelée écart-type de la série par rapport à la valeur centrale v_c .

Si $v_c = x$ alors s_{vc}^2 est appelée la variance de la série et s_{vc} est appelée écart-type de la série.

Parfois on a recours au calcul du coefficient de variation ; défini par le rapport f_{vc}

$$f_{vc} = \frac{s_{vc}}{v_c}$$

Ce coefficient est sans unité et peut facilement être interprété.

O2 - Dispersion dans un intervalle inter-quartile

Soit par exemple à étudier la dispersion des éléments de la série dans l'intervalle inter-quartile $[Q_{1/4}, Q_{3/4}]$, il s'agit de savoir de quel côté se trouve plus d'éléments de la série.

En effet le milieu de cet intervalle est la médiane m_e , on adopte alors la règle de décision suivante :

R.D.D. :

- $$\left\{ \begin{array}{l} * \text{ Si } Q_{3/4} - m_e > m_e - Q_{1/4}, \text{ alors il y a plus d'éléments de la série} \\ \quad \text{à droite de la médiane dans l'intervalle } [Q_{1/4}, Q_{3/4}] \text{ qu'à gauche} \\ * \text{ Si } Q_{1/4} - m_e < m_e - Q_{3/4}, \text{ alors il y a plus d'éléments de la série} \\ \quad \text{à gauche de la médiane dans l'intervalle } [Q_{1/4}, Q_{3/4}] \text{ qu'à droite.} \end{array} \right.$$

Comparaison entre médiane et médiale

- La médiane de la série est un élément qui partage celle-ci en deux sous-séries de même effectif cumulé $n/2$
- La médiale de la série est un élément qui partage celle-ci en deux sous-séries de même masse cumulée $M_t/2$ où M_t est la masse totale de la série.

Pour la comparaison des deux éléments de la série, nous avons à choisir l'une des deux procédures suivantes :

1. Calculer la masse cumulée de la médiane :

$$M_{m_e} = \sum_{i: x_i < m_e} x_i \quad \text{et la comparer à } \frac{M_t}{2}$$

2. Calculer l'effectif ou la fréquence cumulée de la médiale :

$$N_{Mle} = \sum_{i: x_i < Mle} 1 \text{ et la comparer à } \frac{n}{2}$$

Résultat

1. La médiane est toujours inférieure ou égale à la médiale.
2. Ils sont égaux pour une série égalitaire : celle-ci accorde la même importance (même masse) aux classes de poids les plus faibles qu'à celles les plus forts.

O3 : Etude de la concentration de la série statistique

Il s'agit d'étudier la répartition de la série des masses et de voir quelle partie de la série est la plus lourde : est-ce la partie de la série de plus faible poids ou celle de plus fort poids.

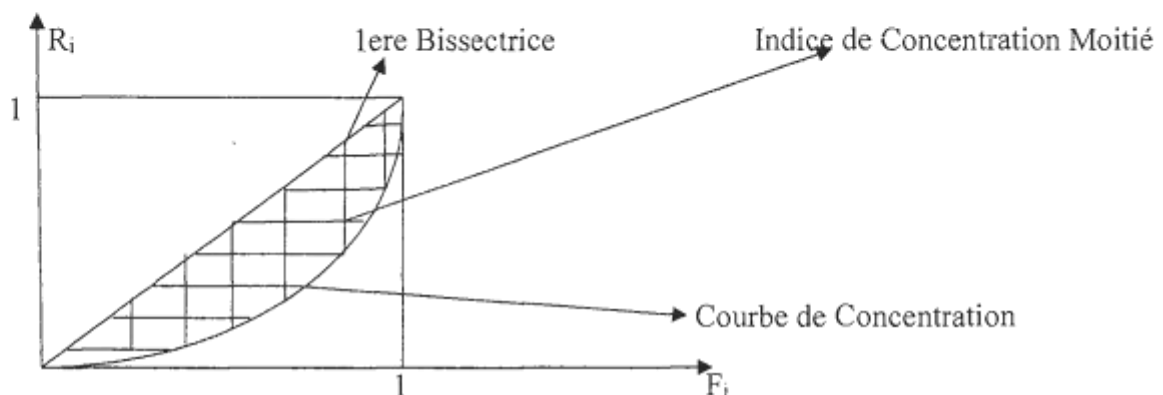
Pour cela, et pour répondre à ce besoin, on définit la courbe de concentration de la série et on calcule l'indice de concentration ou de GINI.

1. COURBE DE CONCENTRATION

Elle définit par l'ensemble des couples

$$\{(F_x, R_x) : x \in I\mathbb{R}\}$$

Elle est située dans le carré de côté 1, au dessous de la première bissectrice $R_x = F_x$



Remarques

1. Pour justifier le tracé précédent, remarque que :
 - la courbe est croissante au sens large
 - utiliser les deux points particuliers médiane et Médiale
 - elle passe par les deux points particuliers (0,0) et (1,1)
2. Nous ne pouvons faire qu'un tracé approché par :

$$\{(F_i, R_i)/i = 1, \dots, s\}$$

2- Indice de concentration

Utilité : Il sert à répondre à la question précédente :

Quelle partie de la série est la plus lourde : est-ce la partie de la série de plus faible poids ou celle de la série de plus fort poids.

Formule de calcul :

$$I_c = \text{Aire de la partie hachurée} = 1 - \sum_{i=0}^s f_i(R_i + R_{i-1})$$

Propriété

1. $0 \leq I_c \leq 1$
2. Si $0 \leq I_c \leq 0.2$ série est proche d'une série égalitaire ; $Mle \approx me$
3. Si $I_c > 0.5$ la série est loin d'une série égalitaire et $Mle > me$

c.- CARACTERISTIQUES DE FORME

Deux Objectifs :

O1.- Déterminer des coefficients qui caractérisent la **forme** de la courbe des fréquences. Cela permettra d'étudier la symétrie ou l'asymétrie de la distribution des effectifs ou des fréquences et de voir quel est le côté de la série droit ou gauche qui contient le plus grand nombre d'éléments de la série.

O2.- Etudier l'aplatissement de la série pour tester sa normalité

A/ Introduction : les moments

Soit q un entier naturel.

Définitions

d1.- On appelle moment d'ordre q de la s.s.s. (x_1, x_2, \dots, x_n) et on note $m_q(x)$, la quantité :

$$m_q(x) = \frac{1}{n} \sum_{i=1}^n x_i^q$$

d2.- On appelle moment centré d'ordre q de la s.s.s. (x_1, x_2, \dots, x_n) et on note $\mu_q(x)$, la quantité :

$$\mu_q(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^q$$

Remarque – Si $m_1(x) = \bar{x}$; $\mu_1(x) = 0$; $\mu_2(x) = s^2$

B/ Les Coefficients d'asymétrie

D3. On dit qu'une s.s.s. (x_1, x_2, \dots, x_n) est symétrique s'il existe un nombre réel **a** tel que les éléments de la série se correspondent deux à deux par symétrie par rapport à **a** (i.e même effectif ou même fréquence).

Remarque – Dans la pratique, il n'existe pas de série symétrique ; mais une série peut se rapprocher plus ou moins d'une série symétrique.

B1.- Coefficient d'asymétrie de YULE : on appelle Coefficient d'asymétrie de **YULE** la quantité :

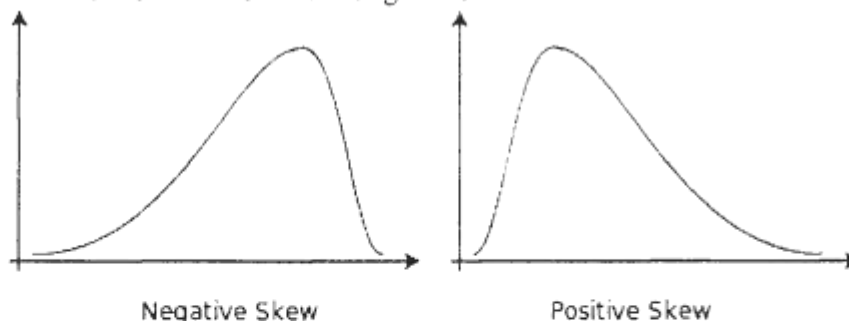
$$c(x) = \frac{Q_{3/4} - me(x) - me(x) - Q_{1/4}}{Q_{3/4} - Q_{1/4}}$$

B1.- Coefficient d'asymétrie de FISHER : on appelle Coefficient d'asymétrie de **FISHER** la quantité :

$$\gamma_1(x) = \frac{\mu_3(x)}{s_x^3(x)}$$

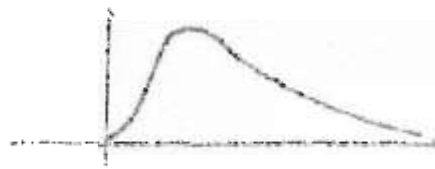
Forme de la distribution

- Un coefficient **positif** indique une distribution décalée à gauche de la médiane, et donc une queue de distribution étalée vers la droite.
- Un coefficient **négatif** indique une distribution décalée à droite de la médiane, et donc une queue de distribution étalée vers la gauche.



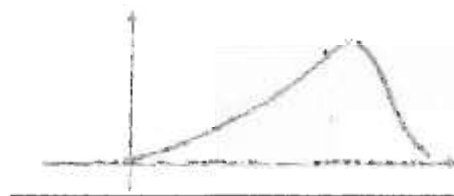
Par exemple :

Cas 1



La courbe des fréquences est oblique à gauche et on a : $c(x) > 0$ et $\gamma_1(x) > 0$.

Cas 2



La courbe des fréquences est oblique à droite et on a : $c(x) < 0$ et $\gamma_1(x) < 0$.

Cas 3

La courbe des fréquences est symétrique par rapport à la médiane ou le mode et on a :

$c(x) = 0$ et $\gamma_1(x) = 0$.

C.-Aplatissement : Coefficient d'aplatissement β_2 de PEARSON

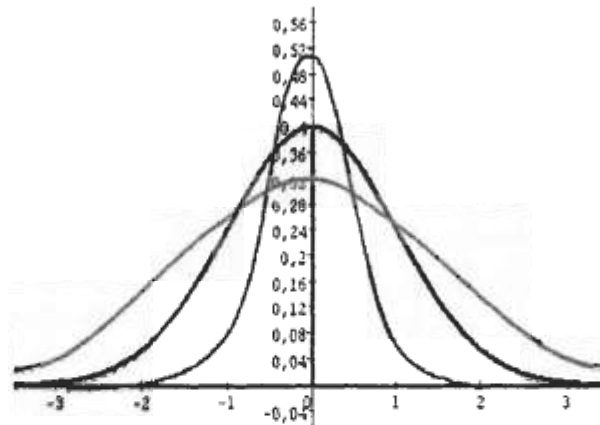
Le coefficient d'aplatissement d'une distribution mesure si elle est "pointue" (recentrée autour de sa moyenne) ou au contraire étalée. Il est défini par :

$$\beta_2(x) = \frac{\mu_4(x)}{s_x^4(x)}$$

appelée Coefficient d'aplatissement β_2 de PEARSON

où μ_i est le i -ème moment de la variable aléatoire. Le coefficient est nul pour une loi normale, négatif pour une distribution étalée (on dit parfois plasticurtique), positif pour une distribution pointue (on dit parfois leptocurtique). En statistique, un coefficient négatif s'interprète par une population peu homogène, un coefficient positif par une population homogène.

Appelée aussi **kurtosis** du grec « courbure » de la série



La kurtosis d'une loi normale (de Gauss) est égale à 3 ; ce qui conforte la loi normale dans son rôle de « loi étalon ».

Un coefficient β_2 de PEARSON positif traduit une distribution leptokurtique (distribution qui s'élève assez haut puis retombe assez brutalement).

Une distribution normale est **mésokurtique** (celle entre les deux) et une distribution à « queues épaisses », dont le coefficient est négatif.