Module : Statistique descriptive et Probabilité.

Programme :

Chap 1. Notions fondamentales et distributions statistiques à une dimension.

Chap 2. Caractéristique de tendance centrale.

Chap 3. Caractéristique de dispersion et de forme.

Chap 4. Analyse Combinatoire et calcul des probabilités.

J. Fourastier Dunod, Stat et Proba Tom 6. M336 Tome 1 et 2

CHAP1 : Notions fondamentales et distributions statistiques à une dimension.

1. La statistique et les statistiques :

Au singulier, le mot "statistique" désigne un ensemble de techniques d'interprétation mathématique qui permettent d'élaborer des modèles probabilistes à partir de données réelles. Ces techniques visent à rassembler, à décrire et à analyser des données.

Au pluriel les "statistiques" représentent les données se rapportant à des faits qu'on souhaite observer (Statistiques sociales, économiques, ... etc).

La statistique descriptive est une branche de la statistique qui développe des méthodes scientifiques pour recueillir, représenter et traiter les données observées.

2. Démarche à suivre :

Toute étude statistique passe impérativement par trois étapes :

1ère étape : Collecte des données

2ème étape : Etablissement de ~~données~~ tableaux et représentations graphiques

3ème étape : ~~Sommaire~~ Utilisation et Calculer du paquet de données pour quelques caractéristiques ou valeurs typiques qui synthétisent au mieux l'ensemble (de la tendance centrale, de la dispersion, ..., e.t.c).

3. Collecte des données :

Dans le cadre d'une étude particulière, le recueil des

informations statistiques peut se faire de différentes façons. Il peut-être issu d'un relevé exaustif concernant la population à étudier : On parle alors de "recensement" (tous les individus de la population sont examinés).
Il peut-être issu d'un "sondage" : Seul un échantillon de cette population est questionné (En pratique ; c'est la situation la plus courante).

I. Vocabulaire élémentaire de la statistique :

- **Population et individu (Univers et unité statistique) :**
  - On appelle unité ou individu, (noté ω), tout élément ou objet examiné dans le cadre d'une étude statistique.
  - On appelle population statistique, notée Ω, toute collection d'unités statistiques élémentaires constituant les objets à étudier.

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

  - On appelle échantillon statistique tout ensemble d'individus prélevé dans une population. En termes ensemblistes, un échantillon est une partie de Ω.
- **Variable (ou caractère) statistique :**
  - Notion de variable statistique :

On appelle variable (ou caractère) statistique une application (ou relation) définie sur une population statistique étudiée notée Ω et à valeurs dans un ensemble M, appelé ensemble des modalités (chaque individu a une unique modalité). L'ensemble des observations élémentaires est l'ensemble des modalités. Cette relation entre la population Ω et l'ensemble des modalités peut-être formalisée ainsi :

variable $X : \Omega \longrightarrow M$

$\omega_i \longmapsto X(\omega_i) = m_i$

(individu) (observation) ou modalité.

- *Type de variable statistique :*

Les variables statistiques peuvent être classées selon leur type (ou nature).

a) *Variable Qualitative :* Une variable statistique est dite qualitative lorsque l'ensemble M des modalités n'est pas un ensemble de nombres. On dit alors que la variable (ou le caractère) n'est pas mesurable. Les observations ne sont pas numériques.

- Une variable qualitative est dite "*Ordinale*" lorsque l'ensemble des modalités peut être ordonné (*exple* : cadre supérieur, cadre moyen, ..., e.t.c).

- Une variable qualitative est dite "*nominale*" lorsque l'ensemble des modalités ne peut être ordonné (*exple* : célibataire, marié, ... e.t.c).

b/ *Variable quantitative :* Une variable statistique est dite "*Quantitative*" (ou numérique) lorsque l'ensemble des modalités est un ensemble de nombres.

- Une variable quantitative est dite "*discrète*" (ou discontinue) lorsque seul un nombre fini ou dénombrable de valeurs peuvent être prises.
- Une variable quantitative est dite Continue lorsque toutes les valeurs prises sont dans un intervalle réel.

c. Passage d'un type de variable à l'autre :

Il est possible de quantifier les modalités d'1 variable qualitative par un système de codage (exple : Niveau de formation : niveau I, niveau II, ..., e.t.c).

A l'inverse, on peut également passer d'1 variable quantitative à une variable qualitative par un système de regroupement.

5° Distribution d'1 variable statistique :

Soit X une variable statistique qualitative ou quantitative, et soient $x_1, x_2, \dots, x_k$ les différentes modalités de la variable X.

a/ Effectifs (ou fréquences absolues) :

- On appelle "effectif" d'1 modalité donnée x_i de la variable X, et on note n_i le nombre de fois où cette modalité est observée.

- On appelle effectif total de la population (ou taille de la population), on note "n" la somme des "k" effectifs particuliers "n_i" correspondant à chaque modalité soit :

$$n = n_1 + n_2 + \dots + n_k = \sum_{i=1}^{k} n_i$$

- On appelle distribution "statistique" d'une variable (ou bien Série statistique univariée) l'ensemble des k couples $\{(x_i, n_i) \; ; \; i = 1, 2, \dots, k\}$.

b/ Fréquences :

On appelle fréquence de la modalité x_i de

la ~~variable X~~ modalité x_i ~~de la~~ variable X, et on la note f_i le rapport de l'effectif particulier n_i correspondant à la modalité x_i et de l'effectif n de la population observée, soit :

$$f_i\,(\%) = \frac{n_i}{n}\,(\cdot 100\%)$$

Ce rapport permet de connaître la ~~proportion~~ proportion d'individus présentant une modalité particulière de la variable par rapport à l'ensemble des individus.

On a toujours : $0 \leq f_i \leq 1$ et $\sum_{i=1}^{k} f_i = 1$

6°. Organisation des données dans des tableaux statistiques :

- En pratique, les données ainsi que les résultats numériques relatifs à leurs traitement statistique, sont présentés sous forme de tableaux.
La forme de ces tableaux est variable, mais un certain nombre de principes fondamentaux déterminent leur construction.
- On appelle tableau élémentaire (le tableau des données brutes) le tableau qui précise, pour chaque individu de la population la modalité correspondante de la variable étudiée.
- On appelle "tri à plat" l'opération de comptage des individus par modalité.
- On appelle tableau de "dénombrement" le tableau à deux colonnes x_i , n_i

- On appelle tableau de "traitement" le tableau ~~d~~ obtenu en complétant le tableau de dénombrement par les résultats relatifs aux différents calculs statistiques, la forme d'1 tableau de traitement peut différer sensiblement selon la nature des modalités.

Exples:

1°/ Variable Qualitative:

Situation familiale	Effectifs n_i	Fréquences f_i
x_1: Célibataire.	$n_1 = 35$	$f_1 = 0,35$ ou 35%
x_2: Divorcé.	$n_2 = 20$	$f_2 = 0,20$
x_3: Marié.	$n_3 = 30$	$f_3 = 0,30$
x_4: Veuf.	$n_4 = 15$	$f_4 = 0,15$
Total: $\sum$	$n = 100$	1

2°/ Variable Quantitative discrète:

N^{bre} d'enfants x_i	Effectifs n_i	Fréquences f_i
$x_1 = 0$	$n_1 = 10$	$f_1 = 0,2$
$x_2 = 1$	$n_2 = 20$	$f_2 = 0,4$
$x_3 = 2$	$n_3 = 05$	$f_3 = 0,1$
$x_4 = 3$	$n_4 = 15$	$f_4 = 0,3$
Total $\sum$	$n = 50$	1

3°.

3/ Variable Quantitative Continue :

e_{i-1}	$[e_{i-1}, e_i[$ classe	Centre de la classe x_i	n_i Effectifs	f_i Fréquences
0				
	$[0, 10[$	5	15	0,15
10				
	$[10, 40[$	25	30	0,3
40				
	$[40, 60[$	50	25	0,25
60				
	$[60, 100[$	80	30	0,3.
100				
	Σ		$n=100$	1

- e_{i-1} : l'extrémité inférieure de la classe.
- x_i : Centre de la classe $x_i = \frac{e_{i-1} + e_i}{2}$
- a_i = Amplitude de la classe (longueur) $a_i = e_i - e_{i-1}$

7°. Effectifs et fréquences cumulés :

On suppose que la variable étudiée est quantitative et que les différentes modalités sont ordonnées par ordre croissant.

a. Effectifs cumulés croissants et décroissants

1°/ Cas de variable statistique discrète :

On appelle effectif cumulé croissant jusqu'à la valeur x_i de la variable X, et on le note N_i la somme des effectifs associés aux valeurs inférieures ou égales à x_i.

2°/ Cas de variable statistique continue :

On appelle effectif cumulé croissant jusqu'à l'extrémité de classe $[e_{i-1}, e_i[$ et on note N_i la somme des effectifs associés à la classe $[e_{i-1}, e_i[$ et des classes précédentes.

On a : $N_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^{i} n_k$

L'effectif décroissant cumulé $N_i' = \sum_{k=i+1}^{k} n_k = n - N_i$.

b/ Fréquences Cumulées croissantes et décroissantes :

Les effectifs Cumulés croissants N_i permettent de connaître le nombre d'individus présentant une valeur de la variable inférieure ou égale à une valeur observée (ou à une extrémité de classe donnée).

Les fréquences cumulées croissantes jusqu'à une valeur observée i de la variable X (ou une extrémité d'1 classe donnée) est :

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{k=1}^{i} f_k$$

Pour les fréquences cumulées décroissantes F_i' :

$$F_i' = \sum_{j=i+1}^{k} f_j = 1 - F_i \quad \text{et} \quad F_i' = \frac{N_i'}{n}$$

Exples

Nbre d'enfants x_i	n_i	N_i	N_i'	f_i	F_i	F_i'
0	10	0	40	0,25	0	1
1	16	10	30	0,4	0,25	0,75
2	7	26	14	0,175	0,65	0,35
3	3	33	7	0,075	0,825	0,175
4	3	36	4	0,075	0,9	0,1
5	0	39	1	0	0,975	0,025
6	1	39	1	0,025	0,975	0,025
		40	0		1	0
$\sum$	40	–	–	1	–	–

3°. Les représentations graphiques : L'objet de la représentation graphique est de représenter l'ensemble des observations en

fonction de la nature de la variable. Il existe plusieurs représentations.

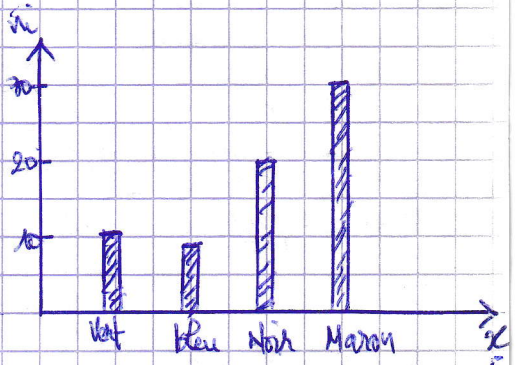
a) Variable qualitative :

1°) Diagramme en tuyaux d'orgue (ou à bandes) :

C'est un graphique qui à chaque modalité x_i associe un rectangle de base constante et de hauteur proportionnelle à l'effectif n_i ou f_i

Exple :

Couleur des yeux	Vert	bleu	Noir	Maron
Effectif n_i	10	8	20	30



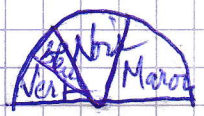
2°) Diagrammes à secteurs (Circulaires ou Semi-circulaires) :

C'est un graphique qui divise un cercle (ou un demi-cercle) en des portions angulaires dont les mesures sont proportionnelles aux effectifs n_i ou f_i. Pour une modalité x_i d'effectif n_i ou f_i, l'angle α_i correspondant est donné par :

$\alpha_i = \frac{n_i}{n} \times 360^\circ = f_i \times 360^\circ = f_i \times 2\pi$ pour le cercle.



$\alpha_i = \frac{n_i}{n} \times 180^\circ = f_i \times 180^\circ = f_i \times \pi$ pour le demi-cercle.



Remarque : En général, les représentations graphiques utilisées par des variables qualitatives restent valables pour des variables quantitatives.

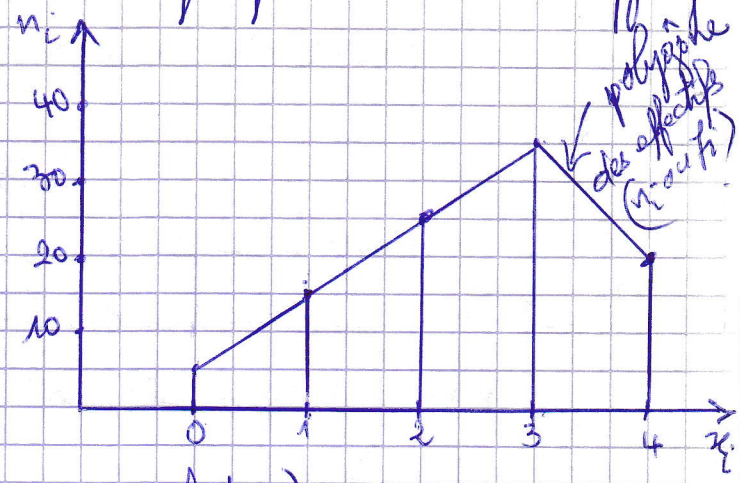
b) Variable quantitative discrète :

1°) **Diagramme en bâtons :** On associe à chaque modalité un

segment de droite dont la hauteur est proportionnelle à l'effectif n_i ou f_i.

Exple :

Nbre d'enfants x_i	0	1	2	3	4
n_i	5	15	25	35	20



2°) Diagramme Cumulatif (Courbe cumulative) :

Fonction de répartition :

On appelle fct de répartition d'1 variable statistique X, qu'on note par F_X, l'application de l'ensemble $\mathbb{R}$ dans $[0,1]$ qui pour une valeur observée x de X précise la proportion des individus de la population dont le caractère est $\leq x$

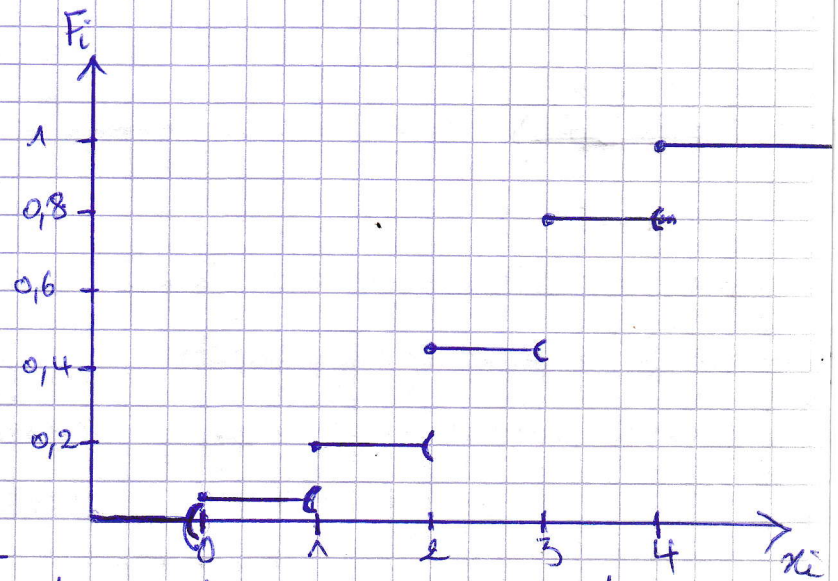
$$F_X : \mathbb{R} \longrightarrow [0,1]$$

$$x \longmapsto F_X(x) = \underbrace{P[X \leq x]}_{\text{Proportion}} \quad \text{avec } F_X(+\infty) = 1$$

et $F(-\infty) = 0$

Exple :

x_i	n_i	f_i	F_i
0	5	0,05	0
1	15	0,15	0,05
2	25	0,25	0,2
3	35	0,35	0,45
4	20	0,20	0,80
			1
$\sum$	n=100	1	



La courbe des F_i est dite courbe Cumulative (en escaliers).

c/ Variable quantitative Continue :

1°) **Histogramme :** On appelle histogramme un diagramme de rectangles contigus d'aires proportionnelles aux n_i ou f_i et de bases déterminées par les 2 extrémités de chaque classe $[e_{i-1}, e_i[$. L'histogramme est la surface de l'ensemble des rectangles.

Exple :

La classe $[e_{i-1}, e_i[$	Effectif n_i	Fréquence f_i	$F_i \uparrow$	$F_i \downarrow$
			0	1
$[0,5[$	4	0,4	0,4	0,6
$[5,10[$	2	0,2	0,6	0,4
$[10,15[$	3	0,3	0,9	0,1
$[15,20[$	1	0,1	1	0
Σ	$n=10$	1		



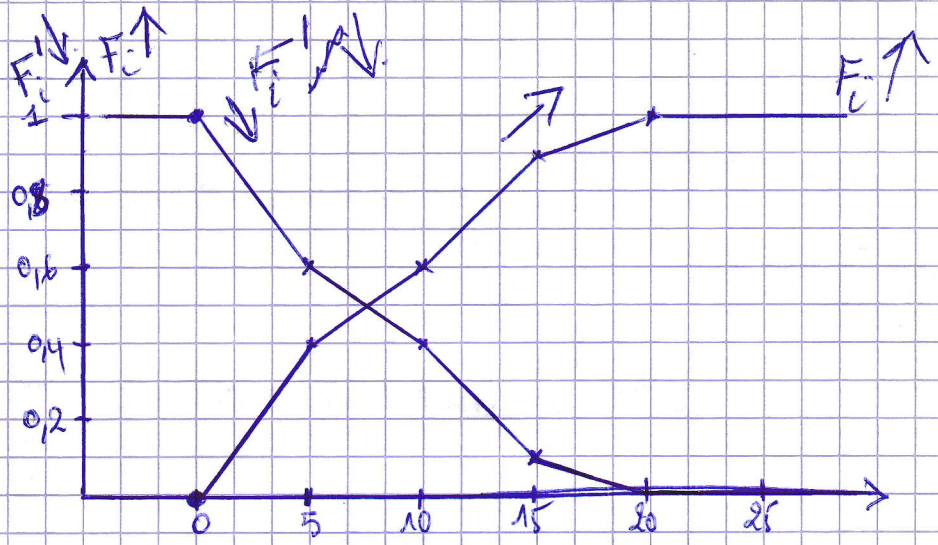
Le polygone des effectifs n_i (ou des fréquences f_i) :

C'est celui qui s'obtient en joignant le milieu de chaque segment d'un rectangle avec le milieu du segment de rectangle suivant. Souvent On ajoute deux fausses classes de même effectif n_i (nulles) aux deux extrémités de l'histogramme pour fermer le polygone et pour rendre égale la surface contenue à l'intérieur du polygone avec celle de l'histogramme.

2°) **Courbe cumulative :** On appelle courbe cumulative croissante la représentation graphique de la fonction de répartition de la variable X. Cette courbe.

Cumulative croissante est la ligne polygonale joignant dans un système d'axes orthogonaux les pts d'abscisses e_{i-1} (extremité inferieure) de la classe et d'ordonnée F_i.

Exple :



Chap. 2: Les caractéristiques de tendance centrale, de dispersion et de forme.

2.0 Introduction:

Le 1er chapitre nous a permis de comprendre comment on "rangeait" les données et comment on pouvait "les représenter".

Grâce aux tableaux et aux graphes, on obtient une première "image" de la distribution des données. Cependant, cette image dépend de l'interprétation faite par l'utilisateur. On cherche donc à résumer les informations par des indicateurs numériques bien choisis.

L'objet de ce chapitre est de présenter quelques uns de ces indicateurs, qu'on appelle caractéristiques de tendance centrale (de position).

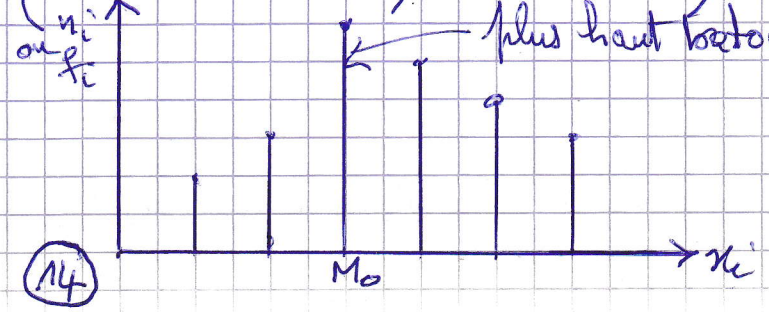
2.1. Le mode (valeur dominante):

a. Cas d'1 variable quantitative discrète:

Déf°: On appelle "mode" d'1 variable quantitative discrète, et on le note "M_0" la modalité la plus fréquente (c.à.d. correspondant à l'effectif maximal).

Rq°: Il peut arriver que l'effectif maximal se répète deux fois (ou plusieurs fois) pour différentes valeurs de la variable X. On parle alors de distribution à 2 modes (bimodale) ou à plusieurs modes (multimodale ou pluri-modale).

Exple°:



b. Cas d'1 variable quantitative continue :

Déf : On appelle "mode" de la variable statistique continue X, on le note Mo la valeur de la variable associée à l'effectif (n_i ou f_i) le plus élevé par unité d'amplitude.

Détermination : La détermination est faite à partir du tableau de dénombrement. Elle s'effectue en deux temps :

1°) Détermination de la classe modale.

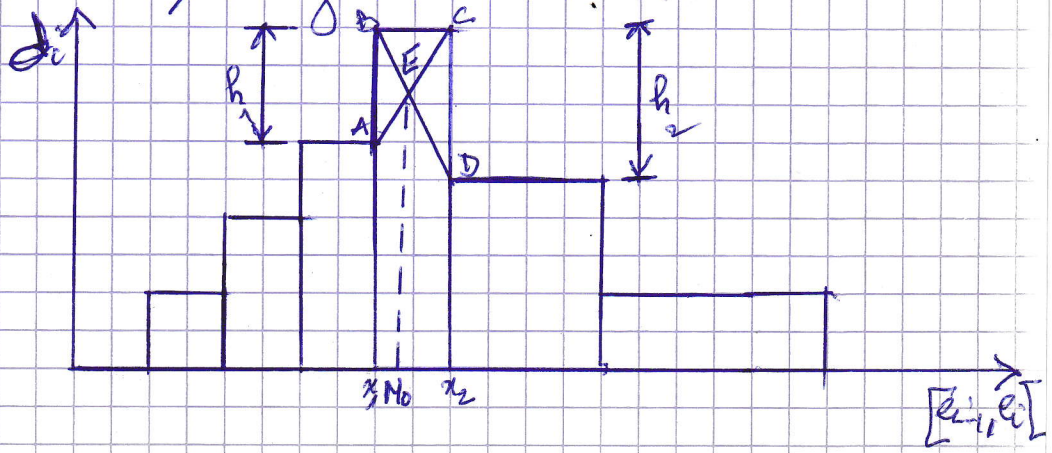
2°) Détermination du mode (valeur particulière de cette classe modale)

1°) **classe modale :** On appelle classe modale d'1 série statistique continue X, la classe associée à la densité maximale (densité notée $d_i = \frac{n_i \text{ ou } f_i}{a_i}$ avec a_i l'amplitude de la classe $[e_{i-1}, e_i[$, elle représente le nbre d'individus (effectif) par unité d'amplitude).

En pratique, on distingue deux cas :

- **Cas général classes d'amplitudes inégales :**

La détermination de la classe modale passe alors nécessairement par le calcul des densités. La classe modale correspond à la classe associée à la plus grande densité.



- Le cas particulier : Classes de mêmes amplitudes :

La classe modale est la classe d'effectif (n_i ou f_i) maximal.

Détermination du mode :

Le mode est situé dans la classe modale, une première estimation peut donc être obtenue en prenant le centre de la classe modale.

Une bonne estimation du mode peut être obtenue à partir de l'histogramme graphiquement ou bien par calcul

Graphiquement voir le graphe de la page précédente

Par calcul :

$[x_1, x_2[$: classe modale

$$M_0 \text{ (L'abscisse du pt E)} = \frac{x_1 \cdot h_2 + x_2 h_1}{h_1 + h_2}$$

avec : $h_1 = d_{i_{max}} - d_{i-1}$

$h_2 = d_{i_{max}} - d_{i+1}$

2.2. Les moyennes :

a. La moyenne arithmétique : On suppose dans cette section que X est une variable quantitative définie sur une population composée de n individus.

- Données non groupées :

Soient $x_1, x_2, \dots, x_n$ les n observations de la variable X.

On appelle moyenne arithmétique simple des n valeurs $\{x_1, x_2, \dots, x_n\}$ ou encore moyenne arithmétique de la variable X et on la note "$\overline{x}$" le rapport défini par :

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{\sum_{i=1}^{n} x_i}{n}.$$

- Données groupées :

On appelle moyenne arithmétique pondérée des k valeurs $\{x_1, x_2, \dots, x_k\}$ affectées des k coefficients (de pondération) $\{n_1, n_2, \dots, n_k\}$ ou encore moyenne arithmétique de la variable X, et on note "$\bar{x}$" le rapport défini par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n_1 + n_2 + \dots + n_k}$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{k} n_i x_i$$

$$= \sum_{i=1}^{k} f_i \cdot x_i$$

Remarque : "x_i" désignent les centres des classes si X est une variable continue

b/ Moyenne géométrique G :

$$G = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \dots x_k^{n_k}} = \prod_{i=1}^{k} x_i^{f_i}$$

c/ Moyenne Harmonique : H.

$$H = \frac{1}{\frac{1}{n}\left(\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_k}{x_k}\right)} = \frac{1}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

d/ Moyenne quadratique Q :

$$Q = \sqrt{\frac{1}{n} \sum_{i=1}^{k} n_i x_i^2} = \sqrt{\sum_{i=1}^{k} f_i x_i^2} = \left(\frac{1}{n} \sum_{i=1}^{n} n_i x_i^2\right)^{1/2}$$

Remarque : on a : $x_{Min} \leq H \leq G \leq \bar{x} \leq Q \leq x_{Max}$

2.3 La médiane :

Définition : On appelle médiane d'1 distribution, et on note "Me" la valeur de la variable partageant les observations classées par ordre croissant en deux groupes de même effectif.

1. Cas d'1 variable discrète :

a. Données non groupées :

Pour déterminer la médiane, on ordonne, dans un premier temps, les valeurs observées de la plus petite à la plus grande :

- Lorsque le nbre d'observations n est impair, la médiane "Me" est la valeur de la série ordonnée située à la position $\frac{n+1}{2}$.
- Lorsque le nombre d'observations n est pair, la médiane "Me" se trouve à l'intérieur de l'intervalle médian compris entre les deux valeurs centrales situées aux positions : $\frac{n}{2}$ et $\frac{n}{2}+1$. Il est d'usage de choisir pour médiane le centre de l'intervalle médian.

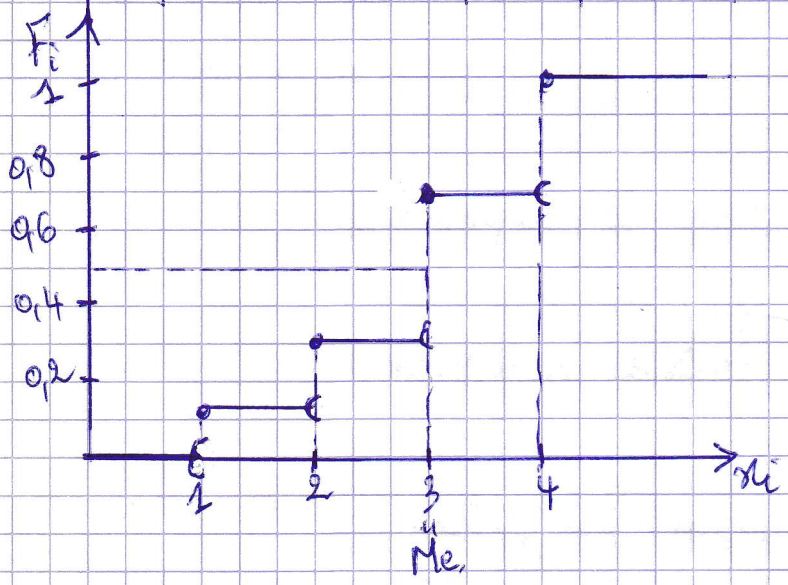
b/ Données groupées :

La détermination de la médiane "Me" peut-être faite

à partir des effectifs cumulés (N_i) soit à partir des fréquences cumulées (F_i). Elle peut-être calculée ou tirée à partir du graphe ou diagramme cumulatif.

Exple :

	x_i	n_i	N_i	f_i	F_i
	1	14	14	0,14	0,14
	2	18	32	0,18	0,32
Me	3	37	69 $\frac{n}{2}=50$	0,37	0,69 $-0,5$
	4	31	100	0,31	1
	$\sum$	100	/	1	



2. Cas d'1 variable Continue :

La détermination se fait à partir des N_i ou F_i. La Me peut-être obtenue par calcul ou graphiquement à partir du diagramme cummulatif.

Exple :

e_{i-1}	$[e_{i-1}, e_i[$	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
10				0	0
	$[10,20[$	15	0,15		
20				15	0,15
	$[20,30[$	20	0,2		
30				35	0,35
	$[30,40[$	45	0,45	80	0,80
40					
	$[40,50[$	20	0,20		
50				100	1
—	$\sum$	100			

classe médiane



Par le calcul :

$$Me = a + \left[(b-a) \cdot \frac{0,5 - F(a)}{F(b) - F(a)}\right]$$

($0,5$: $F(Me)$)

$$= 30 + \left[10 \cdot \frac{0,5 - 0,35}{0,8 - 0,35}\right] = \boxed{33,333}$$

$[a,b[$ c'est la classe médiane.

$F(a)$ et $F(b)$: les fréquences cumulés croissantes associés à $[a,b[$.

Ou bien : $Me = a + \left[(b-a) \cdot \frac{n/2 - N(a)}{N(b) - N(a)}\right]$

$$= 30 + \left[10 \cdot \frac{50 - 35}{80 - 35}\right] = \boxed{33,333}$$

20

2.4 Les quantiles :

a. Définition : On appelle quantiles d'ordre α de la variable statistique X, et on le note "x_α" la valeur de la variable x telle que α % des valeurs observées soient inférieures ou égales à x_α.

Si F est la f^{et} de répartition associée à X, on a $F_X(x_\alpha) = \alpha\ \%$.

b. Principaux Quantiles :

1. Les Quartiles : les quartiles sont les trois valeurs x_{25}, x_{50} et x_{75} de la variable X, qui partagent la population en 4 parties contenant chacune 25% des effectifs.

Les Quartiles sont notés Q_1, Q_2 et Q_3 avec $Q_1 = x_{(25)}$, $Q_2 = x_{(50)}$ et $Q_3 = x_{75}$. Le 2ème quartile est la médiane.

Q_1, Q_2 et Q_3 sont associés respectivement aux fréquences cumulés 25%, 50% et 75%

2. Les déciles : Les déciles, notés $D_1, D_2, \dots, D_9$ sont les 9 valeurs de la variable X qui

partagent la population en dix parties contenant chacune 10% des effectifs.

Aux neuf déciles $D_1, D_2, \ldots, D_9$ sont associés, dans l'ordre les fréquences cumulées 10%, 20%, ..., 90%.

3. Les centiles :

Les centiles, notés $C_1, C_2, \ldots, C_{99}$ sont les 99 valeurs de la variable X qui partagent la population en 100 parties contenant chacune 1% des effectifs.

Aux 99 centiles $C_1, C_2, \ldots, C_{99}$ sont associés les fréquences cumulés 1%, 2%, ..., 99%.

C. Détermination d'un quantile d'ordre α :

La méthode est celle utilisée pour déterminer la médiane. En particulier, lorsque la variable est continue, les quantiles sont estimés à partir de la courbe des F_i ou calculés par la formule déjà donnée.

Chapitre 3. Les caractéristiques de dispersion.

L'objet de ce chapitre est de présenter quelques "indicateurs d'écart" permettant d'évaluer la dispersion des observations autour des valeurs centrales.

I/ Ecarts élémentaires :

On suppose dans cette section que X est une variable quantitative définie sur une population composée de n individus.

Soient $x_1, x_2, \dots, x_n$ les n observations de la variable X.

a. Etendue : $e = x_{max} - x_{min}$.

b. Ecarts interquartiles : b.1 Intervalle interquartiles :

- On appelle "intervalle interquartiles" l'ensemble des valeurs de la variable X situées entre le 1er quartile Q_1 et le dernier quartile Q_3. $I_Q = [Q_1, Q_3]$
- On appelle "Intervalle interdécile", l'ensemble

des valeurs de la variable X situées entre le 1er décile D_1 et le dernier décile D_9.

$$I_D = [D_1, D_9].$$

- On appelle "Intervalle intercentile", l'ensemble des valeurs de la variable X situées entre le 1er centile C_1 et le dernier centile C_{99}.

$$I_C = [C_1, C_{99}].$$

b.2. Ecarts interquantiles:

C'est la longueur des intervalles interquantiles.

. Ecart interquartile:

noté $\boxed{e_Q = Q_3 - Q_1}$ Il concerne 50% des observations.

. Ecart interdécile: noté $\boxed{e_D = D_9 - D_1}$.

Il concerne 80% des observations.

. Ecart intercentile: noté $e_C = C_{99} - C_1$.

Il concerne 98% des observations.

C. Ecarts arithmétiques:

1. Ecart absolu moyen par rapport à la moyenne :

On appelle ecart absolu moyen (ou écart arithmétique) par rapport à la moyenne arithmétique d'1 variable statistique X, noté "$e_{\bar{x}}$" le nombre defini par :

$$\boxed{e_{\bar{x}} = \frac{1}{n} \sum_{i=1}^{k} n_i \cdot |x_i - \bar{x}|}$$

2. Ecart absolu moyen par rapport à la médiane :

$$\boxed{e_{Me} = \frac{1}{n} \sum_{i=1}^{k} n_i \cdot |x_i - Me|}$$

Remarque : $\boxed{e_{Me} \leqslant e_{\bar{x}}}$.

2°/ Variance et écart-type :

On suppose dans cette section que X est une variable quantitative définie sur une population composée de n individus.

Soient $x_1, x_2, \dots, x_n$ les n observations de la variable X.

a) Variance :

Les données non groupées :

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \left[\frac{1}{n}\sum_{i=1}^{n}x_i^2\right] - \bar{x}^2$$

(formule développée de la variance.)

Les données groupées :

$$V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i-\bar{x})^2 = \sum_{i=1}^{k} f_i (x_i-\bar{x})^2$$

$$= \left[\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2\right] - \bar{x}^2$$

$$= \mathcal{Q}^2 - \bar{x}^2$$

On a donc : La variance = Moyenne des carrés des écarts à la moyenne.

Remarques

- La moyenne arithmétique est la valeur de c qui minimise $\frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2$, par suite on a : $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \leq \frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2 \quad \forall c \in \mathbb{R}$.

- La variance est définie à partir des carrés des écarts à la moyenne, il en résulte qu'elle ne s'exprime pas dans la même unité que la variable X. Ainsi si X est exprimée en (DA), Var(X) sera exprimé en

$(DA)^2$, unité qui ne correspond à aucune mesure courante, raison pour laquelle on utilise à sa place, sa racine carrée, l'écart-type.

b/ Ecart-type (Ecart quadratique moyen) :

On appelle écart-type de la variable statistique X, noté $\sigma(X)$ (Sigma de X) ou σ_X la racine carrée de la variance $\sigma(X) = \sqrt{V(X)}$.

L'écart-type est homogène par rapport à la variable (même unité de mesure). Il est d'autant plus grand que la dispersion des observations autour de la moyenne

3°/ Notion de dispersion relative :

Coefficient de variation :

On appelle coefficient de variation de la distribution de X, et on note $Cv(X)$ le rapport défini par : $\boxed{Cv(X) = \frac{\sigma(X)}{\bar{x}}}$

Comme il peut être donné en pourcentage, il sera appelé "indice".

Souvent un coefficient de variation $\leq 15\%$

indique une bonne homogénéité de la distribution.

Le Coefficient est utilisé pour comparer la dispersion de plusieurs distributions.

Si A et B sont 2 distributions et si $CV_A \leq CV_B$ alors A est plus homogène que B. (meilleure)

II Les caractéristiques de forme : Asymétrie et Aplatissement :

Introduction : La courbe de distribution des fréquences (n_i, f_i ou d_i) est la représentation graphique des fréquences (n_i, f_i ou d_i) en fct des valeurs de la variable X (polygone).

Cette Courbe caractérise la forme d'une distribution.

La forme de cette courbe (ou forme de la distribution) peut-être décrite de manière satisfaisante par deux caractéristiques essentielles :

- L'assymétrie qui dépend du nombre de valeurs de la variable situées de part et d'autre d'une valeur centrale (mesure

par les coefficients d'asymetrie).

- L'aplatissement qui est lié à la dispersion plus ou moins importante des observations.

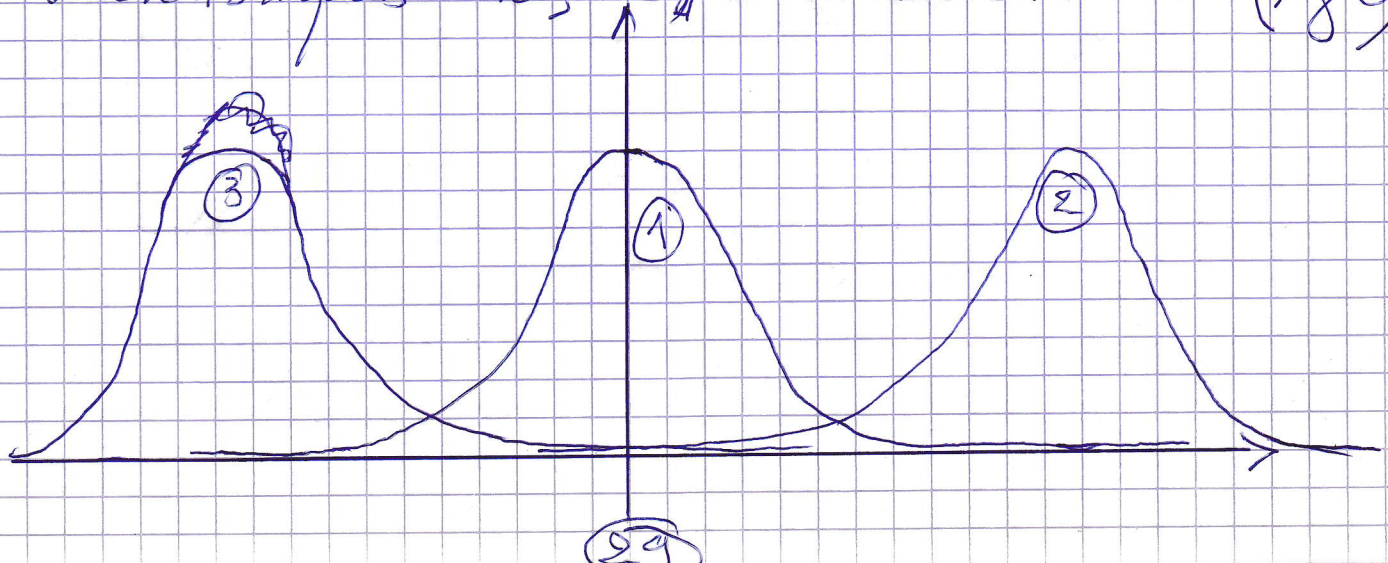
1. Asymétrie

a. Déf: Une distribution statistique est symétrique si les observations, repérées par leurs fréquences sont également dispersées de part et d'autre d'une valeur centrale.

On choisit généralement les trois valeurs centrales suivantes pour repérer l'asymétrie:

M_o, M_e et $\bar{x}$.

Dans une distribution symétrique, les trois caractéristiques M_o, M_e et $\bar{x}$ sont confondues (fig ①)



Une courbe non symétrique (asymétrique) est dite oblique à droite ou étalée à gauche si :

$\bar{x} \leqslant Me \leqslant Mo$ (fig ②)

Une courbe asymétrique est dite oblique à gauche ou étalée à droite si $\bar{x} \geqslant Me \geqslant Mo$ (fig ③).

Remarque : Si la distribution est asymétrique, la moyenne qui est la plus sensible aux valeurs extrêmes, est la plus éloignée du Mo, elle suit l'étalement de la courbe.

La médiane Me, moins sensible aux valeurs aux valeurs extrêmes, est en position intermédiaire.

b) Les Coefficients d'asymétrie :

On retient trois coefficients, connus par les noms de leurs auteurs Yule, Pearson et Fisher.

Les coefficients ne sont généralement valables que si la distribution contient un nombre assez élevé d'observations, et qu'elle ne présente pas plusieurs modes.