Sections A &B

Année Universitaire 2013-14

Semestre S1

Enseignant: DRISS TOUIJAR

Filière : Sciences Economiques et Gestion

Module : Méthodes Quantitatives I

Elément : STATISTIQUE I

Attention !

N'essayez pas de comprendre le cours en lisant tout (e) seul(e) Ce document.

Par contre, je vous recommande vivement d'assister à toutes les séances en espérant mieux cerner le programme de statistique I

INTRODUCTION GENERALE

3

la Statistique et Les Statistiques

– Le mot statistique désigne à la fois un ensemble de données d'observations et l'activité qui consiste dans leur recueil, leur traitement et leur interprétation.

–D'une façon plus précise :

- On désigne par Les statistiques, un ensemble de données ou d'informations relatives à un phénomène ou à un processus donné;

  exemple : la population marocaine en 2012, les naissances au Maroc en 2010, l'évolution des entreprises, des emplois…

4

- Par contre La statistique est en général, un ensemble de méthodes scientifiques qui servent à décrire et à analyser des données. Elles nous permettent aussi de tirer des conclusions et de prendre des décisions et aussi de faire des prévisions.

✓ En ce qui concerne ce semestre, on se contentera d'étudier la méthode descriptive : c'est une méthode qui vise à décrire des ensembles nombreux ; d'où l'appellation:

statistique descriptive.

5

Elle a pour objet de présenter les données sous forme de tableaux et de graphiques et de les résumer en quelques valeurs numériques appelée

caractéristiques.

6

Domaines d'application

- Actuellement, la Statistique est considérée comme l'un des meilleurs outils de la recherche scientifique. En effet, on fait appel aux méthodes statistiques dans presque tous les secteurs de l'activité humaine :
- Agronomie, astronomie, balistique, biologie, démographie, Économie, épidémiologie, gestion, médecine, météorologie, physique, psychologie, sciences politiques ... ,etc.

7

PROGRAMME DU SEMESTRE

- **PARTIE 1 : SERIES SIMPLES**
  - **Chapitre 1** : *Tableaux statistiques et représentations graphiques*
  - **Chapitre 2** : *Les paramètres de tendance centrale*
  - **Chapitre 3** : *Les paramètres de dispersion et de Concentration*

8

PROGRAMME DU SEMESTRE

- **PARTIE 2 : LES SERIES DOUBLES**
  - *Ajustement linéaire et Corrélation.*
- **PARTIE 3 : LES INDICES**
  - *Les Indices élémentaires et synthétiques*

9

BIBLIOGRAPHIE

Titre	Auteurs	Code
statistique descriptive	B. Py	Stat 29
Introduction à la statistique	J.P Bélisle ;J. Desrosiers	Stat 20
statistique descriptive	B. Grais	Stat 69
statistique descriptive	Hamid El Farouki	Stat 59
statistique descriptive	Benhmida et Touijar	-

10

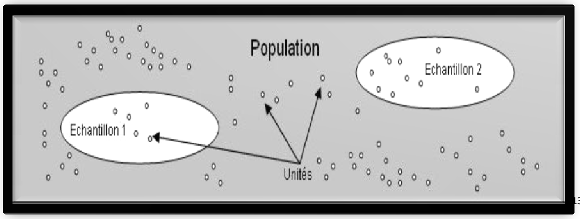
PARTIE 1
SERIES SIMPLES

CHAPITRE 1 :
Tableaux statistiques et représentations graphiques

- **I) DEFINITIONS :**
- ***Définition 1**: La population est un ensemble d'objets ou de personnes sur lesquelles porte une étude.*
- *Exemples de populations :*
- -La population du Maroc à la date du recensement de 2004.
- -L'ensemble des sociétés SARL à Fès en 2013.

12

- ***Définition 2** : Les éléments qui composent une population sont appelés des Individus (ou unités Statistiques). Un sous-ensemble d'une population est appelé Échantillon.*



- ***Définition 3** : La taille d'une population est le nombre d'individus qui la composent.*
- ***Définition 4** : Un caractère est un critère relatif auquel on observe les individus d'une population.*

- A chaque individu, on attribut un ou plusieurs caractères qui peuvent être soit **quantitatifs** (s'ils sont mesurables; exemple : salaire, nb d'enfants par ménage...) ou **qualitatifs** (sinon; exemple : sexe, état matrimonial…).

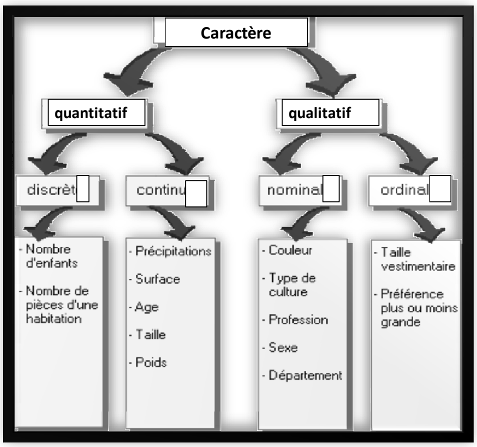
- Une valeur que peut prendre un caractère s'appelle **modalité**.

- Un caractère **qualitatif** peut être soit :
  - **Ordinal** : si ses modalités peuvent être naturellement ordonnées exemple : satisfaction plus ou moins grande après l'achat d'un produit.
  - **Nominal** : si ses modalités **ne** peuvent être naturellement ordonnées exemple : état matrimoniale.

- On appelle **variable statistique**, un caractère **quantitatif**.
  - On distingue deux sortes de variables statistiques:

- **Les variables statistiques discrètes** (notées: **v.s.d**.) : se sont des variables dont l'ensemble des modalités est un ensemble discret (la variable ne peut prendre que des valeurs isolées d'un intervalle).
  - Exemple : Pour le nombre d'enfants par ménage l'ensemble des modalités peut être {0, 1, 2, 3, 4}.
- **Les variables statistiques continues** (**v.s.c**.) : dans ce cas, l'ensemble des modalités est continue; la variable peut prendre toutes les valeurs d'un intervalle.
  - **Exemple :** Salaire, âge, taille, poids …etc.

- Tableau Récapitulatif



- **Exemple :** On observe, au cours d'une semaine, **20** machines selon le nombre de pièces défectueuses produites :

- **8, 16, 9, 33, 14, 5, 3, 7, 10, 7, 9, 9, 3, 8, 3, 3, 5, 14, 8, 7.**

→On l'appelle série brutes.

→L'effectif total de la population est donc

n =20.

- En classant ces nombres par ordre croissant, on obtient la série ordonnée :
- **3, 3, 3, 3, 5, 5, 7, 7, 7, 8, 8, 8, 9, 9, 9, 10, 14, 14,16, 33.**
- On obtient les **K** modalités après regroupement des observations :
- **3, 5, 7, 8, 9, 10, 14, 16, 33 ⇨ (k=9) < (n=20)**

19

- $\rightarrow$ la modalité 3 a pour effectif $n_1=4$
- $\rightarrow$ la modalité ***5*** a pour effectif $n_2=2$
- $\rightarrow$ la modalité ***7*** a pour effectif $n_3=3$
- $\rightarrow$ la modalité ***8*** a pour effectif $n_4=3$
- $\rightarrow$ la modalité ***9*** a pour effectif $n_5=3$
- $\rightarrow$ la modalité ***10*** a pour effectif $n_6=1$
- $\rightarrow$ la modalité ***14*** a pour effectif $n_7=2$
- $\rightarrow$ la modalité ***16*** a pour effectif $n_8=1$
- $\rightarrow$ la modalité ***33*** a pour effectif $n_9=1$
- ***Remarque : 4+2+3+3+3+1+2+1+1=20***

20

- ***Définition*** : *L'effectif* n_i *d'une modalité* x_i *est le nombre d'individus ayant cette modalité. L'effectif total (ou taille) d'une population, noté* n *, est le nombre d'individus qui composent cette population.*
- *On a donc :*

$$\boxed{\sum_{i=1}^{k} n_i = n}$$

21

- ***Définition*** : *On appelle fréquence de la modalité* x_i, *la proportion des individus présentant cette modalité. On écrit :* $\boxed{f_i = \frac{n_i}{n}}$ $;\ i = 1, \ldots, k$
- ***Remarque*** :

$$\boxed{\sum_{i=1}^{k} f_i = 1}$$

- En %:

$$\boxed{f_i\% = f_i \times 100 \Rightarrow \sum_{i=1}^{k} f_i\% = 100}$$

22

•***Exemple :***

$$\boxed{f_1 = \frac{n_1}{n} = \frac{4}{20} = 0{,}20 \text{ et } \ f_6 = \frac{n_6}{n} = 0{,}05}$$

•**Commentaire :** La proportion des machines ayant produit 3 pièces défectueuses est de 20%; et celle des machines ayant produit 10 pièces défectueuses est de 5%.

•***Définition*** : *On appelle distribution d'un caractère* X*, l'ensemble de couples*

$$\{(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)\}$$

23

•**Remarque :**

–En terme de fréquence, la distribution de X, s'écrit aussi :

$$\boxed{\{(x_1, f_1), (x_2, f_2), \ldots, (x_k, f_k)\}}$$

Exemple : *La distribution des défectuosités des 20 machines est :*

$$\{(03;4), (5;2), (7;3), (8;3), (9;3), (10;1), (14;2), (16;1), (33;1)\}$$

▪*Ou* $\{(3;0{,}20), (5;0{,}10), (7;0{,}15), (8;0{,}15)\ldots\}$

24

– *Pour une meilleure exploitation de la distribution, on a intérêt à la représenter par un tableau statistique .*

• II) TABLEAUX STATISTIQUES

• *Exemple Introductif :* Supposons que l'on ait fait une enquête auprès de 20 femmes selon 9 caractères : Prénom, nom, jour de naissance, mois de naissance, années de naissance, nombre d'enfants, revenu annuel du ménage, ville natale, opinion sur la qualité d'un produit alimentaire pour bébé.

– Ces données ont été reportées sur un bordereau, sous forme d'une matrice de 20 éléments qui comprennent chacun les 9 données concernant une femme.

Cette matrice (série selon plusieurs variables) peut être représentée comme suit :

			Date de naissance						
	Prénom	Nom	Jour	Mois	Année	Nombre d'enfants	Revenu	Ville	Opinion
1	Aline	Michalco	23	1	1953	4	30000	Paris	Mauvaise
2	Mathilde	Crepineau	13	10	1953	0	60000	Montpellier	Très bonne
3	Bénédicte	Avelin	9	3	1953	1	9000	Rouen	Mauvaise
4	Henriette	Tufte	3	2	1953	1	15000	Paris	Mauvaise
5	Danielle	Cron	16	4	1953	2	40000	Marseille	Moyenne
6	Ludivine	Laposte	15	5	1953	2	40000	Marseille	Moyenne
7	Agnès	Roche	2	5	1953	2	10000	Nice	Bonne
8	Rita	Mena	5	5	1953	3	45000	Paris	Moyenne
9	Andrée	Lamiral	22	6	1953	3	80000	Nancy	Passable
10	Pauline	Zatti	20	7	1953	4	50000	Nice	Moyenne
11	Zoé	Foret	9	9	1953	2	60000	Nice	Passable
12	Lola	Marseille	7	7	1953	3	55000	Marseille	Bonne
13	Priscilla	Lounad	3	2	1953	3	85000	Montpellier	Bonne
14	Violaine	Turk	16	4	1953	5	60000	Nice	Moyenne
15	Christine	Dodue	15	5	1953	2	40000	Rouen	Bonne
16	Fabiola	Couic	2	5	1953	1	10000	Nancy	Passable
17	Noelle	Gant	5	11	1953	3	120000	Nice	Mauvaise
18	Rachel	Nol	22	2	1953	3	80000	Paris	Moyenne
19	Sabine	Eboum	20	10	1953	4	18000	Paris	Très bonne
20	Jeanne	Rivière	9	12	1953	5	90000	Marseille	Passable

Dans cette 1ère Partie, on ne considérera qu'un seul caractère à la fois.

Représentation de la distribution d'un caractère X par un Tableau

1) Cas d'un caractère qualitatif :

– Soit la distribution d' un caractère qualitatif X étudié sur une population de n individus :

$$\{(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)\}$$

• Sa représentation par tableau est alors comme suit :

Modalité x_i	Effectif n_i	Fréquence f_i
x_1	n_1	f_1
x_2	n_2	f_2
$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i
$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k
Total	$n=\sum n_i$	$1=\sum f_i$

• **Exemple :** On reprend les 20 femmes selon l'opinion «caractère ordinal » :

• Ma, Tb, Ma, Ma, Mo, Mo, Bo, Mo, Pa, Mo, Pa, Bo, Bo, Mo, Bo, Pa, Ma, Mo, Tb, Pa.

• Une fois classées : Ma, Ma, Ma, Ma, Pa, Pa, Pa, Pa, Mo, Mo, Mo, Mo, Mo, Mo, Bo, Bo, Bo, Bo, Tb, Tb.

• On a donc 5 modalités ($k=5$).

• *La distribution s'écrit :*

$$\{(Ma\,;4),(Pa\,;4),(Mo\,;6),(Bo\,;4),(Tb\,;2)\}$$

• *Le Tableau statistique est comme suit :*

x_i	n_i	f_i
Ma	*4*	*0,2*
Pa	*4*	*0,2*
Mo	*6*	*0,3*
Bo	*4*	*0,2*
Tb	*2*	*0,1*
Total	**n=20**	**1**

2) Cas du caractère quantitatif :

a) Variable Statistique Discrète (v.s.d)

– Soit X le caractère qui désigne le nombre d'enfants par ménage pour les *20 femmes* :

4	0	1	1	2	2	2	3	3	4
2	3	3	5	2	1	3	3	4	5

– La distribution est alors :

$$\{(0,1),(1,3),(2,5),(3,6),(4,3),(5,2)\}$$

31

- Le tableau est alors le suivant :

x_i	n_i	f_i %
0	1	05
1	3	15
2	5	25
3	6	30
4	3	15
5	2	10
Total	20	100

32

Question : *Combien de femmes ont au moins ou au plus 3 enfants ?*

i) Effectifs et fréquences cumulés :

***) *Effectifs et fréquences cumulés croissants*:**

– Soit N_i le $i^{ème}$ effectif cumulé croissant associé à x_i

$$\boxed{N_i = n_1 + n_2 + \cdots + n_i = \sum_{j=1}^{i} n_j}$$

N_i est le nombre d'individus présentant au plus la modalité x_i.

33

$$N_4 = n_1 + n_2 + n_3 + n_4 = \sum_{j=1}^{4} n_j = 15$$

On dit que 15 femmes ont au plus $x_4 = 3$ enfants.

En divisant l'égalité ci-dessus par $n=20$, on obtient la **fréquence cumulée croissante:**

$$F_4 = f_1 + f_2 + f_3 + f_4 = \sum_{j=1}^{4} f_j = 75\%$$

On dit que 75% des femmes ont au plus $x_4 = 3$ enfants.

34

En général, on a :

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{j=1}^{i} f_j$$

***Effectifs et fréquences cumulés décroissants*:**

En sommant cette fois à partir du $i^{ème}$ effectif jusqu'au dernier, on obtient le $i^{ème}$ effectif cumulé décroissant, par exemple :

$$N_{3\searrow} = n_3 + n_4 + n_5 + n_6 = \sum_{j=3}^{6} n_j = 16$$

35

- Le tableau complet est comme suit :

x_i	n_i	N_i	$N_{i\searrow}$	F_i	$F_{i\searrow}$
x_1	n_1	n_1	n	$\boxed{f_1}$	1
x_2	n_2	n_1+n_2	$n_2+n_3+\ldots+n_k$	f_1+f_2	$f_2+f_3+\ldots+f_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_i	$n_1+n_2+\ldots+n_i$	$n_i+n_{i+1}+\ldots+n_k$	$f_1+f_2+\ldots+f_i$	$f_i+f_{i+1}+\ldots+f_k$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	n	n_k	1	f_k
Total	n	—	—	—	—

36

On dit que 16 femmes ont au moins $x_3=2$ enfants.

En divisant l'égalité ci-dessus par $n=20$, on obtient la **fréquence cumulée décroissante:**

$$F_{3\searrow} = f_3 + f_4 + f_5 + f_6 = \sum_{j=3}^{6} f_j = 80\%$$

On dit que 80% des ménages ont au moins $x_3=2$ enfants.

37

- Pour notre exemple, on a :

x_i	n_i	$f_i\%$	N_i	$N_{i\searrow}$	$F_i\%$	$F_{i\searrow}\%$
0	1	5	1	20	5	100
1	3	15	4	19	20	95
2	5	25	9	16	45	80
3	6	30	15	11	75	55
4	3	15	18	5	90	25
5	2	10	20	2	100	10
Total	20	100			Au plus	Au moins

38

- ***Interprétation*** :

✓Il y a 19 ménages (soit 95%) qui ont **au moins** 1 enfant.

✓Il y a 9 ménages (soit 45%) qui ont **au plus** 2 enfants.

- ***Remarque*** :
- $F_{i\searrow} = f_i + f_{i+1} + \cdots + f_k = \underbrace{\sum_{j=1}^{k} f_j}_{=1} - \sum_{j=1}^{i-1} f_j$

$= 1 - F_{i-1}$

39

b) Variable statistique continue (v.s.c)

- Dans le cas d'une v.s.c., les modalités sont regroupées en classes. Soit k le nombre de ces classes :
- $\underbrace{[e_0 , e_1[}_{1^{ère}}$; $\underbrace{[e_1 , e_2[}_{2^{ème}}$; … ; $\underbrace{[e_{i-1} , e_i[}_{i^{ème}}$; … ; $\underbrace{[e_{k-1} , e_k[}_{k^{ème}}$
- Pour la $i^{ème}$ classe, on note :
- $\boxed{a_i = e_i - e_{i-1}}$ l'amplitude de cette classe.
- $\boxed{x_i = \dfrac{e_{i-1} + e_i}{2}}$ le centre de cette classe.

40

N° classe	Les classes	n_i	f_i
1	$[e_0 , e_1[$	n_1	f_1
2	$[e_1 , e_2[$	n_2	f_2
⋮	⋮	⋮	⋮
i	$[e_{i-1} , e_i[$	n_i	f_i
⋮	⋮	⋮	⋮
k	$[e_{k-1} , e_k[$	n_k	f_k
Total	—	n	1

41

Exemple : on relève le revenu des 20 femmes(€)

Revenu X	Effectif n_i
9000	1
10000	2
15000	1
18000	1
30000	1
40000	3
45000	1
50000	1
55000	1
60000	3
80000	2
85000	1
90000	1
120000	1
TOTAL	20

Il est plus commode de regrouper les revenus en classe, par exemple, on choisi 4 classes de même amplitude

Classes Revenus(K€)	effectifs
[0 ; 35[	6
[35 ; 70[	9
[70 ; 105[	4
[105 ; 140[	?
TOTAL	20

42

Complétons notre tableau :

Classes	n_i	f_i %	F_i%	$F_{i\searrow}$%
[0 ; 35[	6	30	30	100
[35 ; 70[	9	45	75	70
[70 ; 105[	4	20	95	25
[105 ; 140[	1	5	100	5
Total	20	100	Au plus	Au moins

- ***Interprétation*** :

✓Il y a F_2=75% des femmes touchent **au plus** e_2= 70 m€. ([e_1,e_2[)

✓Il y a $F_{3\searrow}$ =25% des femmes touchent **au moins** e_2= 70 m€. ([e_2,e_3[)

III) REPRESENTATIONS GRAPHIQUES

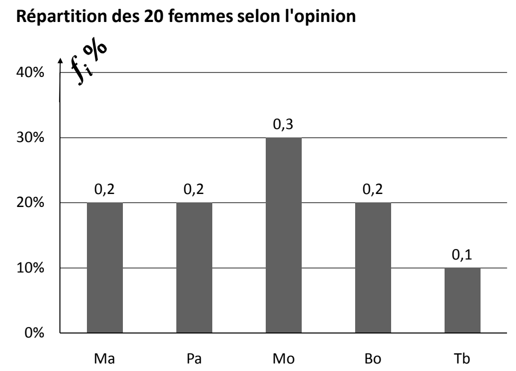
–*Le graphique est une traduction visuelle de l'information qu'elle soit qualitative ou quantitative.*

1) Cas du caractère qualitatif :

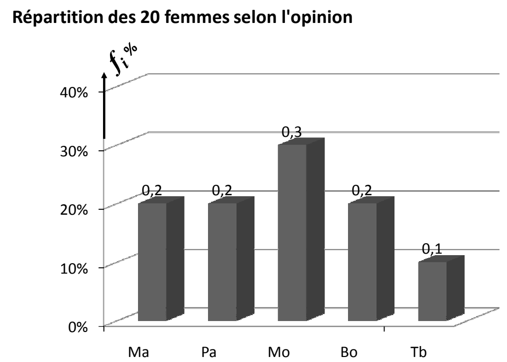
–**A) Le graphique en tuyaux d'orgue**

est formé de rectangles de même base constante et dont les hauteurs sont proportionnelles aux effectifs (ou fréquences) des modalités associées.

•**Exemple :** On reprend les 20 femmes



•**Exemple :** On reprend les 20 femmes



–**B) Le diagramme circulaire**

C'est une représentation en disque. Chaque secteur est proportionnel à l'effectif (ou fréquence) de la modalité associée. En terme d'angle, à x_i on associe l'angle au centre α_i, du secteur i, vérifiant:

$$\boxed{\alpha_i = c\, f_i \quad ; \quad i=1,\ldots,K}$$

Or

$$\sum_{i=1}^{K} \alpha_i = c \sum_{i=1}^{K} f_i \Rightarrow 360° = c \times 1 \Rightarrow \boxed{c = 360°}$$

L'égalité devient:

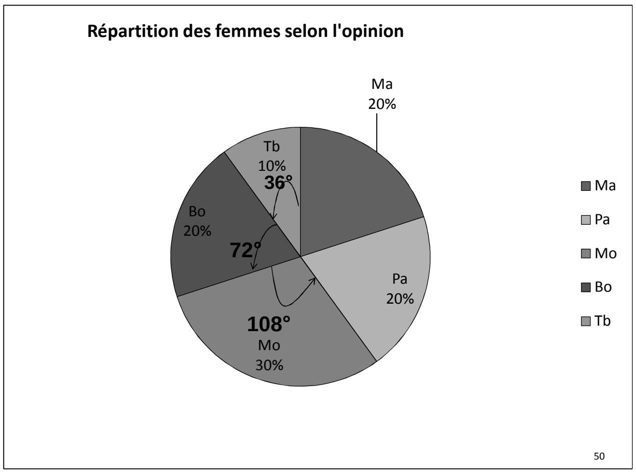
$$\alpha_i = 360 \times f_i \quad ; \quad i=1,...,k$$

Ou

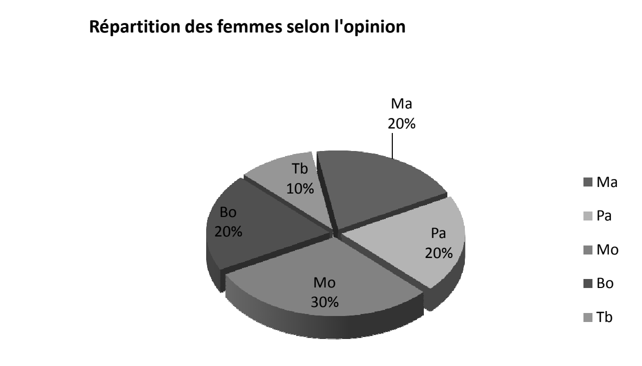
$$\alpha_i = 3,6 \times f_i\ \% \quad ; \quad i=1,...,k$$

x_i	$f_i\%$	$\alpha_i°$
Ma	*20*	*72*
Pa	*20*	*72*
Mo	*30*	*108*
Bo	*20*	*72*
Tb	*10*	*36*
Total	*100*	*360*

49



50



51

2) Cas du caractère quantitatif :

–**A) Les V.S.D.** On utilise deux types de graphiques selon que l'on considère les effectifs (ou fréquences) simple ou les effectifs (ou fréquences) cumulés:

–***i) Diagramme en bâtons***

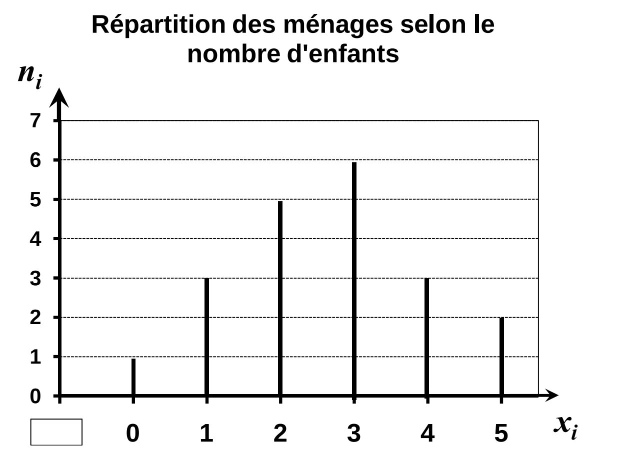
–À chaque modalité x_i on associe un

52

Segment de longueur proportionnelle à l'effectif (ou fréquence).

Exemple : On reprend l'exemple des 20 femmes

53



54

–ii) Courbe Cumulative

–**Définition :** *On appelle fonction de répartition* $F(x)$, *la fonction qui à chaque valeur* x *de* $\mathbb{R}$ *associe* *<u>la proportion</u> d'individus pour lesquels la valeur de la variable* X *est* *<u>inférieure ou égale</u> à* x.

•***Notation :***

$$\boxed{F(x) = P(X \leq x)}$$

55

–***Remarques*** :

–Si $x = x_i$ alors $F(x_i) = f_1 + \cdots + f_i = F_i$.

–Si $x_i \leq x < x_{i+1}$ alors $F(x) = F_i + 0 = F_i$.

–**Conclusion :**

$$\boxed{F(x) = F_i \text{ pour tout } x \text{ tel que } x_i \leq x < x_{i+1}}$$

La représentation graphique de $F(x)$ *est appelée Courbe cumulative, c'est une courbe « en escalier » dont les paliers sont horizontaux, puisque* $F(x)$ *est constante sur chaque intervalle* $[x_i , x_{i+1}[$.

56

–***Propriété de*** F:

•La fonction F est définie sur $\mathbb{R}$ et à valeurs dans $[0 , 1]$.

• $F(x) = 0$ si $x < x_1$

• $F(x) = 1$ si $x \geq x_k$

• $F(-\infty) = 0$ et $F(+\infty) = 1$

• F est constante sur chaque intervalle séparant deux modalités consécutives .

•***Remarque :*** On obtient la courbe cumulative des effectifs en remplaçant les F_i par les N_i .

57

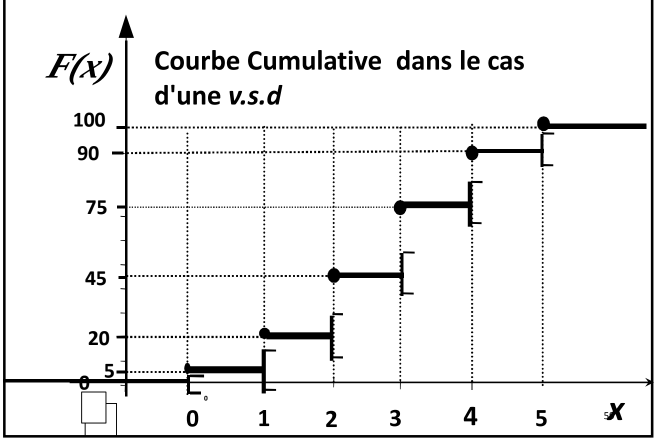
• Pour notre exemple, on a :

x_i	n_i	f_i %	F_i%
0	1	5	5
1	3	15	20
2	5	25	45
3	6	30	75
4	3	15	90
5	2	10	100
Total	20	100	Au plus

02/10/2013

58

Exemple : On reprend les 20 femmes



59

–**B) Les V.S.C.** *On a souvent recourt à trois types de graphiques :*

–*i) Histogramme*

–À chaque classe, on associe un rectangle dont la base est égale à l'amplitude de la classe et dont la hauteur est de telle sorte que sa surface ($S_i = b_i \times h_i$) soit proportionnelle à la fréquence de la classe. La juxtaposition de tous ces rectangles forment un histogramme.

60

La procédure à suivre :

– Toutes les classes ont même amplitude ($a_i = cte = a$; $i=1,\ldots,k$) alors $h_i = f_i$ (ou n_i).

– Au moins une classe a une amplitude différente des autres : dans ce cas on choisi une amplitude de référence a_r (par exemple la plus petite ou la plus répandu) Par suite, on corrige la fréquence des classes différentes en la divisant par l'amplitude associée et en multipliant par a_r :

$$h_i = f_i' = \frac{f_i}{a_i} \times a_r$$

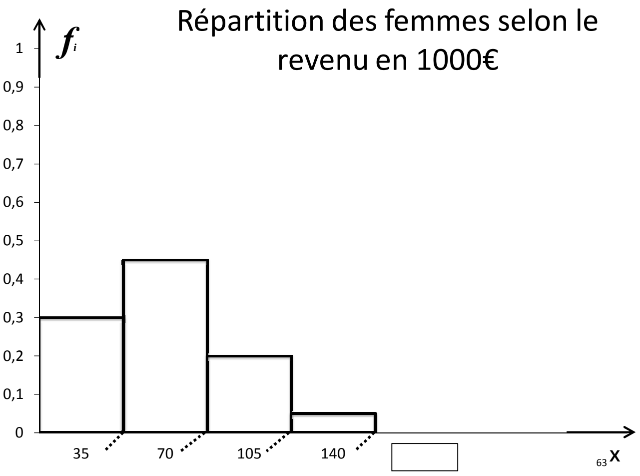
61

Exemple : On reprend les 20 femmes

Toutes les classes ont même amplitude donc pas besoin de corriger les fréquences

Classes	a_i	f_i
[0 ; 35[	35	0,30
[35 ; 70[	35	0,45
[70 ; 105[	35	0,20
[105 ; 140[	35	0,05
Total	a_r=35	1

62



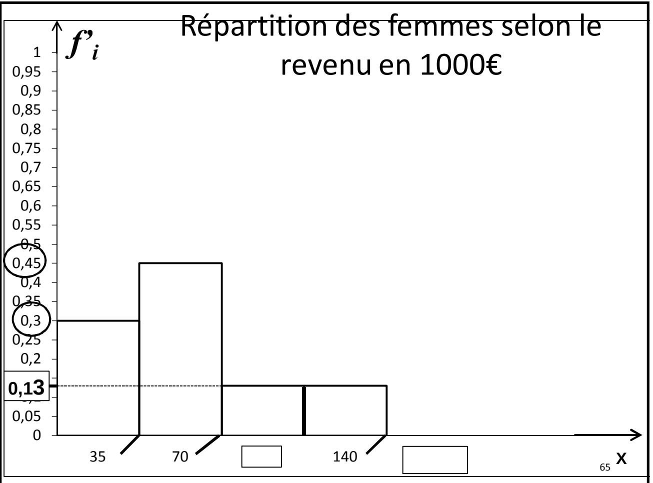
63

Exemple : Si On regroupe les 2 dernières classes

Classes	f_i	a_i	l_i	f_i'
[0 ; 35[	0,30	35	1	0,30
[35 ; 70[	0,45	35	1	0,45
[70 ; 140[	0,25	70	2	0,13
Total	1	a_r=35	---	--

Rem : f_i' s'appelle *densité de fréquence* (sans pourcentage)

64



65

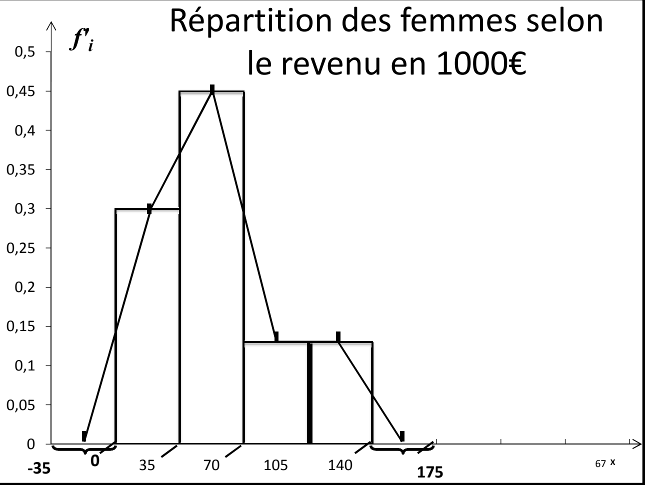
–ii) Polygone de fréquences

– On subdivise l'histogramme en sous rectangles de même base égale à l'amplitude de référence a_r. a_r étant choisie comme la **plus petite** des amplitudes et **vérifiant** :

$$a_i = k a_r \ ; \quad k = 1,2,3,4,\ldots$$

–Après avoir ajouter aux extrémités de l'histogramme deux rectangles fictifs de hauteur nulle et de base a_r ,on joint, par des segments de droites, les milieux des sommets des sous rectangles ainsi obtenus.

66



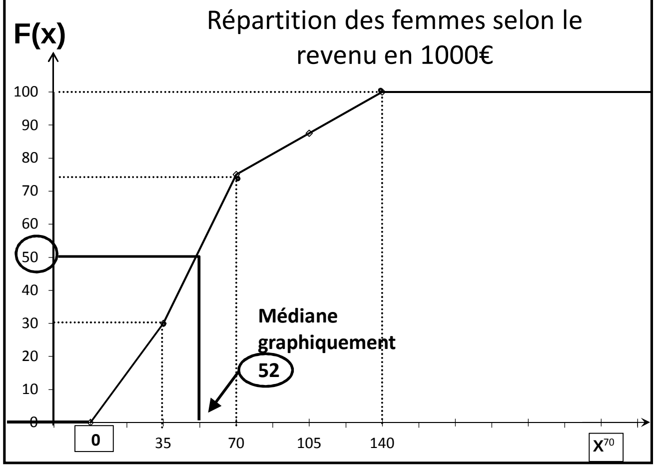
–ii) Courbe Cumulative

–On construit la courbe de la fréquence cumulée en joignant les points (e_i, F_i), où e_i est la borne supérieure de la $i^{ème}$ classe $[e_{i-1}, e_i[$ et F_i est la fréquence cumulée de cette même classe. On note

$$F_i = P(X \leq e_i)$$

Exemple : On reprend les 20 femmes

Classes	f_i%	F_i%
[0 ; 35[	30	30
[35 ; 70[	45	75
[70 ; 140[	25	100
Total	1	Au plus



Chapitre II

- *Les caractéristiques de Tendance centrale*

- *Ici, il s'agit de faire une synthèse de l'information, contenue dans la série brute, par le chiffre; et ce en calculant des paramètres dits de tendance centrale, qui caractérisent l'ordre de grandeur des observations.*
- *Dans ce chapitre, on analysera trois de ces paramètres qui sont :*
  les moyennes, le mode et la médiane.

I) LES MOYENNES

- *(1) La moyenne arithmétique*

- (a) *Définition : La moyenne arithmétique, notée $\overline{x}$, d'une variable statistique X de distribution $\{(x_i, n_i)_{1 \le i \le k}\}$ est la quantité :*

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

- Où, n est la taille de la population, et les x_i sont les modalités dans le cas d'une v.s.d. et les centres des classes dans le cas d'une v.s.c.

- **Exemple 1 :** On reprend l'exemple des 20 femmes selon le nb d'enfants

x_i	n_i	$n_i x_i$	f_i	$f_i x_i$
0	1	0	0,05	0
1	3	03	0,15	0,15
2	5	10	0,25	0,50
3	6	18	0,30	0,90
4	3	12	0,15	0,60
5	2	10	0,10	0,50
Total	20	53	1	2,65

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{6} n_i x_i = \frac{53}{20} = 2,65$$

$$ou \quad \overline{x} = \sum_{i=1}^{6} f_i x_i = 2,65$$

Exemple 2 : Pour les revenus des femmes

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{3} n_i x_i = \sum_{i=1}^{3} f_i x_i = 55125€$$

Classes	f_i	x_i	$f_i x_i$
[0 ; 35[	0,30	17,5	5,25
[35 ; 70[	0,45	52,5	23,625
[70 ; 140[	0,25	105	26,25
Total	1		55,125

(b) *Changement d'origine et d'échelle*

- ***Propriété :*** Soit X une variable statistique de moyenne arithmétique $\overline{x}$. Si Y est une variable statistique telle que $Y=aX+b$, où a et b sont des réels quelconques, alors la moyenne arithmétique de Y est :

$$\boxed{\overline{y} = a\overline{x} + b}$$

Démonstration **:**

$$\bar{y} = \sum_{i=1}^{k} f_i y_i = \sum_{i=1}^{k} f_i (a x_i + b)$$

$$= a \underbrace{\sum_{i=1}^{k} f_i x_i}_{\bar{x}} + b \underbrace{\sum_{i=1}^{k} f_i}_{1} = a\bar{x} + b$$

79

- (c) *Propriétés algébriques de la moyenne arithmétique*
- ***i)*** la moyenne des écarts à la moyenne arithmétique est nulle :

$$\sum_{i=1}^{k} f_i (x_i - \bar{x}) = 0$$

- ***ii)*** La moyenne des carrés des écarts à une constante a est minimale pour *a* = $\bar{X}$

$$\min_{a \in \mathbb{R}} \sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

80

Démonstration :

$$\sum_{i=1}^{k} f_i (x_i - a)^2 = \sum_{i=1}^{k} f_i \left[(x_i - \bar{x}) + (\bar{x} - a) \right]^2$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + \sum_{i=1}^{k} f_i (\bar{x} - a)^2$$

$$+ 2 \sum_{i=1}^{k} f_i (x_i - \bar{x})(\bar{x} - a)$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + (\bar{x} - a)^2 \underbrace{\sum_{i=1}^{k} f_i}_{=1}$$

$$+ 2 (\bar{x} - a) \underbrace{\sum_{i=1}^{k} f_i (x_i - \bar{x})}_{=0}$$

$$= \sum_{i=1}^{k} f_i (x_i - \bar{x})^2 + (\bar{x} - a)^2$$

81

(d) *Propriété de l'agrégation*

- *Soit une population* P *de taille **n**, composée de m sous populations* $P_1, P_2, \ldots, P_m$ *; de tailles respectives* $n_1, n_2, \ldots n_m$ *et de moyennes respectives* $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$ *. Alors la moyenne arithmétique* $\bar{x}$ *de la population* P *est donnée par :*

$$\boxed{\bar{x} = \frac{1}{n} \sum_{i=1}^{m} n_i \bar{x}_i}$$

82

- **Exemple :**

Le salaire moyen des cadres dans l'entreprise E est de 4000 DH.

- Le salaire moyen des cadres masculins est de 4200 DH.
- Le salaire moyen des cadres féminins est de 3000 DH.

1. Quelle est la répartition hommes - femmes des cadres ?

83

- Soit $\bar{x}$ Le salaire moyen global dans E, $\bar{x}_1$ celui des hommes et $\bar{x}_2$ celui des femmes.

On a : $\boxed{\bar{x} = f_1 \bar{x}_1 + f_2 \bar{x}_2}$

- On aboutit alors au système :

$$\begin{cases} f_1 + f_2 = 1 \\ f_1 \bar{x}_1 + f_2 \bar{x}_2 = \bar{x} \end{cases} \Leftrightarrow \begin{cases} f_1 + f_2 = 1 \\ 4200 f_1 + 3000 f_2 = 4000 \end{cases}$$

$$\Rightarrow f_1 = 83{,}3\% \;\; et \;\; f_2 = 16{,}7\%$$

84

I) LES MOYENNES

- *(2) La moyenne géométrique*

(a) Définition : *On appelle moyenne géométrique de la distribution $\{(x_i, n_i)_{1\le i\le k}\}$ que l'on note G, la racine $n^{ème}$ du produit des $x_i^{n_i}$*

$$\boxed{G = \sqrt[n]{\prod_{i=1}^{k}\left(x_i^{n_i}\right)} = \sqrt[n]{x_1^{n_1}\times x_2^{n_2}\times\cdots\times x_k^{n_k}}}$$

où $n = \sum_{i=1}^{k} n_i$

85

➢ C'est plus pratique d'utiliser le logarithme

$$* \;\mathbf{Log}(G) = \mathbf{Log}\left(\left[x_1^{n_1}\times x_2^{n_2}\times\cdots\times x_k^{n_k}\right]^{1/n}\right)$$

$$= \frac{1}{n}\mathbf{Log}\left(\left[x_1^{n_1}\times x_2^{n_2}\times\cdots\times x_k^{n_k}\right]\right)$$

$$= \frac{1}{n}\mathbf{Log}\left(\prod_{i=1}^{k}\left[x_i^{n_i}\right]\right)$$

$$= \frac{1}{n}\sum_{i=1}^{k}\mathbf{Log}\left(x_i^{n_i}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{k} n_i\mathbf{Log}(x_i) = \sum_{i=1}^{k} f_i\mathbf{Log}(x_i)$$

86

- Exemple :calculons la moyenne géométrique de *2 , 12 , 2 , 50 :* $G = \sqrt[4]{2^2\times 12\times 50} = 6{,}999$
- D'une autre façon, calculons *Log(G)* :

$$Log(G) = \sum_{i=1}^{4} f_i Log(x_i) = \frac{1}{4}\sum_{i=1}^{4} n_i Log(x_i)$$

$$= \frac{1}{4}\left[2Log\,2 + Log\,12 + Log\,50\right]$$

$$= 1{,}946$$

D'où $G = e^{1{,}946} = 7$

87

Domaines d'application :

- On utilise la moyenne géométrique dans le calcul du taux d'accroissement moyen et dans le calcul de certains indices statistique.

88

I) LES MOYENNES

- *(3) La moyenne harmonique*

- ***Définition et propriété:*** *La moyenne harmonique , notée H, d'une distribution $\{(x_i, n_i)_{1\le i\le k}\}$ est l'inverse de la moyenne arithmétique de la distribution :* $\left\{\left(\frac{1}{x_i}, n_i\right)_{1\le i\le k}\right\}$

$$\boxed{H = \frac{n}{\sum_{i=1}^{k} n_i\frac{1}{x_i}} = \frac{1}{\sum_{i=1}^{k}\frac{f_i}{x_i}}}$$

où $n = \sum_{i=1}^{k} n_i$

89

- Exemple :calculons la moyenne harmonique de *2 , 12 , 2 , 50 :*
-

$$H = \frac{4}{\left(2\times\frac{1}{2} + \frac{1}{12} + \frac{1}{50}\right)} = 3{,}625$$

90

- **Domaines d'application :**

- On utilise cette moyenne dans le calcul des durées moyennes, dans le calcul des moyennes de rapports et de pourcentages et dans les études du pouvoir d'achat (inverse du MGP)...etc.

91

I) LES MOYENNES

- *(4)* ***La moyenne quadratique***

- ***Définition et propriété :*** *La moyenne quadratique, notée Q , d'une distribution* $\{(x_i, n_i)_{1\le i\le k}\}$ *est la racine carrée de la moyenne arithmétique de la distribution* $\{(x_i^2, n_i)_{1\le i\le k}\}$

$$Q=\sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2}=\sqrt{\sum_{i=1}^{k} f_i x_i^2}$$

92

Domaines d'application :

- La moyenne quadratique intervient dans le calcul de certains paramètres de dispersion.

93

- Exemple :calculons la moyenne
- quadratique de *2 , 12 , 2 , 50 :*

$$Q=\sqrt{\frac{1}{4}(2\times 4+144+2500)}=25{,}749$$

94

(5) Résultat comparatif

Pour une même série statistique, on montre que les quatre moyennes vérifient toujours l'ordre suivant :

$$\boxed{H < G < \overline{x} < Q}$$

95

(6) Conclusion :

- Un inconvénient de la moyenne arithmétique est qu'elle est très sensible aux valeurs extrêmes de la série.
- La moyenne géométrique est peu sensible à ces dernières.
- En ce qui concerne la moyenne harmonique, elle est plus sensible aux plus petites valeurs de la série qu'aux plus grandes.

96

II) LE MODE

- **(1) *Définition :*** *Le mode, noté M_o, d'une série statistique est la valeur de cette série, dont l'effectif (ou la fréquence) est plus grand que les effectifs (ou les fréquences) des valeurs voisines.*

➢ *C'est le plus simple mais le moins utilisé !*

97

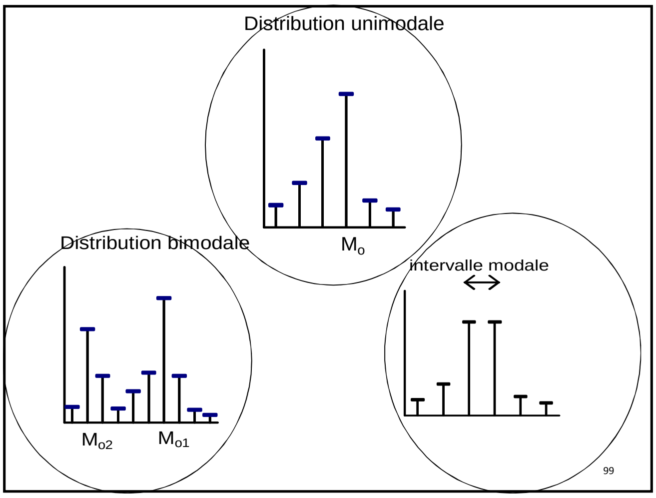
(2) Détermination du mode

- **(a) Cas d'une *v.s.d.***

Dans le cas d'une *v.s.d.* la détermination du mode est immédiate à partir du tableau statistique ou du diagramme en bâtons.

- **Exemples :**
- i) Ci-dessous on donne trois diagrammes en bâtons associés respectivement, à une distribution unimodale (qui a un seul mode), et à une distribution bimodale (qui a deux modes), et à une distribution qui a un intervalle modal.

98



99

Exemple 1 : On reprend l'exemple des 20 femmes selon le nb d'enfants

x_i	n_i	f_i
0	*1*	*0,05*
1	*3*	*0,15*
2	*5*	*0,25*
3	*6*	*0,30*
4	*3*	*0,15*
5	*2*	*0,10*
Total	*20*	*1*

M_o = 3 enfts

100

- **(b) Cas d'une *v.s.c.***
- Le mode se trouve dans la classe modale, c'est la classe qui correspond à la plus grande fréquence corrigée.
- On peut démontrer que l'expression algébrique du mode est comme suit :

$$\boxed{M_0 = e_{i-1} + a_i \frac{h_i - h_{i-1}}{2h_i - (h_{i-1} + h_{i+1})}}$$

101

$$\begin{cases} \text{où } \ [e_{i-1}, e_i[: \text{est la classe modale} \\ h_i = f_i' : \text{est la fréquence corrigée de la classe} \\ \quad \text{modale (c'est la plus élevée des fréquences)} \\ a_i : \text{est l'amplitude de la classe modale} \end{cases}$$

- **Exemple :** On reprend les 20 femmes

$$M_0 = e_1 + a_2 \frac{h_2 - h_1}{2h_2 - (h_1 + h_3)}$$
$$= 35 + 35 \times \frac{0{,}45 - 0{,}3}{0{,}90 - (0{,}3 + 0{,}13)}$$
$$= 46{,}17 \ m€$$

102

Exemple :

Classes	f_i	f'_i
[0 ; 35[	0,30	0,30
[35 ; 70[	0,45	0,45
[70 ; 140[	0,25	0,13
Total	1	--

Classe modale

III) LA MEDIANE

• (1) *Définition :* *La Médiane, notée **M**, d'une série statistique, est la valeur de la série qui partage la population en deux parties d'effectifs égaux. Par conséquent, on aura autant d'observations inférieures à **M** que d'observations supérieures à **M**.*

• ***(2) Détermination de la médiane***

• (a) **Cas d'une série brute**

• Soit la série ordonnée (par ordre croissant) de n observations : $x_1, x_2, ..., x_n$.

✓ Si n est impaire, alors la valeur médiane est l'observation qui occupe le rang $(n+1)/2$.

✓ Si n est paire, on ne peut plus déterminer exactement la médiane, mais on a un intervalle médian

$$\left[x_{n/2} \; ; \; x_{(n/2)+1}\right]$$

• ***(b)*** **Cas d'une distribution**

• ***i)*** **Cas d'1 VSD**

Soit X une *v.s.d.* de distribution $\{(x_i, f_i)_{1 \le i \le k}\}$ Pour déterminer sa médiane, on utilise les fréquences cumulées croissantes F_i.

• - **Procédure à suivre**

➢ Si $\forall i \; F_i \neq 0,5$; autrement dit, si aucune fréquence cumulée F_i n'est égale à $0,5$, dans ce cas la médiane est la modalité x_i qui correspond à la plus petite fréquence cumulée dépassant strictement $0,5$.

➢ S'il existe une modalité x_i pour laquelle $F_i = 0,5$, dans ce cas on parle d'un intervalle médian : $[x_i , x_{i+1}]$.

Exemple 1 : On reprend l'exemple des 20 femmes selon le nb d'enfants

x_i	f_i%	F_i%
0	5	5
1	15	20
2	25	45
3	30	75
4	15	90
5	10	100
Total	100	Au plus

$F_3 = 0,45 < 0,5 < F_4$

D'où $\boxed{M=3 \text{ enfts}}$

• ***ii)*** **Cas d'une *v.s.c.***

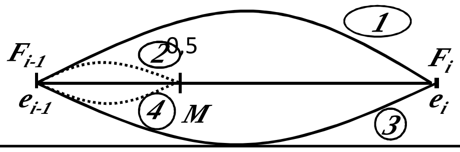
✓Dans le cas continue, la médiane est toujours unique : c'est la valeur qui partage exactement la population en deux parties égales. En d'autres termes, M est la solution de l'équation :

$$\boxed{F(M) = 0,5}$$

où F est la fonction de répartition de X.

✓On a deux méthodes pour déterminer la médiane :

- **(α) Détermination graphique** : -La médiane correspond à l'abscisse du point de la courbe cumulative qui admet pour ordonnée la valeur *0,5* (ou 50%). (Voir Graphique de l'exemple)
- **(β) Détermination par interpolation** : -D'après le tableau ou la courbe cumulative, on détermine la classe contenant la médiane M ; c'est la classe $[e_{i-1}, e_i[$ telle que, $F_{i-1} \leq 0,5 < F_i$; puis on détermine M par interpolation linéaire. donc on a :



109

$$\boxed{M = e_{i-1} + a_i \times \frac{0,5 - F_{i-1}}{f_i}}$$

où

$$\begin{cases} [e_{i-1}, e_i[: \text{est la classe médiane} \\ a_i : \text{étant son amplitude} \\ f_i : \text{est sa fréquence} \\ F_{i-1} : \text{est la fréquence cumulée de} \\ \quad\quad \text{la classe précédente} \end{cases}$$

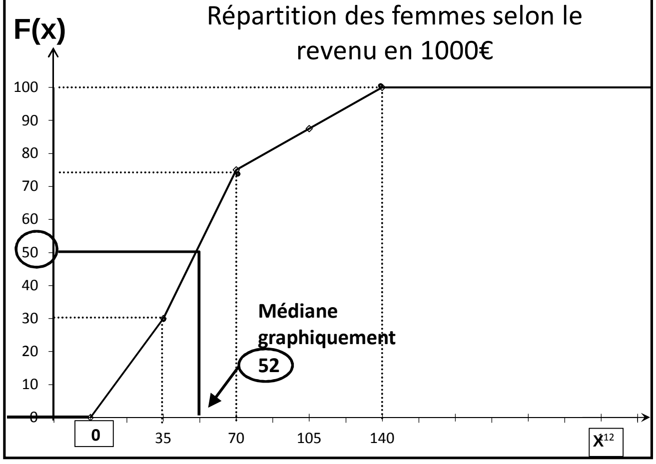
110

Exemple: $M = e_1 + a_2 \times \frac{0,5 - F_1}{f_2}$

Classes	f_i %	F_i%
[0 ; 35[	*30*	*30*
[35 ; 70[	*45*	*75*
[70 ; 140[	*25*	*100*
Total	*1*	*Au plus*

$$M = 35 + 35 \times \frac{50 - 30}{45} = 50,56\, m€$$

111



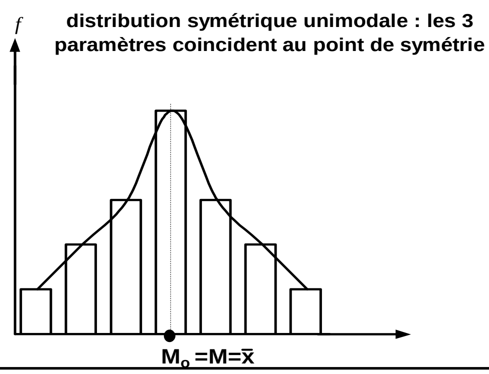
112

Conclusion Générale

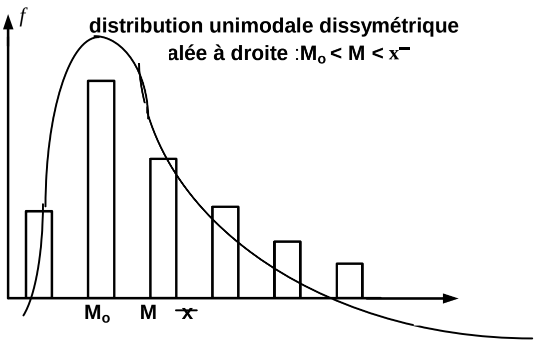
- Population hétérogène : La moyenne arithmétique est préférable à la médiane et au mode.
- Présence des valeurs aberrantes (ou extrêmes) : La médiane est préférable.
- Résultats d'un concours: la note médiane est la plus significative.
- Démographie : « L'espérance de vie » est conseillée pour comparer des pays en voie de dvpt. Pour un seul, la durée médiane ou le mode (âge le plus fréquent à la mort) sont utilisés.

113

IV) POSITIONS RELATIVES DES 3 PARAMÈTRES M_O , M et $\overline{X}$

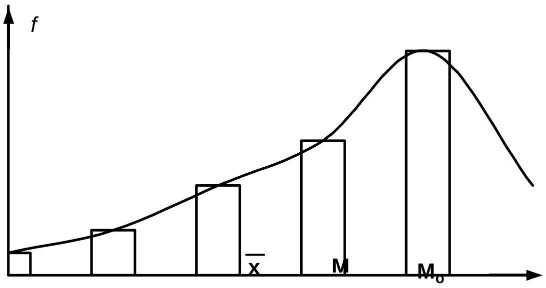


114



115

distribution unimodale dissymétrique étalée à gauche : **$\bar{x}$< M < Mo**



116

- **Remarque :** Pour des distributions « **peu** » dissymétriques, on a la relation suivante :

$$\bar{x} - M_0 \approx 3(\bar{x} - M)$$

117

Chapitre III

- *Les caractéristiques de Dispersion et de Concentration*

118

Introduction

- Les paramètres de dispersion servent à mesurer la dispersion des observations autour d'une tendance centrale.
- On considère deux catégories de paramètres de dispersion :
- *1- Les écarts simples :*
  - *étendue- écart interquantile.*
- *2-L'écart-type, la variance et le coefficient de variation*.

119

I)- Les écarts simples

1) L'étendue:

Définition: *L'étendue, noté **e**, est la différence entre la plus grande et la plus petite observation.*

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

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

Exemple : $e = x_6 - x_1$ = 5-0 = 5 enfts

120

Cas d'une *v.s.c*

$$\boxed{e = e_k - e_0}$$

Exemple: $e = e_3 - e_0$ =140-0= 140 *m€.*

Limites *:* il est très sensible aux fluctuation d' échantillonnages.

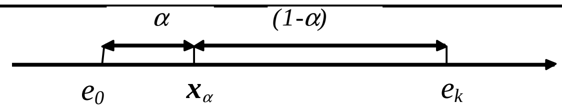
2) Ecart interquantile: Désormais, On ne s'intéressera qu'aux V.S.C.

a) **Les quantiles**: Généralisation de la médiane.

121

Définition: *Soit α un nombre réel compris entre 0 et 1. On dit qu'une valeur x_α du caractère X est un quantile d'ordre α, si :*

$$\boxed{F(x_\alpha) = P(X \leq x_\alpha) = \alpha}$$



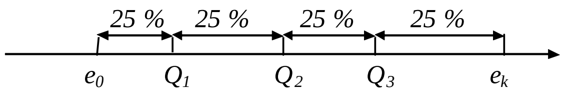
- On cite quatre types de quantiles :
  - quartiles
  - quintiles
  - déciles
  - centiles

122

Valeurs de α	Nom de x_α
$i/4$; $i \in \{1, 2, 3\}$	i ème quartile
$i/5$; $i \in \{1, 2, \ldots, 4\}$	i ème quintile
$i/10$; $i \in \{1, 2, \ldots, 9\}$	i ème décile
$i/100$; $i \in \{1, 2, \ldots, 99\}$	i ème centile

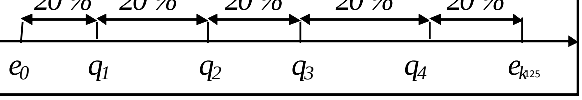
123

- *1)* **QUARTILES**: Soient Q_1, Q_2 et Q_3 les trois quartiles. Par définition :
- $Q_1 = x_{1/4}$, $\Rightarrow P(X < Q_1) = 1/4 = 25\%$
- $Q_2 = x_{1/2}$, $\Rightarrow P(X < Q_2) = 2/4 = 50\%$
- $Q_3 = x_{3/4}$, $\Rightarrow P(X < Q_3) = 3/4 = 75\%$
- Donc Q_1, Q_2 et Q_3 partagent la population en quatre parties de même effectif 25 % chacune.



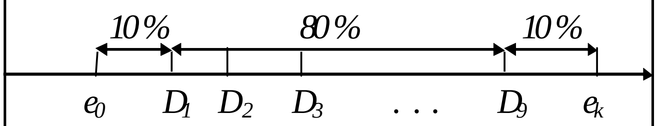
124

- *2)* **QUINTILES**: Soient q_1, q_2, q_3 et q_4 Ces 4 quintiles. On a :
- $q_1 = x_{1/5}$, $\Rightarrow P(X < q_1) = 1/5 = 20\%$
- $q_2 = x_{2/5}$, $\Rightarrow P(X < q_2) = 2/5 = 40\%$
- $q_3 = x_{3/5}$, $\Rightarrow P(X < q_3) = 60\%$
- $q_4 = x_{4/5}$, $\Rightarrow P(X < q_4) = 80\%$
- Donc q_1, q_2, q_3 et q_4 partagent la population en cinq parties d'effectif =20 % chacune.



125

- *3)* **DECILES**: Soient D_1, D_2, …, D_9 Ces 9 déciles. Ils partagent la population en dix parties d'effectif =10 % chacune.



- Par exemple : $P(X < D_5) = 1/2 = 50\%$; d'où D_5 est la médiane :

$$\boxed{D_5 = Q_2 = M}$$

126

La détermination des quantiles:

i) **Détermination Graphique :** elle est pratiquement la même que celle de la médiane, il suffit de remplacer *0,5* par α.

ii) **Détermination par Interpolation** :

$x_\alpha \in [e_{i-1} ; e_i[$, elle correspond à la plus petite fréquence cumulée dépassant strictement α

$$x_\alpha = e_{i-1} + a_i \frac{\alpha - F_{i-1}}{f_i}$$

127

Exemple : On reprend les 20 femmes

Classes	f_i %	F_i%
[0 ; 35[	*30*	*30*
[35 ; 70[	*45*	*75*
[70 ; 140[	*25*	*100*
Total	*1*	*Au plus*

128

Exemple : Pour les vingt femmes :

$$Q_1 = e_0 + a_1 \frac{25 - F_0}{f_1} = 0 + 35\frac{25-0}{30} = 29{,}17\, m€$$

$F_0 \leq 25 < F_1$

$F_2 \leq 75 < F_3$

Q_2=M* *=50,56 m€

$$Q_3 = e_2 + a_3 \frac{75 - F_2}{f_3} = 70 + 0 = 70\, m€$$

129

Exemple : Pour les vingt femmes :

$$D_1 = e_0 + a_1 \frac{10 - F_0}{f_1} = 0 + 35 \times \frac{10}{30} = 11{,}67\, m€$$

$F_0 < 10 < F_1$

$F_2 \leq 90 < F_3$

D_5=M* *=50,56 m€

$$D_9 = e_2 + a_3 \frac{90 - F_2}{f_3} = 70 + 70\frac{90-75}{25} = 112\, m€$$

130

Ecart interquantile: *Ce sont des paramètres de dispersion, donnés par la différence entre le premier et le dernier quantile :*

i) **Ecart interquartile:** $\Delta Q = Q_3 - Q_1$.

ii) **Ecart interquintile:** $\Delta q = q_4 - q_1$.

iii) **Ecart interdécile:** $\Delta D = D_9 - D_1$.

Exemple :

ΔQ = 40,83 m€

ΔD = 79,33 m€

131

c) Les écarts relatifs: *Afin de comparer des distributions différentes, on peut utiliser :*

i)Ecart interquartile relatif : $\Delta Q_r = \dfrac{\Delta Q}{M}$

ii)Ecart interdécile relatif : $\Delta D_r = \dfrac{D_9}{D_1}$

Exemple :

ΔQ_r =40,83/50,56 = 0,808= 80,8%

ΔD_r =112/11,67= 9,6

Les moins riches des 10% les plus riches ont un revenu supérieure ***à 9,6*** fois le revenu des plus riches des 10% les plus pauvres.

132

II)- Ecart-type et variance

1) La variance :

- **Définition**: *La variance, $V(X)$, d'une distribution statistique $\{(x_i, n_i) 1 \le i \le k\}$, est donnée par :*

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

133

- **Remarque**
  - La variance est aussi notée σ_x^2
  - Pour une variable statistique continue les x_i sont les centres des classes.

2) L'écart type :

Définition: *L'écart type de X, noté σ_X, est la racine carrée de la variance :*

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

134

•Cas VSD

Sachant que $\bar{x} = \sum_{i=1}^{3} f_i x_i = 55{,}125\ m€$

Classes	f_i	x_i	$(x_i-\bar{x})^2$	$f_i(x_i-\bar{x})^2$
[0 ; 35[	0,30	17,5	1415,641	424,692
[35 ; 70[	0,45	52,5	6,891	3,101
[70 ; 140[	0,25	105	2487,516	621,879
Total	1	---	---	1049,672

135

$$V(X) = \sum_{i=1}^{3} f_i (x_i - \bar{x})^2 = 1049{,}672$$

$$\Rightarrow \sigma_X = \sqrt{1049{,}672} = 32{,}3986\ m€$$

3) Formule développée de la variance

La variance s'écrit aussi comme la différence entre la moyenne des carrés et le carré de la moyenne :

$$V(X) = \overline{x^2} - \bar{x}^2$$

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

136

Classes	n_i	x_i	$n_i x_i$	$n_i x_i^2$
[0 ; 35[	6	17,5	105	1837,5
[35 ; 70[	9	52,5	472,5	24806,25
[70 ; 140[	5	105	525	55125
Total	20	-	1102,5	81768,75

137

Exemple :

$$V(X) = \overline{x^2} - \bar{x}^2 = \frac{81768{,}75}{20} - (55{,}125)^2$$

$$= 4088{,}438 - 3038{,}766 = 1049{,}672$$

$$\Rightarrow \sigma_X = \sqrt{1049{,}672} = 32{,}399\ m€$$

4) Changement de variable :

Si Y=X+b alors V(Y)=V(X)

Si Y=aX alors $V(Y)=a^2 V(X)$

138

4)Variances intra et inter populations

Soit une population ***P*** *de taille* ***n****, composée de 2 sous populations* P_1 *et* P_2*; de tailles respectives* n_1 *et* n_2 *de moyennes respectives* $\bar{x}_1$ *et* $\bar{x}_2$ *et de variances respectives* σ_1^2*,* σ_2^2*. Alors la variance* σ^2 *de la population* ***P*** *est donnée par:*

$$\sigma^2 = \underbrace{\left(f_1\sigma_1^2 + f_2\sigma_2^2\right)}_{\overline{\sigma_i^2}\ \text{intra pop}} + \underbrace{\left(f_1(\bar{x}_1-\bar{x})^2 + f_2(\bar{x}_2-\bar{x})^2\right)}_{\sigma^2_{\bar{x}_i}\ \text{inter pop}}$$

Où $\bar{x} = f_1\bar{x}_1 + f_2\bar{x}_2 = \frac{1}{n}\left(n_1\bar{x}_1 + n_2\bar{x}_2\right)$

139

5)Coefficient de variation

Définition : *Le coefficient de variation* ***Cv*** *d'une variable statistique positive de moyenne* $\bar{X}$ *et d'écart-type* σ_X *est le rapport :*

$$\boxed{Cv = \frac{\sigma_X}{\bar{x}}}$$

Exemple : le taux d'inflation dans 2 pays est respectivement *{15, 25, 35} et {150, 250, 350}.* On trouve le même taux d'inflation.

$$\boxed{Cv_A = 8{,}16/25 = 0{,}33 = Cv_B = 81{,}6/250}$$

Les paramètres de Concentration :

- La notion de concentration est apparentée à celle de dispersion, mais elle traduit le phénomène de répartition de « la masse globale » de la grandeur étudiée :salaires, chiffre d'affaires, consommation d'un produit, superficie cultivable…
- L'étude de la concentration ne concerne que les variables statistiques continues dont les modalités sont positives et leur somme a un sens. On introduit dans ce chapitre les paramètres de concentration, pour apprécier la différence entre une répartition égalitaire de la masse globale (une répartition idéale) et la répartition réellement observée.

142

- Soit *X* une variable statistique continue. Dont les valeurs sont positives et regroupées en *k* classes $[e_0, e_1[$, $[e_1, e_2[$, $\cdots$, $[e_{k-1}, e_k[$ d'effectifs respectifs $n_1, n_2, \cdots, n_k$, et de centres respectifs $x_1, x_2, \cdots, x_k$. On note :

1. $s_i = n_i x_i$: la masse du caractère *X* dans la classe $[e_{i-1}, e_i[$ *(ou relative à* x_i *)*
2. $S = \sum_{i=1}^{k} s_i = \sum_{i=1}^{k} n_i x_i$:est la masse globale du caractère.

143

$S_i = \sum_{j=1}^{i} s_j = \sum_{j=1}^{i} n_j x_j$ est la masse du caractère possédée par les individus présentant une valeur inférieure à e_i (ou la masse cumulée associée à la classe $[e_{i-1}, e_i[$).

$g_i = \frac{s_i}{S} = \frac{n_i x_i}{S} = \frac{f_i x_i}{\bar{x}}$ est la masse relative de la classe $[e_{i-1}, e_i[$ (ou la proportion de la masse globale dans $[e_{i-1}, e_i[$).

$G_i = \frac{S_i}{S}$ est la masse cumulée relative associée à la classe $[e_{i-1}, e_i[$ (ou la proportion de la masse globale possédée par les individus présentant une valeur inférieure à e_i).

144

Remarques :

- i) On a, $0 \leq G_i \leq 1$ pour tout i.
- ii) Soit x une valeur quelconque du caractère, $G(x)$ est la proportion de la masse globale possédée par les individus présentant une valeur inférieure à x, et on a $G(e_i) = G_i$.

145

Exemple : On considère la distribution statistique des salaires par heure de travail en *DH*, de 110 salariés d'une entreprise :

Salaires en DH/h	n_i	x_i	$s_i = n_i x_i$	$g_i = \frac{s_i}{S}$	G_i
[10 , 12[	15	11	165	0,092	0,092
[12 , 15[	35	13,5	472,5	0,264	0,356
[15 , 20[	45	17,5	787,5	0,441	0,797
[20 , 25[	10	22,5	225	0 ,126	0,923
[25 , 30[	5	27,5	137,5	0,077	1
Total	110	------	S=1787,5	------	------

146

•***(1) Définition :*** *la médiale de la série statistique X est la valeur du caractère qui partage la masse globale en deux parties égales. On la note M_l, et on a :*

$$G(M_l) = 0,5 = 50\ \%.$$

(2) Détermination de la médiale

La procédure de la détermination de la médiale est similaire à celle de la médiane en remplaçant les F_i par les G_i.

147

$$M_l = e_{i-1} + a_i \times \frac{0,5 - G_{i-1}}{g_i}$$

L'écart absolu médiale médiane, noté ΔM est un indicateur de concentration

$$\Delta M = M_l - M \quad et \quad \Delta M_r = \frac{\Delta M}{e}$$

- Si $\Delta M_r = 0$ alors $M = M_l$ donc on a une distribution parfaitement égalitaire
- Plus ΔM_r est grand plus la concentration est forte, et inversement.

148

Revenons à notre exercice :

$0,356 \leq 0,5 < 0,797$ donc $M_l \in [\,15, 20\,[$

$$M_l = 15 + 5 \times \frac{0,5 - 0,356}{0,441} = 16,633 \approx 16,63\ DH/h.$$

$\Delta M = M_l - M = 16,63 - 15,56 = 1,07\ DH/h.$

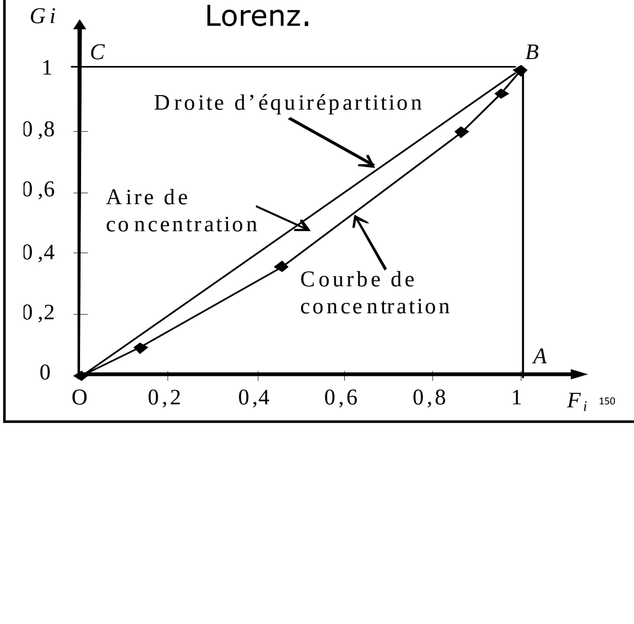
III) Courbe et indice de concentration

•(1) Courbe de concentration

Dans un repère orthonormé, on trace les points de coordonnées (F_i, G_i) et on les joint par des segments de droite.

149

• la courbe ainsi obtenue est appelée courbe de concentration ou courbe de Lorenz.



150

•(2) Indice de concentration

•(a) Définition

- *L'indice de concentration ou indice de Gini, que l'on note I_C est donné par :*

$$I_C = 1 - \sum_{i=1}^{k} f_i (G_{i-1} + G_i)$$

- $I_C = 1 - \sum_{i=1}^{5} f_i (G_i + G_{i-1})$
- *= 1- 0,872 = 0,128 = 12,8 %*

151

Analogie :

Distribution $\{(x_i, n_i)_{1\le i\le k}\}$	Distribution $\{(x_i, n_i x_i)_{1\le i\le k}\}$
n_i	$s_i = n_i x_i$
n	$S = \Sigma\, n_i x_i$
f_i	g_i
F_i	G_i
M_e	M_l
$F(x)$	$G(x)$

152

Salaires	f_i	F_i	$G_i = \frac{S_i}{S}$	$f_i(G_{i-1}+G_i)$
[10 , 12[	0,136	0,136	0,092	0,013
[12 , 15[	0,318	0,454	$\boxed{0,356}$	$\boxed{0,142}$
[15 , 20[	0,410	0,864	0,797	$\boxed{0,473}$
[20 , 25[	0,091	0,955	0,923	0,157
[25 , 30[	0,045	1	1	0,087
Total	1	——	——	$\boxed{0,872}$

FIN

153

PARTIE 2

SERIES DOUBLES

Ajustement linéaire et Corrélation

- **I) TABLEAUX DE CONTINGENCE**
- *(1) Exemple introductif*
- On désire étudier la répartition de *50184* salariés selon l'ancienneté et le salaire mensuel.
- Posons :

$$\begin{cases} X : \text{la v.s.c. désignant l'ancienneté en année} \\ Y : \text{la v.s.c. désignant le salaire mensuel (en DH)} \end{cases}$$

- Les données sont représentées par le tableau suivant :

155

X en année \ Y en CDH	[0 , 5[	[5 , 15[	[15 , 30[	[30 , 60[	Total
[0 , 1[	2890	1836	102	0	4828
[1 , 5[	9044	7378	884	204	17510
[5 , 15[	5746	10404	1564	374	18088
[15 , 35[	1666	6018	1666	408	9758
Total	19346	25636	4216	986	50184

Source :statistiques de la délégation régionale du centre Nord : octobre 1984 en se basant sur le recensement de la population et de l'habitat de 1982 : milieu urbain

156

- Si on ne considère que la 1ère ligne *Y* et la dernière ligne (Total) on obtient la répartition des *50184* salariés selon le salaire (c.à.d. la distribution *marginale* de *Y*)

Y	[0 , 5[	[5 , 15[	[15 , 30[	[30 , 60[	Total
$n_{\bullet j}$	19346	25636	4216	986	50184

- De même si on ne considère que la 1ère colonne *(X)* à la dernière colonne (Total) on obtient la répartition des *50184* salariés selon l'ancienneté : distribution *marginale* de (*X*)

X	$n_{i\bullet}$
[0 , 1[	4828
[1 , 5[	17510
[5 , 15[	18088
[15 , 35[	9758
Total	50184

- La ligne (Total) et la colonne (Total) sont appelées « ***marges*** ».

157

- **Cas des données individuelles :**
- Exemple introductif

Soient *X* et *Y* deux variables statistiques définies sur une même population de taille *n*. On note par (x_i, y_i) ; $i = 1,2,\ldots,n$ le couple de valeurs de *X* et *Y* pour le i^{eme} individu.

l=k=n

	y_1	y_2		y_i		y_n	TOTAL
x_1	1	0		0		0	1
x_2	0	1		0		0	1
x_i	0	0		1		0	1
x_n	0	0		0		1	1
TOTAL	1	1		1		1	*n*

158

- Le service des études économiques de la société α veut mesurer l'incidence de la modulation de la pression marketing (variable **X**: explicative) sur la vente de flacons de parfums (variable **Y**:expliquée). Il enregistre, alors, les ventes y_i (en milliers de flacons) ainsi que les dépenses publicitaires x_i (en milliers de DH)dans 5 zones :

x_i	5	6	9	12	18
y_i	25	30	35	45	65

159

- On cherche à étudier la liaison pouvant exister entre les variables *X* et *Y*. Pour ce, on représente dans un repère orthogonal les points (x_i, y_i). L'ensemble de ces points s'appelle «***nuage de points***». La forme de ce nuage nous renseigne sur la nature de la liaison entre *X* et *Y* et le type de courbe qui ajustera le mieux, ce nuage. On s'intéresse au cas où cette courbe est une droite (*ajustement linéaire ou droite de régression*).

160

I) Ajustement linéaire

- Pour ajuster un nuage de points, on choisi parmi toutes les courbes connues (exponentielle, logarithme, droite...) celle qui passera le plus proche de tous les points du nuage.
- Dans le cas où le nuage a une forme allongée, il est évident que la droite est la plus appropriée. On parle alors d'un *ajustement linéaire*.
- Pour déterminer l'équation de la droite d'ajustement on utilise, très souvent, la méthode des moindres carrés (**M.M.C**.).

161

- Cette méthode consiste à déterminer l'équation d'une droite telle que la somme des carrés des distances entre les points du nuage et cette droite soit minimale.

(1) Droites d'ajustement

- On applique la M.M.C. pour déterminer les deux droites de régression :

a) La droite de régression de *Y* en *X*: est utilisée pour expliquer *Y* par *X*. Cette droite est notée $D_{Y/X}$ et a pour équation $y = a\,x + b$.

Où

$$a = \frac{Cov(X,Y)}{V(X)} \quad et \quad b = \bar{y} - a\bar{x}$$

162

Où

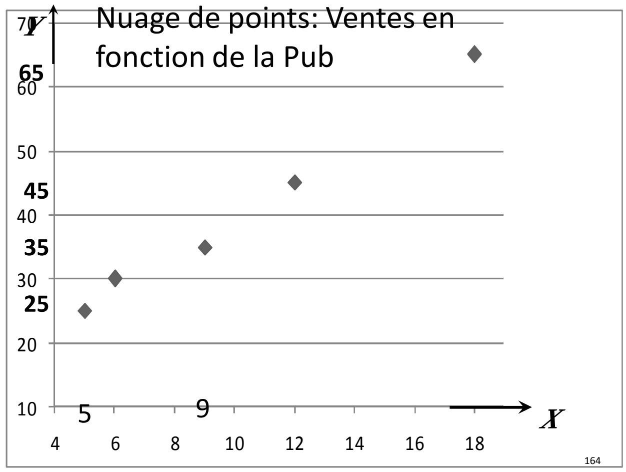
$$Cov(X,Y)=\left(\frac{1}{n}\sum_{i=1}^{n}x_i y_i\right)-\left(\bar{x}\,\bar{y}\right) \quad et$$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \quad et \quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$$

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

Revenons à notre **exemple** introductif

163



164

i	x_i	y_i	$x_i\, y_i$	x_i^2	y_i^2
1	5	25	125	25	625
2	6	30	180	36	900
3	9	35	315	81	1225
4	12	45	540	144	2025
5	18	65	1170	324	4225
TOTAL	**50**	**200**	**2330**	**610**	**9000**
∑/n	**10**	**40**	**466**	**122**	**1800**

165

La forme du nuage est plus ou moins allongée. On peut donc estimer y à partir de x grâce à la droite de régression de Y en X:

$$\bar{x}=\frac{1}{5}\sum_{i=1}^{5}x_i=\frac{50}{5}=10mDH$$

$$\bar{y}=\frac{1}{5}\sum_{i=1}^{5}y_i=\frac{200}{5}=40mflacons$$

$$V(x)=\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)-\left(\bar{x}^2\right)=\frac{610}{5}-100=22$$

166

$$Cov(X,Y)=\overline{xy}-\bar{x}\,\bar{y}$$

$$=\left(\frac{1}{5}\sum_{i=1}^{5}x_i y_i\right)-\left(\frac{1}{5}\sum_{i=1}^{5}x_i\right)\times\left(\frac{1}{5}\sum_{i=1}^{5}y_i\right)$$

$$=\frac{1}{5}\times 2330-10\times 40=66$$

$$\Rightarrow \quad a=\frac{66}{22}=3 \quad et \quad b=40-3\times 10=10$$

167

- La droite $D_{Y/X}$ a pour équation:

$$\boxed{y=a\,x+b}$$

$$\underbrace{y}_{milliers\ flacons}=\underbrace{3}_{flacon/DH}\times\underbrace{x}_{mDH}+\underbrace{10}_{milliers\ flacons}$$

- $a=3$ est la pente de la droite, cela signifie que ***1 DH*** supplémentaire investi en pub permet de vendre ***3*** flacons de plus.

168

- **b)** <u>La droite de régression de *X* en *Y* :</u> elle est utilisée pour expliquer *X* par *Y*. Cette droite a pour équation $D_{X/Y}$: $x = a'y + b'$
- ***Où***

$$\boxed{a' = \frac{Cov(X,Y)}{V(Y)} \quad \text{et} \quad b' = \overline{x} - a'\,\overline{y}}$$

- **Remarque** : Généralement, on trace les deux droites $D_{X/Y}$ et $D_{Y/X}$ dans un même repère. Dans ce cas, l'équation de $D_{X/Y}$ s'écrit :

$$y = \frac{1}{a'}x - \frac{b'}{a'} \qquad \text{où } a' \neq 0.$$

169

Revenons encore une fois à notre exemple

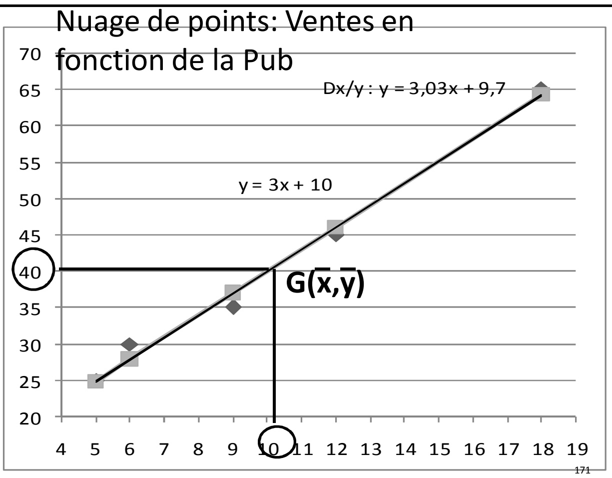
$$V(y) = \left(\frac{1}{5}\sum_{i=1}^{5} y_i^2\right) - \left(\overline{y}^2\right) = 1800 - 1600$$

$$= 200 \quad et \quad a' = \frac{66}{200} = 0{,}33$$

$$b' = 10 - 0{,}33 \times 40 = -3{,}2$$

$$\boxed{D_{X/Y} : \; y = 3{,}03 \times x + 9{,}7}$$

170



REMARQUES:

- $|a| < (1 / |a'|)$; d'où la pente de $D_{Y/X}$ est plus petite que celle de $D_{X/Y}$; donc $D_{Y/X}$ plus horizontale que $D_{X/Y}$.
- Plus les deux droites de régression *D* et *D'* sont proches l'une de l'autre, c'est à dire que l'angle entre $D_{Y/X}$ et $D_{X/Y}$ est petit plus la liaison linéaire entre *X* et *Y* est forte.

172

- **II) Coefficient de corrélation linéaire**
- Dans le cas d'une liaison linéaire entre *X* et *Y*, on mesure le degré de cette liaison par le coefficient de corrélation linéaire.
- (1) **Définition** : *Le coefficient de corrélation linéaire associé à X et Y, que l'on note r, est donné par :*

$$\boxed{r = \frac{Cov(X,Y)}{\sqrt{V(X).V(Y)}} = \frac{Cov(X,Y)}{\sigma_X . \sigma_Y}}$$

- r est sans dimension

173

- (2) Propriétés
- (a) Le coefficient r a le même signe que *Cov(X,Y)*, ce qui donne le signe de la liaison **:**
- \- Si $r < 0$, alors la liaison linéaire est <u>négative</u>.
- \- Si $r > 0$, alors la liaison linéaire est <u>positive</u>.
- \- Si $r = 0$, alors la liaison linéaire est <u>nulle</u>.
- (b) $-1 \leq r \leq +1$ **:**
- i) Pus la valeur absolue de r est proche de 1 et plus la corrélation (ou la liaison) linéaire entre X et Y est <u>forte</u>.
- ii) Pus la valeur absolue de r est proche de 0 et plus la corrélation linéaire entre X et Y est <u>faible</u>

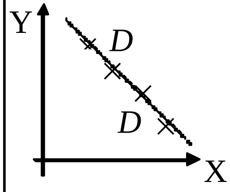
174

Valeurs de r pour interprétation:

- **0 ≤ |r| < 0,40 : faible ou quasi absence de corrélation**
- **0,40 ≤ | r | < 0,60 : moyenne corrélation**
- **0,60 ≤ | r | < 0,80 : bonne corrélation;**
- **0,80 ≤ | r |≤ 1 : corrélation élevée.**

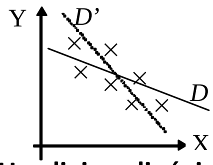
175

r = -1



Une liaison linéaire totale et négative.

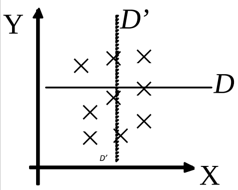
-1 < r < 0



Une liaison linéaire relative et négative.

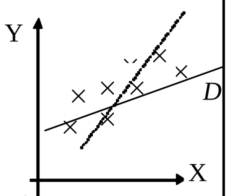
176

r = 0



Indépendance totale.

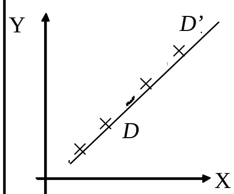
0 < r < 1



Une liaison linéaire relative et positive.

177

r = 1



Une liaison linéaire totale et positive.

178

Revenons encore une fois à notre exemple

$$\boxed{r = \frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{66}{66{,}33} = 0{,}99}$$

Conclusion:

r est positif et sa valeur est très proche de *1*, donc il y a une très forte corrélation linéaire entre *X* et *Y* et est positive. C'est ce que l'on a constaté graphiquement par la construction des droites *D* et *D'*.

179

PARTIE 3

LES INDICES :

• Introduction :

– Le concept d'indice est défini par la comparaison de toutes les observations à l'une d'elle (la 1ère par exemple), et ce dans le temps et/ou dans l'espace.

•Un indice peut être élémentaire (1seul article: prix du pain, des oranges…) ou synthétique (plusieurs articles : produits alimentaires, indices de capitalisation: MASI et MADEX…)

181

I) Indices élémentaires

•(1) **Définition** : *Soit X une grandeur simple prenant les valeurs x_0 et x_1 aux dates t_0 et t_1 . On appelle indice simple (ou élémentaire) de X entre les dates t_0 et t_1, le quotient :*

$$I_{t_1/t_0}(X)=\frac{x_{t_1}}{x_{t_0}}=\frac{x_1}{x_0}$$

182

Remarque : Souvent, on écrit et on lit :

$$I_{1/0}(X)=\frac{x_1}{x_0}\times 100$$

Indice de X à la date 1, base 100 à la date 0

Exemple : Arrivées de touristes aux frontières du Maroc (en milliers)

Avril 2000	Av. 2001	Av.2002
250	260	205

Base100 en 2000 (Avril)

183

On calcule l'indice du nombre de touristes aux frontières en 2001 base 100 en 2000

$$I_{01/00}(X)=\frac{260}{250}\times 100=104$$

Le **taux de variation** en 2001 est

$$\Delta_r X = I_{01/00}(X)-100=+4\%$$

184

(2) **Propriétés des indices simples** :

a) Identité : $I_{0/0}=100$

b) Réversibilité : $I_{t/0}=\left(I_{0/t}\right)^{-1}\times 100^2$

Remarque : b) s'écrit aussi

$$I_{t/0}\times I_{0/t}=100^2$$

Exemple :

$$I_{00/01}(X)=\frac{x_0}{x_1}\times 100=\frac{250}{260}\times 100=96{,}15$$

185

On vérifie que :

$$I_{01/00}\times I_{00/01}=104\times 96{,}15\cong 10000$$

c) Circularité :

$$I_{t_n/t_0}=\frac{I_{t_n/t_{n-1}}\times I_{t_{n-1}/t_{n-2}}\times\cdots\times I_{t_1/t_0}}{100^{(n-1)}}$$

Exemple :

$$I_{02/00}(X)=\frac{x_2}{x_0}\times 100=\frac{205}{250}\times 100=82$$

186

$$Et \quad I_{02/01} \times I_{01/00} = \left(\frac{205}{260}\times 100\right)\times 104 =$$

$$78{,}8\times 104 = 8200 \Rightarrow \frac{I_{02/01}\times I_{01/00}}{100} = 82$$

Remarque : si X=Y×Z ; alors

$$I_{1/0}(X) = \frac{I_{1/0}(Y)\times I_{1/0}(Z)}{100}$$

187

II) Indices synthétiques

•Soit X une grandeur complexe, composée de k variables simples :

$$X^1, X^2, \ldots, X^k.$$

Notons

$$\boxed{I^j_{t/0}(X) = \frac{x^j_t}{x^j_0}\times 100}$$

L'indice élémentaire de la variable X^j à la date t, base 100 à la date 0

188

(1) Définition : *L'indice synthétique $\mathfrak{I}_{t/0}$, est ~~un~~ nombre qui résume la série des k indices simple ($I^1_{t/0}, I^2_{t/0}, \ldots, I^k_{t/0}$).*

Exemple : I.G.P. (Produits alimentaires, Produits non alimentaires)

189

(2) Différentes formules d'indices synthétiques

a) Indice de laspeyres : *$\mathcal{L}_{t/0}$ est la moyenne arithmétique des indices simples $I^j_{t/0}$ pondérés par les coefficients α^j_0 (calculés à l'année de base et désignent l'importance relative de X^j)*

$$\mathcal{L}_{t/0} = \sum_{j=1}^{k}\alpha^j_0 I^j_{t/0} = \sum_{j=1}^{k}\alpha^j_0 \frac{x^j_t}{x^j_0}\times 100$$

$$où \sum_{j=1}^{k}\alpha^j_0 = 1$$

190

b) Indice de Paasche : *$\mathcal{P}_{t/0}$ est la moyenne harmonique des indices simples $I^j_{t/0}$ pondérés par les coefficients α^j_t (calculés à la date courante t)*

$$\mathcal{P}_{t/0} = \frac{1}{\sum_{j=1}^{k}\frac{\alpha^j_t}{I^j_{t/0}}} = \frac{100}{\sum_{j=1}^{k}\alpha^j_t \frac{x^j_0}{x^j_t}}$$

191

c) Indice de Fisher : *$\mathcal{F}_{t/0}$ est la moyenne géométrique de laspeyres et de paasche :*

$$\mathcal{F}_{t/0} = \sqrt{\mathcal{L}_{t/0} \times \mathcal{P}_{t/0}}$$

Souvent, on a

$$\mathcal{P}_{t/0} \le \mathcal{F}_{t/0} \le \mathcal{L}_{t/0}$$

192

(3) Propriétés des indices synthétiques

a) La circularité n'est vérifiée par aucun des trois indices.

b) L'indice de Fisher est le seul à pouvoir vérifier la propriété de réversibilité; En effet

$$\mathcal{F}_{0/t} = \sqrt{\sum_{j=1}^{k} \alpha_t^j \frac{x_0^j}{x_t^j} \times 100 \times \frac{1}{\sum_{j=1}^{k} \alpha_0^j \frac{x_t^j}{x_0^j}} \times 100}$$

$$= \sqrt{(1/\mathcal{P}_{t/0}) \times 100^2 \times (1/\mathcal{L}_{t/0}) \times 100^2}$$

193

$$= \frac{1}{\sqrt{\mathcal{L}_{t/0} \times P_{t/0}}} \times 100^2 = \frac{1}{\mathcal{F}_{t/0}} \times 100^2$$

$$\Rightarrow \quad \mathcal{F}_{t/0} \times \mathcal{F}_{0/t} = 100^2$$

194

(4)Les indices utilisés dans la pratiques

a)Indice des Prix :

i) L' indice de laspeyres des Prix

- Soient p_0^j et q_0^j respectivement le prix et la quantité de l'article *j* consommé à l'année de base *0.* et soient p_t^j et q_t^j ceux de l'année courante *t*. Alors l'indice de laspeyres des prix, noté $\mathcal{L}_{t/0}(p)$, est :

$$\mathcal{L}_{t/0}(p) = \sum_{j=1}^{k} \alpha_0^j \frac{p_t^j}{p_0^j} \times 100; \text{ où } \alpha_0^j = \frac{p_0^j q_0^j}{\sum_{j=1}^{k} p_0^j q_0^j}$$

195

D'où :

- $$\mathcal{L}_{t/0}(p) = \frac{\sum_{j=1}^{k} p_t^j q_0^j}{\sum_{j=1}^{k} p_0^j q_0^j} \times 100$$

✓**Remarque**

le produit $p_0^j \, q_0^j$ n'est autre que la valeur de l'article *j* consommé à l'année de base *0*.

196

$\mathcal{L}_{t/0}(p)$ est donc la moyenne arithmétique des indices des prix des ***k*** articles pondérés par leur valeur globale à la date 0. Il décrit l'évolution du prix d'un panier de consommation dont les quantités sont choisies à l'année de base.

- *ii) L' indice de Paasche des Prix*

- C'est la moyenne harmonique des indices des prix des ***k*** articles pondérés par leur valeur globale (α_t^j) à la date courante t.

197

$$\mathcal{P}_{t/0}(p) = \frac{100}{\sum_{j=1}^{k} \alpha_t^j \frac{p_0^j}{p_t^j}}; \text{ où } \alpha_t^j = \frac{p_t^j q_t^j}{\sum_{j=1}^{k} p_t^j q_t^j}$$

D'où
$$\mathcal{P}_{t/0}(p) = \frac{\sum_{j=1}^{k} q_t^j p_t^j}{\sum_{j=1}^{k} q_t^j p_0^j} \times 100$$

Il décrit l'évolution du prix d'un panier de consommation dont les quantités sont choisies à l'année courante

198

b) Indice des Quantités :

i) L'indice de laspeyres des Quantités

- Cette fois les prix sont constants et les quantités sont variables :

$$\mathcal{L}_{t/0}(q) = \sum_{j=1}^{k} \alpha_0^j \frac{q_t^j}{q_0^j} \times 100 = \frac{\sum_{j=1}^{k} p_0^j q_t^j}{\sum_{j=1}^{k} p_0^j q_0^j} \times 100$$

199

ii) L'indice de paasche des Quantités

$$\mathcal{P}_{t/0}(q) = \frac{100}{\sum_{j=1}^{k} \alpha_t^j \frac{q_0^j}{q_t^j}} = \frac{\sum_{j=1}^{k} q_t^j p_t^j}{\sum_{j=1}^{k} q_0^j p_t^j} \times 100$$

c) Indice de Dépenses (ou valeurs)

$$\mathcal{D}_{t/0} = \frac{\sum_{j=1}^{k} p_t^j q_t^j}{\sum_{j=1}^{k} p_0^j q_0^j} \times 100 = \frac{\text{Dépenses Totales à la date t}}{\text{Dépenses Totales à la date 0}}$$

200

Posons $d = p \times q$; alors :

$$\mathcal{L}_{t/0}(d) = \mathcal{P}_{t/0}(d) = \mathcal{D}_{t/0}$$

On a aussi

$$\mathcal{D}_{t/0} = \frac{\mathcal{L}_{t/0}(p) \times \mathcal{P}_{t/0}(q)}{100} = \frac{\mathcal{L}_{t/0}(q) \times \mathcal{P}_{t/0}(p)}{100}$$

Et on a encore :

$$\mathcal{D}_{t/0} = \frac{\mathcal{F}_{t/0}(p) \times \mathcal{F}_{t/0}(q)}{100}$$

201

•Exercice1: Le C.A. de la filière boulangerie pâtisserie d'une grande surface a été multiplié par 1,5 entre 1999 et 2001 et se répartissait ainsi

Année	1999	2001
Pain	50%	25%
Pâtisserie	30%	40%
Gâteaux	20%	35%

•Les indices des prix(base100 en **98**) ont suivi l'évolution suivante

Année	1999	2001
Pain	100,8	104,7
Pâtisserie	100,8	103,9
Gâteaux	100,9	107,7

202

Questions:

1- Calculer $\mathcal{P}$(p) base 100 en 1999.

2- Calculer l'indice du chiffre d'affaire en01/99

3- En déduire un indice de quantité 01/99.

Réponse

1. **Calculons d'abord les indices simples** base 100 en 99; Circularité

$$I^j_{01/98} = \frac{I^j_{01/99} \times I^j_{99/98}}{100} \Rightarrow I^j_{01/99} = \frac{I^j_{01/98}}{I^j_{99/98}} \times 100$$

203

Année	$I^j_{01/99}$	C^j_{01}	α_t^j	$\alpha_t^j / I^j_{01/99}$
Pain	**103,9**	25	**0,25**	**0,002**
Pâtisserie	**103,1**	40	**0,40**	**0,004**
Gâteaux	**106,7**	35	**0,35**	**0,003**
Total	**----**	**100**	**1**	**0,009**

$$\mathcal{P}_{01/99}(p) = \frac{1}{\sum_{j=1}^{k} \left(\alpha_t^j / I^j_{01/99} \right)} ; où\ \alpha_t^j = \frac{C^j_{01}}{\sum C^j_{01}}$$

204

- D'où

$$\mathcal{P}_{01/99}(p) = \frac{1}{0{,}009} = 111{,}1$$

2- l'indice du chiffre d'affaire en 01/99 :

$\mathcal{V}_{01/99}$ = 150 (par énoncé).

3- indice de quantité ; on a :

$$\mathcal{L}_{t/0}(q) = \frac{\mathcal{V}_{01/99}}{\mathcal{P}_{01/99}(p)} \times 100 =$$

$$\frac{150}{111{,}1} \times 100 = 135$$

205