SOMMAIRE

Statistique descriptive

GENERALITES :

I. Définitions :

- On appelle statistique la méthode scientifique qui vise à observer, collecter, analyser des données quantitatives.
- La statistique descriptive est la partie de la statistique qui sert à décrire un phénomène, c-à-d de mesurer, classer les mesures, présenter ces mesures par quelques indicateurs de manière à donner une idée simple et rapide d'un phénomène étudié.

Les statistiques se sont des données chiffrées relatives à un phénomène étudié.
EX : des statistiques du chômage.

II. Apport de la statistique aux économistes :

La statistique est un outil indispensable tant aux théoriciens qu'aux praticiens de l'économie.

1. La statistique est utile aux théoriciens :

- Elle permet de mettre en évidence (révéler) l'existence d'interdépendance entre différents phénomènes économiques. EX : M=P*T
- Elle permet de tester la validité d'une hypothèse théorique.

Investissement = f (revenu) =0.76R+124

Revenu → Consommé
Revenu → Epargné
Epargné → thésaurisé
Epargné → Investi

2. La statistique est utile aux praticiens de l'économie :

- La statistique permet aux entrepreneurs de mieux contrôler la gestion de leurs entreprises.
- Elle permet également au pouvoir public de mieux définir leurs politiques économique, fiscale, monétaire et d'emploi.

III. Les limites de la méthode statistique :

Pour éviter des erreurs d'interprétation due à une mauvaise utilisation statistique, il faut savoir :

1. La statistique s'intéresse au grand nombre, elle ignore les cas particuliers.
2. La résultante d'un grand nombre d'informations peut être différente de la sommation de ces différentes informations.

   *comportement collectif # sommation des comportements individuels
3. Quand on étudie un phénomène on n'est jamais certain que l'on dispose de toues les informations le concernant.

4. Il ne faut pas oublier que la statistique n'est qu'un outil au service de l'économiste, ce qui nous oblige de ne jamais, oublier de faire une analyse économique des résultats.

- Les mêmes causes # les mêmes effets.
- Les corrélations mêmes très parfaites ne signifient pas toujours qu'il y a interdépendance entre les phénomènes étudiés.

IV. Le vocabulaire utilisé en statistique :

1. *Population statistique* :

Ensemble sur lequel porte l'étude
Ex : Age des étudiants de 1ére année : l'ensemble étudié c'est l'âge.

2. *Unité statistique :*

Une population se compose d'éléments chaque élément est appelé unité statistique.
EX : la population d'étudiants : l'unité statistique est un étudiant.

3. *Caractère statistique :*

C'est le critère retenu pour étudier une population

Il peut être → quantitatif → Continu
quantitatif → discontinu, discret
Il peut être → Qualitatif

- ✓ Un caractère est dit quantitatif lorsqu'il est mesurable
  - Continu : c'est un caractère qui peut prendre toutes les valeurs d'un intervalle donné.

    EX : « âge »

  - Discontinu : c'est un caractère qui ne peut prendre que quelques valeurs dans un intervalle donné

    EX : « le nombre des frères, Ménage »

- ✓ Un caractère est dit qualitatif lorsqu'il n'est pas mesurable

  EX : la nationalité, les catégories sociales professionnelles.

4. *Modalité statistique : « de caractère» :*

On appelle une modalité les différentes situations possibles d'un caractère.
EX : caractère « sexe » : modalités possibles : M/F
Caractère « état matrimonial » : 4 modalités possibles : célibataire/marié/divorcé/veuf.

5. *Effectifs (fréquences absolues) :*

C’est le nombre d’unités statistiques relatif à une modalité donnée :

45Age	Effectifs
17-18 18-19 19-20	200 350 50
total	600

Effectifs

6. *Fréquence relative :*

C’est la part des effectifs d’une modalité.

EX : 200/600=33/100 est la fréquence relative de première modalité

7. Série statistique :

Distribution de fréquences, distribution de statistiques ou tableau statistique, c’est un tableau qui nous donne l’ensemble des valeurs mesurant le caractère.

EX :

sexe	Effectifs
Masc. Fém.	200 100
total	300

Nombre d’enfants	Arbre de ménages
2 3 4 5	18 28 10 4
total	60

Série simple.

Salaires (dh)	Effectifs
[40-60[[60-70[[70-80[	10 25 05
total	40

Série avec des classes.

8. Classes :

On appelle classe un groupement de valeurs du caractère selon des intervalles qui peuvent être égaux ou inégaux.

Pour chaque classe on peut définir :

- Une limite inférieure
- Une limite supérieure
- Intervalle de classe (amplitude)= limite (sup)- limite (inf)
- Centre de classe = [limite (sup) + limite (inf)]/2

NB : « [40-60[« signifie qu’on comptabilise les salariés qui gagnent entre 40 et 60DH, en incluant ceux qui gagnent 40 DH et excluant ceux qui gagnent 60Dh.

V. **Quelque symboles mathématiques utilisés :**

1. Les valeurs du caractère = x1, x2,…, xi,…, xn

Notes	Nbre d'étudiants
x_1	10 x_1
x_2	25 x_2
x_3	12 x_3
x_4	4 x_4

2. Les effectifs sont symbolisés par : $x_1, x_2,\dots, x_i,\dots, x_n$

$x_1, x_2,\dots, x_i,\dots, x_n = N$ =effectif total

3. Fréquence relative :

$F_i =$ effectif de la modalité i / effectif total

4. L'opérateur somme ($\sum$)

- Notation : n variables

$x_1 + x_2 + \dots + x_i + \dots. + x_n = \sum_{i=1}^{n} x_i$

Propriétés :

$$\sum_{i=1}^{n} a x_i = a \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} (a + x_i) = \sum_{i=1}^{n} a + \sum_{i=1}^{n} x_i = n a + \sum_{i=1}^{n} x_i$$

5. L'opération de produit : ($\prod$)

- Notation : le produit de x variable s'écrit :

$X_1.x_2.x_3\dots.x_n = \prod_{i=1}^{n} x_i$

Propriété :

$$\prod_{i=1}^{n} a = a^n \qquad \prod_{i=1}^{n} a x_i = a^n \prod_{i=1}^{n} x_i$$

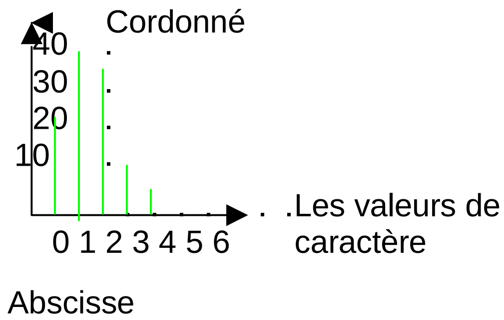
CHAPITRE I : LA REPRESENTATION GRAPHIQUE

L'intérêt d'un graphique c'est de synthétiser des informations statistiques d'une maniéré imagée, c'est à dire globale.

I. Le diagramme en bâtons :

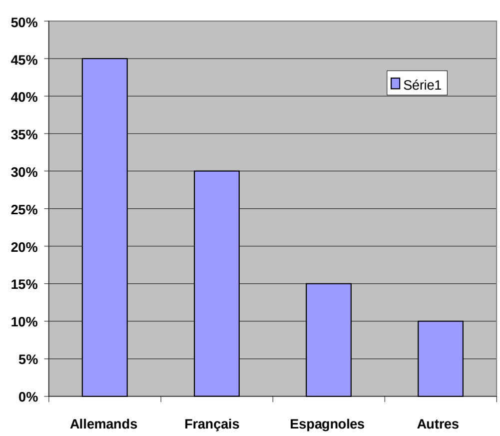
On s'en sert pour représenter des séries à caractère discret.

Nombre d'enfants	Nombre de ménage
0	25
1	42
2	38
3	15
4	6
5	
Total	128



II. Le tuyau d'orgue :

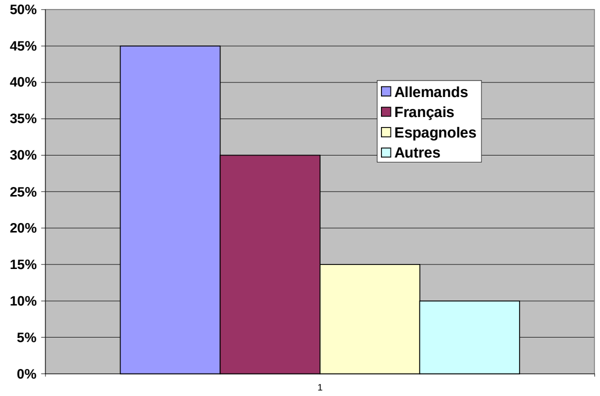
On se sert de ce graphique pour représenter des séries à caractère qualitatif
EX : La population à une station balnéaire est composée de :
Allemands : 45%
Français : 30%
Espagnoles : 15%
Autres : 10%



III. Le diagramme :

Il permet de représenter des séries de caractères ou les observations sont regroupées en classe.

a. Cas ou les intervalles de classe sont égaux :



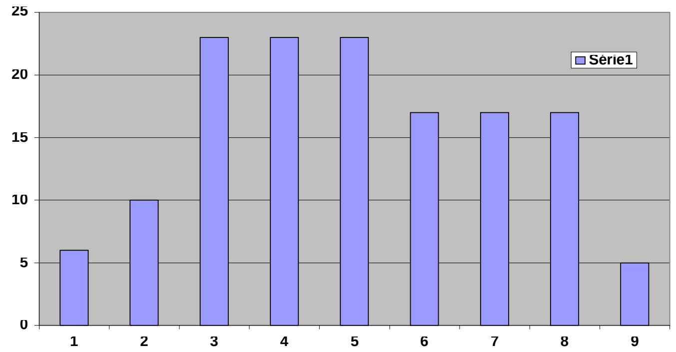
Remarque :

1) Lorsque une des limites de classe n'est pas précisée dans un tableau il convient de prendre comme intervalle de classe le même que celui de la classe suivante ou précédente.

2) La surface des rectangles est proportionnelle à leur effectif.

b. Cas ou les intervalles de classe ne sont pas égaux :

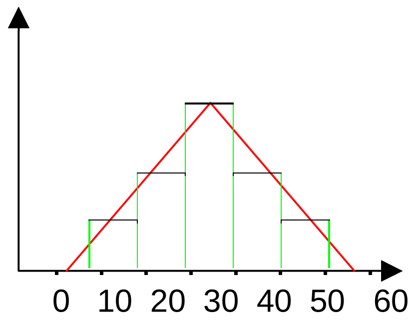
EX : Répartition de population selon leurs salaires.



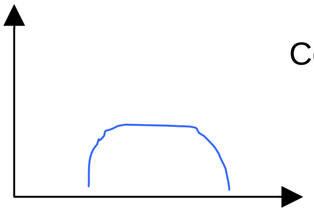
Pour tracer l'histogramme, on commence par corriger les effectifs.

IV. Le polygone des fréquences :

Il permet de donner une image plus lisse du phénomène que l'histogramme. On l'obtient en joignant les milieux des sommes des rectangles de l'histogramme.



Remarque :
1) La surface sous le polygone = la surface de l'histogramme.
2) Lorsqu'il y a un très grand nombre de classe, l'intervalle de classe devient de plus en plus petit et le polygone de fréquences se transforme en cours de fréquence.



V. La courbe de cumulation (courbe des f cumulés) :

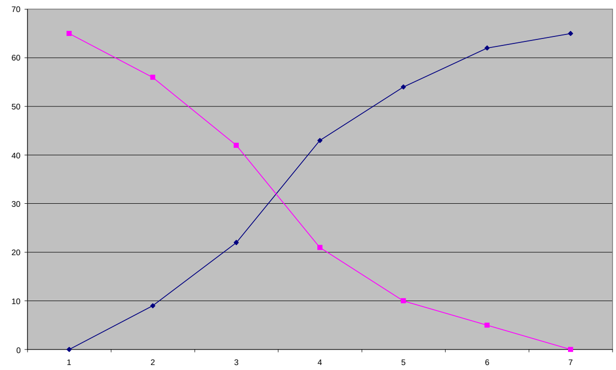
Elle permet de connaître le nombre d'observations supérieures ou inférieures à une valeur donnée.
Les 2 types de courbes de cumulation :

- Courbe cumulative croissante : permet de connaître le nombre d'observations inférieures à une valeur donnée.
- Courbe cumulative décroissante : il permet de connaître le nombre d'observations supérieures à une valeur donnée.

a) Cas d'une variable continue :

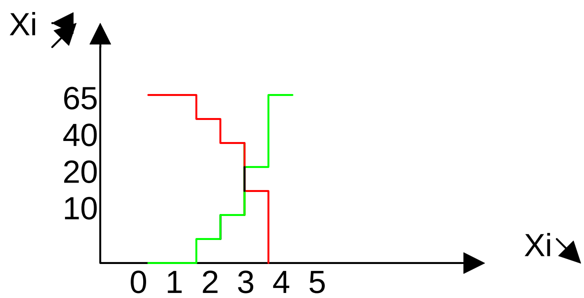
Salaire	xi	Xi cumulés ↗	Xi cumulés ↘
[10-20[[20-30[[30-40[[40-50[[50-60[[60-70[	9 13 22 10 7 4	9 22 44 54 61 65	65 56 43 21 11 4
Total	65	Moins de la borne supérieure	Plus de la borne inférieure

Remarque :
On obtiendrait le même graphique si on remplace les fréquences absolues par les fréquences relatives (les pourcentages)
Courbe cumulée décroissante
Courbe cumulée croissante



b) Cas d’une variable discrète (discontinue)

NB d’enfants (xi)	NB de ménage	Xi cumulés ↗	Xi cumulés ↙
1	5	5	65
2	10	15	60
3	30	45	50
4	20	65	20
Total	65	<=xi	>=xi

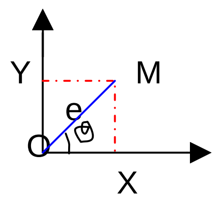


VI. Le diagramme polaire :

On l’utilise pour représenter des séries chronologiques c’est à dire des séries ou les observations seront à des temps réguliers.

a) Les principes des coordonnées polaires : un point M dans l'espace est parfaitement repéré :

- Si on connaît ses coordonnées cartésiennes (x, y).
- Si on connaît ses coordonnées polaires (e, o).

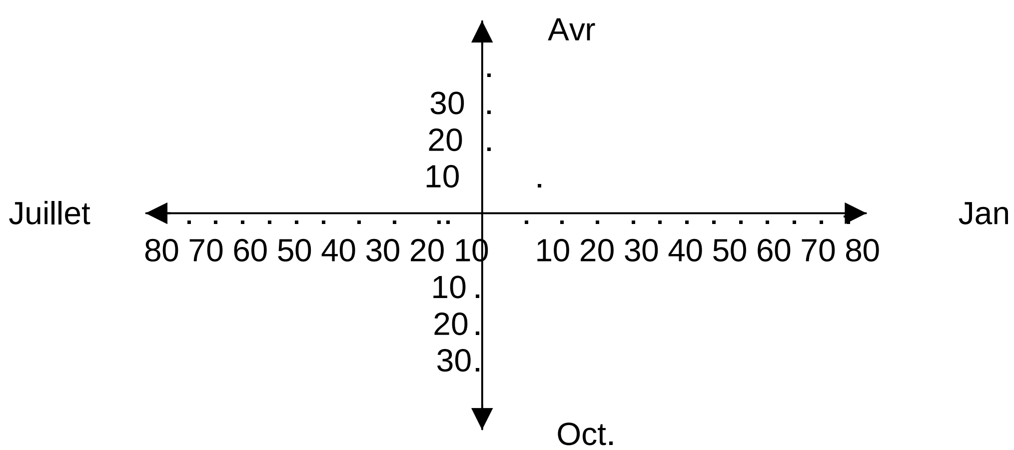


b) Le diagramme polaire :

Soit la série chronologique suivante : chiffre d'affaire mensuel

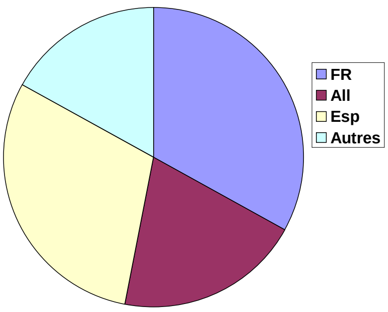
Année	1999	2000
Janvier	55	65
Février	53	75
Mars	65	72
Avril	50	40
Mai	43	42
Juin	41	38
Juillet	35	32
Août	30	34
Septembre	34	38
Octobre	40	40
Novembre	45	33
décembre	55	45

L'idée est de présenter chaque mois par un axe, nous aurons donc 12 axes, chaque axe faisant avec son voisin un angle.



VII. Les graphiques à secteurs :

On les utilise pour représenter une série exprimée en pourcentages.
EX : Pourcentage de touristes.



CHAPITRE II : LES PRANCIPALES CARACTERISTIQUES D'UN SERIE

INTRODUCTION

Avec la représentation graphique nous avons vu comment synthétiser une série avec image.

Dans ce chapitre nous allons voir comment synthétiser une série par quelques chiffres. Ces nombres sont appelés caractéristiques d'une série.

Soit les série suivantes :

Serie1 : 78-79-80-83
Série2 : 60-70-80-90-100
Série3 : 1-1-1-1-396

Les séries ont toutes la moyenne 80 même si elles sont très différentes les unes que les autres. Les valeurs de la $1^{ére}$ série sont proches de la moyenne alors que celles de la 3éme sont éloignées de la moyenne.

Il y a donc nécessité, pour résumer une série de données de la présenter en 2 types de caractéristiques :

- les caractéristiques de valeurs centrales.
- les caractéristiques de dispersion.

SECTION 1 : Les Caractéristiques de Valeur Centrale :

I. LES MOYENNES

A- La moyenne arithmétique :

A-1 Définition

Etant donnée n observations qu'on va appeler X1,X2 ,X3,……Xi,…Xn on appelle une moyenne arithmétique simple le nombre $\overline{X}$

$$\overline{X} = \frac{\text{Somme de toutes les observations}}{\text{Le nombre d'observations}}$$

$$\overline{X} = \frac{x_1 + x_2 + \ldots\ldots + x_i + \ldots\ldots + x_n}{n}$$

$$\boxed{\overline{X} = \frac{\sum_{i=n}^{n} x_i}{n}}$$: Une moyenne arithmétique simple

Lorsque les observations sont groupées c'est-à-dire que l'on observe

N1 fois X1
N2 fois X2

La moyenne arithmétique s'écrit :

$$\overline{X} = \frac{x_1 + x_1 + \ldots \quad + x_2 + x_2 + \ldots \ldots + \ldots}{n_1 + n_2 + \ldots + n_n}$$

$$\boxed{\overline{X} = \frac{\sum_{i=n}^{n} n_i x_i}{\sum_{i=1}^{n} n_i}}$$ Une moyenne arithmétique pondérée

A-2 Application

Exercice1 : soit la série de notes suivante : 2-6-12-10-12-10-10-6

$$\overline{X} = \frac{2+6+12+10+12+10+10+6}{8} = \frac{68}{8}$$

$$\overline{X} = 8,5$$

Exercice2 : soit la série des notes de l'exercice qui peut être présentée de la manière suivante :

$$\overline{X} = \frac{68}{8} = 8,5$$

$$\overline{X} = \frac{\sum_{i=} n_i x_i}{\sum n_i}$$

Notes xi	**Effectifs ni**	**ni xi**
2	1	2
6	2	12
10	3	30
12	2	24
total	8	68

Exercice3 : soit les série suivante : répartition selon l'age

$\overline{X} = \frac{3155}{88} = 35{,}85$ Années

Moyenne de l'age ou l'age moyen

age	**Ni**	**Centre de classe xi**	**ni xi**
$[20-25[$	8	22,5	180
$[25-30[$	10	27,5	275
$[30-35[$	20	32,5	650
$[35-40[$	25	37,5	937,5
$[40-45[$	15	42,5	637,5
$[45-50[$	10	47,5	475
TOTAL	88		3155

<u>a-3 Méthode des simplifications des calculs</u>

Lorsque les calculs sont compliqués, on peut les simplifier en précédant à un changement de variable

Par changement d'échelle : Tout variable Xi peut s'écrire : Xi= a X'i

a= nouvelle échelle Xi= nouvelle variable

Ex

X_i		a	*	X'_i
24	=	1	*	24
36	=	1	*	36

a=1

$X_i = X'_i$

X_i		a	*	X'_i
24	=	6	*	4
36	=	6	*	6

a = 6 $X'_i = 4$

a =6 $X'_i = 6$

par changement d'origine et d'échelle : tout variable Xi peut s'écrire

$$\boxed{x_i = x_0 + ax'_i}$$

X_0 = nouvelle origine a : n.échelle X'_i : n. variable

Ex :

Xi		X_0		a		X'i
14	=	4	+	2	*	5
22	=	4	+	2	*	9

Si on pose $x_i = x_0 + a x_i' \Rightarrow x_i' = \dfrac{x_i - x_0}{a}$

La moyenne arithmétique :

$$\overline{X} = \frac{\sum n_i x_i}{\sum n_i}$$

$$= \frac{\sum n_i (x_0 + a x_i')}{\sum n_i}$$

$$= \frac{x_0 \sum n_i + a \sum n_i x_i'}{\sum n_i}$$

$$\overline{X} = x_0 + a \frac{\sum n_i x_i'}{\sum n_i}$$

$$\overline{X} = x_0 + a \overline{X_i'} \; avec \; \overline{X_i'} = \frac{\sum n_i x_i'}{\sum n_i}$$

X_0= n origine

a: n échelle

$\overline{x_i'}$: n variable

$$\boxed{\overline{X} = x_0 + a \overline{x_i'}}$$

On utilise cette relation pour simplifier les calculs de la manière suivante

On prend pour X_0 la valeur de caractère la plus fréquente

On prend « a » l'intervalle des classes lorsque les classes sont égaux

Application :

Calculez la moyenne avec changement du variable

x_0 = 37,5 c'est le centre de classe modale

a= 5

$x'_i = (x_i - x_0)/5$

$$\overline{X_i'} = \frac{\sum n_i x_i'}{\sum n_i}$$

$\overline{X}$ = 37,5+5(-29/88)=35,8 ans

Age	effictifs	xi	x'$_i$= (xi- x_0)/a	ni*x'$_i$
20-25	8	22,5	-3	-24
25-30	10	27,5	-2	-20
30-35	20	32,5	-1	-20
35-40	25	37,5	0	0
40-45	15	42,5	1	15
45-50	10	47,5	2	20
total	88			-29

a-4 calcul de la moyenne arithmétique à l'aide des fréquences relatives

$$\overline{X} = \frac{\sum n_i x_i}{\sum n_i} = \frac{n_1 x_1 + n_2 x_2 + \ldots\ldots + n_n x_n}{\sum n_i}$$

$$\overline{X} = \frac{n_1 x_1}{\sum n_i} + \frac{n_2 x_2}{\sum n_i} + \ldots\ldots\ldots + \frac{n_n x_n}{\sum n_i}$$

$$= f_1 x_1 + f_2 x_2 + \ldots\ldots\ldots + f_n x_n$$

$\frac{n_i}{\sum ni}$: fréquence relative

d'où : $\overline{X} = f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n$

$$\overline{X} = \sum f_i x_i$$

$\overline{X} = 12{,}7$

x_i	Ni	f_i	$f_i x_i$
10	5	0,125	1,25
11	8	0,20	1,6
12	10	0,25	2,5
13	12	0,30	3,6
14	5	0,125	0,75
	40		12,7

B- La moyenne géométrique :

b-1 Définition

Étant donnée n observations connues individuellement ($x_1, x_2, x_3, \ldots, x_n$) on appelle moyenne géométrique simple de ces n observations la grandeur G t.p :

$$G = \sqrt[n]{X1.X2.\ldots Xn} = (X1.X2\ldots.Xn)^{1/n}$$

$$G = \left(\prod_{i=1}^{i=n} x_i \right)^{1/n}$$

b-2 calcul de G

lorsque les observations sont groupées ; chaque pondéré Xi sera pondéré par l'effectif correspondant, la moyenne géométrique s'écrit :

$$G = \sqrt[n]{X1.X1.X1*X2.X2.X2*X3*X3\,X3} \iff G = \sqrt[n]{X1^{n1}.X2^{n2}.X3^{n3}\ldots.Xn^{nn}}.$$

N= $n_1+n_2+\ldots..+n_n$

calculer G est plus facile en passant par le logarithme, en effet.

G= $\sqrt[n]{X1.X2....Xn} = (X1.X2.X3......Xn)^{1/n}$

log G = 1/n log (X1.X2…..Xn)

[

$$\text{Log G} = \frac{\sum \log Xi}{\text{Ni}}$$

La moyenne géométrique pondérée

$$G = \sqrt[n]{x_1^{n_1}.x_2^{n_2}........x_n^{n_n}}$$

$$G = (x_1^{n_1}.x_2^{n_2}.......x_n^{n_n})^{1/n}$$

$$\log G = \frac{1}{n}\log(x_1^{n_1}.x_2^{n_2}.......x_n^{n_n}) = \frac{\log(x_1^{n_1}.x_2^{n_2}.....x_n^{n_n})}{n}$$

$$= \frac{n_1.\log x_1 + n_2\log x_2 + ...+ n_n \log x_n}{n}$$

$$\boxed{\log G = \frac{\sum n_i \log x_i}{\sum n_i}}$$

Application : calculer la moyenne géométrique

x_i	n_i	log x_i	n_i log x_i
2 6 10 12	1 2 3 2	0,301 0,772 1 1,158	0,301 1,556 3,0 2,158
Total	8		7,316

$$\log G = \frac{7,316}{8} = 0,9145$$

$$G = 10^{0,9145} = 8,2$$

C- la moyenne harmonique :

c-1 Définition

Étant donnée n observations connues individuellement $x_1, x_2, x_3 \ldots x_n$ *on appelle moyenne hormique le nombre H tel que :*

$$\frac{1}{H} = \frac{1/x_1 + 1/x_2 + \ldots + 1/x_n}{n} = \frac{\sum 1/x_i}{n}$$

$$\boxed{H = \frac{n}{\sum 1/x_i}}$$

moyenne harmonique simple.

Si les observations sont groupées la moyenne harmonique s'écrit :

$$\frac{1}{M} = \frac{n_1 . 1/x_1 + n_2 . 1/x_2 + \ldots + n_n . 1/x_n}{n_1 + n_2 + \ldots + n_n} = \frac{\sum n_i \, 1/x_i}{\sum n_i}$$

Moyenne harmonique pondérée

$$\boxed{H = \frac{\sum n_i}{\sum n_i \, 1/x_i}}$$

c-2 Application

x_i	n_i	$1/x_i$	$n_i.1/x_i$
2	*1*	*0,5*	*0,5*
6	*2*	*0,166*	*0,332*
10	*3*	*0,1*	*0,2*
12	*2*	*0,083*	*0,166*
total	*8*		*1,298*

$$\frac{1}{H} = \frac{\sum x_i \, 1/x_i}{n_i} = \frac{1,298}{8}$$

$$\Leftrightarrow H = \frac{8}{1,298} = 6,16$$

c-3 Remarque

$$\frac{1}{H} = \frac{\sum n_i \, 1/x_i}{\sum n_i} = \frac{\sum n_i . X_i}{\sum n_i} \; avec X_i = \frac{1}{x_i}$$

L'inverse de la moyenne = moyenne des inverses

D -La moyenne quadratique :

Définition : Etant donné n observations connues individuellement X1 ; X2 ;…..xn

$$Q^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} \Leftrightarrow \frac{\sum x_i^2}{n}$$

$$\boxed{\Leftrightarrow Q = \sqrt{\frac{\sum x_i^2}{n}}}$$

moyenne quadratique simple

si les observations sont groupées, la moyenne quadratique s'écrit :

$$Q^2 = \frac{n_1 . x_1^2 + n_2 . x_2^2 + \ldots + n_n . x_n^2}{n_1 + n_2 + \ldots + n_n}$$

$$Q^2 = \frac{\sum n_i . x_i^2}{\sum n_i} \Leftrightarrow Q = \sqrt{\frac{\sum n_i . x_i^2}{\sum n_i}}$$

moyenne quadratique pondérée

Application :

$$Q^2 = \frac{\sum n_i . x_i^2}{\sum n_i} = \frac{664}{8} = 83$$

$$\Leftrightarrow Q = \sqrt{83} = 9.1$$

xi	*Ni*	*Xi²*	*Ni. Xi²*
2	*1*	*4*	*4*
6	*2*	*36*	*72*
10	*3*	*100*	*300*
12	*2*	*144*	*288*
total	*8*		*664*

$$\boxed{Q^2 = \frac{\sum n_i . x_i^2}{\sum n_i} = \frac{\sum n_i . X_i}{\sum n_i} \text{ avec } X_i = x_i^2}$$

Carré de la moyenne = la moyenne des carrés

Généralisation de la notion moyennes :

d.1- moyenne d'ordre r

on appelle moyenne d'ordre r la quantité M_r tel que :

$$M_r = \left(\frac{x_1^r + x_2^r + \dots\dots\dots\dots\dots + x_n^r}{n} \right)^{1/n}$$

$$M_r^r = \frac{x_1^r + x_2^r + \dots . + x_n^r}{n}$$

Si r= 1 $M_1 = \dfrac{x_1 + x_2 + \dots + x_n}{n} \Rightarrow M_1 = \bar{X}$

si r= 2 $M_2^2 = \dfrac{x_1^2 + x_2^2 + \dots\dots + x_n^2}{n} \Rightarrow M_2^2 = Q^2 \Rightarrow M_2 = Q$

si r= -1 $M_{-1}^{-1} = \dfrac{x_1^{-1} + x_2^{-1} + \dots\dots + x_n^{-1}}{n} \Rightarrow \dfrac{1}{M_{-1}} = H^{-1} = \dfrac{1}{H} \Rightarrow M_{-1} = H$

si r= ε → 0. $\boxed{M_{\varepsilon \to 0} = G}$

d.2- le classement des moyennes : les inégalités entre les moyennes :

On démontre que les moyennes s'ordonnent selon la valeur de r c-à-d que si : $r_1 < r_2 \Rightarrow M_{r_1} < M_{r_2}$ Ce qui nous donne : $M_{-1} < M_0 < M_1 < M_2$

$H < G < \bar{X} < Q$

Dans notre exemple, on trouve : 6,16 <8,2 <8,5 < 9,11.

d-3 Le choix d'une moyenne :

En théorie, aucune moyenne n'est meilleure que l'autre. L'utilisation de telle moyenne dépend du problème posé.

Exemple :

Ex1 : Soit un petit jardin sous forme de rectangle, le propriétaire ne peut se souvenir que d'un seul chiffre.

9

4 4

9

S'il veut entourer son champs de fil de fer il a intérêt à se souvenir de la moyenne arithmétique car le périmètre est lié à la somme des différents côtés.

S'il veut mettre de l'engrais à son jardin, il a intérêt à se souvenir de la moyenne géométrique

$$\overline{X} = \frac{9+4+9+4}{4} = 6{,}5 ; G = \sqrt{9*4} = 6$$

moyenne arithmétique du périmètre =26 =6,5 * 4 ≠ 6 * 4
moyenne géométrique : surface =36 =6*6 ≠ 6,5 * 6,5

Généralités :

D'une manière générale, on retient la moyenne arithmétique quand les variables s'additionnent, et on utilise la moyenne géométrique lorsque les variables se multiplient.

Ex2 : Une voiture parcourt 100Km/h, puis 160Km/h à 80Km/h.

$$Vitessemoy = \frac{distoncetotale}{tempstotal} = \frac{100+160}{\frac{100}{50}+\frac{160}{80}} = \frac{100+160}{100 . \frac{1}{50} + 160\frac{1}{80}}$$

$$\Leftrightarrow MH = \frac{\sum n_i}{\sum n_i . \frac{1}{x_i}}$$

La vitesse moyenne est égale à la moyenne harmonique des vitesses pondérées par les distances.

Ex3 : Une voiture roule pendant une heure à 50 Km/h puis 3hà 80Km/h.

$$Vitesse.moy = \frac{distoncetotal}{tempstotal} = \frac{(1\times 50) + (3\times 80)}{1+3}$$

$$\overline{X} = \frac{\sum n_i x_i}{\sum n_i}$$

La vitesse moyenne est égale donc à la moyenne arithmétique des vitesses pondérées par le temps.

Ex 4 : Une grandeur S_0 a augmenté sur 3 années, d'abord de 10% puis de 15% et 30% pour le 3éme année.

Quel est le taux moyenne de croissance ?

1ère année : S_0 devient $S_1=S_0 + (S_0*10/100)$ ➔ $S_1 = S_0(1+0{,}10) = 1{,}10S_0$
2éme année S_1 devient S2 = S1 +0,15S1 ➔S1*1,15➔ (S1*(1+0,15))
3éme année S2 devient S3 = S2 +0,3S2 = 1,3S2➔ (S2*(1+0,3))
$S3 = S0 1{,}1 \times 1{,}15 \times 1{,}3$

Moyenne géométrique $G = \sqrt[3]{1{,}1 \times 1{,}15 \times 1{,}3} = 1{,}1804$

Remarque: le taux de croissance moyenne est 18.04%

Ex 5 : Un étudiant a obtenu les notes suivantes : 8-10-12 on veut calculer la moyenne des écarts entre les notes et la moyenne arithmétique.

$$\overline{X} = \frac{8 + 10 + 12}{3} = 10$$

Ecart type à la moyenne

8-10 = -2

10-10 =0

12-10 =2

moyenne arithmétique des écarts = (-2+0+2)/3

moyenne arithmétique des écarts = 0

On retrouve ici une des propriétés des moyennes arithmétiques :

$$\sum (x_i - \overline{X}) = 0$$

Démonstration : $\sum x_i - \sum \overline{x} = \sum x_i - n\overline{x} = \sum x_i - n\frac{\sum x_i}{n} = 0$

Si on veut calculer la moyenne des écarts, il vaut mieux calculer la moyenne quadratique

$$Q^2 = \frac{(-2)^2 + (0)^2 + (2)^2}{3} = \frac{8}{3}$$

$$Q = \sqrt{\frac{8}{3}} = 1{,}6$$

II. *La médiane (Me)*

b-1- Définition :

On appelle médiane d'une série classée par ordre croissant ou décroissant, la valeur du caractère qui partage en deux parties égales les effectifs.

C'est la valeur du caractère telle que la moitié des effectifs lui est supérieure et l'autre lui est inférieure.

b-2- Calcul de ME :

Cas d'une variable discrète

Si la série a un nombre impair de terme

75 62 57 12 18 $\Leftrightarrow$ Me =57

Si la série a un nombre pair $\Leftrightarrow$

12 25 32 44 52 69

Intervalle Médian [32-44]

On prend le centre de l'intervalle comme la médiane :

Cas d'une série de classes :

Salaires	Effectifs	Effectifs cumulés
10-15	9	9
15-20	25	34
20-25	32	66
25-30	16	82
Total	82	

Le calcul de la médiane se fait en 3 étapes :

<u>1ére étape</u> : on repère le rang de la médiane. Rang = 82/2 = 41

$$\text{Rang} = \frac{\sum ni}{2}$$

<u>2éme étape</u> : on repère la classe de Me :

Il s'agit de trouver la classe à laquelle appartient le 41 $^{\text{éme}}$ individu, pour cela on classe les individus par ordre croissant des salaires, ce qui revient à construire la colonne des effectifs cumulés.

Me $\in$ [20-25], on peut calculer avec plus de précision Me en faisant une interpolation linéaire.

<u>3éme étape</u> : l'interpolation linéaire :

On connaît les salaires des 34 individus 20

On connaît les salaires des 66 individus 25

Le 41 $^{\text{éme}}$ individus c'est le 7éme individus que je rencontre dans la classe 20 -25, son salaire sera obligatoirement égal à 20 + supplément que l'on calcule par interpolation.

En supposant que les 32 individus de la classe 20-25 sont répartis d'une manière uniforme dans la classe

20-25 puis sont séparés par la même quantité de salaire

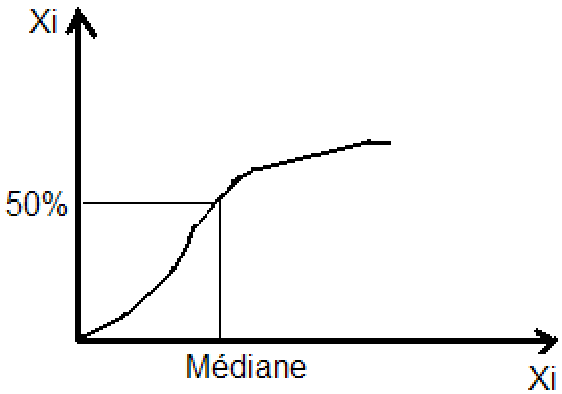
On raisonne alors de la manière suivante :

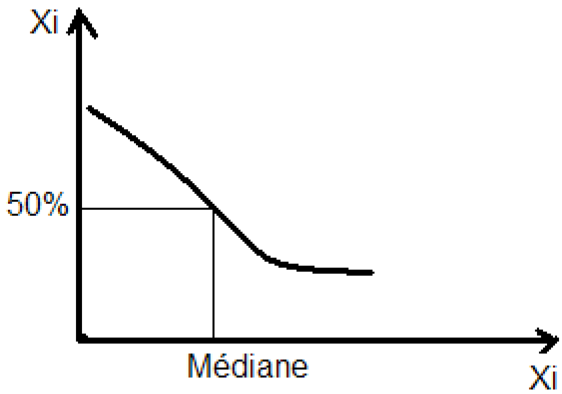
Si pour 32 individus nous avons un écart de salaire de 5 DH

Pour 1 individu → 5/32
Pour 7 → individus 5/32 * 7 = 1.09 DH
Me=20+1.09 =21.09
La moitié des effectifs gagnent plus de 21,09 DH et l'autre moitié gagne (moins de 21,09 DH)

b-3- Détermination graphique de la médiane :

Courbe cumulative



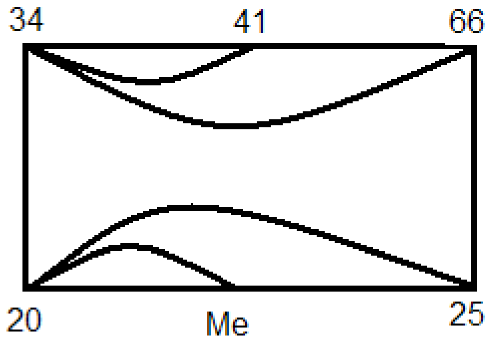


b-4-Remarque :

Salaire	Xi	Xi ↗
10 – 15	9	9
15 - 20	25	34
20 - 25	32	66
25 - 30	16	82

Total Xi =82

Méthode rapide d'interpolation :



$$\frac{Me - 20}{25 - 20} = \frac{41 - 34}{66 - 34} \Leftrightarrow Me = \frac{7 \times 5}{32} + 20 \approx 21$$

2. le 41 éme individu normalement la médiane devrait se situer entre le 41 éme et le 42 éme, mais on convient lorsque les effectifs sont nombreux de prendre (N / 2)

III. *Le Mode :*

C'est la valeur du caractère le plus fréquent.

A- Calcul Mode :

1- Cas d'une variable discrète :

Xi	ni
3	3
14	18
21	7
42	4

Mo =14
Série
Uni modal

Xi	Ni
2	4
17	16
33	15
39	16
51	8

Mo = 17

Mo = 39
Série bimodale

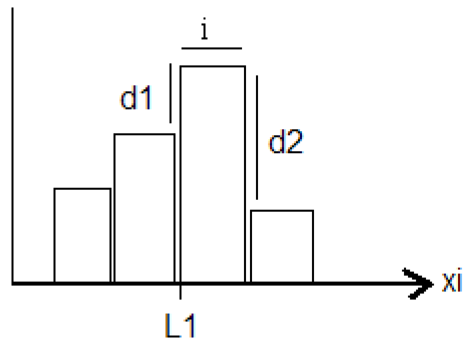
Série plurimodale (série à plusieurs modes)

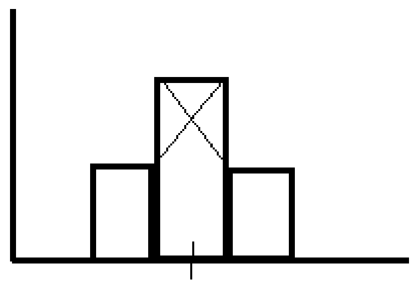
2-Cas d'une série de classe :

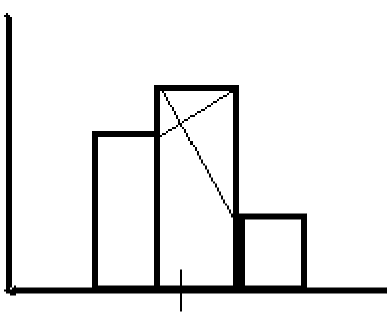
Salaires	ni
10 – 15	9
15 – 20	25
20 – 25	32
25 - 30	16
Total	82

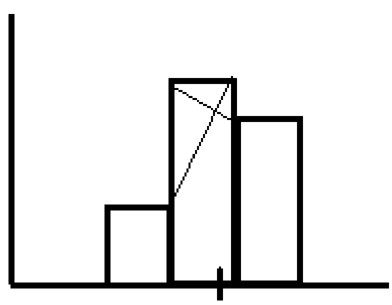
-Nous avons une classe modale : 20 – 25
- On peut prendre comme mode le centre de classe 22,5
- On peut chercher à obtenir le mode avec plus de précision :

1/ Par Méthode graphique : Elle consiste d'abord à construire l'histogramme









N.B : Ne pas oublier, lorsqu' on construit l'histogramme de corriger les effectifs.

2/ Par la méthode algébrique :

$$\boxed{Mo = L1 + [d1.\ I / (d1 + d2)]}$$

$$Mo = 20 + \frac{(32 - 25) *}{(32-25) + (32 - 16)}\ 5$$

L1 : Limite Inférieure de classe modale
d1 : La différence entre les effectifs de la classe modale et les effectifs de classe précédente
d2 : La différence entre les effectifs de classe modale et les effectifs de classe suivante
i : L'intervalle de la classe modale

IV. VI- Le choix d'une caractéristique de tendance centrale :

A : Les conditions de Yule :

1 ér conditions : Une modalité caractéristique doit être : définie de façon objective. (2 personnes différentes doivent trouver le même résultat)
2 éme conditions : Tenir compte de toutes les observations
3 éme conditions : être facile à comprendre
4 éme conditions : être facile à calculer
5 éme conditions : Doit se prêter au calcul algébrique

B : Comparaison des différentes caractéristiques de tendance centrale :

1-La moyenne :

Elle répond parfaitement aux conditions de Yule ; c'est pour cela qu'elle est la caractéristique la plus utilisée, mais il y a des cas ou il faut lui préférer la médiane quand elle risque d'être influencé des valeurs extrêmes.

EX:

Notes	Xi	Ni * Xi
1	1	1
16	2	32
17	5	85
18	2	36
	10	154

X = 154 / 10 = 15,4
X = 153 / 9 = 17

2-La médiane :

Elle ne satisfait pas les conditions de yule.
En effet, la valeur de la médiane ne change pas quand on augmente la valeur d'une observation qui lui est inférieure

15 22 <u>34</u> 41 60 122 <u>34</u> 41 110
1 2 <u>34</u> 41 60

3-Le mode :

Ne remplit pas les conditions de Yule, mais il y a des cas ou il est utile, en particulier quand on cherche la valeur la plus typique d'une série :
Ex : un vendeur de chaussures ne va pas stocker des chaussures de pointure moyenne, mais va stocker les chaussures les plus vendues.

SECTION 2 : <u>Les Caractéristiques de Dispersion</u>:

Partons de 3 séries

Série 1 : 9 11 $\bar{X} = 10$

Série 2 : 5 15 $\bar{X} = 10$

Série 3 : 1 19 $\bar{X} = 10$

Les 3 séries ont la même moyenne : 10 et portant ils sont différents l'unes des autres.

Dans la 1ère série ; les valeurs du caractère sont proches de la moyenne. La moyenne est représentative.
Dans la 3ème Série les valeurs du caractère sont éloignées de la moyenne. Il faut donc lorsqu'on résume une série, indiquer par un nombre si les valeurs sont proches ou éloignées de la valeur centrale.
Ce nombre est appelé caractéristiques de dispersion.

I. L'intervalle de variation ou l'étendue :

C'est la différence entre la plus grande valeur du caractère et la plus petite.
L'intervalle de variation = Val MAX – Val MIN
$\Delta = 2$ Série 1 $\Delta = 10$ série 2 $\Delta = 18$ Série 3
Etendu ou intervalle de variation n'est pas un indicateur toujours fiable, car il dépend des valeurs extrêmes qui prouvent être fausses ou aberrantes.
EX :
17…………….18……………20…………60……….Age
1000 étudiants
$\Delta = 3$

$\Delta = 60 - 17 = 43$

II. L'intervalle inter quartile :

A- Définition des quartiles :

On appelle 1ér quartile Q1 la valeur du caractère tel que : 25% des observations lui sont inférieurs et 75% lui sont supérieurs. 25% < ; 75%>

2éme quartile Q2= Me 50% < 50%>

3émé quartile Q3= 75%< 25%>

B- Définition inter quartile :

On appelle inter quartile : Q3 – Q1 différence entre 1ér quartile et 3éme quartile.

N.B : Intervalle Inter quartile contient 50% des observations

C- Application :

N= 82

Rang : 82/4 =20 ,5

Classe : [15-20]

Interpolation : 15+ Δ

Salaires	Effectifs
10-15	9
15-20	25
20-25	32
25-30	16
Total	82

Ecart I. Inter quartile

Q3 – Q1
=24,3 - 17,3
= 7DH

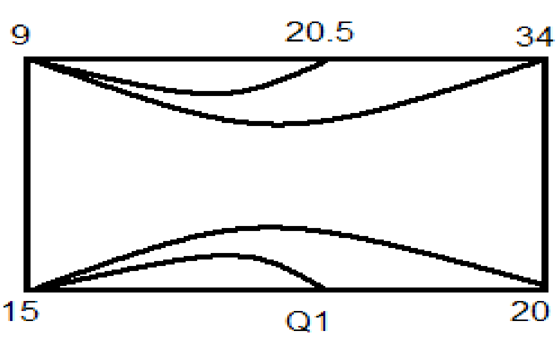
Ni Cum
9
34
66
82

Interprétation : Si 25 individus Augmentation de 5 DH
Si 01 Individu Augmentation 5/25 DH

(20,5 - 9) = 11,5 5/25 * 11,5

Donc Q1 = 15 + 5/25 *11,5 = **17,3 DH**

2éme Méthode :



Q1 -15 = 20.5 - 9
20 - 15 34 - 9

Calcul de Q3
Rang : 82*3/4 =61,5
Classe = [20-25]
Interpolation : si 32 individus augmentation de 5 DH

01 Individu Augmentation de 5/32

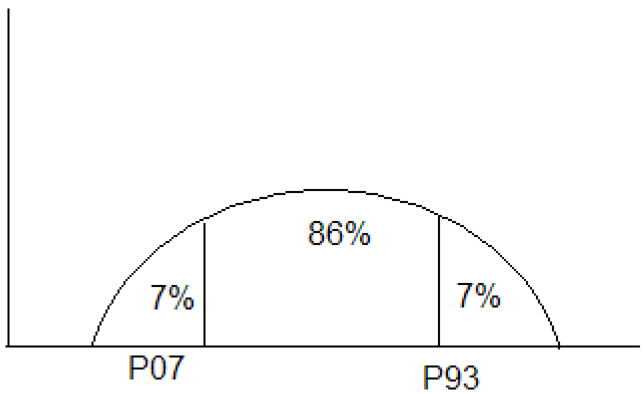
(61,5 – 34) = 27,5 individus Augmentation 5/32 *27,5

Donc Q3 = 20+ [(5/32) *27,5]
Signification : 24,3dh c'est le salaire tel que 75% gagnent plus de 24,3 et 25% gagnent moins de 24,3 DH.
Inter. Inter quartile : 7 DH = Q3-Q1
Signification : pour 50% des effectifs l'écart Maximum de salaire est de 7 DH

D – Remarque :

1- Les déciles : valeur du caractère que 10 % des observations ont une valeur qui est inférieure à D1 et 90% des observations ont une valeur qui est supérieure à D1.
On appelle 9 éme décile de 9 la valeur du caractère tel que 90% des observations lui sont inférieures, et 10% des observations lui sont supérieures. L'intervalle inter décile D9 - D1 contient 80% des observations

2- Les percentiles :
On appelle percentiles P1 la valeur du caractère telle que un pourcent (1%) des observations ont une valeur inférieure à P1 et 98% ont une valeur supérieure à P1.
Pour le statisticien KELLY pour supprimer les valeurs aberrantes il suffit de calculer l'intervalle inter percentile P_{93} –P_{07} qui contient 86% des observations.



L'écart absolu moyen :

A- Définition : On appelle écart absolu moyen que l'on désigne par la moyenne arithmétique des écarts absolus entre les valeurs du caractère et la moyenne arithmétique.

$$\boxed{C_a = \sum ni \left| xi - \bar{x} \right| / \sum ni}$$

B- Application : soit le tableau suivant :

Poids	ni	xi	ni * xi	$\lvert xi - \bar{x} \rvert$	$ni \lvert xi - \bar{x} \rvert$
55-60	12	57,5	690	10,25	123
60-65	17	62,5	1062,50	5,25	89,25
65-70	36	67,5	2430	0,25	9
70-75	24	72,5	1740	4,75	114
75-80	11	77,5	852,50	9,75	107,25
	100		**6775**		**442,5**

$C_a = 442.5 / 100 = 4.42$ Kg $\qquad \bar{X} = 67.75$ Kg

Signification : Ca = 4.42 Kg signifie qu'en moyenne, chaque individu s'éloigne de la moyenne (67.75 Kg) de 4.42 Kg.

Remarque : Pour dire si une dispersion est grande ou non, pour comparer deux séries entre elles, on se sert de $\boxed{\text{l'indice de dispersion relatif} = Ca / X * 100}$

Exemple :

Poids de filles	Poids des garçons
$\bar{X}$ =52 Kg	$\bar{X}$ =68 Kg
Ca= 2 Kg	Ca = 17 Kg
2/52 *100= 3.8%	17/68 * 100 = 25%
Dispersion Faible	dispersion plus importante

IV- La variance et l'écart type :

A- Définition :

On appelle une variance la moyenne arithmétique des carrés des écarts entre les valeurs du caractère et la moyenne arithmétique.

$$\boxed{\sigma^2 = \sum ni (xi - \bar{x})^2 / \sum ni}$$

On appelle écart-type (ou écart quadratique moyen) la racine carré de 6^2

$$\sigma = \sqrt{\sum ni(xi - \bar{x})^2 / \sum ni}$$

B- Application :
Le même tableau précédent

$(xi-\bar{X})^2$	$ni*(xi-\bar{X})^2$
105,0625	1260,75
27,5625	468,5625
0,0625	2,25
22,5625	541,50
95,0625	1045,6875
	3318,75

$$\sigma = \sqrt{\sum ni(xi - \bar{x})^2 / \sum ni} = \sqrt{3318.75/100} = 5.76$$

Signification : En moyenne chaque individu s'écarte du poids moyen (67.5 kg) de 5.76 kg.

C- Remarque :
Si on veut savoir la valeur de dispersion on utilise le cœfficient de variation = $\sigma/\bar{X}$
Ex :
$\bar{X}$ =67.75 Kg $\sigma/\bar{X}$ =(5.76/67.75) *100= 8.5%

Ex 2 :
Soient 2 modèles d'ampoules électrique dont on a relevé les durées de vie.

Modèle 1 : Durée de vie moyenne 1400 H.
Modèle 1 : Durée écart-type =100 H
Modèle 2 : Durée de vie moyenne 1800 H.
Modèle 2 : Durée écart-type = 250 H

Modèle I Modèle II

$6/\bar{X} = 100/1400 = 7\%$ $250/1800 * 100 = 14\%$

Le modèle I est plus faible que le modèle II

Formule développée :

Donc $\sigma = \sqrt{\frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}^2}$

Poids	ni	xi	xi	ni * xi
55-60	12	57,5	330625	39675
60-65	17	62,5	390625	66406,25
65-70	36	67,5	455625	164025
70-75	24	72,5	525625	126150
75-80	11	77,5	600625	66068,75
	100			**462325**

$$\sigma^2 = \frac{462325 - (67.75)^2}{100} \approx 33.19$$

$$\sigma = \sqrt{33.19} = 5.76$$

SECTION III : Les Caractéristiques de Concentration

La concentration ne s'applique qu'à des séries statistiques ou la concentration de la variable a un sens

EX : on peut parler de la concentration de revenus, concentration foncière

Autres EX : on ne peut pas parler de concentration d'âge

On peut déterminer la concentration soit algébriquement soit graphiquement

I. La détermination algébrique de la concentration

Cette détermination nécessite la connaissance de la « médiale »

Notion de la médiale (Ml)

A- La médiale

Si dans une série on désigne par xi la valeur du caractère, par ni les effectifs, la médiale est la valeur du caractère qui partage en deux parties égales le produit cumulé de ni xi.

Si xi désigne un salaire

Ni désigne le nombre de salariés

Le produit cumulé des ni xi représente la totalité des salaires Versés $\sum nixi$

C'est-à-dire la masse salariale.
La médiale, c'est le salaire tel que la moitié de la masse salariale a servi à payer une partie qui touche moins de cette Médiale et l'autre moitié de la masse s a servi à payer les gens qui touchent plus de cette Médiale.

B- Mesure de la concentration

ΔM sert à mesurer la différence entre ML et ME :

$\boxed{\Delta M = ML - ME}$

* Si ΔM = 0 cela veut dire que ML =ME
C'est-à-dire l'individu qui est au milieu l'effectif est en même temps celui qui est placé tel que la moitié de la masse salariale a été versée à des gens qui touchent moins que lui, et l'autre moitié à des gens qui reçoivent plus que lui, on a donc une distribution égalitaire concentration est nulle
* Si Δm $\neq$ 0 cela indique qu'il y a une concentration
* Si Δm est faible par rapport à l'intervalle de variation la concentration est faible
* Si $\frac{\Delta m}{\text{Inter variation}}$ est important, la concentration est forte

C- application

salaire	ni	xi	nixi	nixi
10-15	8	12.5	112.5	112.5
15-20	25	17.5	437.5	550
20-25	32	22.5	720	1270
25-30	16	27.5	440	1710
total	**82**		**1710**	

ΔM= ML – ME
Calcule de la ML :
Rang = 1710/2=855
Classe [20.25]
Interpolation linéaire

720 $\rightarrow$ 5dh
1dh $\rightarrow$ 5/720dh
(855-550) =3055 $\rightarrow$ 5/720*305dh

Donc ML= 20+5/720*350

ML = 22.12dh $\}\Longleftrightarrow$ ΔM = ML - ME
= 22 ,12 - 21,09 $\approx$ 1dh

ΔM/inter varia = 1/20=5%$\Rightarrow$ concentration faible
L'intervalle de variation
«Étant égale à : (30-10)=20
Signification ML = 22.12 dh

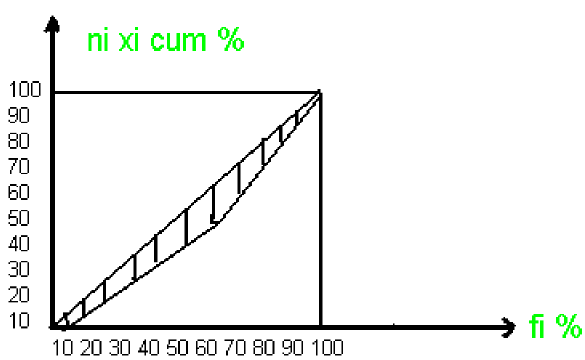
C'est le salaire tel que la moitié de la masse salariale a servi à payer des gens qui gagnent moins que 22.12 dh et l'autre moitié de la masse salariale a servi à payer les gens qui gagnent plus que 22.12 dh

II. La détermination graphique de la concentration la courbe de Lorentz GINI

A- la graphique de GINI

GINI propose de mesurer la concentration en mettant en abssices les fréquences cumulées en%, et en ordonnées ni xi cumulés en %

salaire	ni	Fi%	Fi% *n	xi	nixi	**Nixi%**	**Nixi%cum**
10-15	9	11	11	12.5	112.5	6.6	**6.6**
15-20	25	30.5	41.5	17.5	437.5	25.6	**32.2**
20-25	32	39	80.5	22.5	720	24.1	**74.3**
25-30	16	13.5	100	27.5	440	25.7	**100**
total	82	100			1710		



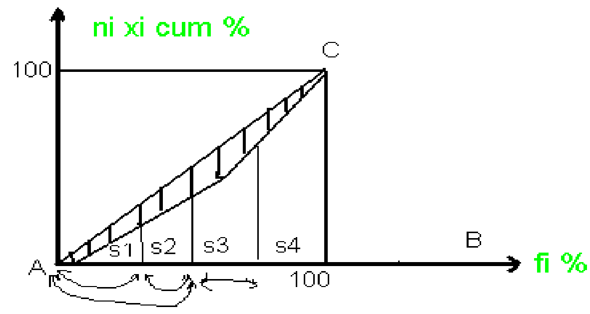
/ : Diagonal de l'égalité
: Aire de concentration

Remarques :

1) si 10% de la population touchent 10% du revenu, 20% de la population touchent 20% du revenu. Dans le cas d'une répartition égalitaire du salaire, l'aire de concentration serait confondue avec diagonal.
2) Dans le cas d'une repartions illégalitaire parfaite des salaires, (comme dans le cas théorique ou 0.1% de la population toucherait 99.99% de la masse salariale : la courbe

B)-Le coefficient de Gini :

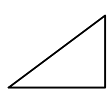
Gini propose de calculer la concentration à l’aide de coefficient suivant :



$$C = \frac{\text{Aire de concentration}}{\text{Aire du triangle ABC}}$$

$$C = \frac{\text{Aire de G}}{5000(100*100/2)}$$

On peu estimer l’aire de concentration de la manière suivant :
Aire de concentration = 5000-(S1+S2+S3+S3)



B S=1/2 a*b

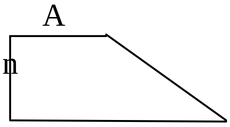
A

S1 = ½(116.6)
S2= (41.5-11)/2(6.6+32.2)
S3= (80.5-41.5)/2(32.2+74.3)
S4 = (100-80.5)/2(74.3+100)

$\sum$ Si = 4404

A

n



S = n/2(a+b)

b

Remarque : 0<c<1

c = 0 Concentration élevé

c = 1 Concentration faible

Donc $\boxed{c=5000-4404/5000 \approx 0.12}$

C à d les gens sont pareils

CHAPITRE III :LES SERIES A DOUBLE ENTREES : REGRESSION LINEAIRE (CORRELATION)

I- notion de tableau de contingence :

A. une distribution statistique double

C'est une distribution ou l'observation s'effectue selon 2 caractères.

EX : Répartition des étudiants selon la taille et l'âge

Répartition des logements selon le nbre de pièces et superficie

superficie / nbr de piece	10-30	30-50	50-70	70-80	total
1	3	1			
2	1	14	3		18
3		1	7	4	12
4			10	7	17
5			6	6	6
total	4	16	20	17	57

B. distributions marginales

Ce sont les distributions relatives à la seul variable X ou Y

a- la répartition des logements selon le nombre de pièces (X)

Nbre de pièces (x)	Nbre de logement
1	4
2	18
3	12
4	17
5	6
total	57

Cette distribution qui concerne la seule variable x est appllée distribution marginale (marginal car on la trouve à la marge du tableau statistique)

On peut calculer la moyenne de cette distribution, (et sa signification est le nbre de pièces moyenne par logement)

Moyenne appelée moy.marginale notée

b- la répartition des logements selon superficie :

superficie y	*Nbre de logements*
10-30	**4**
30-50	**16**
50-70	**20**
70-80	**17**
total	**57**

Cette distribution qui concerne la seule variable ‘ y’ est appelée distribution marginale on peut calculer la moyenne (qui exprime la surface moy des logements) appllée moy.marginal notée $\bar{\bar{Y}}$

C. *Les distributions conditionnelles :*

On appelle distribution Conditionnelle la distribution ou l’on a posé une condition sur l’une des variables.
Ex : Réparation de logements de 30-50m
Cette distribution est appelée Distribution Conditionnelle parce que l’on ne s’intéresse qu’aux logements qui satisfont la condition de surface 30-50 m^2.
On peut calculer la moyenne de cette distribution (c-a-d le nombre moyen de pièces des logts de 30-50 m^2) on appelle cette moyenne : moyenne conditionnelle.
Dans cet exercice on calcule
Remarque il existe autant de distributions conditionnelles relatives au caractère x que le caractère y a de modalités

II- généralisation du tableau de contingences :

x y	Y_1	Y_2	………	Y_j	………	Y_m	total
X_1	X_{11}	X_{12}	………	X_{1j}	………	X_{1m}	$X_{1.}$
X_2	X_{21}	…	………	X_{2j}	………	X_{2m}	$X_{2.}$
…	…		………	…	………	…	…
X_i	X_{i1}	X_{i2}	………	X_{ij}	………	X_{im}	$X_{i.}$
…	…	…	………	…	………	…	…
X_k	X_{k1}	X_{k2}	………	X_{kj}	………	X_{km}	$X_{k.}$
total		$x_{.2}$		$x_{.j}$		$x_{.m}$	$x_{..}$

$x_1\, x_2\, \ldots\, x_k =$ les modalités de x
$y_1\, y_2\, \ldots\, y_k =$ les modalités de y
x_1 .effectifs pour la $1^{ére}$ modalités de x et pour toutes les modalités de y

La distribution marginale de X :

X(xi)	$X_{i.}$
X_1	$X_{1.}$
X_2	$X_{2.}$
.	.
.	.
X_i	$X_{i.}$
X_k	$X_{k.}$
Total	$X_{..}$

La distribution marginale de y :

y(xi)	Xj.
y_1	$X_{.1}$
y_2	$X_{.2}$
.	.
.	.
y_i	$X_{.i}$
y_m	$X_{m.}$
Total	X..

Distribution conditionnelle relatif à X et à Y

Dist. Conditionnelle relative à X

X	X_{ij}
X_1	X_{1j}
X_2	X_{2j}
.	.
.	.
X_i	X_{ij}
X_k	X_{kj}
Total	X.j

Dist. Conditionnelle relative à Y

y	X_{ij}
y_1	X_{i1}
y_2	X_{i2}
.	.
.	.
y_i	X_{ij}
y_m	X_{im}
Total	Xi.

III- La régression linéaire

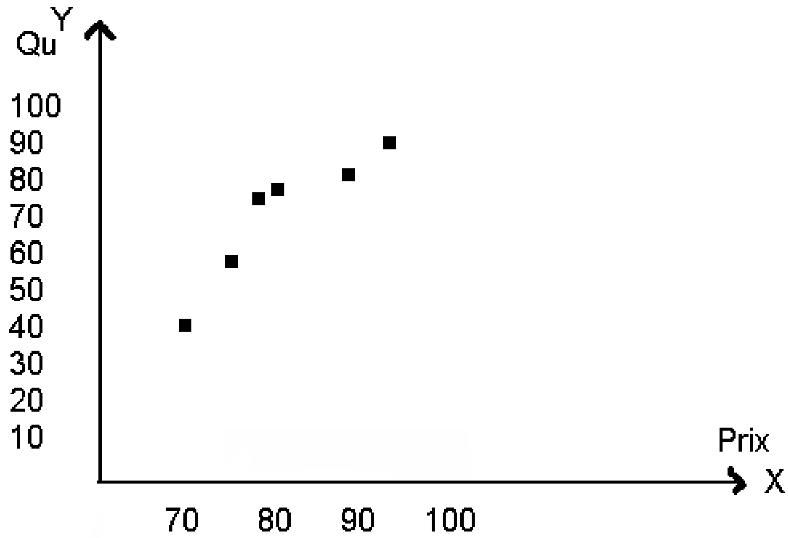
A. Présentation du problème :

Soit le tableau suivant :

qu	42	51	60	62	74	83	Total
70							1
75		1					1
77							1
80				1			1
86							1
93						1	1
Total	1	1	1	1	1	1	6

Ce tableau est un tableau de contingence ou les observations sont connues individuellement, on peut présenter plus simplement ce tableau de la manière suivante :

PRIX	QU
70	42
75	57
77	60
80	62
86	74
93	83
Total	



Nous avons un ensemble de points « un nuage statistique »qui nous indique que les prix est les quantités évoluent selon la même tendance.
Il est possible de schématiser ce nuage :
-Par une fonction simple : la fonction linéaire (Droite) qui sont inconnus et qu'il faudra trouver.

a=pente de droite
b=ordonnée à l'origine

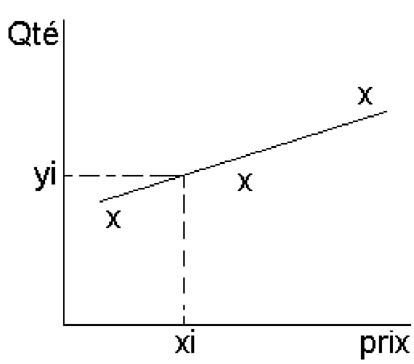
Une telle droite est appellée droite de régression D(x)
A=coefficient de régression
La régression c'est le fait de relier y à x par une fonction
Calcule des paramètres de la droite de régression :

B.

la méthode des moindres carrés

Notion de moindres carrés :

Partons d'un nuage statistique théorique :



- Il s'agit de résumer ce nuage par une droite.
- Soit y'= ax+b l'équation de la droite recherchée.
- Pour toute valeur de x (x_i) nous avons une valeur réellement observée y'.
- Pour toute valeur x_i, nous avons une valeur calculée sur la droite y'.
- Pour toute une valeur xi, nous avons une erreur d'estimation égale à $| y_i – y'_i |$.

- La droite de régression idéale doit être de telle manière que la somme des erreurs d'estimation doit être la plus faible possible, $\sum | y_i - y'_i |$ doit être minimum.
- Pour éviter les valeurs absolues, on convient de calculer les carrés des erreurs. La droite de régression doit être telle que :

$\sum (y_i - y'_i)^2$ minimum, et on appelle cela la condition des moindres carrés.

C. Calcul des paramètres de la droite de régression.

Il s'agit de trouver y'= ax + b sachant que : $\sum (y_i - y'_i)^2$ min.
Remplaçons y'i par sa valeur ➔ $\sum (y_i - (ax_i+b))^2$ min.
Posons $\sum (y_i - ax ; - f)^2 = Z$ (a , b).
Pour que Z soit minimum, il suffit d'annuler (rendre nul) les dérivés de ce polynôme par rapport à 'a' et par rapport à 'b'.

1 – Calcul de b :

Supposons 'a' est connu, et dérivons par rapport a 'b' et 'a'.
dZ / db = 2 [$\sum (y_i - ax ; -b)$] (-1) = 0

$Z = U^2 \Longrightarrow Z' = 2UU'$

$U' = (y_i - ax ; -b)$

$\sum$ [yi – ax ; -b) = 0
$\sum$yi – a$\sum$xi – nb = 0

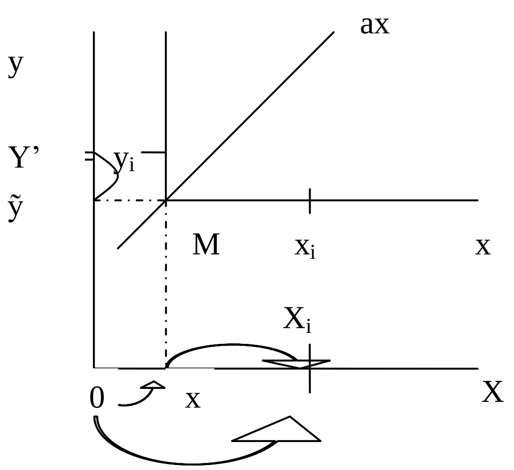
Divisons par n, on obtient ($\sum$yi / n – a$\sum$xi / n –b = 0 $\boxed{\tilde{y} - a\bar{x} - b}$

Donc :

$\boxed{b = \tilde{y} - a\bar{x}}$

La droite de régression passe donc par le point moyen ($\bar{x}$, ỹ).

2 – Calcul des a :



Le paramètre a Que nous cherchons correspond à la pente de la droite de régression qui passe par le point moyen M ($\bar{x}$; ỹ).
Procédons un changement d'origine, et prenons comme nouvelle origine le point moyen M(x' ; ỹ), les nouvelles cordonnées deviennent :

$$\begin{cases} X_i = x_i - \bar{x} \\ Y_i = y_i - \tilde{y} \end{cases}$$

La droite de régression a pour équation y' = ax

La condition des moindres carrée s'écrit ;

$\sum (y_i - \tilde{y}_i)^2$ min

$\sum (y_i - y'_i)^2 = \sum (y_i - ax_i)^2$ min

Dérivons par rapport à 'a' : $2 [\sum (y_i - ax_i)] (-X_i) = 0$

$[\sum (y_i - ax_i)] X_i = 0 \Rightarrow \sum (y_i - a_i) X_i = 0 \Rightarrow \sum x_i y_i - a \sum x_i^2 = 0$

Donc $a = \sum x^i y^i / \sum x^i = \sum (x_i - x')(y_i - \tilde{y}) / \sum (x_i - x')$

3- l'équation de la droite de régression

Dy(x) =

Y = ax + b

$$\begin{cases} a = \sum (x_i - \bar{x})(y_i - \tilde{y}) / \sum (x_i - \bar{x})^2 \\ b = \tilde{y} - a\tilde{y} \end{cases}$$

D – Application:

Prix(x)	Qtés(y)
70	72
75	51
77	60
80	62
86	74
33	83
481	372

Dy (x) a pour équation:

Y = ax + b

$$a = \frac{\sum xi\,xi}{\sum x^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = 481 / 6 = 80$$

$$\bar{y} = 372 / 6 = 62$$

Trouver Dy (x).

$x_i - \bar{x}$	$y_i - \tilde{y}$	$(x_i - \bar{x})(y_i - \tilde{y})$	$(x_i - \bar{x})^2$
-10	-20	200	100
-5	-11	55	25
-3	-2	6	9
0	0	0	0
6	12	72	36
13	21	273	169
		606	339

a = 606 / 339 = 1.79

b = 62 – (1.73)80

b = -81

Donc

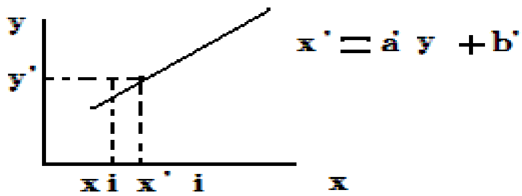
$Dy_{(x)}$ a pour équation :

$\boxed{y = 1.79x - 81}$

La loi de l'offre pour ce bien

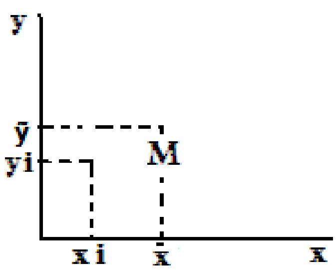
IV- la corrélation linéaire :

Dans le paragraphe précédent, nous avions estimé y en fonction de x, et nous avions obtenu la droite de régression $Dy_{(x)}$
On peut pour le même nuage statistique estimer x en fonction de y, et trouver la droite de régression Dx(y) lui aura pour équation.



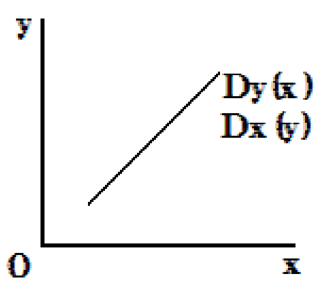
Pour toute yi, nous avons une valeur observée x_i.
Pour toute y_i, nous avons une valeur estimée sur la droite x'_i
Pour toute y_i, nous avons une erreur d'estimation égale à $| x_i - x'_i$
Dx(y) idéale est tel que : $\sum | x_i - x'_i |$ minimum ou encore $\sum (x_i - x'_i)^2$ minimum
En procédant de la même manière que dans le paragraphe précédent, on trouve l'équation de Dx(y).
$X = a'y + b'$

$$\begin{cases} a' = \dfrac{\sum x_i\, y_i}{\sum y_i^2} \\ b' = \bar{x} - a'\,\tilde{y} \end{cases}$$



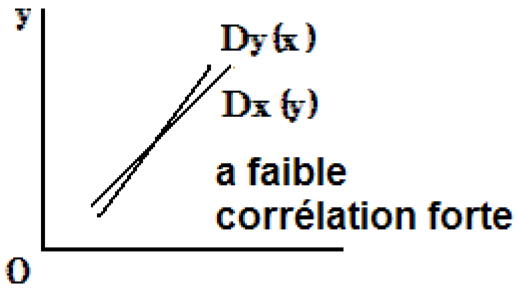
Dans le référentiel XMY nous obtenons 2 droites :
Soit y = ax pour Dy(x)
Soit x = a'y pour Dx(y)
Ou encore y = 1/a' x
4 cas peuvent se produire :

1er cas : les 2 droites sont confondues

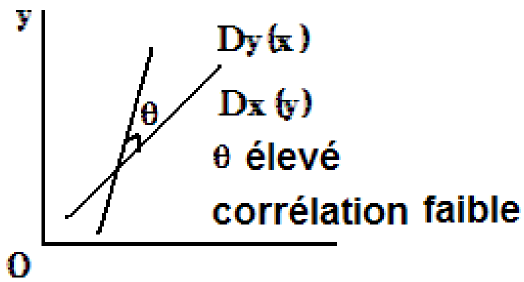


$$\left.\begin{array}{l} Y = ax \\ X = a'y \\ Y = 1/y'x \end{array}\right\} \Leftrightarrow a = 1/a' \Leftrightarrow aa' = 1$$

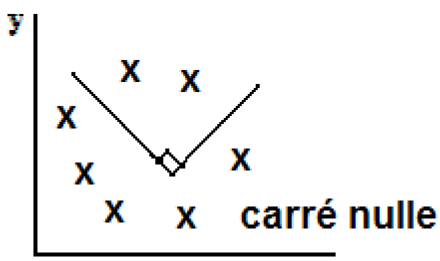
2éme cas : les 2 droite font entre elles un angle très faible :



3éme cas : les 2 droite font entre elles un angle élevé :



4éme cas : les 2 variables sont indépendantes l'une de l'autre :



Si on appelle coeff de corrélation la Quantité r tel que : r^2 = a . a', on peut écrire :

- Si r = ±1 on a une corrélation parfaite.
- Si r = +1 on a une corrélation parfaite positive.
- Si r = -1 on a une corrélation parfaite.

Corr. positive : c à d les variables varient dans le même sens.

- Si r = -1 = corrélation parfaite négative.

C à d les deux phénomènes varient en sens inverse.
Par exemple Prix et Quantité

- Si 0 < r < 1 = la corrélation est positive, elle est d'autant plus forte que l'on se rapproche de 1.
- Si -1 < r < 0 = la corrélation est négative, et elle est d'autant plus forte que l'on se rapproche de -1.
- Si r = 0 = corrélation nulle.

Application : calculer le coefficient de corrélation d'une autre façon (existe-t-il un lien entre y et x).

Prix	Qté	$x - \bar{x}$	$y - \tilde{y}$	$(x_i - \bar{x})(y_i - \tilde{y})$	$(x_i - \bar{x})^2 \; (y_i - \tilde{y})^2$
70	42				
75	51				
77	60				
80	62				
86	74				
33	83				
			606	339	1110

$r^2 = a.\,a' = \frac{606}{339} \times \frac{606}{1110}$ donc r = 0.98

$$a = \frac{\sum x_i y_i}{\sum x^2_i} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{606}{339} = 1.79$$

$$a' = \frac{\sum x_i y_i}{\sum y^2_i} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} = \frac{606}{1110} = 0.545$$

On a une très forte corrélation car r = 0.975 tend vers 1

⃠Remarque : lorsqu'on écrit $r^2 = a.\,a'$ ➔ r = racine a .a', nous avons une expression très positif. Comment trouver alors le signe d'une corrélation ?

Réponse : le sens de la corrélation est donnée par le signe de a et a'.

- Si a et a' sont >0➔ le produit a.a' >0➔ corrélation positive.
- Si a et a' sont <0➔ le produit a.a'>0 ➔corrélation négative.

On peut dire d'une corrélation qu'elle est très satisfaisante à partir 0.86.

On peut dire d'une corrélation qu'elle parfaite à partir de 0.96.

IV – formule facilitant les calculs :

1/ calcul de a :

$a = \frac{\sum (x_i - \bar{x})(y_i - \tilde{y})}{\sum (x_i - x')^2} = \frac{N}{D}$, $N = \sum x_i y_i - \tilde{y} \sum x_i - \bar{x} \sum y_i + \sum \bar{x}\,\tilde{y}$

Or $\bar{x} = \frac{\sum x_i}{N}$ ➔ $\sum x_i = n\bar{x}$

$\tilde{y} = \frac{\sum y_i}{N}$ ➔ $\sum y_i = n\,\tilde{y}$

On remplace : $N = \sum x_i y_i - \cancel{\tilde{y} n \bar{x}} - n \bar{x}\tilde{y} + \cancel{n \bar{x}\tilde{y}}$

$$\boxed{N = \sum x_i y_i - n\bar{x}\bar{y}}$$

$D = \sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2$

$= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$

$$\boxed{D = \sum xi^2 - n\bar{x}^2}$$

Donc $a = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2}$

Formule développée

xi	y_i	$X_i\ y_i$	x_i^2

x' ỹ

2 – calcul de r :

r2= a.a' $a = \dfrac{\sum x_i y_i - n\bar{x}\tilde{y}}{\sum x_i^2 - n\bar{x}^2}$

$a' = \dfrac{\sum x_i y_i - n\bar{x}\tilde{y}}{\sum y_i^2 - n\tilde{y}^2}$

Donc $r = \sqrt{a * a'}$

V – Autre formule de r :

$$r = \frac{[\sum (x_i - \bar{x})(y_i - \tilde{y})]^2}{\sum (x_i - \bar{x})^2 \sum (y - \tilde{y})^2}$$

Or $б^2x = \dfrac{\sum (x_i - \bar{x})^2}{n}$ ➔ $\sum (x_i - \bar{x})^2 = nб^2x$

$б^2y = \dfrac{\sum (y_i - \tilde{y})^2}{n}$ ➔ $\sum (y_i - \tilde{y})^2 = nб^2y$

Donc $r == \dfrac{\sum [(x_i - \bar{x})(y_i - \bar{y})]^2}{n^2 . \delta^2 x . \delta^2 y}$

$$\frac{\sum (x_i - \bar{x})(y_i - \tilde{y})}{n.бx.бy}$$

Si on appelle : covariance de x et de y l'expression :

$Cov\,(xy) = \dfrac{\sum (x_i - \bar{x})(y_i - \tilde{y})}{n}$

r s'écrit : $r = \dfrac{Cov\,(xy)}{бx.бy}$

CHAPITRE IV : ANALYSE DES SERIES CHRONOLOGIQUES.

I – Généralités :

A.

Définition :

Une série chronologique est une série où les observations de la variable sont faites à des intervalles réguliers de temps.

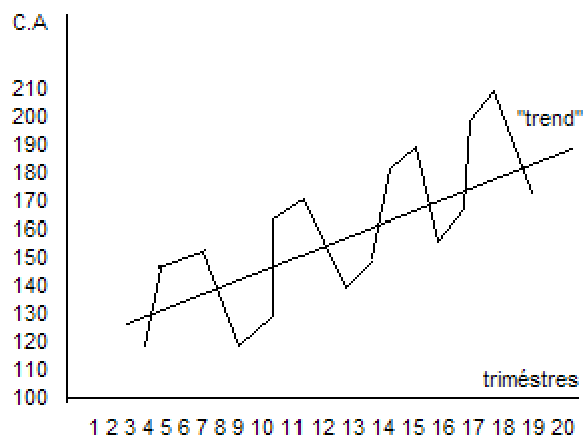
B.

les différentes composantes d'une série chronologique.

Soit la série chronologique suivante : Evolution trimestrielle du chiffre d'affaire d'une entreprise

trimètres	1	2	3	4
1998	120	148	155	120
1999	130	162	169	132
2000	144	178	186	145
2001	157	196	210	160

Représentation graphique de la série :



L'examen d'une série chronologique révèle l'existence de différences composantes :
Un mouvement de tendance longue (à long terme), appelée « trend ».
Un mouvement saisonnier qui est les variations saisonnières.
Des variations accidentelles : ce sont des variations imprévisibles dues à des circonstances exceptionnelles.

C.

intérêt d'une analyse d'une série chronologique :

L'analyse des séries chronologiques permet de séparer le mouvement de long terme du mouvement saisonnier, ce qui nous permettra de faire des calculs de prévision.

II – l'analyse de la tendance longue : « trend »

Déterminer le trend, cela revient à « lisser » la série pour éliminer les variations saisonnières, cette technique de « lissage » de la série est appelée Ajustement. Les 2 méthodes d'ajustement les plus utilisés sont :

- La méthode des moyennes mobiles.
- L'ajustement analytique.

A. la méthode des moyennes mobiles :

Elle consiste à diviser un nuage statistique en « sous – nuages » comprenant chacune (n–1) données du sous nuages précédent, et à remplacer chaque sous nuage par un point tel que : x'i = médiane des x_i – y_i = moyenne des valeurs y_i.

B. Opérations sur les matrices :

1 – matrices transposées :

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & -1 & 5 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 \\ 3 & -1 \\ 4 & 5 \end{bmatrix}$$

2 – L'addition :

$$\begin{bmatrix} 1 & -1 & 3 \\ 2 & 4 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 4 & 3 \\ 3 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 6 \\ 5 & 3 & 0 \end{bmatrix}$$

$(aij) + (b_ij) = (a_ij + b_ij)$

Propriétés :

- commutativité
- association
- élément neutre
- élément symétrique $\quad a_{ii} = 0_{(n\,;p)}$ la matrice nulle

$t\,(a+b) = t_a + t_b$

3- Multiplication par un réel :

$$3 * \begin{bmatrix} 1 & -1 & 3 \\ 2 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -3 & 9 \\ 6 & 12 & 3 \end{bmatrix}$$

CHAPITRE V :POPULATIONS ET ECHANTILLONS, RECENSEMENTS ET SONDAGES

Les journaux, la télévision, les revues nous inondent constamment de graphiques, de tableaux et de statistiques de toutes sortes, dans différents domaines :

Ces présentations peuvent parfois nous induire en erreur volontairement ou non.

Il nous faut donc développer un esprit critique et savoir interpréter ces informations.

I. Quelques termes de base :

La population cible est l'ensemble de tous les objets que l'on étudie.

Une unité statistique est un objet de cette population.

Un échantillon est une partie choisie d'une population.

Le nombre d'objets composant une population ou un échantillon est appelé sa **taille.**

Lorsque l'on veut connaître certaines **caractéristiques** d'une population, on dit qu'on **enquête** sur la population.
Une enquête peut être réalisée auprès de toute la population ou sur un échantillon.

Un recensement est une enquête réalisée auprès de toute la population.

Un sondage est une enquête réalisée sur un échantillon.

1. Étude portant sur la langue maternelle des Québécois:

   la population est l'ensemble des Québécois
   et la caractéristique est la langue maternelle.

2. Étude portant sur la durée des ampoules électriques produites à l'usine X.

   La population est constituée des ampoules électriques produites à l'usine X
   et la caractéristique étudiée est la durée des ampoules.

3. Une compagnie pharmaceutique veut vérifier un nouveau vaccin contre une certaine maladie. On administre ce produit à 50 patients atteints de la maladie.

   La population est formée de tous les gens atteints de la maladie,
   l'échantillon est formé des 50 patients à qui on a administré le médicament et la caractéristique étudiée est la réponse au médicament.

Les coûts élevés et les délais trop longs, reliés à un recensement, sont les principales raisons qui nous amènent à utiliser un sondage puisque la taille d'un échantillon est beaucoup plus petite que celle de la population.
Au Canada, il y a un recensement tous les cinq ans. Le dernier date de 1996.

EXERCICES

I OBJECTIFS VISES :

1. construction d'un tableau statistique :
2. distinguer une variable quantitative d'une variable qualitative
3. représentation graphique des variables quantitatives discrètes et continues
4. calcul et interprétation des caractéristiques de tendance centrale :
   - moyenne.
   - médiane
   - mode
   - quartiles
5. calcul et interprétation des caractéristiques de dispersion :
   - variance
   - écart type
   - coefficient de variation

Exercice 1 :

Dans une entreprise de 80 salariés on a enregistré les salaires mensuels suivants :

- 54 salariés gagnent 6 000 dirhams ou plus ;
- 34 salariés gagnent 8 000 dirhams ou plus ;
- 20 salariés gagnent 10 000 dirhams ou plus ;
- 8 salariés gagnent 12 000 dirhams ou plus ;

1. Présenter ces données dans un tableau avec des classes de même amplitude en sachant qu'aucun salarié ne gagne plus de 14 000 DH.
2. Calculer la moyenne et donner sa signification.
3. Calculer la médiane et donner sa signification.
4. Calculer le mode graphiquement, algébriquement et donner sa signification.
5. Combien gagnent les 20% des salariés les mieux payés.

Exercice 2 :

La répartition des salariés d'une entreprise de confection selon leurs gains mensuels (en milliers de dirhams) se présente comme suit :

Gains mensuels	effectifs
[4-6[	25
[6-8[	40
[8-12[	58
[12-18[	27
[18-20[	6
20 et plus	4

1. déterminer graphiquement le salaire modal
2. calculer le coefficient de variation
3. calculer l'étendue
4. calculer algébriquement et graphiquement la médiane.

Exercice 3 :

La répartition par âge d'une population d'un centre de vacances est comme suit :

Classe d'age (en années)	effectifs
0-5	16
5-15	42
15-25	44
25-35	40
35-45	30
45-55	32
55-60	15
60-75	36
75-100	15

1. tracer l'histogramme de cette distribution
2. calculer l'écart type et donner sa signification
3. on désire rajeunir cette population en invitant au centre des vacances des personnes de la classe [25-35[.combien faudrait-il en faire venir pour que la moyenne de la population soit de 35 ans.

Exercice 4 :

Dans une commune urbaine, on a relevé la répartition en pourcentages de 10 000 contribuables selon le montant des impôts payés.

Classes d'impôts	Fréquences relatives en pourcentages
1-3	8
3-6	12
6-L2	20
L2-12	26
12-18	F6
18-22	10
22-30	6

1. Trouver les valeurs manquantes de ce tableau sachant que la moyenne est égale à 11,42
2. tracer la courbe cumulative croissante
3. déterminer graphiquement et algébriquement l'impôt médian. donner sa signification
4. quel est le pourcentage des contribuables qui paient un impôt annuel supérieur à 20 000dh ?cela représente combien de personnes ?

Exercice 5 :

Soit la distribution statistique suivante qui donne la répartition des propriétaires terriens selon la superficie des terres cultivables dans une certaine région agricole :

Superficie des terres en hectares	Nombre de propriétaires
2-4	24
4-8	36
8-14	22
14-20	18
20-40	14
40-100	6

Partie I :

1. préciser le caractère étudié et préciser sa nature.
2. donner la signification de du centre de la $2^{ème}$ classe.
3. déterminer rapidement la médiane et donner sa signification
4. déterminer algébriquement le mode et donner sa signification
5. calculer la superficie moyenne et l'écart type. Que peut on conclure ?
6. déterminer le 1_{er} et le $9^{ème}$ décile et donner leurs significations

Partie II :

1. déterminer graphiquement la concentration foncière dans cette région agricole, Calculer l'indice de GINI
2. déterminer algébriquement la concentration
3. déterminer graphiquement le pourcentage des propriétaires dont la superficie des terres est inférieure à la médiale.

Exercice 6 :

Pendant 9 années les bénéfices d'une entreprise ont augmenté :

- de 4% par an pendant les 3 premières années.
- de 7% par an pendant les 4 années suivantes.

De 10% par an pendant les 2 dernières années de la période considérée.

Quelle est l'augmentation moyenne des bénéfices de cette entreprise sur les 9 années ?

Exercice 7 :

Le tableau suivant donne la répartition des salaires mensuels des cadres d'une entreprise :

Salaires en 1000DH	Nombre des cadres
6-8	50
8-10	70
10-16	80
16-22	50
22-30	50
30-34	80
34-38	20
total	400

1. préciser le caractère étudié et sa nature
2. représenter graphiquement cette distribution, tracer le polygone des fréquences
3. déterminer rapidement :
   - le salaire médian des cadres donner sa signification.
   - Le $3^{ème}$ quartile (Q3). donner sa signification.
4. donner graphiquement le salaire modal des cadres.
5. calculer le salaire moyen des cadres.
6. Calculer le coefficient de variation et donner sa signification
7. Pour motiver davantage ses cadres, l'entreprise décide une augmentation générale des salaires de 20%. Calculer la nouvelle moyenne et le nouveau coefficient de variation.

II OBJECTIFS VISES :

1. Calcul de la fonction linéaire
2. calcul et commentaire du coefficient de corrélation
3. interprétation des distributions marginales
4. interprétation des distributions conditionnelles

Exercice 8 :

Une entreprise a présenté ses dépenses de publicité et ses chiffres pour les 6 dernières années dans le tableau suivant (en 10^6 DH)

Dépenses de publicité	Chiffre d'affaires
2	10
4	16
10	50
14	120
18	140
24	210

1. L'entreprise pense qu il y'a un lien entre dépenses de publicité (X) et le chiffre d'affaire(Y).pouvez vous le confirmer ?
2. établir par la méthode des moindres carrés la relation liant le chiffre d'affaires et les dépenses de publicité
3. combien l'entreprise peut-elle espérer réaliser comme chiffre d'affaireS avec des dépenses de publicité de 30 ?

Exercice 9 :

On a observé une population en retenant 2 caractères : le nombre d'enfants(X) et la taille du logement (Y).les résultats sont les suivants :

Nombre de pièces / Nombre d'enfants	2	3	4	Total
1	22	15	9	46
2	7	38	22	67
3	0	7	30	37
Total	29	60	61	150

1. calculer le nombre moyen d'enfants et le nombre moyen de pièces des logements.
2. calculer $\bar{x}^2$ et donner sa signification
3. calculer $\bar{y}_3$ et donner sa signification
4. on se propose de voir s'il existe un lien entre le nombre d'enfants et la surface des logements. Confirmer

Exercice 10 :

Le tableau suivant donne la répartition des salariés d’une entreprise de bâtiment selon le nombre d’enfants à charge X et les salaires mensuels perçus y en milliers de DH

Nombre de pièces Y Nombre d’enfants X	1-3	3-5	5-9
1	4	8	**<u>16</u>**
2	6	12	24
3	**<u>3</u>**	6	12
4	2	4	8

1. donner la distribution marginale de la variable X
2. donner la distribution conditionnelle de la variable Y liée à la modalité 4 de X.
3. que signifient les valeurs 16 et 3 soulignée dans le tableau
4. vérifier de deux manières différentes que les deux variables sont indépendantes. Dites dans ce cas à est égal le coefficient de corrélation linéaire : r (sans le calculer.
5. calculer la variance marginale de Y.

Exercice 11 :

Une étude réalisée dans un club sportif concernant le poids et la taille de 124 adhérents a fourni les informations suivantes :

poids en Kg Y taille en mètres X	50-60	60-65	65-75	75-80
1,60-1,70	12	7	6	4
1,70-1,75	?	6	8	3
1,75-1,80	9	8	<u>8</u>	4
1,80-1,90	?	<u>7</u>	5	6
1,90-2,00	3	5	3	3

1. compléter le tableau sachant qu’il y a 27 adhérents qui mesurent entre 1.70met 1.75m.
2. quels sont les caractères étudiés ? Quelle est leur nature ?
3. que signifient les chiffres 7 et 8soulignés dans le tableau
4. quelle est la moyenne du poids des adhérents ? Comment appelle-t-on cette moyenne ?
5. quelle est la taille moyenne des adhérents ? Comment appelle-t-on cette moyenne ?
6. en désignant par X la taille et par Y le poids calculer et donner la signification de $\bar{y}_2$
7. donner sans la calculer la signification de $\bar{x}_3$

Exercice 12 :

Une entreprise commerciale a présenté ses ventes xi et ses frais de publicité yi au cours du premier semestre de l’année 2003 comme suit (en 1000 DH)

Mois	Ventes	Frais de publicité
Janvier	40	1.1
Février	30	0.8
Mars	42	1.2
Avril	46	1.4
Mai	44	1.3
juin	38	1.1

1. déterminer une fonction linéaire qui donne le montant des ventes lorsqu’on connaît les frais de publicité.
2. quel serait le montant des ventes si les frais de publicité atteindront 3500DH.
3. déterminer s’il y a ou non une liaison entre les ventes et les frais de publicité.