Clp I ELEMENTS DE LA TERMINOLOGIE

I-Définition statistique

La statistique est une science qui élabore ou besoins est des méthodes de collecte, d'analyse quantitative et qualitative des donnes ou informations pour une prise de décision. Utilisées au pluriel, les statistiques désignent les chiffres ou les nombres.

II- La terminologie de base

1. La population statistique

La population statistique est l'ensemble des faits d'individus ou de phénomène ou d'objets sur lequel porte une étude statistique.

Exemple : si notre intention est de faire une étude sur les étudiants de ……T C2-B…… alors la population statistique est les étudiants de ……T.C.2.-B……

1- Unité statistique (individu)

C'est l'élément constitutif de la population statistique.

Exemple : Un individu des étudiants ……de TC2-B…… sur lesquelles on effectue l'étude statistique.

2- Caractère statistique

C'est une propriété distinctive de l'unité statistique à laquelle on s'intéresse (religion, le sexe, la taille).

3- La modalité statistique

C'est l'ensemble des valeurs ou situations possible qu'on attribue avec caractère statistique.

Exemple : le sexe a deux modalités (féminin et masculin)

On distingue 2 types de caractère statistique :

- ✓ Le caractère qualitatif : les modalités ne sont pas chiffrables, mesurables, quantifiable.
- ✓ Le caractère quantitatif : c'est le caractère dont les modalités sont mesurables (âge, taille, poids …)

4- Variable statistique

C'est la valeur qu'on attribue à un caractère statistique. La variable statistique est note **xi**. On distingue deux types de variable :

- La variable continue : les modalités sont en intervalle.
- La variable discrète : dont les modalités prennent des valeurs isolées ou entières.

5- Effectif

C'est le nombre d'unité statistique possédant une modalité donnée. Il est noté n_i, l'effectif est également appelé la **Fréquence absolue.**

6- La fréquence

C'est la proportion d'unité statistique possédant une modalité spécifique ou particulière note fi.

La fréquence se calcule comme suite :

$$f_i = \frac{ni}{N} * 100$$

7- Enquête statistique

C'est une opération statistique destinée à la collecte des données d'informations. Elle se fait par le biais des questionnaires.

Lorsque l'enquête est faite sur toutes les unités de la population statistique on parle d'enquête **exhaustive ou recensement**. Mais lorsqu'elle est faite sur une partie ou une portion de cette population on parle alors de **sondage.**

8- Le dépouillement ou le pointage

C'est l'opération qui consiste à dénombrer ou à compter les observations ou données recueillies à l'issus d'une enquête statistique. Il se fait par l'intermédiaire du pointage.

Chp 2 TABLEAUX STATISTIQUES ET REPRESENTATION GRAPHIQUES

Section 1 TABLEAUX STATISTIQUES

I. Définition

Le tableau statistique ou distribution statistique est un tableau dans lequel sont consignées les données ou observations recueillies à l'issue ou au terme d'une enquête statistique.

II. Types de tableaux statistiques

On distingue deux types de tableaux statistiques :

I. Tableau statistique simple

Ce tableau permet de faire une description selon un seul caractère.

Exemple : la répartition des étudiants selon leur sexe

sexe	Pointage	n_i	$d°$
Féminin		10	88
Masculin		31	272
Total	—	41	360

41 → 360°
n_i → ?

II. Tableau statistique à double entrée ou tableau de contingence :

Il permet de faire une description selon deux caractères simultanément

Exemple : la répartition des étudiants selon leur sexe et la série du BAC 2

SERIE DE BAC 2 / SEXE	~~A~~ C	D	G_2	Total
Féminin	2	8		10
Masculin	7			31
total	9	32		41

NB : Pour qu'un tableau statistique ait de la valeur, il lui faut un titre simple, clair et précis permettant de décrire un phénomène concret en un lieu et à une date déterminée.

Section 2 : LES REPRESENTATIONS GRAPHIQUES

La représentation graphique des données chiffrées se fait par le biais des figures géométriques (points, lignes, surfaces …). On distingue les graphiques d'illustration et les courbes cumulatives.

1. Les graphiques d'illustration

a) Caractère qualitatif

On retiendra, pour illustrer une distribution de caractère qualitatif, les diagrammes à secteur (circulaire ou semi-circulaire), le diagramme à tuyaux d'orgue ou le diagramme à barres.

Exemple : répartition selon le sexe

sexe	Pointage	n_i	degré
Féminin		10	88
Masculin		31	272
Total	-	41	360

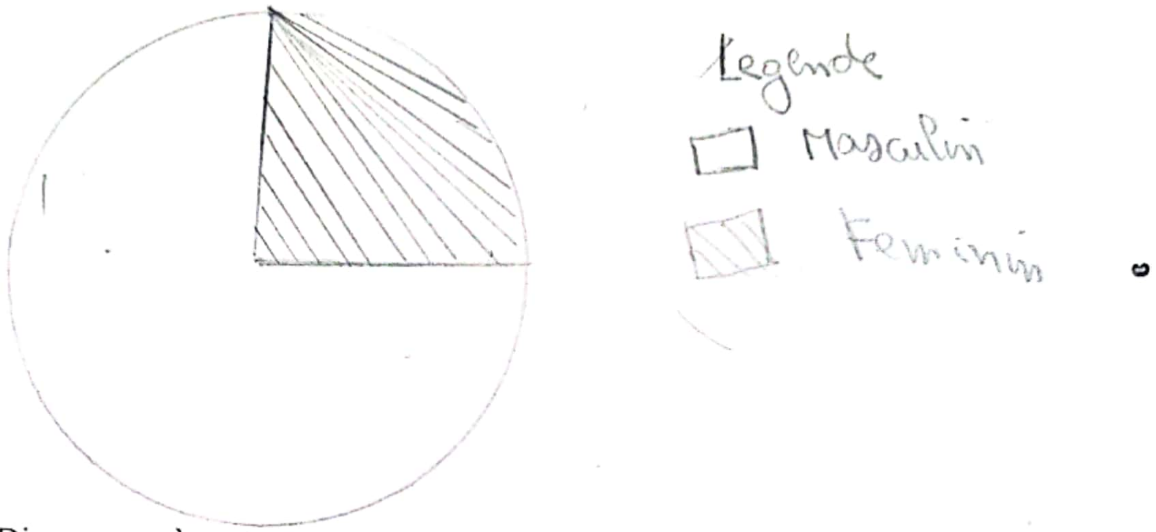


Diagramme à secteur

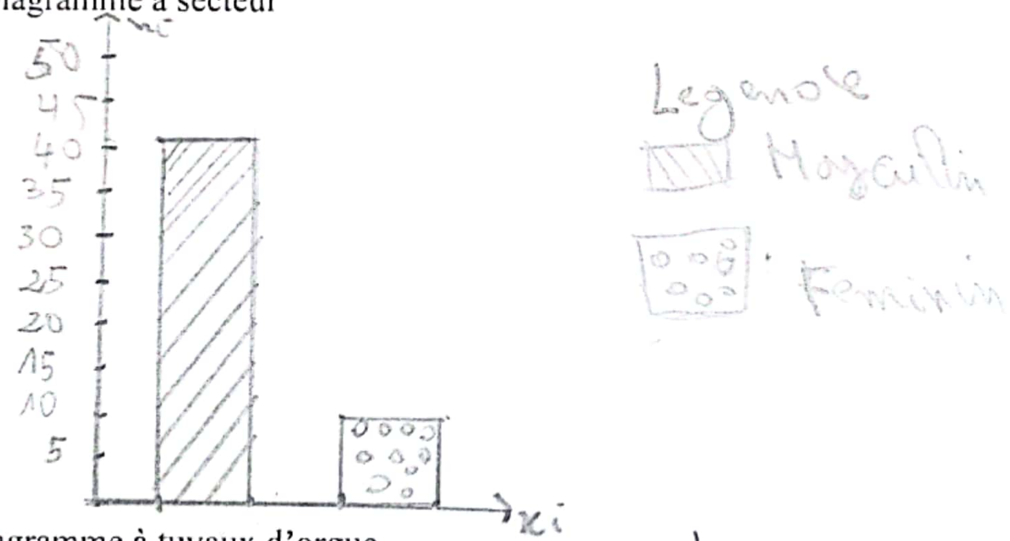


Diagramme à tuyaux d'orgue

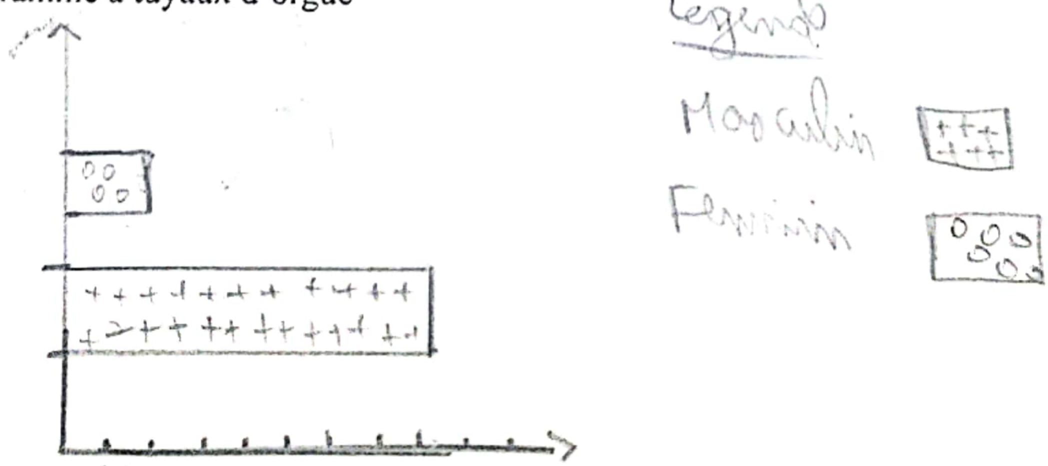


Diagramme à barres

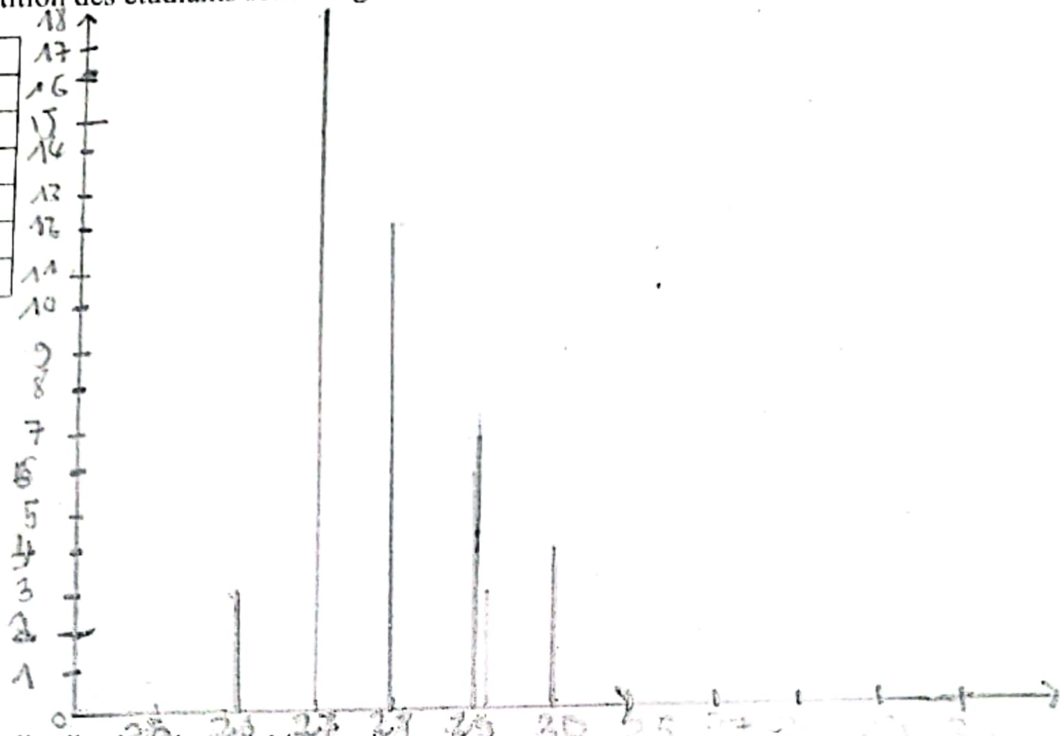
b) Caractère quantitatif

- Cas d'une distribution de variable discrète

Le diagramme à bâton est le graphique d'illustration d'une distribution de variables discrètes

Exemple : répartition des étudiants selon l'âge

Age	n_i
26	3
27	18
28	12
29	7
30	4
Total	



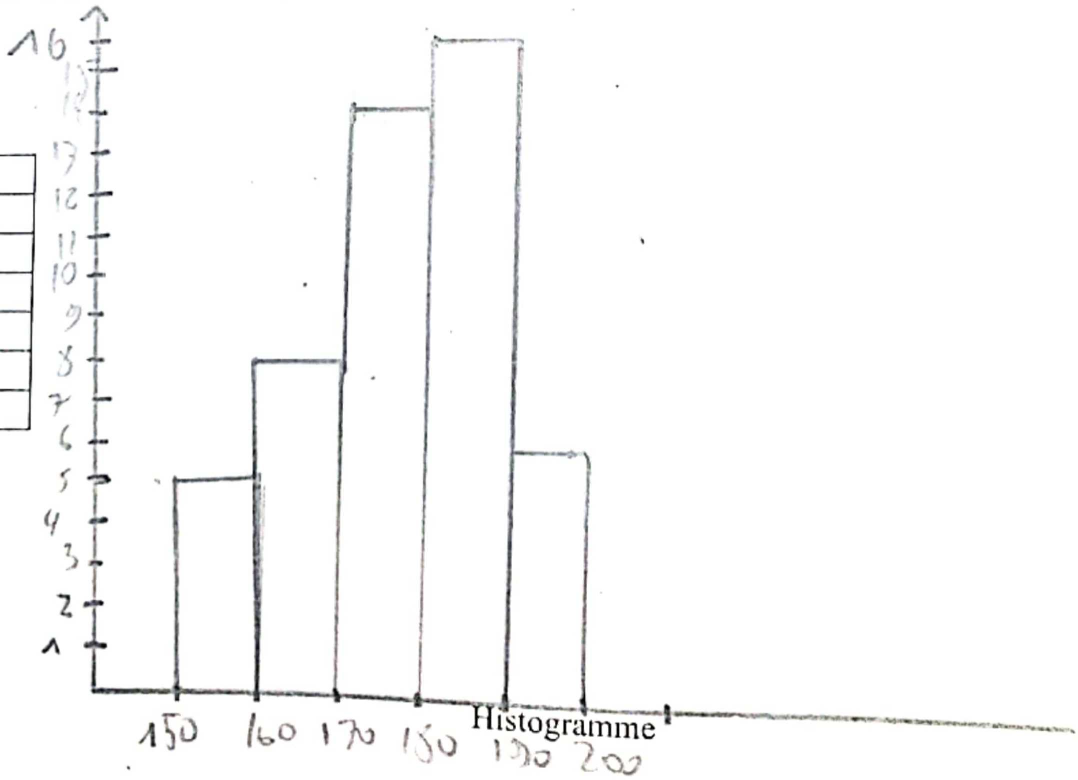
- Cas d'une distribution de variable continue

L'histogramme est le graphique d'illustration d'une distribution de variables continues.

✓ Variable continue à amplitude égales

Exemple : Répartition des étudiants selon la taille en cm

Taille	n_i	k_i
150 – 160	5	10
160 - 170	8	10
170 - 180	14	10
180 - 190	16	10
190 - 200	6	10
Total	43	



Histogramme

✓ Variables continue à amplitudes inégales

Dans ce cas, il faut préalablement corriger les effectifs ou les fréquences avant la construction de l'histogramme. La correction des effectifs se fait par le biais de la formule suivante :

$$n'_i = \frac{ni}{ki} * C$$

Avec n' : effectif corrigé et C comme constante (l'amplitude qui se répète le plus ou la plus petite amplitude).

Exemple : Répartition des étudiants selon la note en statistique

Notes	n_i	k_i	n_i'
0 - 4	12	4	6
4 - 6	10	2	10
6 - 10	18	4	9
10 - 12	15	2	15
12 - 14	8	2	8
14 - 20	9	6	3
Total			

2. Courbes cumulatives

On distingue la courbe des effectifs cumulés croissants et la courbe des effectifs cumulés décroissants, ou la courbe des fréquences cumulées croissantes et la courbe des fréquences cumulées décroissantes.

Notes	n_i	N↗	N↘
0 - 4	2	2	20
4 - 8	3	5	18
8 - 12	6	11	15
12 - 16	5	16	9
16 - 20	4	20	4
Total	20	–	–

A(4,2) ; A'(0,20)

B(8,5) ; B'(4,18)

C(12,11) C'(8,15)

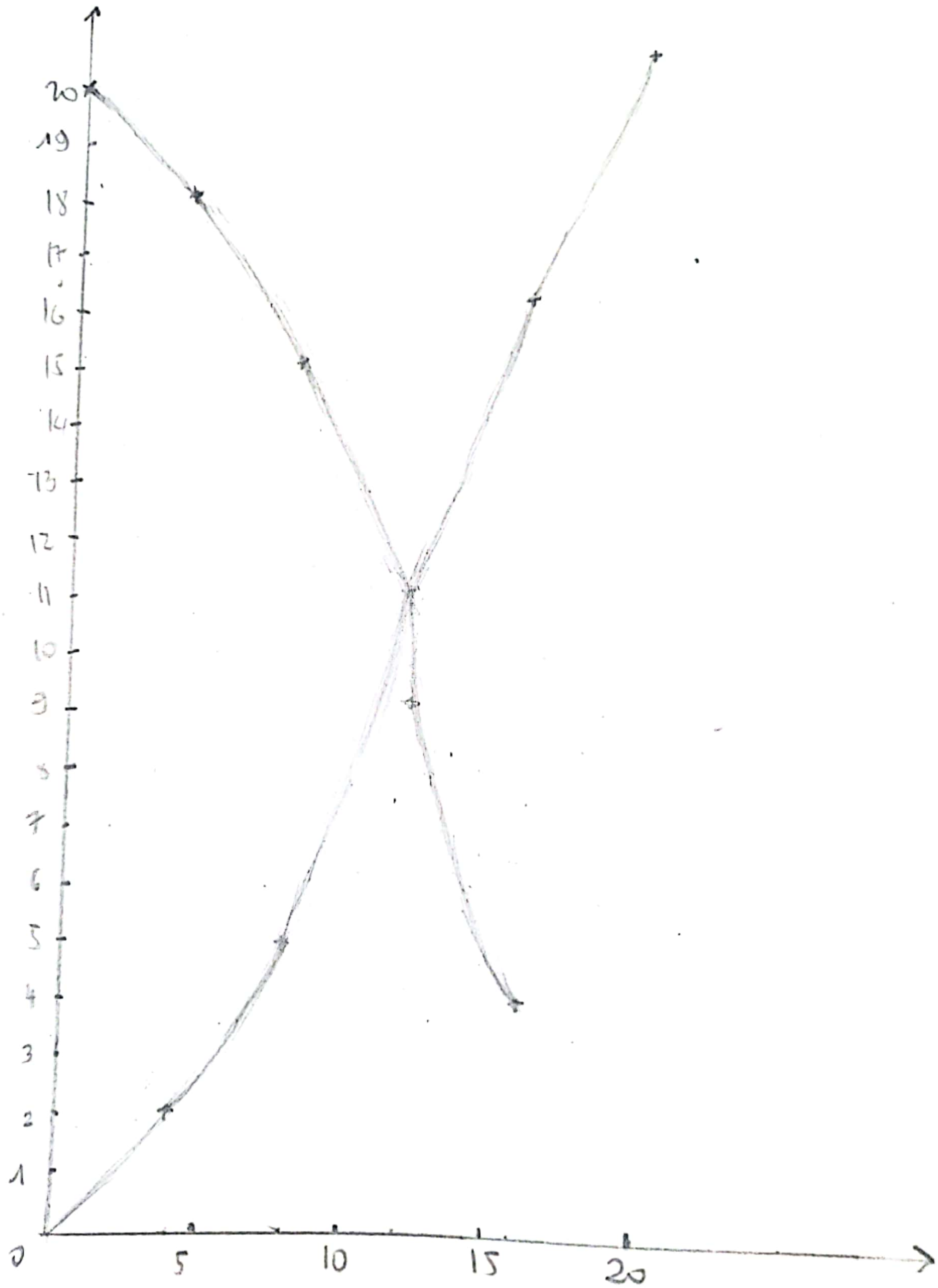
D(16,16) D'(12,9)

E(20,20) E'(16,4)

$$Me = 8 + 4\,\frac{10-5}{6} = 8 + \frac{20}{6} = 8 + \frac{10}{3} = \frac{24+10}{3} = \frac{34}{3}$$

$= 11,333$

tracer la courbe



LES TRAVAUX DIRIGES ET ENCADRES

EXERCICE 1

Dans un établissement d'enseignement supérieur, une étude statistique sur un échantillon d'étude a donné les résultats suivants :

NB : M = région Maritime ; P = région des Plateaux ; C = région centrale ; K = région de la Kara ; S = région des Savanes

NOM	REGION	AGE	FILIERE
Mlle Alice	M	21	SD
Mlle Laure	P	22	AC
Mlle Camy	P	22	AC
M Kevin	M	21	CI
Mlle Fatou	M	24	AC
Mlle Grace	C	24	SD
Mlle Paula	C	21	CI
M Eunok	M	23	CI
Mlle Mira	M	22	AC
Mlle Lize	M	23	CI
M Abdel	K	21	CI
M Daniel	K	23	CI
Mlle aicha	P	20	SD
Mlle Aude	M	24	SD
M Ben	K	21	SD
M Joel	M	23	AC
Mlle Agnes	K	22	CI
Mlle Flore	C	23	SD
Mlle Claire	M	22	AC
Mlle Léa	M	22	CI

NOM	REGION	AGE	FILIERE
M Didier	M	20	CI
M Yves	M	23	CI
Mlle Eliora	M	22	AC
Mlle anny	P	22	AC
M Guy	C	24	CI
M Olivier	M	24	SD
Mlle Ella	M	22	SD
M Ricky	K	21	AC
M Victor	C	22	CI
Mlle Eva	C	24	SD
M Bill	M	23	AC
M Will	S	21	CI
Mlle Anne	M	22	CI
Mlle Emilie	S	20	AC
Mlle Jane	S	22	CI
Mlle Gise	P	24	CI
Mlle Marte	C	20	SD
M Huges	C	21	CI
Mlle Fati	M	20	CI
Mlle Nina	M	21	CI

QUESTION

1- Préciser la population statistique et l'unité statistique et a taille de l'échantillon
2- Identifier tous les caractères étudiés et leur nature respective
3- Etablir la distribution des étudiants selon l'âge puis représenter
4- Etablir la distribution des étudiants selon l'âge et la filière choisie
5- Etablir la distribution des étudiants selon la région puis représenter
6- Etablir la distribution des étudiants selon le sexe et la région

EXERCICE 2

L'entreprise commerciale le DELIC pour des besoins d'études de sa cellule d'étude a organisé une enquête auprès de sa clientèle. Chaque client devait déclarer s'il est une entreprise ou un ménage, le montant de ses achats du mois de décembre 2000 auprès de l'entreprise le DECLIC et donner son appréciation de l'accueil réservé au client. Le tableau 2 présente les déclarations des 20 premiers clients.

No Client	01	02	03	04	05	06	07	08	09	10
Type	1	2	1	2	1	1	2	1	1	2
Achats	52	50	55	61	51	54	50	58	61	64
Accueil	1	2	3	1	3	2	2	1	2	3

No Client	11	12	13	14	15	16	17	18	19	20
Type	2	1	1	2	2	1	2	1	2	1
Achats	63	57	57	63	60	50	64	62	55	53
Accueil	1	2	1	1	2	3	3	2	1	1

NB : se servir du dictionnaire des variables

1- Donner la répartition des clients selon l'appréciation de l'accueil et représenter
2- Donner la répartition des clients selon le montant des achats. Faire trois classes de mêmes amplitudes entre 50 mille et 65 mille
3- Donner la répartition des clients selon m'appréciation de l'accueil et du type de client.

Dictionnaire des variables

No client : c'est le numéro d'ordre du client

Type : renseigne sur le type du client. Il vaut

- 1 si le client est un ménage
- 2 si le client est une entreprise

Achats : c'est le montant des achats réalisés par le client auprès de l'entreprise le DECLIC en décembre 2000 en milliers de francs CFA.

Accueil : renseigne sur la façon dont le client apprécie l'accueil. il vaut

- 1 si l'accueil est très bon
- 2 si l'accueil est acceptable
- 3 si l'accueil est mauvais

Chp 3 LES CARACTERISTIQUES NUMERIQUES

INTRODUCTION

Les tableaux statistiques, les graphiques d'illustration et les courbes cumulatives donnent une idée sommaire de la série statistique mais la vraie physionomie d'une distribution statistique est donnée par les paramètres chiffrés, quantifiables d'où les caractéristiques numériques.

On distingue 4 caractéristiques numériques :

- ✓ Les caractéristiques de tendance centrale ou de position
- ✓ Les caractéristiques de dispersion
- ✓ les caractéristiques de forme
- ✓ Les caractéristiques de concentration

SECTION I : LES CARATERISTIQUES DE TENDANCE CENTRALE OU DE POSITION

On distingue trois principales :

- La moyenne
- La médiane
- Le mode

A- LA MOYENNE

1- Définition

La moyenne est la valeur de la variable noté $\bar{x}$ qui en même temps qu'elle donne une idée générale sur la distribution statistique, peut remplacer toutes les variables sans toutefois modifier l'image objective du phénomène à étudier.

2- Formule de la moyenne

Données / Moyennes	Données Simples	Données regroupées
Moyenne Arithmétique	$\bar{x} = \frac{\sum xi}{N}$	$\bar{x} = \frac{\sum xini}{\sum ni}$

3- Exemple de calcul de la moyenne arithmétique

Age (xi)	ni	xini
18	3	
19	5	
20	7	
21	2	
22	1	
Total	18	

- Données Simple

Déterminer la moyenne d'un étudiant qui a obtenu les notes suivantes en statistique :

12- 19-15-07-14

$$\bar{x} = \frac{\sum xi}{N} = \frac{12+19+15+7+14}{5} = 13,4$$

- Données regroupé,

  Cas d'une variable discrète

  Exemple : la répartition des étudiants selon l'âge

Age (xi)	ni	xini
18	3	54
19	5	95
20	7	140
21	2	42
22	1	22
Totale	18	353

$$\bar{x} = \frac{\sum xini}{\sum ni} = \frac{18\times 3+19\times 5+7\times 20+21\times 2+22}{18}$$

$$\bar{x} = 19,61$$

- Donnée regroupée cas d'une variable continue

Taille (cm)	ni	xi	Xini
150-160	3	155	465
160-170	5	165	825
170-180	7	175	1225
180-190	2	185	370
190-200	1	195	195
Total	18	875	3080

$$\bar{x} = \frac{\sum xini}{\sum ni} =$$

$$\bar{x} = 171,11$$

Soit un intervalle [a, b [$\rightarrow$ $xi = \frac{a+b}{2}$

$$\bar{x} = \frac{\sum xini}{\sum ni} =$$

Idée generale d'une distribution = moyenne

Propriété de la moyenne arithmétique

- A effectif égaux, la moyenne arithmétique pondéré est égale à la moyenne arithmétique simple
- La somme des écarts d la variable par rapport à la moyenne arithmétique est nulle

B- La médiane

1- Définition

Calculer à partir des données rangées, la médiane est la valeur de la variable « Me » qui occupe la position centrale divisant ainsi la distribution statistique en deux parties égale.

2- Médiane des données simples

- **Cas d'un nombre impair d'observation**

Après avoir rangé les données par ordre croissant ou décroissant, la médiane est celle qui occupe la position centrale.

Exemple : 1 – 13 – 15 – 9 – 3 – 7 – 19

→ 1 – 3 – 7 – 9 – 13 – 15 - 19 Me = 9

- **Cas d'un nombre pair d'observation**

Dans ce cas après avoir rangé les données la médiane est la demi-somme des deux valeurs qui occupent la position centrale.

→ 1 – 3 – 7 – 9 – 13 – 15 $Me = \frac{7+9}{2} = 8$

3- Médiane des données regroupée

- **Cas d'une variable discrète**

Dans ce cas, la médiane est la valeur de la variable qui correspond au premier effectif cumulé croissant strictement supérieur à la moitié des observations ou à la première fréquence cumulée croissante strictement supérieur à 50%

Age (xi)	ni	N ↑
18	4	4
19	6	10
20	3	13
21	2	15
22	1	16
Total	16	-

$\frac{\Sigma ni}{2} = 8$

Me = 19

- **Cas d'une variable continue**

La médiane s'obtient par la formule suivante :

$$\mathbf{M_e = L_1 + k\frac{\frac{1}{2}\sum n_i - N_1}{n_e}}$$

Avec :

L_1 : borne inférieure de la classe médiante

K : amplitude de la classe médiante

N_1 : Effectif cumulé qui précède celui de la classe médiante

n_e : Effectif de la classe médiante

NB : la classe médiante est celle qui correspond au premier effectif cumulé croissant strictement supérieur à la moitié des observations ou à la première fréquence cumulée croissante strictement supérieur à 50%

Exemple :

La répartition selon la taille en cm

taille (cm)	ni	$N\uparrow$	ki
150-160	3	3	10
160-170	5	8	10
170-180	7	15	10
180-190	2	17	10
190-200	1	18	10
Total	18	-	10

$Mo \in [170, 180[$

$Mo = 170 + 10 \frac{7-5}{7-5+7-2}$

$Mo = 170 + 10 \frac{2}{7}$

$Mo = 170 + \frac{20}{7}$

$Mo = 170 + \frac{20}{7}$

$Mo = 172,85$

$\frac{1}{2}\sum ni = 9$

$Me = L_1 + k \frac{\frac{1}{2}\sum ni - N_1}{ne}$

$Me = 170 + 10 \frac{9-8}{7} = 171,428$

*même si l'amplitude est pas exacte partout on calcule pas ni et on corrige pas les amplitudes

* si $[150, 160[$ était la classe Me $N_1 = 0$

Interprétation :

50% 50%

Moins performant Me Plus performant

Les 50% des agents les moins élancés ont une taille comprise entre 150 cm et Me et les 50% des agents les plus élancés ont une taille comprise entre Me et 200 cm.

C- Le mode

1- Définition

Le mode ou moyenne de fréquence ou encore le dominant est la valeur de la variable la plus observée ou la plus fréquente de la distribution. Le mode est noté Mo. On ne calcule que le mode des données regroupés

2- Mode d'une variable discrète

Dans ce cas, le mode est la valeur de la variable qui correspond à l'effectif le plus élevé ou à la fréquence la plus élevée de la distribution

Exemple : la répartition selon l'Age

Age (xi)	ni
18	3
19	8
20	12
21	4
22	5
Total	32

$M_o = 20$

3- Mode d'une distribution continue

Le mode s'obtient par la formule suivante :

$$\boldsymbol{M_o} = \mathbf{L_1} + \mathbf{k}\frac{(n_o - n_1)}{(n_o - n_1) + (n_o - n_2)}$$

L_1 : Borne inférieure de la classe modale

K : amplitude de la classe modale

n_o : effectif de la classe modale

n_1 : effectif qui précède celui de la classe modale

n_2 : effectif qui suit celui de la classe modale

NB : La classe modale est celle qui correspond à l'effectif le plus élevée ou à la fréquence la plus élevée de la distribution

- Cas d'une distribution à amplitude égale

Exemple : Répartition selon les notes en statistique

note	ni	k
0 - 4	3	4
4 - 8	5	4
8 - 12	8	4
12 - 16	4	4
16 - 20	1	4
Total	21	-

$$Mo = 8 + 4\left(\frac{8-5}{8-5+8-4}\right)$$

$$Mo = 8 + 4\left(\frac{3}{7}\right)$$

$$Mo = 8 + \frac{12}{7} \Rightarrow 9,714$$

- . Cas d'une distribution à amplitude inégale

Dans ce cas, on doit préalablement corriger les effectifs ou la fréquence avant la détermination de la classe modale

note	ni	k	ni'
0-4	10	4	5
4-6	10	2	10
6-10	24	4	12
10-12	15	2	15
12-14	8	2	8
14-20	9	6	3
Total	76	-	-

$$ni' = \frac{ni}{k} \times C$$

$$Mo = 10 + 2\,\frac{15-12}{15-12+15-8}$$

$$Mo = 10 + 2\left(\frac{3}{10}\right)$$

$$Mo = 10 + \frac{3}{5}$$

$$Mo = \frac{53}{5}$$

$$Mo = 10,6$$

SECTION II : *LES AUTRES CARACTERISTIQUES DE TENDANCE CENTRALE*

I- LES QUARTILES

Ce sont les trois valeurs Q_1, Q_2, Q_3 de la variable qui divise la distribution en quatre parties égale. Elle se calcule par la formule suivante :

$$Q_\alpha = L_1 + k\,\frac{\frac{\alpha}{4}\sum ni - N1}{nQ\alpha}$$

Avec α qui peut prendre les valeurs : (1,2 ,3)

Exemple : Répartition des employés selon leur salaire en millier de francs

Salaire	ni	N↗
60-70	2	2
70-80	7	9
80-90	5	14
90-100	9	23
100-110	10	33
Total	33	-

Exemple : Calculer et interpréter le Q_3

II- LES DECILES

Ce sont les neuf valeurs D_1 D_2,..., D9 de la variable qui divise la distribution en dix parties égales. D'où la formule :

$$D_\alpha = L_1 + k\,\frac{\frac{\alpha}{10}\sum ni - N1}{n_{D_\alpha}}$$

Avec α qui peut prendre les valeurs : {1, 2, 3, 4, ... ,9}

Exemple : calculer et interpréter le D_3

III- LES CENTILES OU LES PERCENTILLES

Ce sont les 99 valeurs P_1 P_2……P_{99} de la variable qui divise la distribution en 100 parties égales.

$$\mathbf{P}_{\alpha} = \boldsymbol{L_1} + \mathbf{k}\,\frac{\frac{\alpha}{100}\sum \boldsymbol{ni} - \boldsymbol{N_1}}{\boldsymbol{n_{P_{\propto}}}}$$

Avec α qui peut prendre les valeurs : {1, 2, 3, 4, ... ,99}

Exemple : Calculer et interpréter P_{13}

SECTION III : LES CARACTERISTIQUES DE DISPERTION

I- L'ETENDU

L'étendu noté « e » est la différence entre la valeur maximale et la valeur minimale

e = Xmax - Xmin

II- ECART ABSOLUE

- **Cas des données simples**

$$e(\alpha) = \frac{\sum(xi - \alpha)}{N}$$

Avec α peut prendre des valeurs : $\bar{x}$, Mo, Me

- **Cas des données regroupés**

$$e(\alpha) = \frac{\sum(xi - \alpha)ni}{\sum ni}$$

Avec α peut prendre des valeurs $\bar{x}$, **Mo, Me**

III- LA VARIANCE ET L'ECART TYPE

A- LA VARIANCE

- **Cas de Donnée simple**

La variance noté V(x) est égale à б(x)et se calcule par la formule :

$$V(x) = \frac{\sum(xi - \bar{x})^2}{N} \quad (1)$$

$$V(x) = \frac{\sum x_i^2}{N} - \bar{x}^2 \quad (2)$$

Avec $\bar{x}$ égal à la moyenne de la distribution

- **Cas de données regroupés**

$$V(x) = \frac{\sum(xi - \bar{x})^2 ni}{\sum ni} \quad (1)$$

$$V(x) = \frac{\sum x_i^2 n_i}{\sum n_i} - \bar{x}^2 \quad (2)$$

B- ECART-TYPE

C'est la racine carrée de la variance :

$$\sigma_x = \sqrt{V(x)}$$

NB : on n'interprète pas la variance mais plutôt l'écart-type car cette dernière caractéristique qui mesure la variabilité d'une unité statistique à l'autre.

IV. LE COEFFICIENT DE VARIATION

Cette caractéristique mesure l'homogénéité d'une distribution statistique. Elle est notée CV et se calcule par la formule suivante :

$$CV = \frac{\sigma_x}{\bar{x}} * 100$$

- Si **CV ≤ 33%** alors la distribution est dite homogène
- Si **CV > 33%** alors la distribution est dite hétérogène

I. CARACTERISTIQUEDE CONCENTRATION

A – Notion de médiale

La médiale c'est la valeur de la variable Mle qui divise la masse totale des valeurs en deux parties égales.

Pour une distribution de variable discontinue on utilise la formule :

$$\text{Mle} = L_1 + k\,\frac{0{,}5\sum mi - M1}{me}$$

EXEMPLE

Salaire	ni	xi	mi=xini	Mi↑	Ni ↑
70 – 80	2	75			
80 – 100	4	90			
100 - 120	8	110			
120 - 140	5	130			
140 - 150	1	145			
TOTAL	20				

B - INDICE ELEMENTAIRE

$$ic = \frac{\Delta M}{e} = \frac{Mle - Me}{Xmax - Xmin}$$

C -INDICE DE CONTRATION de LORENTE GINI

$$I_G = 1 - \frac{fi\ Si}{10\ 000}$$

avec $fi = \frac{ni}{\sum ni} * 100.$

$Si = Qi + Q_{i-1}$

Qi = cumul de de qi

$qi = \frac{mi}{\sum mi} * 100$

EXEMPLE

Salaire	ni	xi	mi=xini	fi	qi	QI	Si	fiSi
70 – 80	2	75						
80 – 100	4	90						
100 - 120	8	110						
120 - 140	5	130						
140 - 150	1	145						
TOTAL	20							

Interprétation :

Disparité
Inégalité
Concentration

0 faible 0,5 forte 1

LES TRAVAUX DIRIGES ET ENCADRES

EXERCICE 1

En 2010, un expert de la banque mondiale a réalisé une enquête sur le salaire en milliers de francs de certains agents de la fonction publique d'un pays

37	40	50	46	43	54	49	36	43	44
45	49	44	57	51	52	47	41	50	46
42	39	47	54	41	32	55	52	38	42
39	59	48	53	40	49	35	47	36	40

1- Regrouper les données sous forme d'une distribution continue d'amplitude égale 5 la borne
   Inferieure de la première classe étant 30.
2- Préciser la population statistique, l'unité statistique, le caractère et sa nature puis la taille de l'échantillon.
3- Donner une idée générale de cette distribution
4- De calculer puis d'interpréter la médiane de la distribution
5- Déterminer le salaire de la plupart des agents
6- Déterminer la proportion des agents ayant au moins 45000 FRANCS
7- De combien varie le salaire d'un employé
8- La distribution est-elle homogène ?
9- Déterminer indice élémentaire de cette distribution ?
10- Déterminer l'indice de concentration de Lorentz Gini

Exercice 2

Le Directeur des Ressources Humaines de l'Entreprise « LE DECLIC » a réuni les renseignements suivants relatifs au salaire mensuel en milliers de francs de certains agents.

Salaire	20 - 28	28 - 36	36 - 44	44 - 52	52 - 60	TOTAL
ni	24	36	32	40	28	160

1- Préciser la population statistique, l'unité statistique, le caractère et sa nature.
2- Donner une idée générale de la distribution.
3- Déterminer le salaire mensuel de la plupart des agents.
4- Déterminer le salaire médian puis interpréter.
5- Déterminer la proportion des agents ayant un salaire d'au moins 44 000 francs.
6- Déterminer la proportion des agents ayant un salaire de moins 52 000 francs.
7- Déterminer la proportion des agents dont le salaire est au moins égal à 28 000 francs mais moins de 52 000 francs.

Chp 4 **STATISTIQUE DE DEPENDANCE**

Introduction

L'étude unie dimensionnelle des variables statistiques donne une idée tronquée de la réalité puisque les phénomènes économiques sont les plus souvent liés entre eux. C'est ainsi qu'on ne peut pas étudier :

- La consommation sans le revenu
- L'épargne sans le revenu
- Le chiffre d'affaires sans les dépenses publicitaires
- Le taux de mortalité infantile sans le nombre de médecin dans la région
- La note obtenue en statistique sans le temps de consacré à l'étude de la statistique

Pour étudier l'interaction entre ses phénomènes économique on passe par le biais de la corrélation et de la régression

Pour ces études, on doit clairement préciser la variable expliquée ou la variable dépendante très souvent noté **yi** et la variable explicative ou indépendante très souvent notée **xi**

Le coefficient de corrélation est un coefficient sans unité qui confirme l'existence d'une relation. Entre deux variables, ce coefficient le degré de liaison entre des deux variables. Le coefficient de corrélation est note R_{xy}. **($-1 \leq Rxy \leq 1$)**

I - ETUDE DE LA CORRELATION

Pour étudier la corrélation entre deux variables, on calcule un coefficient appelé **coefficient de corrélation**

1- Définition

Le coefficient de corrélation est un nombre sans unité qui confirme l'existence d'une relation entre les variables statistiques et mesure le degré de cette liaison.

2 – Calcul du coefficient de corrélation de Karl PEARSON

a– formule

$$R_{xy} = \frac{\overline{xy} - \overline{x}\overline{y}}{\sigma_x \sigma_y}$$

b- interprétation

- Si $Rxy < 0$ alors il existe une relation négative donc les deux variables évoluent en sens contraire
- Si $0 < Rxy < 0{,}5$ alors la corrélation est faible
- Si $0{,}5 < Rxy < 0{,}7$ alors la corrélation est moyenne
- Si $|Rxy| \geq 0$ la corrélation est forte

Exemple :

Le tableau représente le chiffre d'affaires (CA) et les dépenses publicitaires (DP) en million de francs sur une période de 5 ans

x_iy_i	432	616	520	854	780	3202
CA y_i	48	56	52	61	60	277
DP x_i	9	11	10	14	13	57
x_i^2	81	121	100	196	169	667
y_i^2	2304	3136	2704	3721	3600	15465

TAF :

1- Identifier la variable expliquée et la variable explicative
2- Calculer le coefficient de corrélation et interpréter

Pour interpréter le coefficient de corrélation, on calcule le coefficient de détermination :

$$D = (Rxy)^2$$

Le coefficient de détermination mesure la part de la variable explicative dans l'explication de la variable expliquée.

D= 0,98* 0,98 = 96%

Ce qui implique que le CA est expliqué à 96% par les dépenses publicitaires

3- Etude de la régression

La régression permet d'établir une relation analytique entre les 2 variables à des fins prévisionnels ou des estimations. On distingue l'ajustement affine et l'ajustement non affine

a. Ajustement affine

Yi = axi + b

✓ **Méthodes des points extrêmes**

Elles consistent à faire l'hypothèse selon lesquelles la droite de régression passe par les deux points extrêmes et que les coordonnées de ces points doivent vérifier l'équation.

Exemple :

Xi	2	5	8	12	15	18
Yi	5	7	15	21	24	30

✓ La méthode du double moyen ou la méthode de Mayer ou des points moyens

Elle consiste à diviser la distribution en 2 parties plus ou moins égales et à retenir comme points stratégiques par lequel la droite de régression passe par les points moyens de ces 2 sous parties.

✓ La méthode des moindres carrées

$yi = axi + b$

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

$$a = \frac{\sum xiyi - n\bar{x}\bar{y}}{\sum xi^2 - n\bar{x}^2}$$

$$a = \frac{\overline{xy} - \bar{x}\bar{y}}{V_x}$$

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

$y = ax + b$

$a = \frac{y_B - y_A}{x_B - x_A}$

$y_B = ax_B + b$

$y_A = ax_A + b$

Interprétation :

b : est la valeur de yi (variable expliquée) lorsque xi = 0

a : est la variation de yi suite à une variation d'une unité de xi

Exemple :

Année	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Producteur	60	70	65	90	80	95	100	110	130	150

TAF :

1- Etablir une régression linéaire entre les deux variables

2- En supposant que l'évolution de la production se poursuit dans le temps :

a) Estimer la production de 1990

b) Déterminer l'année à partir de laquelle la production sera au moins 250 tonnes

SERIES CHRONOLOGIQUES

I- DEFINITION

On appelle séries chronologiques ou séries temporelles ou encore chronique est une suite de données chiffrées et ordonnées décrivant l'évolution des phénomènes économiques dans le temps.

II- ETUDE DES COMPOSANTES D'UNE CHRONIQUE

Soit la répartition du chiffre d'affaires trimestriel de l'entreprise DELTA de 2018 à 2020

TRIMESTRE / ANNEES	I	II	III	IV
2018	10	9	5	12
2019	13	12	8	14
2020	16	15	11	17

1. TENDANCE GENERALE OU LE TREND

a- Méthode graphique :

Elle consiste à représenter le nuage de points dans un repère orthogonal le temps t ou x_i en abscisse et la valeur y_t ou y_i du phénomène ordonné.

b- Méthode mécanique :

C'est la méthode des moyennes échelonnées et des moyennes

t	yt	MOYENNE			
		ECHOMLONNEES		MOBILES	
		M=2	M=3	M=2	M=3
1	10				
2	9				
3	5				
4	12				
5	13				
6	12				
7	8				
8	14				
9	16				
10	15				
11	11				
12	17				

Remarque :

Avec la méthode des moyennes échelonnées, Pour m = 2 la tendance générale ne se fait pas bien voir mais avec m = 3, cela montre aisément que la tendance générale est à la hausse.

Avec les moyennes mobiles, la même chose se fait voir pour m = 2 ou la tendance générale n'apparait pas mais avec m = 3 la tendance générale est à la hausse sauf que la baisse a repris vers la fin.

En somme, pour cet exercice, c'est la méthode des moyennes échelonnées avec une période m = 3 qui permet de faire ressortir la tendance générale.

c- Méthode analytique

Toutes les méthodes d'ajustement sont encore retenues

Ajustement affine : yi = axi + b

- Méthode des points stratégique où nous avons la méthode des points extrême et méthode de la double moyenne ou méthode de MAYER.

- Méthode des moindres carrées qui donne le meilleur estimateur linéaire non biaisé.

Interprétation dans le cas de la série chronologique

$$yi = axi + b \text{ ou } yt = axt + b$$

b : est la valeur de y_t à l'instant t = 0 ou xi = 0

a : est l'accroissement périodique de y_t ou y_i sur la période

Application

Le tableau suivant donne la répartition du chiffre d'affaires (en million de francs) de l'entreprise le DECLIC de 1994 à 2003.

Années	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
x_i	1	2	3	4	5	6	7	855	9	10
y_i	100	105	120	130	145	155	160	170	175	190

xi désigne le rang de l'année et yi le prix du lot puis l'on donne les résultats numériques

$$\sum x_i = 55;\ \sum y_i = 1450;\ \sum xi^2 = 385;\ \sum yi^2 = 218600\ ;\sum xiyi = 8800$$

Pour une analyse des données, on vous demande :

1 - de calculer le coefficient de corrélation entre les deux variables

2 - d'établir par la méthode des moindres carrés la droite de régression de la forme : $yi = axi + b$ puis interpréter les coefficients a et b

3 - En supposant que l'évolution du chiffre d'affaires se poursuive dans le temps

a) Estimer le chiffre d'affaires pour l'année 2006.

b) Déterminer en quelle année le chiffre d'affaires excèdera 300 millions de francs.

c) Déterminer le bénéfice qu'on escomptera en 2008 si les charges sont estimées

DEVOIR DE STATISTIQUES DESCRIPTIVES

Durée : 2 heures

Filière : **L_2 TC**

EXERCICE 1

Soit la répartition suivante

AGES	n_i	$F_i\%$	ECC	ECD
18- 22	a	f	i	m
22- 26	b	25	j	n
26 - 30	120	g	298	222
30 - 34	c	21	k	o
34 – 38	d	h	l	18
Total	e	100		

Reproduire le tableau en complétant les données manquantes a, b, c, d, e, f, g, h, i, j, k, l, m, n, o sur les effectifs simples n_i cumulés croissant, cumulés décroissants, les fréquences $f_i\%$

Avec ECC : Effectif cumule croissant, ECD : Effectif cumulé décroissant

EXERCICE 2

On donne la distribution suivante :

Classe	Effectif n1
{20,30{	100
{30 ;40{	140
{40 ; x{	125
{x ;70{	200
{70 ;100{	180
{100 ;170{	55

1- Calculer X sachant que la médiane est =56,8

2- Quelle est l'idée générale de la distribution

EXERCICE 3

En 2010, un expert de la banque mondiale a réalisé une enquête sur le salaire en milliers de francs de certains agents de la fonction publique d'un pays

37	40	50	46	43	54	49	36	43	44
45	49	44	57	51	52	47	41	50	46
42	39	47	54	41	32	55	52	38	42
39	59	48	53	40	49	35	47	36	40

1- Regrouper les données sous forme d'une distribution continue d'amplitude égale 5 la borne inferieure de la première classe étant 30.

2- Préciser la population statistique, l'unité statistique, le caractère et sa nature puis la taille de l'échantillon.

3- Donner une idée générale de cette distribution

4- Déterminer le salaire de la plupart des agents de la fonction
5- Quel est le salaire médian des agents ?
6- De combien varie le salaire d'un agent à l'autre

Fin