OFPPT

ROYAUME DU MAROC

مكتب التكوين المهني وإنعاش الشغل

Office de la Formation Professionnelle et de la Promotion du Travail

DIRECTION RECHERCHE ET INGENIERIE DE FORMATION

RESUME THEORIQUE
&
GUIDE DE TRAVAUX PRATIQUES

MODULE : STATISTIQUES

SECTEUR : TERTIAIRE

SPECIALITE : COMPTABILITE DES ENTREPRISES

NIVEAU : TECHNICIEN

Document élaboré par :

Mlle Nadia BENHADDOU BAKKIOUI   ISTA Taroudant   DR SMD

Révision linguistique:
-
-
-

Validation :
-
-
-

SOMMAIRE

Module : Statistiques

Durée : 50 H
40% : Théorique
60% : Pratique

OBJECTIF OPERATIONNEL DE PREMIER NIVEAU
DE COMPORTEMENT

COMPORTEMENT ATTENDU

Pour démontrer sa compétence, le stagiaire doit
appliquer les méthodes statistiques.
Selon les conditions, les critères et les précisions qui suivent :

CONDITIONS D'EVALUATION

- A partir des études de cas, mise en situation, consignes du formateur, toute documentation nécessaire ;
- A l'aide de : calculatrice, tableur et logiciel de statistiques.

CRITERES GENERAUX DE PERFORMANCE

- Respect de la démarche de calcul
- Respect des principes de gestion de temps
- Respect des pratiques courantes et des règles établies par l'entreprise
- Exactitude des calculs
- Vérification appropriée du travail.

OBJECTIF OPERATIONNEL DE PREMIER NIVEAU
DE COMPORTEMENT

PRECISION SUR LE COMPORTEMENT ATTENDU	**CRITERES PARTICULIERS DE PERFORMANCE**
A. Comprendre les variables statistiques	o Qualification d'une variable qualitative o Qualification d'une variable quantitative discrète o Qualification d'une variable quantitative continue
B. Réaliser des représentations graphiques	o Représentation correcte des variables quantitatives discrètes o Représentation correcte des variables quantitatives continues
C. Calculer les caractéristiques des distributions	o Calcul et interprétation juste des paramètres de tendance centrale ▪ Mode ▪ Médiane ▪ Quartiles ▪ Moyennes o Calcul et interprétation correcte des paramètres de dispersion ▪ Etendue ▪ Ecart absolu moyen et écart quantile ▪ Variance, écart-type et coefficient de variation
D. Déterminer les liens entre deux variables	o Traitement du cas de deux caractères quantitatifs (coefficient de corrélation linéaire, ajustement par la droite des moindres carrés, rapport de corrélation) o Traitement du cas d'un caractère quantitatif et d'un caractère qualitatif (rapport de corrélation) o Traitement du cas de deux caractères qualitatifs

E. Réaliser des sondages	o Réalisation de sondage simple avec : • estimateur d'une moyenne ou d'une proportion • variance de ces estimateurs • estimateurs de ces variances • algorithmes de tirages
F. Réaliser des enquêtes	o Détermination optimale de l'échantillon o Elaboration du questionnaire o Recueil des données o Dépouillement, codage et saisie o Validation des données o Traitement statistique o Analyse des résultats

OBJECTIFS OPERATIONNELS DE SECOND NIVEAU

Avant d'apprendre à comprendre les variables statistiques, le stagiaire doit :

1- Comprendre la notion des « statistique »
2- Comprendre les objectifs des statistiques

Avant d'apprendre à réaliser les représentations graphiques, le stagiaire doit :

3- Distinguer entre les variables qualitatives et les variables quantitatives
4- Distinguer entre les variables quantitatives discrètes et les variables quantitatives continues
5- Présenter les séries statistiques dans des tableaux

Avant d'apprendre à calculer les caractéristiques des distributions, le stagiaire doit :

6- Réaliser des représentations graphiques
7- Interpréter ces représentations graphiques

Avant d'apprendre à déterminer les liens entre deux variables, le stagiaire doit :

8- représentez les distributions à deux variables dans des tableaux
9- représentez graphiquement ces distributions
10- calculer les caractéristiques des distributions
11- Interpréter ces caractéristiques des distributions

Avant d'apprendre à réaliser des sondages, le stagiaire doit :

12- définir le sondage
13- comprendre les objectifs de la réalisation des sondages
14- calculer les caractéristiques des distributions

Avant d'apprendre à réaliser des enquêtes, le stagiaire doit :

15- définir l'enquête
16- comprendre les objectifs de la réalisation des enquêtes

PRESENTATION DU MODULE

Ce module s'adresse en priorité à aux techniciens comptables des entreprises et aux techniciens spécialisés en gestion des entreprises.

Il répond à trois objectifs fondamentaux :

1) L'acquisition des connaissances : chaque chapitre comprend ainsi une partie Cours détaillée : les formules mathématiques fondamentales, mais aussi les points délicats du cours sont abordés.

2) L'utilisation des connaissances : chaque chapitre comprend des applications nombreuses et variées qui permettent aux stagiaires d'utiliser leurs connaissances.
La plupart de ces applications sont accompagnées d'indications de résultats ou éléments de réponse.

3) L'adaptation des connaissances : des Travaux Pratiques proposés, devront permettre aux stagiaires de mettre en application leurs qualités de raisonnement et d'adaptation face à des problèmes plus longs où de nombreuses connaissances sont exigées.

La masse horaire affectée à ce module est de 50 heures dont 30 heures consacrées aux travaux pratiques.

Module : Statistiques Descriptives
RESUME THEORIQUE

Chapitre I- Les statistiques descriptives :

I- Terminologie :

1. Statistique :

La statistique est une méthode scientifique dont l'objet est de recueillir, d'organiser, de résumer et d'analyser les données d'une enquête, d'une étude o d'une expérience, aussi bien que de tirer les conclusions logiques et de prendre les décisions qui s'imposent à partir des analyses effectuées.

2. Population :

Ensemble d'individus définis par une propriété commune donnée.
Exp : si l'on veut étudier la durée de vie des ampoules électriques fabriquées par une compagnie, la population considérée est l'ensemble de toutes les ampoules fabriquées par cette compagnie.

3. Echantillon :

Sous-ensemble de la population.
Exp : pour établir la durée de vie des ampoules électriques produites par une machine, on peut prélever au hasard un certain nombre d'ampoules - un échantillon- parmi toutes les celles produites par cette machine.

4. Individu ou unité statistique :

Chaque élément de la population ou de l'échantillon.
Exp : dans l'exemple précédant, chaque ampoule constitue un individu ou une unité statistique.

5. La taille :

Représente le nombre d'individus d'un échantillon ou d'une population. Elle est symbolisée par « n » dans le cas d'un échantillon et par « N » dans le cas d'une population.

6. Le caractère :

C'est l'aspect particulier que l'on désire étudier.
Exp : concernant un groupe de personnes, on peut s'intéresser à leur age, leur sexe leur taille…

7. Les modalités :

Les différentes manières d'être que peut présenter un caractère.
Exp 1 : le sexe est un caractere qui presente deux modalités : feminin ou masculin
Exp 2 : quant au nombre d'enfants par famille, les modalités de ce caractere peuvent etre 0,1 2,3…,20.

8. Caractère qualitatif :

Ses modalités ne s'expriment pas par un nombre
Exp : la religion, le sexe, l'opinion…

9. Caractère quantitatif :

Ses modalités sont numériques.
Exp : l'age, la taille, le poids…

10. Caractère quantitatif discret

L'ensemble des valeurs que peut prendre le caractère est fini ou dénombrable. Le plus souvent, ces valeurs sont entières.
Exp :le nombre d'enfant dans une famille, le nombre de téléviseurs par foyer et la pointure des souliers.

11. Caractère quantitatif continu :

Le caractère peut prendre théoriquement n'importe quelle valeur dans un intervalle donné de nombres réels.
Exp : la taille d'un individu, le poids…

12. Série statistique :

L'ensemble des différentes données associées à un certain nombre d'individus.
Exp : la série suivante résulte d'une courte enquête auprès de quelques personnes pour connaître leur age :
18 21 19 19 17 22 27 18 18 17 20 20 23

II- Tableaux statistiques :

A- Cas d'une seule variable :

Le tableau brut se présente sous la forme suivante:

Individu	variable
1	x_1
2	x_2
.	.
.	.
.	.
n	x_n

Le nombre d'individus observé étant en général important, le tableau précédant ne permet pas d'analyser l'information obtenue. Il est donc nécessaire de créer un tableau plus synthétique où les observations identiques (possédant la même modalité) ont été regroupées.

modalité	effectif
C_1	n_1
C_2	n_2
.	.
.	.
.	.
C_k	n_k

Pour une variable qualitative, les modalités ne sont pas mesurables.

Pour une variable quantitative, les modalités sont mesurables. Ce sont

- des valeurs numériques ponctuelles lorsque la variable est discrète
- des intervalles lorsque la variable est continue ou lorsque la variable est discrète et qu'elle comporte beaucoup de modalités.

<u>Application :</u>
Nous étudions une population de 1000 entreprises selon le caractère modalité « forme juridique ».
Les modalités retenues : S.A (Société Anonyme), SARL (Société A Responsabilité Limitée), EI (Entreprise Individuelle), SNC (Société en Nom Collectif).
Leurs effectifs respectifs : 200, 400, 340, 60.
<u>T.A.F :</u>
Présentez cette série dans un tableau.

B- Cas de deux variables :

Le tableau brut se présente sous la forme suivante:

Individu	variable1	variable2
1	x_1	y_1
2	x_2	y_2
.	.	.
.	.	.
.	.	.
n	x_n	y_n

On désire créer un tableau appelé tableau de contingence donnant le nombre d'individus possédant simultanément la modalité i de variable1 et la modalité j de variable2 qui se présentera sous la forme suivante:

		Variable2				
		D_1	. .	D_j	. .	D_r
variable1	C_1	n_{11}	. .	n_{1j}	. .	n_{1r}
	.	.		.		.
	.	.		.		.
	C_i	n_{i1}	. .	n_{ij}	. .	n_{ir}
	.	.		.		.
	.	.		.		.
	C_k	n_{k1}	. .	n_{kj}	. .	n_{kr}

Application:
Dans une entreprise, une enquête statistique a été faite sur 300 employés, et portant sur deux caractères, l'age et la rémunération. Les résultats de l'enquête sont présentés dans les deux tableaux suivants :

Age	n
20 à 25	150
25 à 30	100
30 à 35	200
35 à 40	50

Rémunération en dhs	n
Moins de 1500	200
1500 à 2000	150
2000 à 2500	100
plus de 2500	50

TAF :
Présentez dans un même tableau la distribution de ces deux caractères.

III- Représentations graphiques :

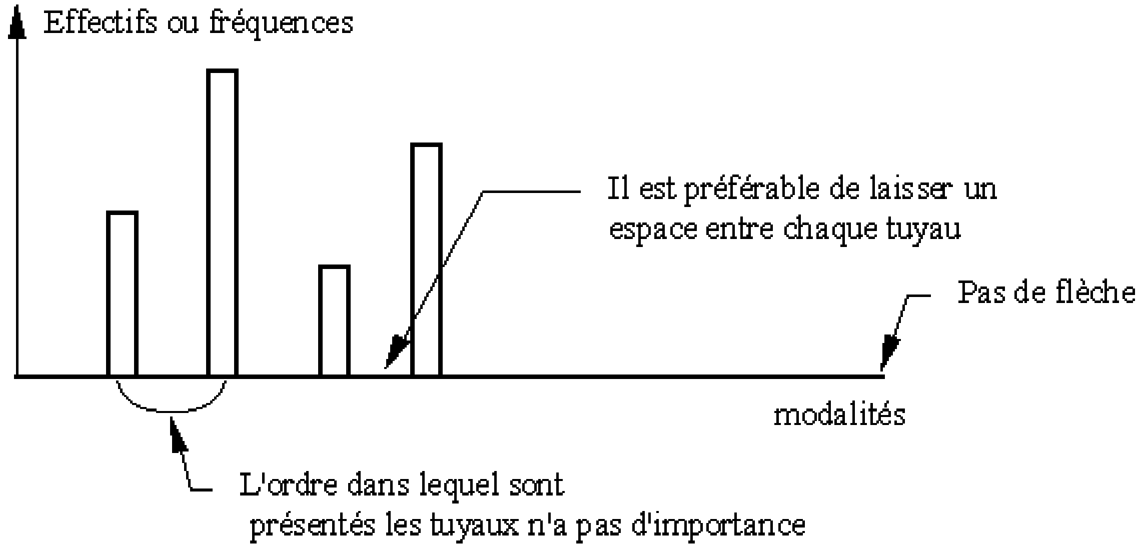
Lorsqu'on observe un caractère sur des individus, on aboutit à un tableau de chiffres peu parlant. L'objectif est de donner une représentation graphique de ce tableau qui permette d'un seul coup d'œil d'avoir une idée de la manière dont se répartissent les individus.

A- Variable qualitative :

A chaque modalité i est associé un effectif ni.
La seule représentation qui nous intéresse est celle des effectifs ni (ou des fréquences ni/n). Suivant la variable observée, de nombreuses représentations plus ou moins informatives peuvent être utilisées. Cependant les 2 plus classiques sont:

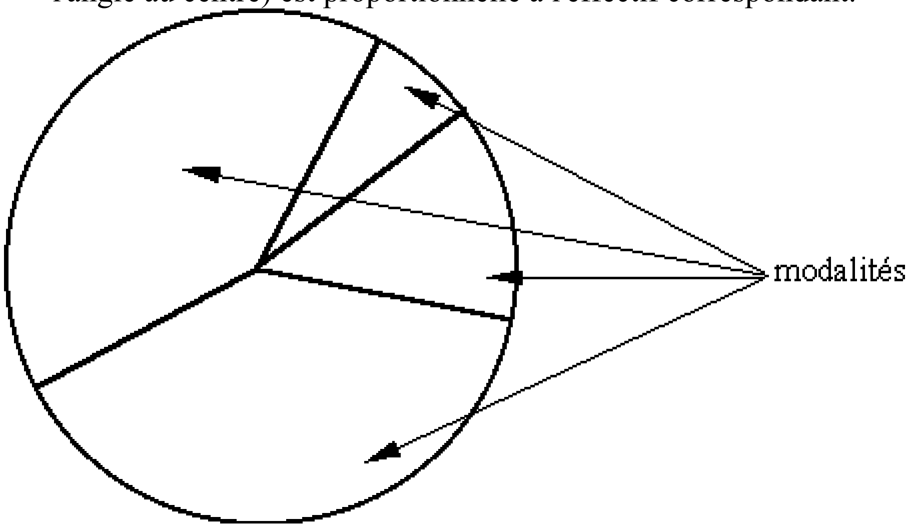
- **Les tuyaux d'orgue (ou diagramme en barre ou diagramme à bandes)**

  - les modalités de la variable sont placées sur une droite horizontale (attention: ne pas orienter cette droite car les modalités ne sont pas mesurables et il n'y a donc pas de relation d'ordre entre elles).
  - les effectifs (ou les fréquences) sont placés sur un axe vertical. La hauteur du tuyau est proportionnelle à l'effectif.



- **<u>les diagrammes à secteurs (ou camemberts)</u>**

- L'effectif total est représenté par un disque.
- Chaque modalité est représentée par un secteur circulaire dont la surface (pratiquement : l'angle au centre) est proportionnelle à l'effectif correspondant.



<u>Application :</u>
La répartition des candidats convoqués pour participer au Test d'Admissibilité à la Formation en Management (TAFEM 1998) pour l'accession à L'Ecole Nationale de Commerce et de Gestion d'Agadir , selon la série du baccalauréat se présente comme suit :

Série du Bac xi	Nombre de candidats ni
Sciences économiques	**250**
Sciences mathématiques	**200**
Sciences expérimentales	**400**
T.G.A	**50**
T.G.C	**100**
Total	**1000**

<u>TAF:</u> représentez cette distribution en Tuyaux d'orgues et Diagramme circulaire.

B- Variable quantitative :

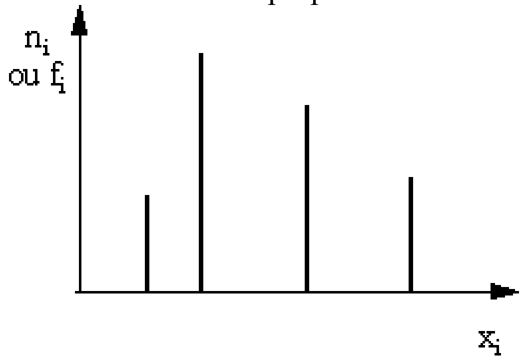
Avant toute tentative de représentation, il y a lieu de distinguer entre variable discrète et variable classée (regroupements en classes).

Deux types de graphiques sont intéressants de représenter:
a) les diagrammes différentiels qui mettent en évidence les différences d'effectifs (ou de fréquences) entre les différentes modalités ou classes.
b) les diagrammes cumulatifs qui permettent de répondre aux questions du style "combien d'individus ont pris une valeur inférieure (ou supérieure) à tant?".

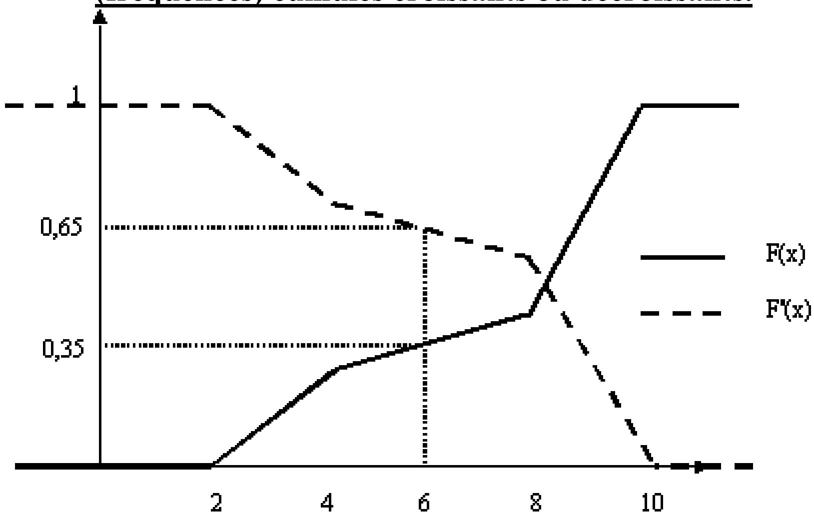
1) Variable discrète

- **Diagramme différentiel : le diagramme en bâtons**

  Les valeurs discrètes xi prises par les variables sont placées sur l'axe des abscisses, et les effectifs (ou les fréquences) sur l'axe des ordonnées. La hauteur du bâton est proportionnelle à l'effectif.



- **Diagrammes cumulatifs : ils permettent de visualiser l'évolution des effectifs (fréquences) cumulés croissants ou décroissants.**

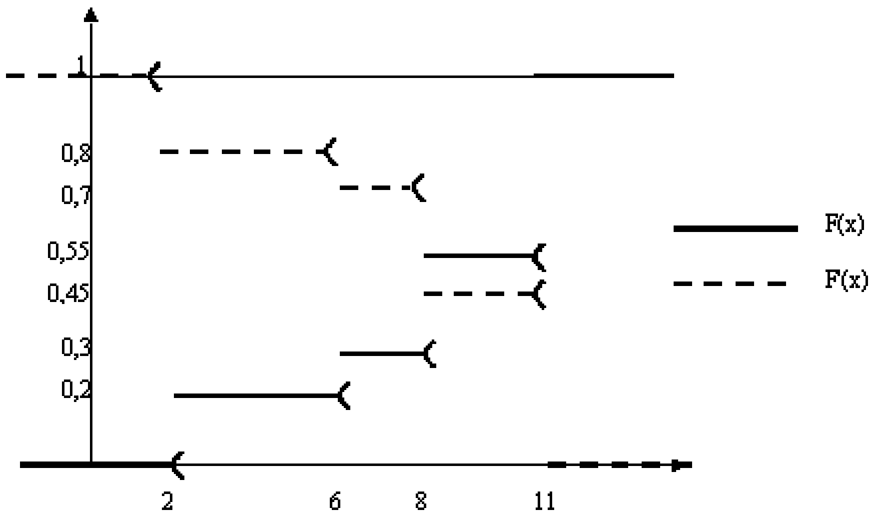


Remarque: les deux courbes sont symétriques par rapport à un axe horizontal d'ordonnée n/2 pour les effectifs, ½ pour les fréquences.
On utilise l'effectif (fréquence) cumulé croissant pour répondre aux questions du style :
Quel est le nombre (%) d'individus dont la valeur du caractère est inférieure ou égale à x ?
On utilise l'effectif (fréquence) cumulé décroissant pour répondre aux questions du style :
Quel est le nombre (%) d'individus dont la valeur du caractère est strictement supérieure à x ?

Se souvenir:

(au plus x) équivalent à ($\leq$ x) donc utiliser N(x) ou F(x)
(plus que x) équivalent à (> x) donc utiliser N '(x) ou F '(x)

Exemple:



- **(au plus 6)** équivalent à **($\leq$ 6)** donc on pourra lire la fréquence cumulée croissante en 6, c-à-d. **F(6) = 0,3**
- **(plus de 6)** équivalent à **(> 6)** donc on pourra lire la fréquence cumulée décroissante en 6, c.à.d. **F '(6) = 0,7**
- **(moins de 6)** équivalent à **(< 6)** équivalent à **($\leq$ 6-□)** où □ est une très faible valeur positive, donc on pourra lire la fréquence cumulée croissante en **6-□**, c.à.d. **F(6-□) = 0,2**

- **(au moins 6)** équivalent à **($\geq$ 6)** équivalent à **(> 6-□)** où □ est une très faible valeur positive, donc on pourra lire la fréquence cumulée décroissante en **6-□**, c.à.d. **F '(6-□) = 0,8**

Application :
Représentez graphiquement la distribution des 50 étudiants en fonction du nombre de personnes par ménage suivante :

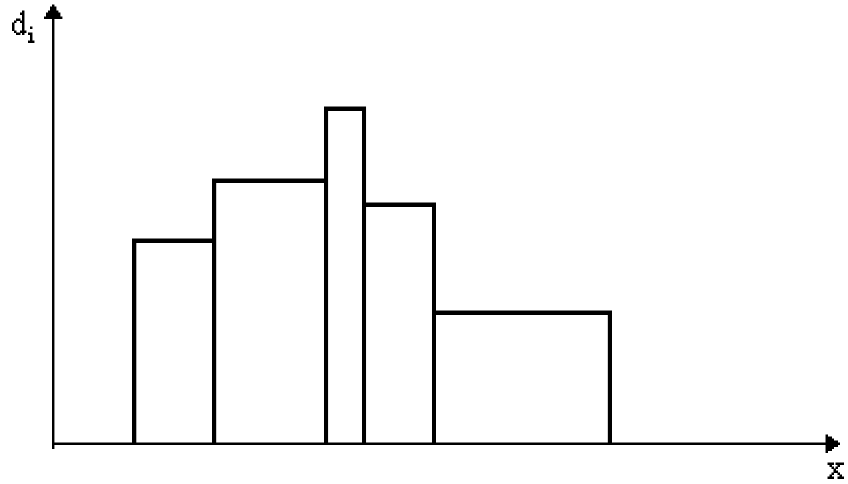
Nombre de personnes par ménage xi	Nombre d’étudiants ni
3	5
4	15
6	15
7	10
8	5
Total	50

2) Variable classée

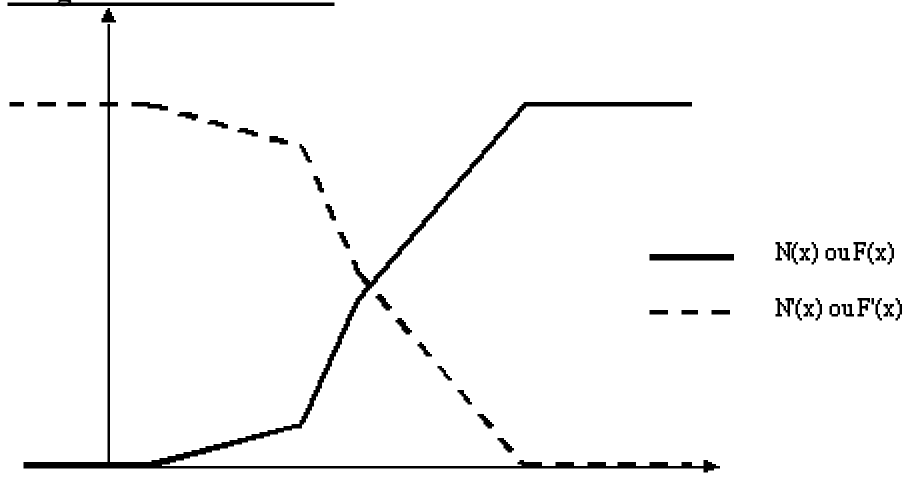
- **Diagramme différentiel : l'histogramme**

C'est un ensemble de rectangles contigus, chaque rectangle associé à chaque classe ayant une surface proportionnelle à l'effectif (fréquence) de cette classe.
Attention: Avant toute construction d'histogramme, il y a lieu de regarder si les classes sont d'amplitudes égales ou inégales.

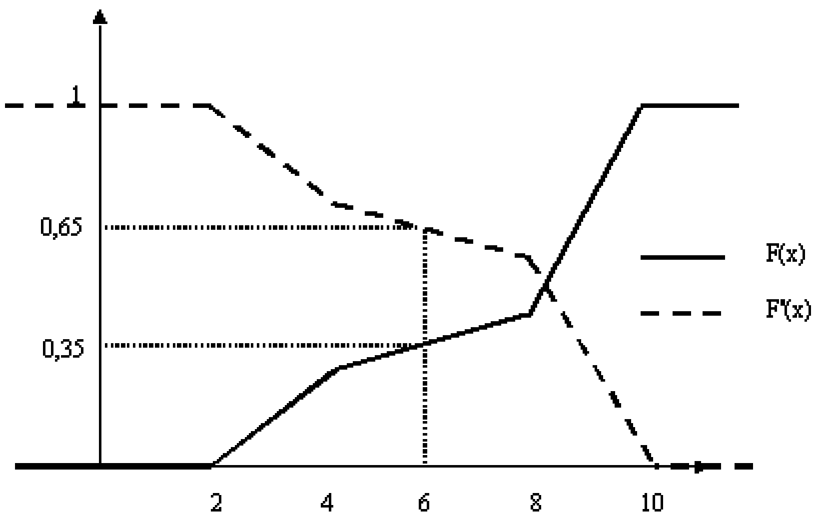
Le cas des classes d'amplitudes égales ne pose aucune difficulté car il suffit de reporter en ordonnée l'effectif (la fréquence).
Dans le cas d'amplitudes inégales on reporte en ordonnée la densité di (effectif divisé par l'amplitude de la classe)



- **Diagrammes cumulatifs**



L'utilisation des courbes est identique au cas discret.
Exemple:



Application :
Représentez graphiquement la distribution de 50 étudiants en fonction de leur taille suivante :

Taille en cm xi	Nombre d'étudiants
150-160	**16**
160-165	**6**
165-170	**12**
170-175	**14**
175-180	**2**
Total	**50**

IV- Caractéristiques de tendance centrale et de position :

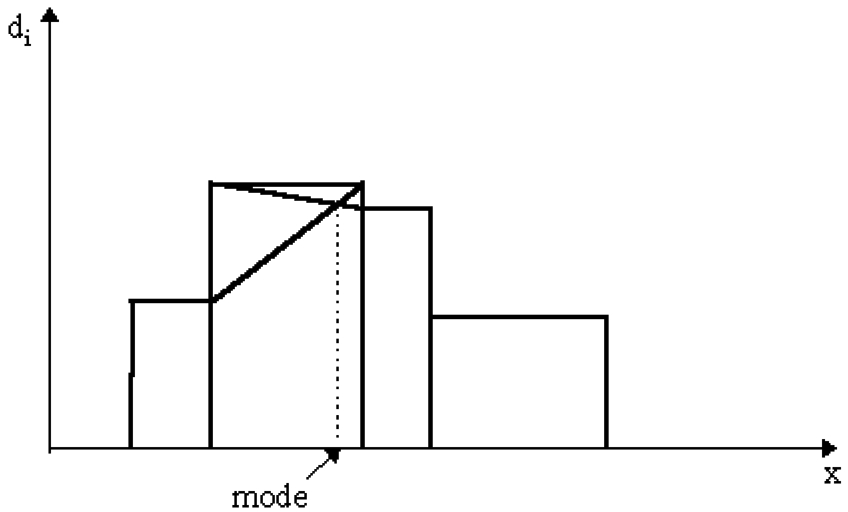
Les caractéristiques de tendance centrale essayent de donner la valeur la plus représentative d'un ensemble de valeurs numériques.

A- Mode :

C'est la valeur observée d'effectif maximum.

Variable discrète: classer les données par ordre croissant. Celle d'effectif maximum donne le mode.
Il est fortement conseillé d'utiliser le diagramme en bâtons pour déterminer le mode. En effet, deux valeurs consécutives xi , xi+1 peuvent avoir le même effectif maximum; on parlera d'intervalle modal [xi , xi+1]. Il peut aussi y avoir un mélange de deux populations qui conduit à un diagramme en bâtons où apparaissent deux bosses; on considérera deux modes. Il est déconseillé, sauf raison explicite, d'envisager plus de deux modes.

Variable classée: la classe modale correspond à la classe ayant l'effectif maximum. Il est fortement conseillé d'utiliser l'histogramme pour déterminer le mode. Comme pour le cas discret, on peut avoir deux classes modales. Toutes les valeurs de la classe pouvant à priori se réaliser, on ne se contentera pas de déterminer la classe modale. Une des valeurs de cette classe sera le mode. Certains auteurs préconisent par simplicité de prendre le centre de la classe modale. Il est préférable cependant de tenir compte des classes adjacentes de la manière suivante:



Application :
Déterminez la valeur modale de la distribution suivante, de 50 étudiants selon leur taille :

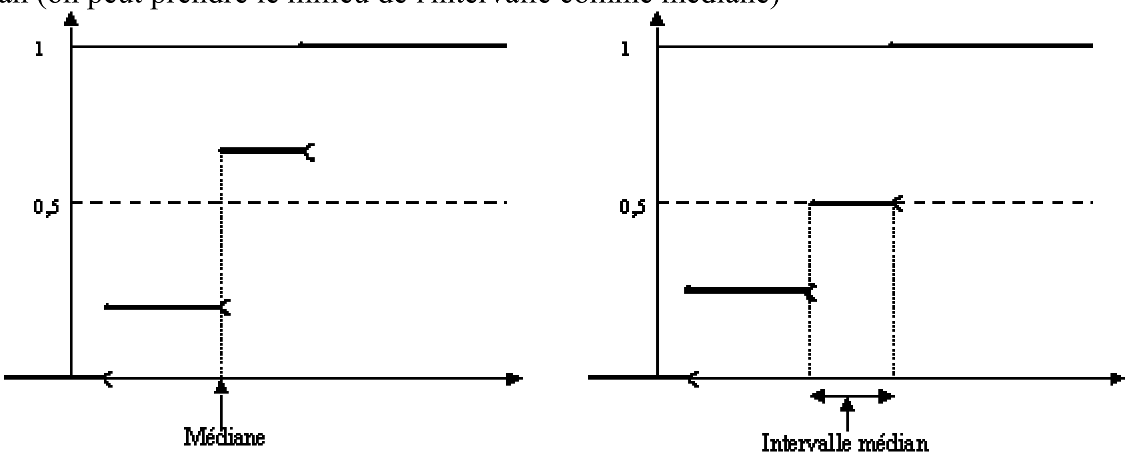
Taille en cm : xi	Nombre d'étudiants : ni
150-160	15
160-170	6
170-175	10
175-180	16
185-200	3
Total	50

Eléments de réponse :
Mo = 173.77 cm

B- Médiane :

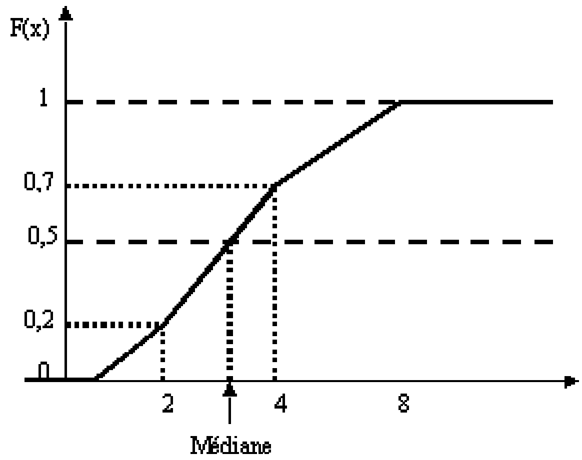
Les valeurs étant rangées par ordre croissant, c'est la valeur de la variable qui sépare les observations en deux groupes d'effectifs égaux.

Variable discrète: la détermination peut s'obtenir à partir du tableau statistique en recherchant la valeur de la variable correspondant à une fonction cumulée égale à n/2 (effectif cumulé) ou ½ (fréquence cumulée). Il est encore plus facile de lire sur les graphiques cumulatifs les abscisses des points d'ordonnée n/2 (effectif cumulé) ou ½ (fréquence cumulée). Si tout un intervalle a pour image n/2 (½ pour la fréquence), on parlera d'intervalle médian (on peut prendre le milieu de l'intervalle comme médiane)



Application :
Soit la série statistique suivante :
19 17 20 18 17 17 20 19 15 16 20 23 22 14 15 24
TAF : Calculez la médiane de cette série
Eléments de réponse :
Me=18.5

Variable classée: l'abscisse du point d'ordonnée n/2 (½ pour la fréquence)se situe en général à l'intérieur d'une classe. Pour obtenir une valeur plus précise de la médiane, on procède à une interpolation linéaire. La valeur de la médiane peut être lue sur le graphique ou calculée analytiquement.



$$\frac{\text{Mé} - 2}{4 - 2} = \frac{0{,}5 - 0{,}2}{0{,}7 - 0{,}2}$$

d'où la valeur de la médiane.
De manière générale, si a et b sont les bornes de la classe contenant la médiane, F(a) et F(b) les valeurs de la fréquence cumulée croissante en a et b, alors

$$\text{Mé} = a + (b - a) \times \frac{0{,}5 - F(a)}{F(b) - F(a)}$$

Application :
Déterminez la valeur médiane de la distribution des tailles suivantes :

Taille en cm xi	**Nombre d'étudiants ni**	**↑N**	**↓N**
150-160	**15**	**15**	**50**
160-165	**5**	**20**	**35**
165-170	**10**	**30**	**30**
170-175	**18**	**48**	**20**
175-180	**2**	**50**	**2**
Total	**50**	**#**	**#**

Eléments de réponse : Me = 167.5

C- Moyenne arithmétique :

Si xi sont les observations d'une variable discrète ou les centres de classe d'une variable

la moyenne arithmétique $\bar{x}$ est égale à $\sum_{i=1}^{k} \frac{n_i x_i}{n} = \sum_{i=1}^{k} f_i x_i$

classée,

La moyenne arithmétique est un paramètre de tendance centrale plus utilisé que les autres de par ses propriétés algébriques:

a) Pour plusieurs populations d'effectifs n1, n2, ..., nk, de moyennes respectives $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$

moyenne globale = moyenne des moyennes

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

b) La moyenne arithmétique conserve les changements d'échelle et d'origine

$$x : (x_i, n_i) \rightarrow y : (y_i = a x_i + b, n_i)$$

$$\bar{x} \rightarrow \bar{y} = a\bar{x} + b$$

Application :
Déterminez la taille moyenne des 50 étudiants dont la distribution par taille se présente comme suit :

Taille en cm xi	Nombre d'étudiants
150-160	16
160-165	6
165-170	12
170-175	14
175-180	2
Total	50

Eléments de réponse :
$\overline{x} = 168.3$ **cm**

D- Moyenne géométrique :

Si xi sont les observations d'une variable quantitative, la moyenne géométrique est égale à

$$G = \sqrt[n]{x_1^{n_1} \times ..... \times x_k^{n_k}}$$

Ce type de moyenne est surtout utilisé pour calculer des pourcentages moyens.
r étant un taux d'accroissement, 1+r est appelé coefficient multiplicateur; et le coefficient multiplicateur moyen est alors égal à la moyenne géométrique des coefficients multiplicateurs.

E- Moyenne harmonique :

Si xi sont les observations d'une variable quantitative, la moyenne harmonique est égale à

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

Il n'est pas évident d'utiliser ce type de moyenne.
Elle intervient lorsqu'on demande une moyenne de valeurs se présentant sous forme de quotient de deux variables x/y (km/h, km/litre,...). Attention, il faut cependant bien décortiquer le problème car il peut aussi s'agir d'une moyenne arithmétique.

Application :
Un cycliste effectue une traversé de 50 kms. Pendant les 20 premiers kms il roulait avec une vitesse constance de km/h, les 15 kms suivants à une vitesse constante de 30 km/h. Du point kilométrique 35 au 55 la vitesse de notre cycliste n'est que de 10 km/h et au-delà du point kilométrique sa vitesse n'est que de 5 km/h.
TAF :
Quelle est la vitesse de ce cycliste sur l'ensemble du parcours ?
Eléments de réponse :
H = 16.67

F- Moyenne quadratique :

Si xi sont les observations d'une variable quantitative, la moyenne harmonique est égale à

$$Q = \sqrt{\frac{n_1 x_1^2 + ... + n_k x_k^2}{n}}$$

G- Quantiles :

Ce sont des caractéristiques de position.
Il y a 1 médiane Me qui sépare les observations en 2 groupes d'effectifs égaux
3 quartiles Q1, Q2, Q3 qui séparent les observations en 4 groupes d'effectifs égaux
9 déciles D1, D2, ..., D9 qui séparent les observations en 10 groupes d'effectifs égaux
99 centiles C1, C2, ..., C99 qui séparent les observations en 100 groupes d'effectifs égaux
La détermination de ces caractéristiques est identique à celle de la médiane.
Les quartiles sont obtenus lorsqu'on a cumulé 25, 50, 75% de la population
Les déciles sont obtenus lorsqu'on a cumulé 10, 20,...., 90% de la population
Les centiles sont obtenus lorsqu'on a cumulé 1, 2,...., 99% de la population
Remarque: la notion de déciles et de centiles n'a de sens que s'il y a beaucoup d'observations et donc essentiellement pour une variable classée.

Application :
Soit la population de 80 salariés classés d'après le niveau de leur salaire journalier.

	Classes en dhs	ni	ni cumulés
1	90 à 100	5	5
2	100 à 110	9	14
3	110 à 120	16	30
4	120 à 130	25	55
5	130 à 140	13	68
6	140 à 150	7	75
7	150 à 160	3	78
8	160 à 170	2	80
Total		80	

TAF : calculez la médiane et les deux quartiles
Eléments de réponse :
Me = 124
Q_1= 110+ (10x6)/16 = 113.7
Q_3= 130+(10x5)/13 = 133.8

V- Caractéristiques de dispersion :

Comme leur nom l'indique, ces caractéristiques essayent de synthétiser par une seule valeur numérique la dispersion de toutes les valeurs observées.

A- Étendue :

C'est la différence entre la plus grande et la plus petite observation

Application :
Quelle est l'étendue de la série statistique suivante :
10 390 395 405 410 1000
Eléments de réponse :
Etendue = 990

B- Intervalle inter-quartile :

C'est la différence entre le troisième et le premier quartile

Application :
Reprenez les données de l'application sur les quartiles et calculez l'intervalle inter-quartile.
Eléments de réponse :
Q3-Q1=20

C- Variance et écart-type :

Si xi sont les observations d'une variable discrète ou les centres de classe d'une variable classée, la variance

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

On a aussi $V = \sum_{i=1}^{k} \frac{n_i x_i^2}{n} - \bar{x}^2$

c.à.d. moyenne des carrés - carré de la moyenne

On utilise plus couramment l'écart type qui est la racine carrée de la variance et qui a l'avantage d'être un nombre de même dimension que les données (contrairement à la variance qui en est le carré)

La variance est un paramètre de dispersion plus utilisé que les autres de par ses propriétés algébriques:

a) Pour plusieurs populations d'effectifs $n_1, n_2, \ldots, n_k$, de moyennes respectives $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, de variances $V_1, V_2, \ldots, V_k$

Variance globale = variance des moyennes + moyenne des variances

$$V = \sum_{i=1}^{k} \frac{n_i}{n} (\bar{x}_i - \bar{\bar{x}})^2 + \sum_{i=1}^{k} \frac{n_i}{n} V_i$$

où $\bar{\bar{x}}$ représente la moyenne des moyennes

b) changement d'échelle et d'origine

$$x : (x_i , n_i) \rightarrow y : (y_i = a x_i + b , n_i)$$

$$V_x \rightarrow V_y = a^2 V_x$$

D- Coefficient de variation :

$$CV = \frac{\sigma}{\bar{x}}$$

C'est un coefficient qui permet de relativiser l'écart type en fonction de la taille des valeurs. Il permet ainsi de comparer la dispersion de séries de mesures exprimées dans des unités différentes

Applications :

App.1- Les séries suivantes représentent la mesure d'un caractère auprès des individus d'une population :

a. 6 1 8 10 5 4 11 3 2 9 7 12 13
b. 19 17 7 1 4 24 15 22 10 13
c. 15 12 17 15 20 15 20 15 15 9 7
d. 21 25 34 10 20 27 14 20 34

Dans chacun de ces cas calculez : la moyenne, la médiane, le mode,la variance, l'écart type et le coefficient de variation.

Eléments de réponse :

- **a.** $\bar{x}$**=7, Me=7, pas de mode, σ²=14, σ=3.74, V=53.4%**
- **b. x=13.2, Me=14, pas de mode, σ²=52.76, σ=7.26, V=55%**
- **c.** $\bar{x}$**=14.5, Me=15, Mo=15, σ²=14.61, σ=3.82, V=26.3%**
- **d. x=22.8, Me=21, deux modes :20 et 34, σ²=59.28, σ=7.70, V=33.8%**

App.2- La distribution suivante représente la répartition de la longueur de pinces d'écrevisse provenant d'une rivière :

Limites	ni
1.02---1.23	**5**
1.24---1.45	**7**
1.46---1.67	**4**
1.68---1.89	**1**
1.90---2.11	**4**
2.12---2.33	**6**
2.34---2.55	**3**
2.56---2.77	**1**

TAF : calculez : la moyenne, la médiane, le mode,la variance, l'écart type et le coefficient de variation.

Eléments de réponse :

$\overline{x}$=1.757, Mo=1.345 (le centre de la classe modale), Me=1.648, σ^2=0.238, σ=0.488, V=27.8%

VI- La concentration :

L'objectif est de mesurer les inégalités dans la répartition d'une variable à l'intérieur d'une population. Cette notion n'a d'intérêt que dans la mesure où les valeurs globales suivantes ont une signification concrète

A- Valeurs globales :

xi représentent les valeurs ponctuelles ou les centres des classes, ni les effectifs correspondants.
Les valeurs globales de la série (xi , ni) sont les quantités gi = ni xi

B- Médiale :

La médiale de la série (xi , ni) est la médiane de la série (xi , gi)

Application :

L'importance quantitative des portefeuilles de titres déposés dans une société de portefeuille « Maroc Invest » en Kdh en 1996.

Importance du portefeuille en kdh	f%	f cumulé	f'%	f'cumulé
Moins de 10.000	**41**	**41**	**2**	**2**
10.000 à 50.000	**37**	**78**	**15**	**17**
50.000 à 100.000	**10**	**88**	**11**	**28**
100.000 à 200.000	**6**	**94**	**13**	**41**
200.000 à 500.000	**4**	**98**	**19**	**60**
500.000 à plus	**2**	**100**	**40**	**100**
Total	**100**	**-**	**100**	**-**

f représentent les pourcentages du nombre total des portefeuilles.
f' représentent les pourcentages de la valeur totale des portefeuilles.
TAF : calculez la médiane et la médiale de cette distribution

Eléments de réponse :

Me = 19730, Ml= 342105 kdh

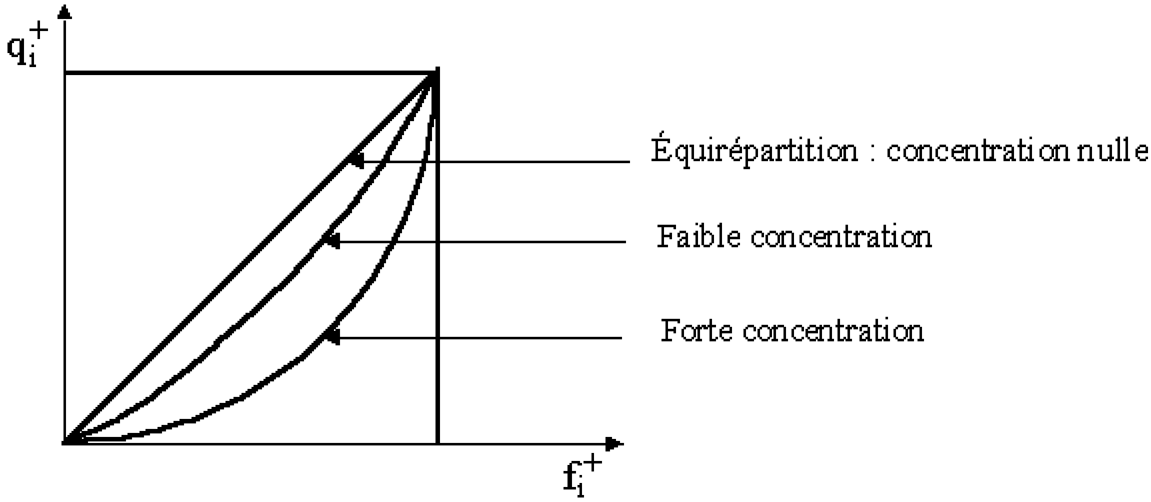
C- Courbe de concentration (ou de LORENZ)

C'est la courbe obtenue en représentant

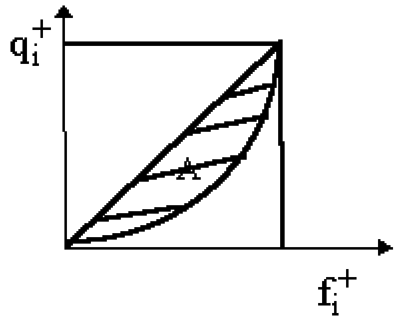
en abscisse, f_i^+ les fréquences cumulées croissantes de la série (x_i , n_i)

en ordonnée, q_i^+ les fréquences cumulées croissantes de la série (x_i , g_i)

L'allure de la courbe permet d'avoir une idée de la concentration



D- Indice de GINI



$$\gamma = 2 A$$

Propriétés:

- $0 < \gamma < 1$
- γ proche de 1 $\Rightarrow$ forte concentration
- γ proche de 0 $\Rightarrow$ faible concentration

Exercice synthétique : (voir TP N°1)

VII- Les indices :

Permettent de mesurer l'évolution d'un phénomène au cours du temps

A- Indices élémentaires :

L'indice d'évolution d'une variable élémentaire y entre la date t0, dite date de référence ou date de base, et la date t, dite date courante est

$$i_{t/t_0} = \frac{y_t}{y_{t_0}}$$

L'indice base 100, c.à.d. exprimé en pourcentage est

$$I_{t/t_0} = i_{t/t_0} \times 100$$

Remarque: Il est toujours préférable d'effectuer les calculs avec i et de donner le résultat en base 100 à la fin des calculs.

On utilise essentiellement l'indice des prix (P), l'indice des quantités ou volumes (Q), et l'indice des valeurs ou dépenses (V = P Q)

Propriétés:

- identité $i_{t/t} = 1$
- réversibilité $i_{t_2/t_1} \times i_{t_2/t_1} = 1$
- circularité $i_{t_3/t_1} = i_{t_3/t_2} \times i_{t_2/t_1}$
- L'indice est étroitement lié au taux de croissance

$$r_{t_1/t_0} = \frac{y_{t_1} - y_{t_0}}{y_{t_0}} = i_{t_1/t_0} - 1$$

i = r +1 est aussi appelé coefficient multiplicateur par les économistes

$$r = 0 \quad \Leftrightarrow \quad i = 1$$

$$r > 0 \quad \Leftrightarrow \quad i > 1$$

$$-1 < r < 0 \quad \Leftrightarrow \quad 0 < i < 1$$

Applications :

App.1- Le prix de la tomate au Maroc a été de 1.5 dhs en moyenne en 1980 et de 2.3 dhs en 1995.
TAF : calculez l'indice élémentaire du prix de la tomate en 1995, base 100 en 1980 et interprétez-le.

Eléments de réponse :

$I_{95/80} = \frac{G_{95}}{G_{80}} = (2.3/1.5) \times 100 = 153.33$

Le prix de la tomate au Maroc a augmenté de 53.33% entre 1980 et 1995

App.2- On savait que le prix du sucre dans un pays X a augmenté de 2.5% entre 1960 et 1975 et de 7.5% entre 1960 et 1995.
TAF : déterminez l'indice élémentaire du prix du sucre en 1995 base 100 en 1975, pour le pays en question.

Eléments de réponse :

$I_{95/75} = \frac{I_{95/75}}{I_{75/60}} = \frac{107.5}{102.5} \times 100 \approx 104.88$

Exercice de synthèse :

Les données concernant l'évolution des prix de plusieurs articles entre les périodes 1995 et 1985, ainsi que leur poids sont groupés dans le tableau suivant :

Prix / Articles	P'_{85}	P'_{95}	α^i
A	36	40	0.15
B	12	15	0.10
C	40	45	0.25
D	15	13	0.05
E	42	50	0.15
F	5	8	0.10
G	30	40	0.05
H	8	10	0.15

TAF: calculez les indices élémentaires des prix des différents articles, puis déterminez l'indice général des prix.

Eléments de réponse :

I95/85 (PA) = 40/36 x 100 = 111.11
I95/85 (PB) = 15/12 x 100 = 125
I95/85 (PC) = 45/40 x 100 = 112.5
I95/85 (PD) = 13/15 x 100 = 86.67
I95/85 (PE) = 50/42 x 100 = 119.05
I95/85 (PF) = 8/5 x 100 = 160
I95/85 (PG) = 40/30 x 100 = 133.33
I95/85 (PH) = 10/8 x 100 = 125

- L'indice des moyennes: $I_{95/85} = \frac{\overline{P_{95}}}{\overline{P_{85}}} = 31.2/26.85 \times 100 = 116.2$

- La moyenne des indices : $I_{95/85}(P) = \sum \alpha^i I_{95/85}^i = 120.9$

B- Indices de LASPEYRES et de PAASCHE

Ce sont des indices synthétiques qui sont des résumés numériques des indices élémentaires lorsqu'on cherche à mesurer l'évolution d'un ensemble de plusieurs produits.
coefficient de pondération ou budgétaire du produit j par rapport à la date t :

$$\alpha_{j,t} = \frac{P_{j,t}Q_{j,t}}{\sum_{j=1}^{n} P_{j,t}Q_{j,t}}$$

a) Indice de Laspeyres des prix

$$\mathcal{L}(P)_{t/t_0} = \sum_j \alpha_{j,t_0} I(P_j)_{t/t_0} = \frac{\sum_j P_{j,t}Q_{j,t_0}}{\sum_j P_{j,t_0}Q_{j,t_0}} \times 100$$

$$= \frac{\text{Dépense de la date courante exprimée en quantités de la date de référence}}{\text{Dépense de la date de référence}} \times 100$$

b) Indice de Laspeyres des quantités

$$\mathcal{L}(Q)_{t/t_0} = \sum_j \alpha_{j,t_0} I(Q_j)_{t/t_0} = \frac{\sum_j P_{j,t_0}Q_{j,t}}{\sum_j P_{j,t_0}Q_{j,t_0}} \times 100$$

$$= \frac{\text{Dépense de la date courante exprimée en prix de la date de référence}}{\text{Dépense de la date de référence}} \times 100$$

c) Indice de Paasche des prix

$$\mathcal{P}(P)_{t/t_0} = \frac{1}{\sum_j \frac{\alpha_{j,t}}{I(P_j)_{t/t_0}}} = \frac{\sum_j P_{j,t}Q_{j,t}}{\sum_j P_{j,t_0}Q_{j,t}} \times 100$$

$$= \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence exprimée en quantités de la date courante}} \times 100$$

d) Indice de Paasche des quantités

$$\mathcal{P}(Q)_{t/t_0} = \frac{1}{\sum_j \frac{\alpha_{j,t}}{I(Q_j)_{t/t_0}}} = \frac{\sum_j P_{j,t}Q_{j,t}}{\sum_j P_{j,t}Q_{j,t_0}} \times 100$$

$$= \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence exprimée en prix de la date courante}} \times 100$$

Application :
Les données concernant l'évolution des prix et des quantités de plusieurs articles entre les périodes 1995 et 1985 :

Prix / Articles	P'_{85}	P'_{95}	Q'_{85}	Q'_{95}
A	36	40	6	7
B	12	15	20	20
C	40	45	13	11
D	15	13	15	15
E	42	50	9	18
F	5	8	25	25
G	30	40	10	9
H	8	10	30	30

TAF : calculez les différents indices synthétiques des prix, des quantités et des valeurs.
Eléments de réponse :
- Indice de Laspeyrs des prix :
$L_{95/85}$ (P) = 125
- Indice de Paasche des prix :
P (P) = 119
- Indice de Laspeyrs des quantités:
$L_{95/85}$ (Q) = 119
- Indice de Paasche des quantités :
P (P) = 134
- indice des valeurs (indice des dépenses totales) :

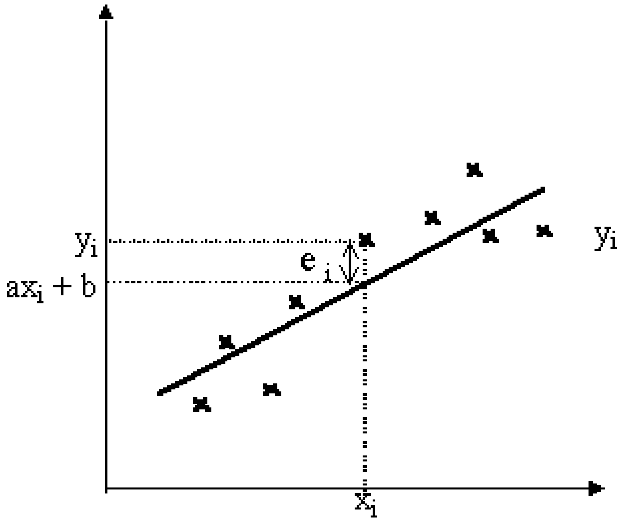
$$D_{95/85} = \frac{\sum P'_{95} Q'_{95}}{\sum P'_{85} Q'_{85}} = 3030/2136 \quad x\ 100 = 142$$

VIII- Régression et corrélation :

Lorsqu'on observe deux variables quantitatives sur les mêmes individus, on peut s'intéresser à une liaison éventuelle entre ces deux variables.
La régression fournit une expression de cette liaison sous la forme d'une fonction mathématique.
La corrélation renseigne sur l'intensité de cette liaison.

A- Ajustement d'un nuage de points à une fonction mathématique :

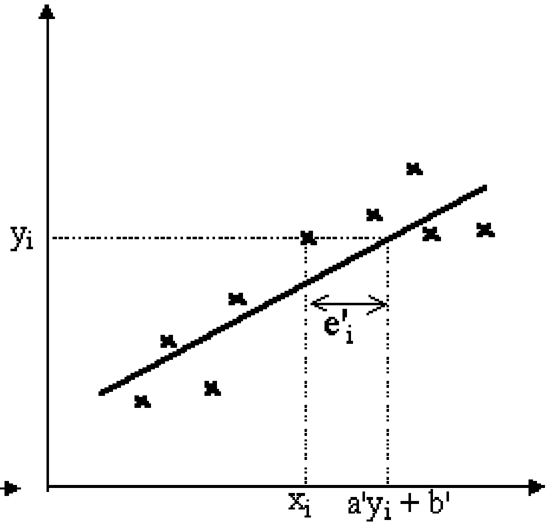
a) Ajustement linéaire par la méthode des moindres carrés
Lorsque le nuage de points (xi , yi) est à peu près rectiligne, on peut envisager d'exprimer la liaison entre x et y sous forme de fonction affine y = ax + b



$$\sum_i e_i^2 = \sum_i (y_i - ax_i - b)^2$$

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

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$



$$\sum_i e_i'^2 = \sum_i (x_i - a'y_i - b')^2$$

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

$$a' = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

b) <u>Ajustement à une fonction exponentielle</u>

Pour ajuster un nuage de points à une courbe exponentielle $y = ba^x$, il suffit de faire le changement de variable Y = ln y , X = x , A = ln a , B = ln b , pour obtenir l'équation Y = AX + B, et d'utiliser ensuite l'ajustement linéaire par la méthode des moindres carrés sur les points (Xi , Yi).

c) <u>Ajustement à une fonction puissance</u>

Pour ajuster un nuage de points à une courbe puissance $y = bx^a$, il suffit de faire le changement de variable Y = ln y , X = ln x , A = a , B = ln b , pour obtenir l'équation Y = AX + B , et d'utiliser ensuite l'ajustement linéaire par la méthode des moindres carrés sur les points (Xi , Yi).

B- Mesure de l'intensité de la relation linéaire entre deux variables :

1) <u>Covariance</u>

$$\mathrm{Cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$\mathrm{Cov}(x,y) > 0 \Leftrightarrow$ x et y varient dans le même sens

$\mathrm{Cov}(x,y) < 0 \Leftrightarrow$ x et y varient en sens contraire

$\mathrm{Cov}(x,y) = \mathrm{Cov}(y,x)$

$\mathrm{Cov}(x,x) = V(x)$

2) Coefficient de corrélation linéaire

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\mathrm{cov}(x,y)}{\sigma(x)\,\sigma(y)}$$

$$-1 \leq r \leq 1$$

$$y = ax + b \quad \Leftrightarrow \quad \begin{cases} r = 1 & \text{si} \quad a > 0 \\ r = -1 & \text{si} \quad a < 0 \end{cases}$$

$|r| = 1 \quad \Leftrightarrow$ relation fonctionnelle linéaire

$r = 0 \quad \Leftrightarrow$ indépendance linéaire

$0 < |r| < 1 \quad \Leftrightarrow$ dépendance linéaire d'autant plus forte que $|r|$ est grand

Attention:

Une forte causalité entre x et y implique une forte relation entre x et y qui n'est pas forcément linéaire; on n'a donc pas obligatoirement une forte corrélation linéaire.
Une forte corrélation linéaire n'implique pas forcément une forte causalité.

3) Droites de régression

Dy/x : y = ax + b avec $a = \dfrac{\mathrm{cov}(x,y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$

Dx/y : x = a'y + b' avec $a' = \dfrac{\mathrm{cov}(x,y)}{V(y)}$ et $b' = \bar{x} - a'\bar{y}$

La position des deux droites de régression l'une par rapport à l'autre donne un renseignement sur l'intensité de la relation linéaire:

* droites de régression confondues $\Leftrightarrow \quad aa' = 1 \quad \Leftrightarrow$ relation fonctionnelle linéaire
* droites de régression perpendiculaires dont une de pente nulle $\Leftrightarrow \quad aa' = 0 \quad \Leftrightarrow$ indépendance linéaire
* Plus les droites sont proches, plus la relation linéaire est importante

Relations intéressantes:

$r^2 = aa'$

$$r = a\frac{\sigma(x)}{\sigma(y)} = a'\frac{\sigma(y)}{\sigma(x)}$$

Application :

Les séries statistiques simples de deux variables continues X et Y se présentent comme suit :

Individus	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
X	**2**	**12**	**13**	**7**	**6**	**3**	**12**	**10**	**9**	**7**	**4**	**2**	**10**	**6**	**3**
Y	**22**	**2**	**4**	**14**	**15**	**19**	**7**	**8**	**10**	**11**	**16**	**18**	**11**	**12**	**21**

TAF : après avoir élaboré un tableau de contingence, en adoptant des classes d'amplitudes égales à 4 unités pour la variable X et des amplitudes à 5 unités pour la variable Y, il vous est demandé d'apprécier la liaison qui existe entre ces deux variables.

Eléments de réponse :

Y **X**	**2 – 7**	**7 – 12**	**12 – 17**	**17 – 22**	**$n_{.j}$**
2 – 6	**0**	**0**	**2**	**3**	**5**
6 – 10	**0**	**3**	**2**	**0**	**5**
10 – 14	**3**	**2**	**0**	**0**	**5**
$n_{i.}$	**3**	**5**	**4**	**3**	**15**

Les équations des droites d'ajustement linéaire :
-l'ajustement linéaire de Y à X : Y= a.X + b = -1.37 X+ 22.79
-l'ajustement linéaire de X à Y : X = a.Y + b = -0.56 Y+14.62
- coefficient de corrélation r : r = -0.87 → Forte liaison linéaire négative entre les deux variables.

IX- Séries chronologiques :

Ce sont des séries d'observations échelonnées dans le temps. L'objectif de l'étude des séries chronologiques est double:

- analyse d'un phénomène temporel en mettant en évidence essentiellement la tendance générale et les fluctuations saisonnières
- élaboration d'un modèle permettant de faire de la prévision à court terme

A- Décomposition des chroniques :

L'évolution dans le temps d'un phénomène résulte de plusieurs facteurs :
- le Trend ou Tendance : T. C'est le mouvement de longue période que l'on considère le plus souvent comme une droite (tendance linéaire)
- les cycles : C. C'est une alternance de mouvements croissants et décroissants de moyen terme.
- les variations saisonnières : S. On estime qu'il y a une composante saisonnière dans une série, si, chaque année, à la même période, il se produit une variation du phénomène d'au moins 25% par rapport à la valeur moyenne.
- le résidu ou aléa : ε. C'est un événement exceptionnel impossible ou difficile à estimer.

L'évolution d'une variable X peut alors s'exprimer comme suit :
(1) $X= T+C+S+\varepsilon$ ou (2) $X= T.C.S.\varepsilon$
Le modèle additif (1) suppose que chaque composante apporte une contribution pure à l'évolution observée.
Le modèle multiplicatif (2) montre que chaque composante amplifie les autres et traduit l'interdépendance entre les composantes.

B- La détermination du Trend :

1) Ajustement linéaire par la méthode des moindres carrés

La droite de régression de Y par rapport au temps t donne pour chaque t une valeur Tt

2) Lissage par moyennes mobiles d'ordre k (k = nombre d'observations dans un cycle)

temps	variable	moyennes mobiles d'ordre 3	moyennes mobiles d'ordre 4
1	y1		
2	y2	(y1 + y2 + y3)/3	
3	y3	(y2 + y3 + y4)/3	(y1/2 + y2 + y3 + y4 + y5/2)/4
4	y4	(y3 + y4 + y5)/3	(y2/2 + y3 + y4 + y5 + y6/2)/4
5	y5	(y4 + y5 + y6)/3	(y3/2 + y4 + y5 + y6 + y7/2)/4
6	y6	(y5 + y6 + y7)/3	
7	y7		

les moyennes mobiles donnent pour chaque t (mis à part les valeurs extrêmes) une valeur Tt

Application :

La société BMT a pour activité la vente de système d'alarme. Le caractère porteur de ce marché lui a permis sur les cinq dernières années d'enregistrer les ventes suivantes en KDH :

Années	N –4	N –3	N –2	N –1	N
Chiffre d'affaires	71697	90574	94550	125257	138150

TAF :estimez la prévision des ventes pour l'année N+1 en utilisant la méthode des moindres carrés.

Eléments de réponse :

soit x le rang de l'année et y le chiffre d'affaires

	xi	yi	xiyi	xi²
	1	71697	71697	1
	2	90574	181148	4
	3	94550	283650	9
	4	125257	501028	16
	5	138150	690750	25
Sommes	15	520228	1728272	55
Moyennes	3	104046		

a=16759 et b=53769

le chiffre d'affaires y s'exprimerait donc en fonction du rang x de l'année :

y=16759x + 53769

Pour l'année N+1 (rang 6), la prevision serait la suivante : y=16759 x 6+53769 = 154323 kdh

C- Analyse de la composante saisonnière :

1) modèle additif

- calcul des différences Yt - Tt = St + At
- calcul des coefficients saisonniers bruts S'j : pour chaque saison j, S'j = moyenne des différences de la saison j
- calcul des coefficients saisonniers $S_j = S'_j - \overline{S}'$

2) modèle multiplicatif

- calcul des rapports Yt / Tt = St . At
- calcul des coefficients saisonniers bruts S'j : pour chaque saison j, S'j = moyenne des rapports de la saison j
- calcul des coefficients saisonniers $S_j = S'_j / \overline{S}'$

D- Analyse de la composante aléatoire

1) modèle additif

At = Yt - Tt - St

2) modèle multiplicatif

At = Yt / (Tt . St)

E- Désaisonnalisation :

Pour exprimer ce qu'aurait été le mouvement brut sans l'influence saisonnière, on utilise la série corrigée des variations saisonnières Y* (ou Ycvs)

1) modèle additif

Y*t = Yt - St

2) modèle multiplicatif

Y*t = Yt / St

F- Série Ajustée

Cette série est utilisée pour représenter ce qu'aurait été le phénomène en l'absence de phénomènes aléatoires

1) modèle additif

$\hat{Y}_t$ = Tt + St

2) modèle multiplicatif

$\hat{Y}_t$ = Tt . St

F- Prévision à court terme:

Lorsque le trend est obtenu par la méthode des moindres carrés, il est possible d'obtenir une prévision postérieure à l'intervalle d'étude (à condition de rester dans des limites raisonnables), en utilisant le modèle précédent. Pour une date x correspondant à un coefficient saisonnier Sx , la tendance vaut Tx , et la prévision est donc donnée par Tx + Sx en modèle additif ou Tx . Sx en modèle multiplicatif

Application :

La société Jihane fabrique des jouets en plastique. Son activité a un caractère saisonnier très marqué. On dispose des données suivantes relatives aux années N-2, N-1 et N :

	N – 2	N – 1	N
Trimestre 1	18912	25052	27635
Trimestre 2	28362	37579	41440
Trimestre 3	33098	43837	48357
Trimestre 4	14178	18789	20718
Total	94550	125257	138150

TAF :

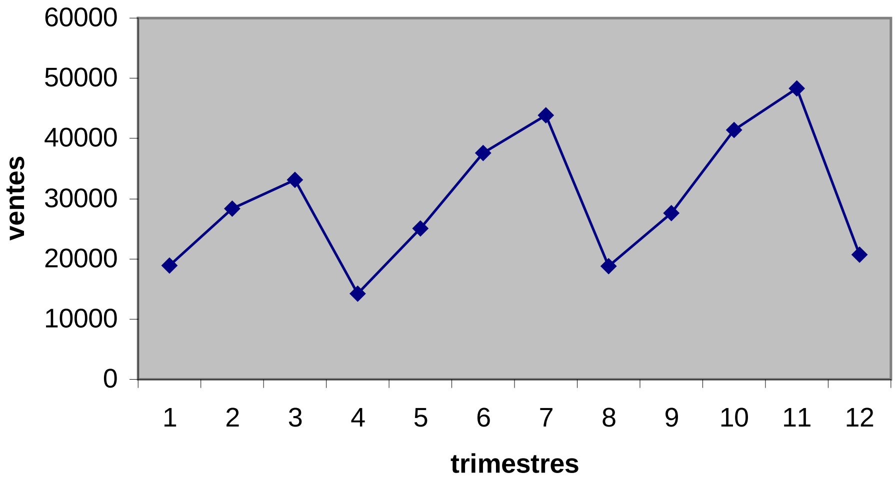
1. Représentez graphiquement cette série statistique

2. Calculez les coefficients saisonniers de cette série.

3. Déterminez la série corrigée des variations saisonniers

4. Quelles sont les prévisions pour les années N+1, N+2, N+3 et N+4 ?

Eléments de réponse :



2.

	Trimestr1			Timestre2			Trimestre3			Trimestre4		
	y_t	y'_t	y_t/y'_t	y_t	y'_t	y_t/y'_t	y_t	y'_t	y_t/y'_t	y_t	y'_t	y_t/y'_t
N – 2	18912			28362			33098	24405	1.36	14178	26325	0.54
N-1	25052	28819	0.87	37579	30738	1.22	43837	31637	1.39	18789	32443	0.58
N	27635	33490	0.83	27635	34296	1.21	48357			20718		
Coeff saisonniers			0.85			1.215			1.375			0.56

Coefficient saisonnier 1[er] trimestre = (0.87+0.83)2 = 0.85

3.

Trimestre t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	18912	28362	33098	14178	25052	37579	43837	18789	27635	41440	48357	20718
Coeff.sais.	0.85	1.215	1.375	0.56	0.85	1.215	1.375	0.56	0.85	1.215	1.375	0.56
Série corrigée	21013	22690	25460	25778	27836	30063	33721	34162	30706	33152	37198	37669

4. la prévision de la tendance nécessite un ajustement de la série corrigée des variations saisonniers (les moyennes mobiles).
Droite d'ajustement de y'_t => y'_t = 1391x + 21228
On obtient les prévisions suivantes pour la tendance :

Trimestre	**13**	**14**	**15**	**16**
Prévision	**39311**	**40702**	**42093**	**43484**

Prévisions des ventes des trimestres 13,14,15 et 16 (N+1, N+2, N+3 et N+4)

Trimestre	**13**	**14**	**15**	**16**
Prévision de la tendance	**39311**	**40702**	**42093**	**43484**
Coeff. Saisonn.	**0.85**	**1.215**	**1.375**	**0.56**
Prévisions des ventes	**33414**	**49453**	**57878**	**24351**

Chapitre II. Réalisation des enquêtes

Enquête : Investigation auprès d'une population donnée pour obtenir des réponses précises à des questions sur un marché (enquête par téléphone, enquête postale, enquête par Internet..)

I- Détermination optimale d'un échantillon

Echantillon : fraction représentative d'une population ou d'un univers statistique sur lequel porte une étude. Tous les membres de la population considérés doivent avoir la même chance d'être choisis.

A. Méthodes d'échantillonnage :

Il existe différentes manières d'extraire un échantillon d'une population. Nous ne verrons que les deux pratiques les plus courantes :

1- Echantillon aléatoire :

Tous les individus d'une population possèdent au départ des chances égales de faire partie de l'échantillon. On effectue un choix au hasard.

2- Echantillon stratifié :

On divise en strates le population et on tire au hasard dans chaque strate homogène, les éléments obtenus dans chaque strate sont combinés pour obtenir le résultat final.

3- Tirage par quota :

Il consiste à reconstituer une population mère miniaturisée, au sein de l'échantillon. L'échantillon est considéré comme représentatif de la population mère.

Exp : dans une population donnée, il y a 49% de femmes et 51% d'hommes ; on définit les quotas qui permettront d'obtenir un échantillon comprenant 49% de femmes et 51% d'hommes.

B. Détermination optimale de la taille de l'échantillon :

Exp : un calcul financier prévisionnel a un chef de produit que sa nouvelle marque doit obtenir une part de marché d'au moins 15%, s'il veut dégager un bénéfice. Une étude est menée auprès de s acheteurs potentiels. Le chef de produit fait pari qu'une part de marché de 20% est tout à fait probable. Il se donne une marge de fluctuation de ± 3 points autour de ce chiffre. Il veut organiser un test qui simule un achat réel, en présentant les principales marques du marché. Combien faudra-t-il interroger de consommateurs potentiels pour vérifier la prévision,

Formule de calcul : $\boxed{n=\frac{z^2 p\, q}{e^2}}$

avec :
n : taille de l'échantillon nécessaire
z : valeur fournie par la table de la loi normale ; elle varie selon le risque d'erreur que l'on accepte pour généraliser les résultats. L'usage est de retenir 5% soit une valeur de z=1.96
p : pourcentage prévu de consommateurs qui achètent la nouvelle marque, soit ici 20%
q =1-p : pourcentage de consommateurs qui choisissent une autre marque , ici 80%.
e: marge de fluctuation (précision) acceptée pour généraliser les résultats : ici ± 3 points de part de marché, soit 0.03.

Résultats :

$$n=\frac{(1.96)^2(0.2)(0.8)}{(0.03)^2}=683$$

II- Elaboration du questionnaire

A- Définition :

Instrument de collecte de l'information. Il est fondé sur un recueil de réponses à un ensemble de questions posées généralement à un échantillon représentatif d'une population.

B- Finalités :

- Recueillir des informations auprès des personnes concernées par le sujet à traiter
- Dresser le portrait d'une réalité à un moment précis dans le temps
- Evaluer les effets d'une action
- Réaliser un sondage sur un échantillon important

C- Domaine d'application :

Tout type de sujet

Analyse de l'existant	Critique de l'existant	Diagnostic	Elaboration et choix de solutions	Mise en œuvre	Suivi et ajustement

D- Caractéristiques :

- Le questionnaire implique généralement le choix d'un échantillon de la population concernée
- La standardisation du questionnaire est nécessaire : il est présenté à tous les interlocuteurs sous la même forme, avec les mêmes modalités
- Le questionnaire est un instrument pré-testé : il doit être mis à l'essai avant d'être utilisé pour vérifier sa pertinence
- Le questionnaire permet d'obtenir trois catégories d'informations :
  - Les faits, les attitudes, les attentes, les opinions…
  - Les caractéristiques associées aux répondants (sexe, âge, fonction…)
  - Les informations reliées à l'administration du questionnaire (date, lieu, groupe de répondants, etc…)
- Le questionnaire doit être accompagné en amont par une communication sur les objectifs et l'utilité du questionnaire, et en aval par une communication sur les résultats obtenus.

E- Mode d'emploi :

Démarche en 8 étapes :

- Définition de la problématique
- Définition de la population
- choix du type de questionnaire. Il existe deux types de questionnaires : Le questionnaire auto-administré où le sujet répond lui même et le questionnaire administré individuellement complété par l'enquêteur lui même lors d'un entretien individuel.
- Formulation des questions. Les questionnaires possèdent en général à la fois des questions ouvertes et fermées :
- conception du questionnaire
- Pré-test du questionnaire : Il consiste à vérifier si le questionnaire fonctionne ou si certaines modifications s'imposent en termes de contenu et de forme
- Codification des résultats. Réaliser une matrice de données à double entrée :

  *Chaque ligne correspond à un "répondant"

  *Chaque colonne correspond à une variable ou information demandée

Questions fermées : A l'aide d'un code numérique ou alphanumérique, on transforme l'information dans un format qui la rend exploitable

Questions ouvertes : Il faut à posteriori développer une liste de codes pour identifier les diverses réponses des interlocuteurs

Exemple :

Questions	1			2		3		4			5			…		n		
Réponses	1	2	3	O	N	1	2	1	2	3	1	2	3	…	…	…	…	…
Question1																		
Question2																		
Question3																		
…																		
Question n																		

- Analyse et interprétation des résultats. L'analyse a pour but de résumer les données recueillies de façon à répondre aux questions soulevées par la problématique abordée.

Démarche en 3 étapes

- L'analyse quantitative

Il s'agit grâce au calcul statistique d'analyser les informations recueillies, en se plaçant du point de vue précis des objectifs de l'enquête.

Deux grandes catégories d'approche statistique sont généralement utilisées :

- Les statistiques descriptives :
  Utilisation des mesures de tendance centrales (moyenne, médiane, mode), ainsi que des indices de dispersion autour de ces mesures (écart type, interquartile…)

- Les statistiques déductives :
  Utilisées pour rechercher des rapports significatifs entre des variables (corrélation). Elles permettent de faire ressortir des liaisons que l'on n'avait pas soupçonnées lors du lancement de l'enquête

- L'analyse qualitative

Elle privilégie les aspects socio-économiques et psychologiques des résultats. Elle vise à l'interprétation des réponses fournies.

- Le rapport d'enquête

Il fournit une série de tableaux accompagnés de commentaires sur les points les plus importants. ; il est structuré de la manière suivante :

- La présentation de l'enquête qui comprend ;
- La présentation des résultats qui concerne ;
- Les conclusions .

Chapitre III. Réalisation des sondages

Quelques définitions :

Sondage : Etude d'une partie d'une population considérés directement ou après redressement, comme représentative de la population totale. Les résultats obtenus sont rapportés à la totalité de cette population.
Le sondage s'oppose au recensement qui est l'étude exhaustive de toutes les unités d'un ensemble .
Base de sondage : liste ou fichier regroupant l'univers étudié et permettant le tirage au sort des unités de l'échantillon.
La statistique : toute mesure calculée à partir des données échantillonnales
Paramètre : toute mesure calculée à partir de l'ensemble des données de la population.
Estimation : le procédé par lequel on cherche à déterminer la valeur d'un paramètre d'une population.
Estimateur : la statistique utilisée pour effectuer l'estimation ; c'est une variable aléatoire.
Valeur estimée : la valeur que prend l'estimateur une fois l'échantillon tiré ; c'est une valeur de la variable aléatoire que constitue l'estimateur.

I- Estimateur d'une moyenne ou d'une proportion

Problématique : Quelle statistique de l'échantillon constituera le meilleur estimateur d'un paramètre de la population ?
Exp : on désire connaître la grandeur moyenne de toutes les femmes âgées de 18 ans ou plus vivant dans une certaine ville. Puisqu'il serait trop long d'étudier toute la population, on procède donc à partir d'un échantillon aléatoire. Mais, puisque les individus de l'échantillon ont été choisis de façon à ce qu'il représente le plus fidèlement possible la population, on est

en droit de penser que la moyenne de l'échantillon peut prendre une valeur proche de la moyenne de la population. Mais la moyenne d'un échantillon choisi aléatoirement dans la population rencontre-t-elle le critère d'un estimateur sans biais ?

A- Espérance mathématique d'une moyenne :

L'espérance mathématique de la moyenne d'un échantillon est un estimateur sans biais de la moyenne de la population à laquelle il appartient :

$$\mathbf{E\,(\bar{X}) = \mu}$$

Exp : soit la population $\{2,3,6,8\}$. Considérons la variable X représentant la moyenne d'un échantillon de taille 2 tiré avec remise. L'ensemble de tous les échantillons possibles auxquels on associe la moyenne est :

		$\bar{X}$
2	2	2.0
	3	2.5
	6	4.0
	8	5.0
3	2	2.5
	3	3.0
	6	4.5
	8	5.5
6	2	4.0
	3	4.5
	6	6.0
	8	7.0
8	2	5.0
	3	5.5
	6	7.0
	8	8.0

D'où la distribution de probabilité suivante :

$\bar{X}$	2.0	2.5	3.0	4.0	4.5	5.0	5.5	6.0	7.0	8.0
Fi $(\bar{X})$	1/16	2/16	1/16	2/16	2/16	2/16	2/16	1/16	2/16	1/16

On a donc : $E(\bar{X}) = (2.0)\ 1/16 + (2.5)\ 2/16 + \ldots. + (8.0)\ 1/16 = 4.75$

De plus la moyenne de la population :

$$\mu = \frac{2+3+6+8}{4} = 4.75$$

B- Espérance mathématique d'une proportion :

La proportion d'individus présentant un caractère particulier dans un échantillon est un estimateur sans biais de la proportion de ces individus dans la population à laquelle appartient l'échantillon.

Exp :

Reprenons l'exemple précédant, considérons cette fois-ci la variable aléatoire P représentant la proportion de nombre impair dans un échantillon de taille 2 tiré avec remise. L'ensemble des résultats possibles est :

		$\overline{P}$
2	2	0/2
	3	1 /2
	6	0/2
	8	0/2
3	2	1 /2
	3	2/2
	6	1 /2
	8	1 /2
6	2	0/2
	3	1 /2
	6	0/2
	8	0/2
8	2	0/2
	3	1 /2
	6	0/2
	8	0/2

D'où la distribution de probabilité suivante :

$\overline{P}$	0	1 /2	1
Fi $(\overline{P})$	9/16	6/16	1/16

On a donc : $E(\overline{P}) = (0)\ 9/16 + (1/\ 2)\ 6/16 + (1)\ 1/16 = 1/4$
De plus la proportion de nombres impairs dans la population est :
$\pi = 1/\ 4$

Estimation ponctuelle d'un paramètre :

L'estimation ponctuelle d'un paramètre consiste en l'évaluation de la valeur du paramètre de la population à l'aide d'une valeur unique prise dans un échantillon. La statistique utilisée comme estimateur doit rencontrer un certain nombre de critères, on a vu celui de l'estimateur sans biais. D'autres caractéristiques existent mais ne font pas notre objectif.
Il importe davantage de connaître les résultats qui suivent :

Signification des termes	Paramètre (population)	Statistique utilisée (échantillon)
Moyenne	μ	$\overline{X}$
Proportion	π	$\overline{P}$

Application :
Soit la population {3,7,12,16,25}. Considérer tous les échantillons de taille 2 pris avec remise dans celle-ci.
1. pour chacun des échantillons, calculez la valeur de la variable aléatoire X
2. calculez E(x)
3. calculez μ, la moyenne de la population
4. comparez les résultats obtenus en b et c

Eléments de réponse :

1.
0.3 5.0 7.5 9.5 14.0 5.0 7.0 9.5 11.5 16.0 7.5 9.5 12.0 14.0 18.5 9.5 11.5 14.0 16.0 20.5 14.0 16.0 18.5 20.5 25.0
2. 12.6
3. 12.6
4. E(x) = μ

II- Variance des estimateurs

On peut s'interroger sur les chances que la valeur estimée, à partir de l'échantillon, égale la valeur du paramètre de l population. Il convient donc de pouvoir faire l'estimation d'un paramètre tout en étant capable d'évaluer les chances qu'à cette estimation de se réaliser. Pour ce faire nous effectuons ce qu'on appelle une estimation pat intervalle de confiance d'un paramètre de la population. Le problème consiste donc à trouver les bornes de cet intervalle.

La moyenne de la variable aléatoire X est : $E(x) = \mu_X = \mu$ et l'écart -type de X est $\sigma_X = \sigma/\sqrt{n}$ (sachant que var (x) = $E(x^2) - [E(x)]^2$)

Si l'échantillon est tiré sans remise dans une population infinie ou très grande avec $n < 0.05N$ ou encore avec remise dan,s la population, quelle que soit la taille de celle-ci, et

$$\sigma_X = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Si l'échantillon est tiré sans remise dans une population finie.

Exp : reprenons l'exemple précédant :

X	2.0	2.5	3.0	4.0	4.5	5.0	5.5	6.0	7.0	8.0
Fi (X)	1/16	2/16	1/16	2/16	2/16	2/16	2/16	1/16	2/16	1/16

On sait que var (x) = $E(x^2) - [E(x)]^2$
Or, on a :
$E(x^2) = (2.0)^2\ 1/16 + (2.5)^2\ 2/16 + \ldots + (8.0)^2\ 1/16 = 25.40$

D'où : var (x) = 25.40 – $(4.75)^2$

De plus $\sigma^2 = \dfrac{(2-4.75)^2 + (3-4.75)^2 + (6-4.75)^2 + (8-4.75)^2}{4} = 5.69$

et $\sigma^2/n = 5.69/2 = 2.84$ où n représente la taille de l'échantillon.

Application :

Un échantillon de taille n est tiré, sans remise, d'une population de taille 350 dont la moyenne et la variance sont respectivement 115 et 169. pour chacune des valeurs suivantes de n, évaluer la variance et l'écart_ type de la variable aléatoire $\bar{X}$:

1. **5**
2. **15**
3. **30**
4. **50**

Eléments de réponse :

1. **33.5 et 5.8**
2. **11.3 et 3.4**
3. **5.2 et 2.3**
4. **2.9 et 1.7**

III- Estimation par intervalle de confiance de μ :

On appelle INTERVALLE DE CONFIANCE un intervalle de la forme $[L_1,L_2]$, ayant une certaine probabilité de contenir la valeur d'un paramètre.

$L_1=\bar{X} - z_{\alpha/2}\,\sigma_{\bar{x}}$ et $L_2=\bar{X} - z_{\alpha/2}\,\sigma_{\bar{x}}$

Où : $z_{\alpha/2}$ est la valeur de la variable z telle que $P(z \leq z_{\alpha/2}) = 1- \alpha/2$, α le risque d'erreur et $\sigma_{\bar{x}}$ l'écart- type de la distribution d'échantillonnage de $\bar{X}$ appelée aussi ERREUR TYPE.

Il convient d'utiliser :

$z_{\alpha/2} = 2.58$ si $\alpha = 1\%$
$z_{\alpha/2} = 1.96$ si $\alpha = 5\%$
$z_{\alpha/2} = 1.65$ si $\alpha = 10\%$

On appelle NIVEAU DE CONFIANCE, noté $1 - \sigma$, la probabilité qu'a l'intervalle de confiance de contenir la valeur du paramètre.

On appelle RISQUE D'ERREUR , noté σ , la probabilité qu'a l'intervalle de confiance de ne pas contenir la valeur du paramètre.

Exp :
La moyenne et l'écart -type du résultat cumulatif d'un échantillon de 36 étudiants d'une université sont 2.6 et 0.3 respectivement. Trouvons un intervalle de confiance à 99% pour la moyenne des résultats cumulatifs de tous les étudiants de cette université. On a donc :

$\bar{X} = 2.6$, $z_{\alpha/2}= z_{1/2\%}=2.58$
Et $\sigma_{\bar{x}} = 0.3/\sqrt{36} = 0.05$
D'où : $L_1 = 2.6 – (2.58)0.05 = 2.47$
Et $L_2 = 2.6 + (2.58)0.05 = 2.73$

Donc : $\mu \in [2.47 ; 2..73]$
Avec un niveau de confiance de 99% , c'est à dire que l 'intervalle [2.47 ; 2..73]
Possède 99% des chances de contenir la moyenne μ du résultat cumulatif des étudiants de cette université.

<u>Application :</u>
Dans une région, on s'intéresse au temps moyen μ, inconnu , que prennent les individus d'un groupe pour se rendre à leur travail. A partir d'un échantillon aléatoire de taille 100, on a obtenu un temps moyen de 12 minutes. Construisez un intervalle de confiance à 90% pour μ, si l'on sait que $\sigma^2 = 9$.
<u>Eléments de réponse :</u>
[11.505 ; 12.495] minutes

Contrôle continu

Durée : 2h

Un professeur d'EPS en charge de deux groupes de filles n'ayant jamais pratiqué le saut à la perche décide de les initier à ce sport en utilisant deux méthodes d'initiation différentes. Les performances réalisées à la fin du cycle d'apprentissage sont les suivantes :

Groupe 1(méthode A) :

2.20	2.35	2.40	1.15	2.35	2.00	2.55	2.05	1.85	2.85
2.65	2.35	1.90	2.70	2.05	1.95	2.15	2.05	2.80	2.45

Groupe 2(méthode B) :

1.80	2.00	1.45	2.05	2.00	1.65
2.05	1.65	1.50	1.60	2.15	2.10

1- construire les histogrammes des deux séries de valeurs en utilisant des classes de largeur 0.2m du type : [1.00-1.20[

2- laquelle de ces deux méthodes semble donner les meilleurs résultats ? répondre à la question tout d'abord d'après les histogrammes puis selon que le critère est :

- moyenne la plus élevée
- médiane la plus élevée
- classe modale la plus élevée
- maximum le plus levée
- minimum le plus élevé
- écart – type le plus faible
- étendue la plus faible
- autres critères ?

3- construire un nouvel histogramme, cette fois uniquement pour le groupe 1, en utilisant des classes de largeur 0.5. le comparer à celui de la question 1. Lequel apporte l'information la plus pertinente ?

Module : Statistiques
GUIDE DES TRAVAUX PRATIQUES

TP 1

Objectifs visés :

- représenter graphiquement une distribution statistique
- étudier la tendance centrale de cette distribution
- étudier la dispersion de cette distribution
- apprécier la forme de cette distribution

Durée du TP :
2h

Description du TP :
Cet exercice permet au stagiaire de maîtriser la représentation graphique d'une distribution à caractère quantitatif continu, de s'entraîner sur le calcul des paramètres de la tendance centrale et de dispersion et également de faire un commentaire en se basant sur la forme de la représentation graphique de la distribution.

Déroulement du TP :
Dans une commune rurale, où aucune exploitation agricole n'atteint 123 Ha. La distribution des 100 exploitants en fonction de la superficie se présente comme suit :

Superficie en Ha : xi	Le pourcentage des propriétaires fonciers :fi
Moins de 5	15
5 – 10	20
10 – 15	15
15 – 20	10
20 – 30	10
30 – 50	12
50 et plus	18
Total	100

Questions :

1- quelle est la population cible ?
quel est le caractère étudié ?
quel est le nombre de modalités ?
2- représentez graphiquement la distribution étudiée (simple et cumulative)
3- déterminez les différentes caractéristiques de tendance centrale
4- qu'en est-il de la dispersion ?
5- est-ce que la répartition des terres au sein de cette commune est équitable ?

Eléments de réponse :
1- population cible : les 100 exploitations
caractère étudié : la superficie ; sa nature : quantitatif continu
nombre de modalités : 7
3-
$\bar{X}$=28.55 Ha
Me = 15 Ha
Mo= 7.5 Ha
4- Etendue = 125 Ha
intervalle interquartile : $[Q_1 ;Q_3]$ = [7.5 ;38.33]
coefficient de variation = 1.04

5- indice de GINI : I_G=0.613
l'indice tend vers 1 plus que vers 0, on dira que la distribution des terres dans cette commune est assez concentrée donc cette distribution est non équitable.

TP 2

Objectifs visés :
- réaliser des représentations graphiques pour des variables quantitatives continues.

Durée du TP :
1h30

Description du TP :
Ce TP permettra au stagiaire de maîtriser la lecture d'un tableau représentant la distribution d'une variable quantitative continue. Il lui permettra également de représenter graphiquement ce genre de variable.

Déroulement du TP :
On considère la distribution définie par le tableau ci-dessus :

Loyer mensuel en DH	Nombre d'appartements
150-179	3
180-209	8
210-239	10
240-269	13
270-299	33
300-329	40
330-359	35
360-389	30
Total	172

Questions :
a- quelles sont les bornes inférieures et supérieures de la 1ere classe ?
b- quelles sont les vraies limites de la 1ere classe ?
c- l'intervalle de classe utilisée est identique pour chaque classe ? quelle est sa taille ?
d- quel est le centre de la 1ere classe ?
e- quels sont les vraies limites de la classe correspondant à l'effectif le plus élevé ?
f- quelles sont les bornes de la classe à l'intérieur de laquelle s'est trouvé recensé un loyer mensuel de 239.50 DH ?
g- construisez un histogramme exprimant les données du tableau.
h- construisez une courbe d'effectifs pour les données du tableau.

Eléments de réponse :

a- 150dh et 179dh
b- 149.50dh et 179.50dh
c- 179.50 – 149.5 = 30
d- 149.5 + 30/2 = 164.50 dh
e- 299.5 dh et 329.50 dh
f- 240 dh et 269 dh

TP 3

Objectifs visés :
- calculer les paramètres de tendance centrale
- interpréter les paramètres de tendance centrale

Durée du TP :
1h30

Description du TP :
Cet exercice permet au stagiaire de maîtriser l'utilisation des formules de calcul des paramètres de tendance centrale.

Déroulement du TP :

Une agence d'urbanisme a effectué une étude sur la structure des familles susceptibles de venir habiter une ville nouvelle dont elle est chargée d'établir le projet. Trois types de familles ont été définis selon la présence et l'activité du conjoint. D'après cette étude, les distributions de fréquences de ces familles selon le nombre d'enfants sont les suivantes :

Nombre d'enfants	Chef de famille…		
	…sans conjoint	… avec femme active	…avec femme inactive
0	33.3	16.2	8.4
1	39.3	26.6	16.4
2	16.6	26.6	25.2
3	6.4	15.6	20.6
4	2.5	9.3	15.3
5	1.1	4.5	12.2
6	0.8	1.2	1.9
7	0.0	0.0	0.0
Total	100.0	100.0	100.0

Les trois types de familles considérès se repartissent en pourcentage comme ci-après :

Total	Chef de famille…		
	…sans conjoint	… avec femme active	…avec femme inactive
100	5.8	52.9	41.2

Questions :

1- déterminez pour chaque type de famille et pour l'ensemble, le mode de la distribution selon le nombre d'enfants.
2- déterminez pour chaque type de famille et pour l'ensemble, la médiane de la distribution selon le nombre d'enfants.
3- calculez pour chaque type de famille et pour l'ensemble, le nombre moyen d'enfants .

Eléments de réponse :

1-

	Ensemble	Chef de famille…		
		…sans conjoint	… avec femme active	…avec femme inactive
Valeur du mode	2 enfants	1 enfant	Intervalle modale : 1 à enfants	2 enfants

2- On retient pour la médiane la valeur M pour laquelle la fréquence cumulée est égale à ½.

3-

	Ensemble	Chef de famille…		
		…sans conjoint	… avec femme active	…avec femme inactive
Nombre moyen d’enfants	2.171	1.120	1.935	2.622

TP 4

Objectifs visés :
- traiter le lien entre variables à caractère quantitatif
- choisir la représentation graphique adéquate pour chaque distribution statistique
- interpréter les représentations graphiques

Durée du TP :
2h30

Description du TP :
Cet exercice permet au stagiaire d'étudier le lien existant entre deux variables à caractère quantitatifs en se basant sur la lecture d'une représentation graphique.

Déroulement du TP :

Au cours de la décennie 1990-2000, les effectifs employés au fond d'une houillère et la production nette de charbon ont évolué de façon suivante :

Année	Effectifs du fond (milliers de personnes	Production nette de charbon (millions de tonnes)
1990	71.3	40.1
1991	65.3	35.8
1992	57.6	32.7
1993	50.4	28.4
1994	47.1	25.7
1995	45.8	25.6
1996	42.4	25.1
1997	38.6	24.4
1998	35.9	22.4
1999	32.7	21.1
2000	30.8	20.7

1- représentez l'évolution de ces deux séries sur deux graphiques à coordonnées arithmétiques présentés l'un au dessous de l'autre façon à mettre en évidence l'existence de covariations éventuelles dans le temps.
2- quels sont les inconvénients de cette présentation ?
3- quel type de graphique permettrait d'y remédier ?
4- tracer le graphique de corrélation correspondant au tableau précédant.
5- comment interprétez-vous ce graphique ?

TP 5

Objectifs visés :
- construire des représentations graphiques adaptées aux variables qualitatives et quantitatives discrètes
- calculer les paramètres de la tendance centrale
- calculer les paramètres de dispersion

Durée du TP :
2h

Description du TP :
Ce TP permet au stagiaire de s'entraîner sur la représentation graphique des variables qualitatives et quantitatives discrètes. Il lui permet également de maîtriser le calcul des paramètres de la tendance centrale et ceux de la dispersion.

Un sondage sur la capacité pulmonaire des individus nous a donné les résultats suivants :

Age	Sexe	Capacité pulmonaire (litre)
54	F	2.94
19	M	4.03
18	F	3.75
26	M	6.04
19	F	4.92
22	M	6.57
18	M	5.28
20	M	5.19
20	F	4.08
18	M	4.68
17	M	5.38
29	M	4.71
17	M	5.20
43	M	4.50
30	M	4.93
18	F	3.92
25	M	6.54
38	M	5.35
19	F	4.21
26	M	5.40
20	M	6.66
18	M	5.14
16	F	3.49
19	M	5.82
20	M	5.25
21	M	4.89
19	M	6.07
19	F	3.82
19	M	6.71
30	M	5.93
24	M	6.22
17	F	3.86

Questions:

1- Construisez une distribution d'effectifs pour chacune des variables
2- donner une représentation graphique pour chacun des cas
3- donnez la mesure de tendance centrale la plus appropriée, pour chacune des variables
4- calculez l'écart type de la distribution de la capacité pulmonaire

Eléments de réponse :
3-
Age : $\bar{x}$= 23.4 ans, sexe: Mo=M, capacité pulmonaire : $\bar{x}$= 4.98 litres
4- 0.93 litres

TP 6

Objectifs visés :
- tracer un nuage statistique
- trouver l'équation de la droite d'ajustement linéaire
- faire des prévisions en se basant sur la droite d'ajustement linéaire
- étudier la corrélation entre deux variables

Durée du TP :
2h30

Description du TP :
Cet exercice permet au stagiaire de faire des prévisions en trouvant la droite d'ajustement linéaire par la méthode des moindre carrés. Il permet également d'étudier la corrélation entre deux variables.

Déroulement du TP :

Des étudiants de 1ere année TCE ont eu les résultats en statistiques et en mathématiques financières (/100):

x (notes de statistiques)	66	64	69	93	80	71	87	73	79	56	47
Y(notes de math.fin.)	72	70	60	94	82	68	86	82	90	55	64

Questions :

1- tracez le nuage statistique
2- ajustez la droite des moindres carrés
3- quelle note de mathématiques financières pouvez-vous prédire à un étudiant de ce niveau qui a eu 75 en statistiques ?
4- calculez le coefficient de corrélation ?

Eléments de réponse :
2- $y= 16.82+0.81x$
3- 77.8
4- 0.845

TP 7

Objectifs visés :
- connaître la terminologie principale des statistiques
- établir des tableaux statistiques
- construire des représentations graphiques
- calculer et interpréter les différents paramètres des distributions

Durée du TP :
18h

Description du TP :
Ce TP est présenté sous forme de QCM. Il couvre presque la totalité des points traités dans ce module. Il pourrait être utilisé comme test de connaissances à la fin de chaque section.

Déroulement du TP :

TERMINOLOGIE ET TABLEAUX STATISTIQUES

1-

Les caractères suivants sont	qualitatifs	quantitatifs
- Le tour de ceinture d'une personne	☐	☐
- Le code postal de l'habitation d'un foyer français	☐	☐
- La superficie d'une exploitation agricole	☐	☐
- Le groupe sanguin d'un individu	☐	☐

2-
Les classes suivantes sont-elles bien définies?

[−∞,0[	[0 ; 100[	[100 ; 300[	[300 ; 600[	[600 ;+∞[	☐ oui	☐ non
moins de 4	entre 4 et 8	entre 8 et 12	entre 12 et 14	plus de 14	☐ oui	☐ non
$X \leq 1$	$1 < X \leq 2$	$2 < X \leq 5$	$5 < X \leq 10$	$X > 10$	☐ oui	☐ non
[0,5 ; 2,5[	[3 ; 4,5[	[5 ; 5,5[	[6 ; 6,5[	[7 ; 7,5[	☐ oui	☐ non

3- La fréquence d'une classe s'obtient en divisant l'effectif de la classe par

L'effectif total	☐
Le nombre de classes	☐
L'amplitude de la classe	☐

4- Le caractère quantitatif discret x admet le tableau de distribution suivant

valeurs	1	2	3	4	5	total
fréquences	10,5%	22,3%	30,4%	23,6%	13,2%	100%

5- Quelle est la fréquence cumulée croissante pour x = 3

☐	☐	☐	☐
67,2%	63,2%	32,8%	30,4%

6- Pour une distribution continue, l'effectif total s'obtient en multipliant l'effectif de chaque classe par le centre de la classe et en ajoutant les nombres ainsi obtenus

☐ vrai	☐ faux

7- Le tableau ci-dessous (notes obtenues par 40 étudiants à un examen de statistique) est un tableau

12	9	7	1	13	18	12	3
4	6	9	14	5	0	6	15
7	10	3	5	9	5	6	9
0	7	13	8	4	4	11	3
10	12	6	5	8	0	1	7

☐ De données ponctuelles	☐ De distribution

8- Les caractères quantitatifs suivants peuvent-ils être considérés comme des variables statistiques continues

le nombre d'accidents du travail survenus dans une PME en 1 an	☐ oui	☐ non
la teneur en aluminium d'un alliage	☐ oui	☐ non

9- Les étudiants de formation continue sont répartis selon leur âge dans le tableau suivant

âge	[20 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[	[40 ; 45[	+ de 45	total
effectifs	38	59	47	24	12	2	182

Quelle limite doit-on donner à la dernière classe si l'on veut que toutes les classes aient la même amplitude

☐ 50	☐ 55	☐ 34

Quel est le centre de la classe [30 ; 35[

☐ 33	☐ 35	☐ 37,5	☐ autre réponse

Quelle est la proportion d'étudiants âgés de moins de 35 ans

☐ 53,3%	☑ 79,12%	☐ 92,31%	☐ 25,82%

10- La fréquence cumulée croissante est définie par

- proportion d'individus dont la valeur du caractère est inférieure à x	☐
- proportion d'individus dont la valeur du caractère est supérieure à x	☐
- ensemble des modalités que peut prendre le caractère	☐
- autre réponse	☐

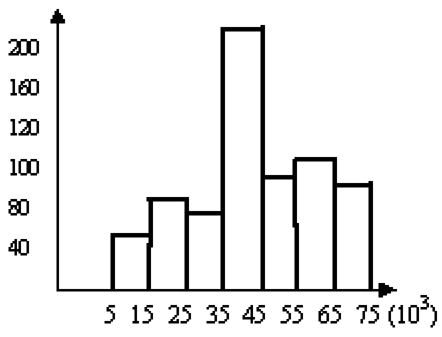
11- On a pu regrouper les individus d'une population par classes dont les centres sont les suivants : 52, 60, 68, 76, 84, 92. Quelle est l'amplitude des classes

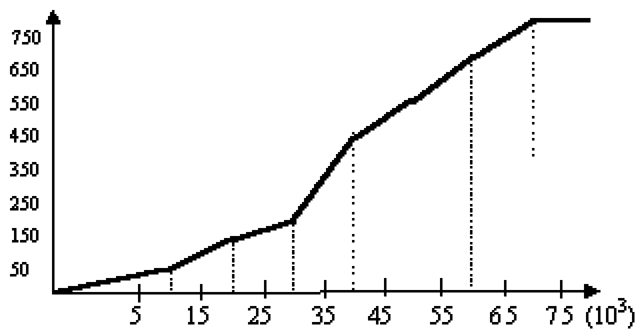
☐ 2	☐ 4	☐ 6	☐ 8	☐ 16

REPRÉSENTATIONS GRAPHIQUES

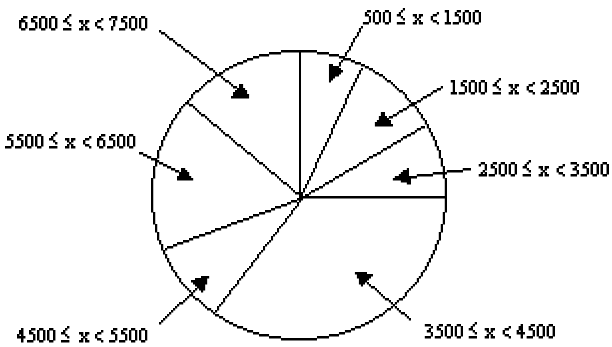
1- A partir du tableau ci-dessous, 3 graphiques ont été établis. Indiquez celui (unique) de ces graphiques qui ne constitue pas une représentation correcte du phénomène

Classe	effectifs	effectifs cumulés
$500 \leq x < 1500$	41	0
$1500 \leq x < 2500$	75	41
$2500 \leq x < 3500$	62	116
$3500 \leq x < 4500$	226	178
$4500 \leq x < 5500$	89	404
$5500 \leq x < 6500$	109	493
$6500 \leq x < 7500$	83	602
total	685	685





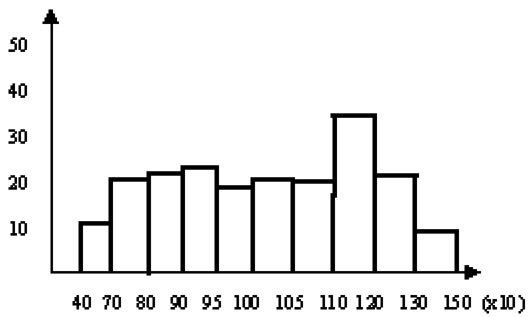
☐ 1	☐ 2

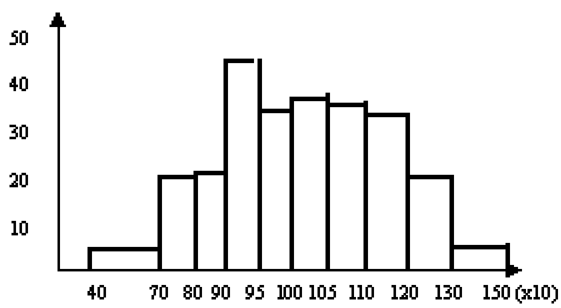


☐ 3

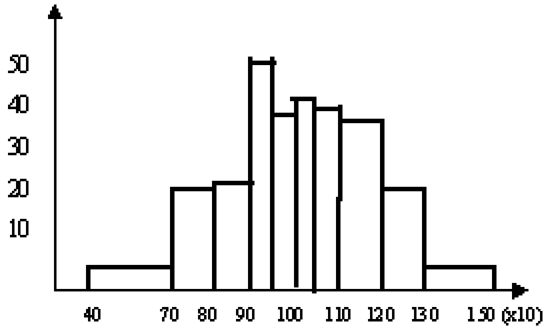
2- Lequel des graphiques ci-dessous correspond à l'histogramme des données suivantes

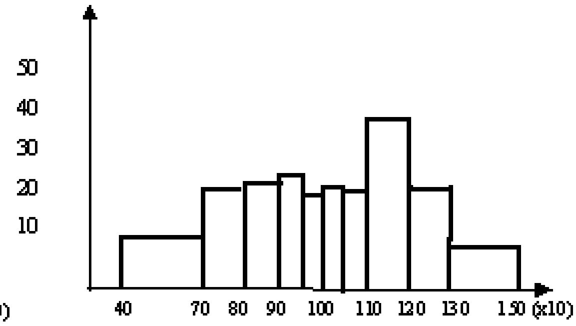
Classe	effectifs	effectifs cumulés
400 ≤ x < 700	11	0
700 ≤ x < 800	21	11
800 ≤ x < 900	23	32
900 ≤ x < 950	24	55
950 ≤ x < 1000	18	79
1000 ≤ x < 1050	20	97
1050 ≤ x < 1100	19	117
1100 ≤ x < 1200	35	136
1200 ≤ x < 1300	21	171
1300 ≤ x < 1500	8	192
total	200	200





☐ 1	☐ 2



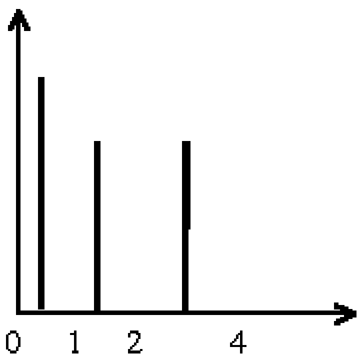


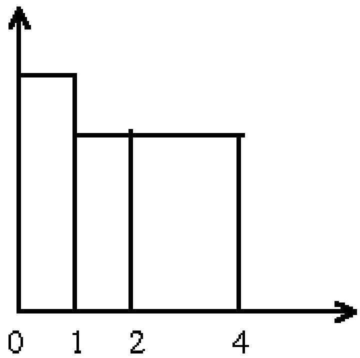
☐ 3	☐ 4

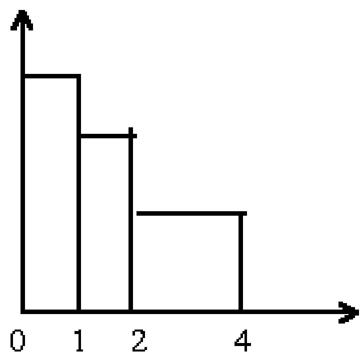
3- Le caractère quantitatif X admet la distribution suivante:

classes	[0 ; 1[	[1 ; 2[	[2 ; 4[
effectifs	40	30	30

Quelle est la représentation graphique des fréquences qui convient?





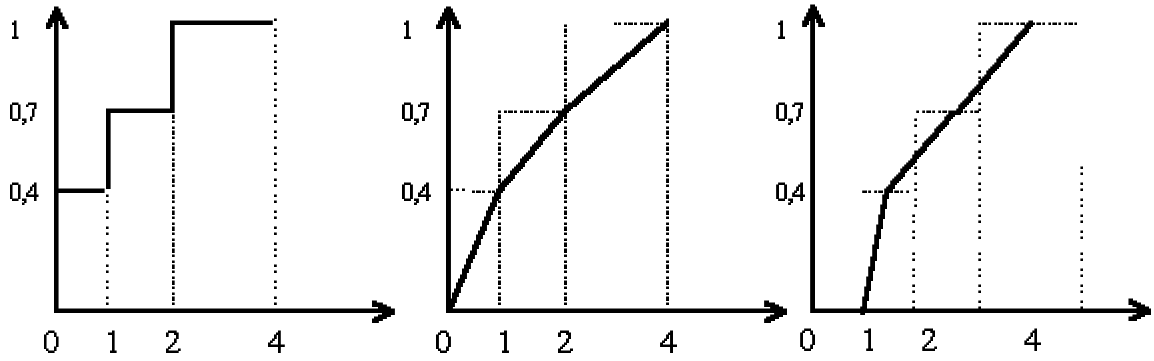


☐ 1	☐ 2	☐ 3
	☐ une autre représentation	

4- Le caractère quantitatif X admet la distribution suivante:

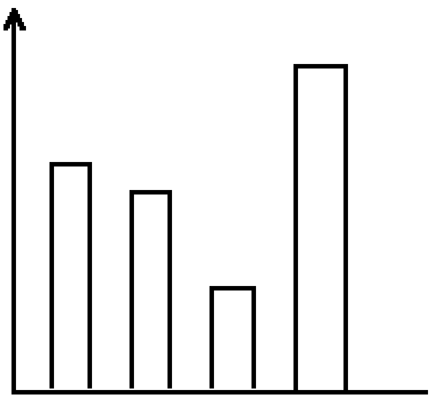
classes	[0 ; 1[	[1 ; 2[	[2 ; 4[
effectifs	40	30	30

Quelle représentation graphique des fréquences cumulées croissantes convient?



☐ 1	☐ 2	☐ 3
	☐ une autre représentation	

5- La représentation graphique ci-dessous est un diagramme



☐	en bâtons
☐	à secteurs
☐	à bandes

6- Un histogramme est une représentation graphique de la distribution des fréquences d'une variable statistique continue

VRAI	☐
FAUX	☐

7- Dans un diagramme à secteurs, la modalité n° 2 du tableau ci-dessous serait représentée par un secteur d'angle

modalités	effectifs
1	30
2	15
3	25
4	30

15 degrés	☐
54 degrés	☐
60 degrés	☐

8- Le tableau suivant donne la répartition des ménages d'une population selon le nombre de véhicules possédés

nombre d'automobiles	0	1	2	3	4 et plus
nombre de ménages	528	2463	906	156	12

9- La représentation graphique qui convient le mieux est

☐ un diagramme en bâtons	☐ un histogramme	☐ une autre représentation

CARACTÉRISTIQUES DE TENDANCE CENTRALE ET DE POSITION

1- Quelle est la moyenne des valeurs ci-dessous

xi	ni
20	58
30	188
40	54

☐ 82,89
☐ 29,87
☐ 30
☐ 30,54

2- La médiane d'une distribution est toujours égale au second quartile

☐ OUI	☐ NON

3- Dans une série statistique, il est possible de déterminer dix déciles

☐ OUI	☐ NON

4- On observe pendant 79 jours ouvrables, le nombre de lettres recommandées émises au cours de la journée, par le service des approvisionnements. L'évolution de ces envois au cours de cette période est fournie dans le tableau suivant. Déterminer le premier et le troisième quartile de cette série d'expéditions quotidiennes de lettres recommandées.

rang	nbre lettres	rang	nbre lettres	rang	nbre lettres	rang	nbre lettres	rang	nbre lettres
1	1	17	6	33	7	49	8	65	11
2	3	18	6	34	7	50	8	66	11
3	3	19	6	35	7	51	9	67	11
4	4	20	6	36	7	52	9	68	11
5	4	21	6	37	7	53	9	69	11
6	5	22	6	38	7	54	9	70	11
7	5	23	6	39	8	55	9	71	11
8	5	24	6	40	8	56	9	72	12
9	5	25	7	41	8	57	9	73	12
10	5	26	7	42	8	58	9	74	12
11	5	27	7	43	8	59	10	75	12
12	6	28	7	44	8	60	10	76	13
13	6	29	7	45	8	61	10	77	13
14	6	30	7	46	8	62	10	78	14
15	6	31	7	47	8	63	10	79	15
16	6	32	7	48	8	64	10		

- ☐ Q1=7 Q3=12
- ☐ Q1=6 Q3=11
- ☐ Q1=7 Q3=10
- ☐ Q1=3,75 Q3=11,25
- ☐ autre réponse

5- Cocher la nature des indicateurs numériques suivants

	Paramètre de position	Paramètre de dispersion	ni l'un ni l'autre
effectif total	☐	☐	☐
3° décile	☐	☐	☐
moyenne géométrique	☐	☐	☐

6- Soit le tableau suivant

modalités	effectifs
employés de service	2
manoeuvres	3
ouvriers	12
ouvriers spécialisés	22
agents de maîtrise	15
employés	28
cadres	13
cadres supérieurs	?

Sachant que la moyenne arithmétique est 12,5 le nombre de cadres supérieurs est
☐ 7
☐ 10
☐ 5
☐ autre réponse

7- Il existe 100 centiles qui partagent une série statistique

☐ OUI	☐ NON

8- On donne la série statistique suivante : 14, 16, 12, 9, 11, 18, 7, 8, 9, 16, 7, 9, 18. La médiane est égale à

☐ 9	☐ 11	☐ 14	☑ 16	☐ 18	☐ [9;18[	☐ [11;18[	☐ autre réponse

9- La moyenne géométrique d'une série statistique est

La racine carrée du produit des valeurs observées	☐
la racine cubique du produit des valeurs observées	☐
la racine n-ième du produit des valeurs observées	☐
le produit des racines n-ième des valeurs observées	☐
le quotient des racines n-ième des valeurs observées	☐
autre réponse	☐

10- Quand les classes d'une série statistique sont d'amplitudes inégales, il faut obligatoirement corriger les effectifs ou les fréquences pour calculer la médiane

☐ OUI	☐ NON

11- La moyenne harmonique d'une série statistique est égale à l'inverse de la moyenne arithmétique des inverses des valeurs

☐ OUI	☐ NON

12- La médiane partage l'histogramme en deux surfaces égales

☐ OUI	☐ NON

13- Soit la série suivante

xi	ni
1	20
2	30
3	15
4	10
5	5
6	2

la moyenne quadratique est égale à	☐ 1,92	☐ 2,78	☐ 357
	☐ 4,86	☐ 5,04	☐ 15
la moyenne géométrique est égale à	☐ 1,87	☐ 2,15	☐ 3,57
	☐ 6,25	☐ autre réponse	
la moyenne harmonique est égale à	☐ 6,25	☐ 215	☐ 1,92
	☐ 1,87	☐ autre réponse	

14- La répartition des célibataires selon leur âge est fournie par le tableau suivant

âge	[15 ; 30[	[30 ; 40[	[40 ; 50[	[50 ; 60[	[60 ; 70[	[70 ; 80[	[80 ; 90[
effectifs	4500	450	400	230	200	?	20

Sachant que l'âge moyen est égal à 28,8 ans, la valeur manquante est

☐ 65	☐ 97	☐ 102
☐ 150	☐ 165	☐ autre réponse

l'âge médian est

☐ 20,4	☐ 22,6	☐ 24,8
☐ 26,7	☐ autre réponse	

CARACTÉRISTIQUES DE DISPERSION

1- Complétez le tableau suivant pour calculer la

	x_i	n_i	$n_i\ x_i$	$n_i x_i^2$
	40	...	...	...
	50	...	5000	...
	...	87	5220	313200
TOTAL	...	...	12740	664000

variance

la variance vaut

☐ 6,293	☐ 7,69	☐ 4341,73	☐ 59,08

2- Calculez le coefficient de variation des données suivantes:

xi	ni
70	91
80	189
90	70

☐ 0,085	☐ 45,64
☐ 0,546	☐ 6,76

3- La synthèse d'un ensemble d'observations relatives à une variable quantitative peut s'effectuer par des paramètres de tendance centrale et de dispersion.
L'une des quatre réponses suivantes n'a rien à voir avec ce type de synthèse:

☐ moyenne et écart-type	☐ fréquence moyenne par unité d'amplitude et mode
☐ médiane et écart-type	☐ variance et mode

4- On observe sur un tronçon d'autoroute, pendant 51 jours, le nombre x de dépannages effectués au cours de la journée. Calculer l'intervalle inter-quartile des observations

rang	nbre dépannages	rang	nbre dépannages	rang	nbre dépannages	rang	nbre dépannages	rang	nbre dépannages
1	1	11	3	21	4	31	4	41	6
2	1	12	3	22	4	32	4	42	6
3	1	13	3	23	4	33	5	43	6
4	1	14	3	24	4	34	5	44	6
5	1	15	3	25	4	35	5	45	6
6	2	16	3	26	4	36	5	46	6
7	2	17	3	27	4	37	5	47	7
8	2	18	3	28	4	38	5	48	8
9	2	19	3	29	4	39	5	49	9
10	3	20	4	30	4	40	5	50	10
								51	11

L'intervalle inter-quartile vaut

☐ 3	☐ 4	☐ 5	☐ 6	☐ autre réponse

5- La variance est toujours positive ou nulle

☐ OUI	☐ NON

6- Une entreprise E possède 3 établissements A, B, C. Les effectifs et les salaires moyens pour les ouvriers , les employés , et les cadres , sont donnés dans le tableau suivant

	A		B		C		E	
	effectifs	salaire moyen	effectifs	salaire moyen	effectifs	salaire moyen	effectifs	salaire moyen
Ouvriers	60	10	180	8	5	10	245	8,5306
Employés	30	20	10	16	30	25	70	21,571
Cadres	10	100	10	90	15	100	35	97,143
Total	100	22	200	12,5	50	46	350	20

La variance intra-établissements est égale à	☐ 129,86	☐ 478,28	☐ 562,51

LA CONCENTRATION

1- Si, pour un caractère quantitatif continu et positif, la médiane est très peu différente de la médiale, alors l'indice de concentration de Gini est peu différent de

☐ 0	☐ 0,5	☐ 1

2- Dans un diagramme de concentration on porte généralement en ordonnées les valeurs des fréquences cumulées des valeurs globales. Comment s'écrivent ces valeurs

☐ $\sum \frac{n_i x_i}{n}$	☐ $\frac{\sum n_i x_i}{nx}$	☐ $\sum \frac{n_i x_i}{\sum n_i x_i}$
☐ $\sum \frac{\sum n_i x_i}{n_i x_i}$	☐ autre réponse	

INDICES

1- Le chiffre d'affaires d'une entreprise a augmenté de 2% par an pendant 2 ans, puis a diminué de 9% par an pendant 4 ans, et a augmenté de 8% par an pendant 3 ans. Quelle est l'augmentation moyenne sur la période

☐ 1%	☐ 9%	☐ 10%	☐ autre réponse

2- Étant donné une population de 50 millions qui a crû au taux de 20% par an, quelle était cette population il y a 12 ans

☐ 38 486 689	☐ 39 424 659	☐ 1 555 318	☐ 5 607 832	☐ autre réponse

3- Une hausse de 80% suivie d'une baisse de 50% revient à

☐ une baisse de 10%	☐ une baisse de 20%	☐ une baisse de 30%
☐ une hausse de 10%	☐ une hausse de 30%	☐ autre réponse

4- Une hausse de 60% suivie d'une baisse de 40% revient à

☐ une hausse de 20%	☐ une baisse de 10%	☐ une hausse de 10%
☐ une baisse de 20%	☐ une baisse de 4%	☐ autre réponse

5- Une grandeur augmente de 10% par an. Au bout de combien d'années aura-t-elle doublé

☐ 11 ans	☐ 11,1 ans	☐ 10 ans	☐ 7,27 ans	☐ 6,23 ans
☐ 1 an	☐ 12,45 ans	☐ 8,27 ans	☐ autre réponse	

6- Le calcul de l'indice de Laspeyres nécessite de pondérer les indices élémentaires par des coefficients budgétaires relatifs

☐ à la période de base	☐ à la période courante

7- Calculez l'indice de Laspeyres des prix de 1998 par rapport à 1990 à partir des données du tableau suivant

Modèle	Quantités		Prix		Ventes	
	1990	1998	1990	1998	1990	1998
Produit A	50	55	18	22	900	1210
Produit B	69	62	23	25	1587	1550
Produit C	96	115	28	25	2688	2875
Total					5175	5635

☐ 108,91	☐ 100,97
☐ 107,85	☐ 99,98

8- Calculez l'indice de Paasche des quantités de 1998 par rapport à 1990 à partir des données du tableau suivant

Modèle	Quantités		Prix		Ventes	
	1990	1998	1990	1998	1990	1998
Produit A	90	99	13	16	1170	1584
Produit B	56	50	18	20	1008	1000
Produit C	78	94	23	21	1794	1974
Total					3972	4558

☐ 109,53	☐ 108,58
☐ 104,81	☐ 105,69

RÉGRESSION LINÉAIRE

1- Pour justifier un ajustement affine (y = ax + b) , on a calculé le coefficient de corrélation linéaire r. Dans les cas suivants, le résultat est

r = 1,22	☐ médiocre	☐ bon	☐ idiot
r = -0,89	☐ médiocre	☐ bon	☐ idiot

2- Quand on ajuste linéairement x et y par la méthode des moindres carrés, on obtient deux droites de régression. L'équation de la droite D de y par rapport à x est

☐ $y = \bar{y} + \frac{\text{cov}(x,y)}{V(x)}(x - \bar{x})$	☐ $x = \bar{x} + \frac{\text{cov}(x,y)}{V(y)}(y - \bar{y})$
☐ $y = \bar{x} + \frac{\text{cov}(x,y)}{V(y)}(y - \bar{y})$	☐ $x = \bar{y} + \frac{\text{cov}(x,y)}{V(x)}(x - \bar{x})$

3- Dans le cas d'indépendance totale, le coefficient de corrélation linéaire est égal à

☐ 0	☐ 1	☐ -1	☐ autre réponse

4- Une valeur élevée du coefficient de corrélation linéaire est signe d'une réelle relation causale, dans le cas

du revenu national et de la consommation finale	☐ OUI	☐ NON
du prix d'un produit et du prix d'un produit substituable	☐ OUI	☐ NON
du nombre d'abonnés au téléphone et des ventes de médicaments contre le stress	☐ OUI	☐ NON
des heures travaillées par les étudiants pour réviser leurs examens et leurs taux de réussite à ces examens	☐ OUI	☐ NON
de la taille des salariés et de leurs salaires	☐ OUI	☐ NON
de la taille des salariés et de leurs poids	☐ OUI	☐ NON
de la température et de l'allongement d'une barre d'acier	☐ OUI	☐ NON

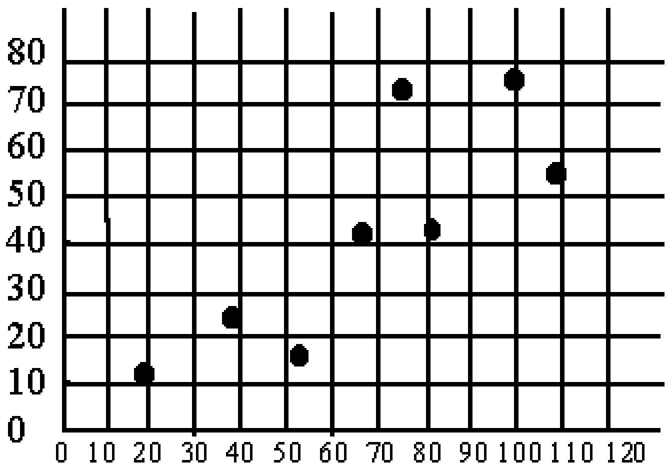
5- Utiliser les calculs effectués dans le tableau ci-dessous pour calculer la covariance entre les variables x et y

i	xi	yi	xiyi	xi²	yi²
1	50	7	350	2500	49
2	60	5	300	3600	25
3	70	6	420	4900	36
4	80	3	240	6400	9
5	90	1	90	8100	1
SOMME	350	22	1400	255000	120

☐ -6300	☐ -28	☐ 28	☐ 308	☐ autre réponse

6- D'après les données et le graphique du tableau ci-dessous, indiquer laquelle des propositions s'applique correctement à ces informations

xi	yi
19	12
52	17
38	25
81	43
109	55
75	73
66	42
100	75



☐ La covariance entre x et y est positive	☐ La covariance entre x et y est négative
☐ on ne peut rien dire à priori sur le signe de la covariance entre x et y	☐ Le concept de la covariance n'est pas pertinent pour analyser statistiquement le phénomène étudié
☐ aucune proposition ne convient	

7- Calculer la pente a de l'équation de régression y = ax + b , pour les données du tableau suivant

i	1	2	3	4	5
xi	10	12	14	16	18
yi	957	939	971	1006	1012

☐ 853,1	☐ 977	☐ 0,09	☐ 8,85	☐ autre réponse

8- Calculer l'ordonnée à l'origine b de l'équation de régression y = ax + b , pour les données du tableau suivant

i	1	2	3	4	5
xi	16	18	20	22	24
yi	462	449	458	378	365

☐ 422,4	☐ -13,25	☐ 756,14	☐ 687,4	☐ autre réponse

SÉRIES CHRONOLOGIQUES

1-On considère la série chronologique

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
1995	10	12	13	14
1996	11	15	16	13
1997	12	17	18	15
1998	13	17	19	16

2- Si une série suit un modèle multiplicatif et qu'on divise les valeurs de la série brute par les valeurs des coefficients saisonniers, on obtient

☐ la série des variations aléatoires ou accidentelles
☐ la série ajustée
☐ la série désaisonnalisée (C.V.S.)
☐ autre réponse

3- Soit la série chronologique suivante, qui suit un modèle multiplicatif

t	1	2	3	4	5	6	7	8	9	10	11	12
yt	47	30	39	14	62	40	50	16	69	50	62	15

Le trend, à la date t = 3, calculé par les moyennes mobiles d'ordre 4 est égal à

☐ 39	☐ 22	☐ 34,38	☐ 68,75	☐ 28,51

La valeur à la même date de la série CVS est

☐ 41,46	☐ 0,98	☐ 37,5	☐ 38,4	☐ 33,9

4- Soit la série chronologique

	Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4
Année 1	20	18	20	22
Année 2	24	22	24	26
Année 3	28,8	26,8	28,8	30,8
Année 4	34,6	32,6	34,6	36,6
Année 5	41,5	39,5	41,5	43,5

La série suit un modèle de type

☐ additif	☐ multiplicatif

5- Soit Yt la série du chiffre d'affaires mensuel d'une entreprise de janvier 1987 à décembre 1991. L'équation du trend est Tt = 3,76 t + 700 ; (t = 1,....,60)
Les coefficients saisonniers sont :

janvier S1 = -16	mai S5 = 11	septembre S9 = - 60
février S2 = -51	juin S6 = 64	octobre S10 = -1
mars S3 = -80	juillet S7 = 0,09	novembre S11 = 62
avril S4 = -81	août S8 = -69	décembre S12 = 222

Sachant qu'on a un modèle additif, une estimation de la valeur future de juin 1993 est

☐ 940,64	☐ 1057,3	☐ 764
☐ 1038,48	☐ 831,7	☐ autre réponse

6- Soit la série chronologique ci-après qui suit un modèle de type additif

	1996	1997	1998
1° trimestre	420	515	500
2° trimestre	615	685	835
3° trimestre	825	1000	980
4° trimestre	540	620	700

- La moyenne mobile d'ordre 4 du 3° trimestre 1997 est

☐ 768	☐ 772	☐ 703	☐ 733	☐ 680

- La valeur du coefficient saisonnier brut S' du 1° trimestre est

☐ 5,15	☐ 48	☐ - 65	☐ - 192	☐ - 109

- Le coefficient saisonnier S du 1° trimestre est

☐ - 109	☐ - 179	☐ -194	☐ - 13

- La valeur de la série CVS au 2° trimestre de l'année 1996 est

☐ 609	☐ 679	☐ 576	☐ 642	

Evaluation de fin de module

Durée : 2h30
Questions : (8 points)
1- qu'est ce qu'on entend par :
- caractère qualitatif ?
- caractère quantitatif ?
- variable statistique discrète ?
- variable statistique continue ?

2- Définissez les termes suivants :
- le mode
- la médiane
- l'étendue
- l'écart type

Exercice 1 (6 points)
En l'année N, les recettes du budget de l'Etat se présentent de la façon suivante (en milliards de HD):
- taxe de la valeur ajoutée (TVA) : 348
- Impôt général sur les revenus(IGR) : 168
- Impôt sur les sociétés (IS) :71
- Taxe sur les produits pétroliers : 54
- Autres impôts : 161
- Recettes non fiscales : 41

Travail à faire :
Représentez graphiquement les recettes du budget de l'Etat en N par deux graphiques adéquats de votre choix.

Exercice 2 : (6 points)
Une série d'observations concernant les notes obtenues à un examen par un groupe de stagiaires de même age a donné les résultats suivants :

Notes	[10,30[	[30,50[	[50,70[	[70,90[	[90,110[	[110,130[	[130,150[	[150,170[	[170,190[
Effectifs	4	17	63	83	72	33	21	5	2

Travail à faire :
Déterminez la note moyenne et calculez l'écart type de la série.

<u>Eléments -critères d'évaluation</u>

Questions :
1- distinguer les différents types de caractères
2- définir les différents paramètres de tendance centrale et de dispersion

Exercice 1 :
- choisir les représentations graphiques correspondantes
- choisir les graduations et les légendes adéquates

Exercice 2 :
- Calculer avec exactitude la moyenne
- Calculer avec exactitude l'écart –type
- Suivre une méthodologie pour le calcul.

Liste des références bibliographiques :

Ouvrage	Auteur	Edition
Probabilités et statistiques	Audet, Boucher, Caumartin et Skeene	Gaëten morin, 1983
Manuel de statistiques descriptives	Omar Rajaâ	El Wataniya, 2001
Mémento pratique sta tistiques	Rachid Boutti	Collection Expertise, 1996
Gestion prévisionnelle et mesure de la performance	Brigitte Doriath et christian Goujet	Dunod, 2002
L'essentiel du marketing	Eric Vernette	Editions d'Organisation, 2002
Statistiques descriptives Niveau technicien	O.F.P.P.T	Mars 1993
www.larrun.iut.bayonne.univ-pau.fr		

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 : <u>des statistiques du chômage.</u>

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 $1^{\text{ére}}$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 / discontinu, discret
→ 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

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

Série avec des classes.

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

Série simple.

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 x_2 x_3 x_4	10 x_1 25 x_2 12 x_3 4 x_4

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

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

3. Fréquence relative :

F_i = effectif de la modalité i / effectif total

4. L'opérateur somme ($\sum$)

- Notation : n variables

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

- Propriétés :

$$\sum_{i=1}^{n} axi = a\sum_{i=1}^{n} xi$$

$$\sum a + xi = \sum a \sum_{i}^{n} xi = n.a + \sum_{i=1}^{n} xi$$

5. L'opération de produit : (Π)

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

$X_1.x_2.x_3\ldots.x_n = \prod_{i=1}^{n} xi$

- Propriété :

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

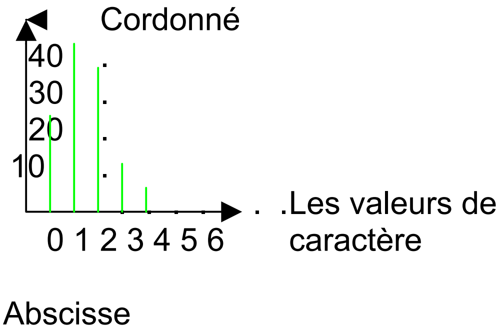
CHAPITRE I : LA REPRESENTATION GRAPHIQUE

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 :

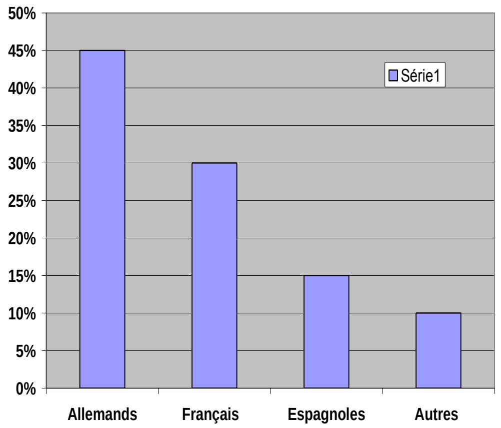
On s'en sert pour représenter des séries à caractère discret.

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



II. Le tuyau d'orgue :

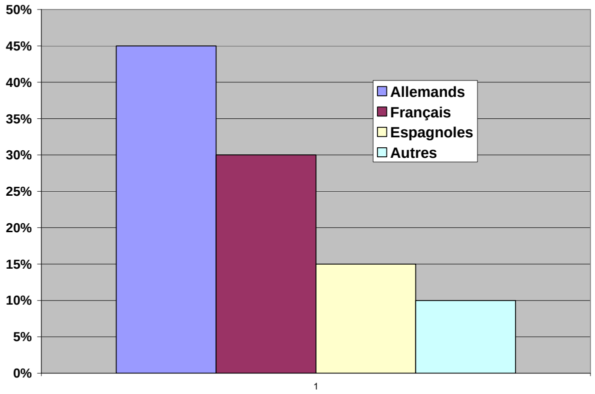
On se sert de ce graphique pour représenter des séries à caractère qualitatif
EX : La population à une station balnéaire est composée de :
Allemands : 45%
Français : 30%
Espagnoles : 15%
Autres : 10%



III. Le diagramme :

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

a. <u>Cas ou les intervalles de classe sont égaux :</u>



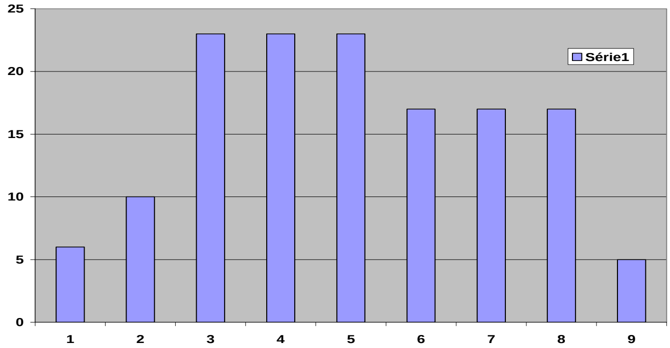
Remarque :

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. <u>Cas ou les intervalles de classe ne sont pas égaux :</u>

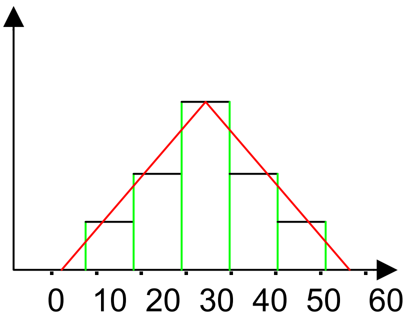
EX : Répartition de population selon leurs salaires.



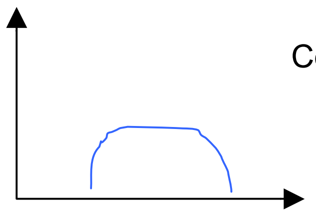
Pour tracer l’histogramme, on commence par corriger les effectifs.

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.



Courbe de fréquences

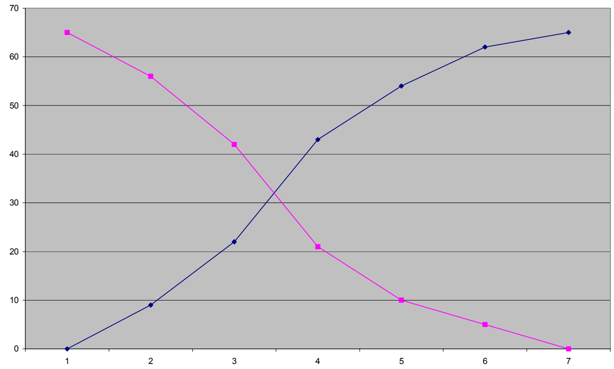
V. La courbe de cumulation (courbe des f cumulés) :

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 :

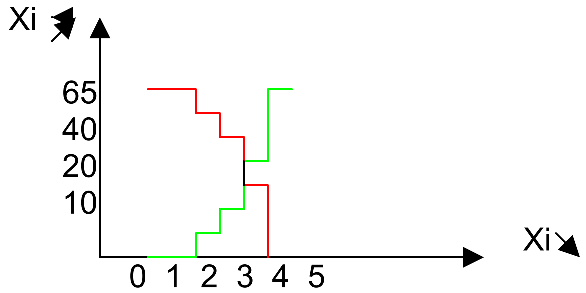
Salaire	xi	Xi cumulés ↗	Xi cumulés ↘
[10-20[	9	9	65
[20-30[	13	22	56
[30-40[	22	44	43
[40-50[	10	54	21
[50-60[	7	61	11
[60-70[	4	65	4
Total	65	Moins de la borne supérieure	Plus de la borne inférieure

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 2 3 4	5 10 30 20	5 15 45 65	65 60 50 20
Total	65	<=xi	>=xi

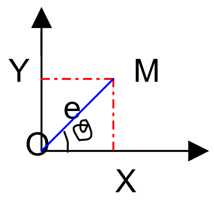


VI. Le diagramme polaire :

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).

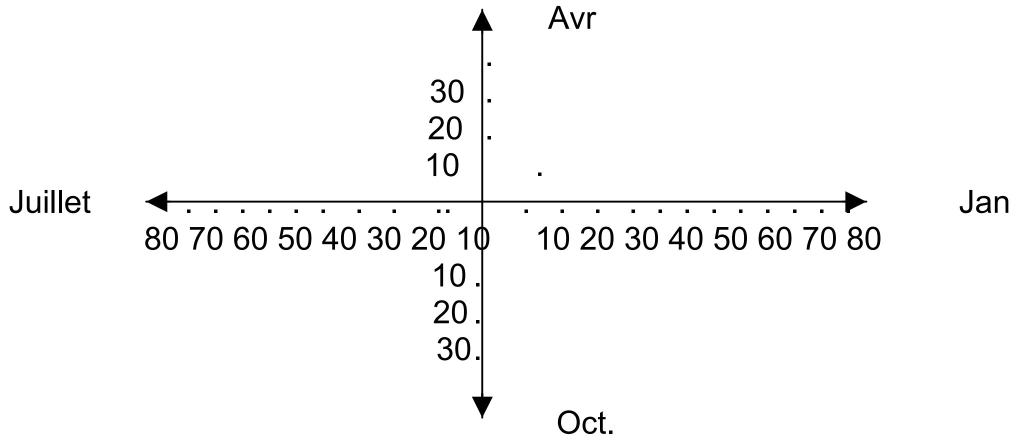


b) Le diagramme polaire :

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

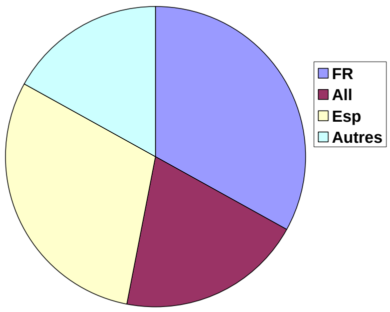
L'idée est de présenter chaque mois par un axe, nous aurons donc 12 axes, chaque axe faisant avec son voisin un angle.



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 :

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

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

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 :

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

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{\sum n_i}$$

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

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

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

Moyenne de l'age ou l'age moyen

a-3 Méthode des simplifications des calculs

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

$$\boxed{x_i = x_0 + a x_i^{'}}$$

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 $x_i = x_0 + ax_i^{'} \Rightarrow x_i^{'} = \frac{x_i - x_0}{a}$

La moyenne arithmétique :

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

$$= \frac{\sum n_i \left(x_0 + ax_i^{'}\right)}{\sum n_i}$$

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

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

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

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

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

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

$\overline{X}$ = 37,5+5(-29/88)=35,8 ans a-4 calcul de la moyenne arithmétique à l'aide des fréquences relatives

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

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

$$= f_i x_i + f_2 x_2 + \ldots\ldots\ldots + f_n x_n$$

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

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

$\overline{X} = \sum f_ix_i$

$\overline{X} = 12,7$

x_i	Ni	f_i	f_ix_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 :

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

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

b-2 calcul de G

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

$G = \sqrt[n]{X1.X1.X1 * X2.X2.X2 * X3 * X3X3} \Leftrightarrow G = \sqrt[n]{X1^{n1}.X2^{n2}.X3^{n3}....Xn^{nn}}$.

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

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)

= 1/n$\left[\log X1 + \log X2..... + \log Xn\right]$

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

La moyenne géométrique pondérée

$$G = \sqrt[n]{x_1^{n_1}.x_2^{n_2}........x_n^{n_n}}$$
$$G = \left(x_1^{n_1}.x_2^{n_2}.......x_n^{n_n}\right)^{1/n}$$

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

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

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

Application : calculer la moyenne géométrique

x_i	n_i	log x_i	n_i log x_i
2	1	0,301	0,301
6	2	0,772	1,556
10	3	1	3,0
12	2	1,158	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 $x_1, x_2, x_3 \ldots\ldots x_n$ *on appelle moyenne hormique le nombre H tel que :*

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

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

moyenne harmonique simple.

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

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

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

Moyenne harmonique pondérée

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

$$\frac{1}{H} = \frac{\sum n_i\ 1/x_i}{\sum n_i} = \frac{\sum n_i .X_i}{\sum n_i} avecX_i = \frac{1}{x_i}$$

L'inverse de la moyenne = moyenne des inverses

D -La moyenne quadratique :

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

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

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

moyenne quadratique simple

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

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

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

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

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

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 :

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

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

Si r= 1 $M_1^1 = \dfrac{x_1 + x_2 + x_3 + ... + x_n}{n} \Rightarrow M_1^1 = \overline{\mathrm{X}}$

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

si r= -1 $M_{-1}^{-1} = \dfrac{x_1^{-1} + x_2^{-1} + ... + x_n^{-1}}{n} \Rightarrow \dfrac{1}{M_{-1}} = H^{-1} = \dfrac{1}{H} \Rightarrow M_1 = H$

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 : $M_{-1} < M_0 < M_1 < M_2$

$$H < G < \overline{\mathrm{X}} < Q$$

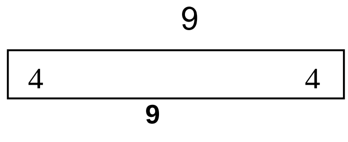
Dans notre exemple, on trouve : 6,16 <8,2 <8,5 < 9,11.

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.



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 3^{éme} 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 = S01,1 × 1,15 × 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 :

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

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 ⇔ Me =57
Si la série a un nombre pair
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

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

<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 :

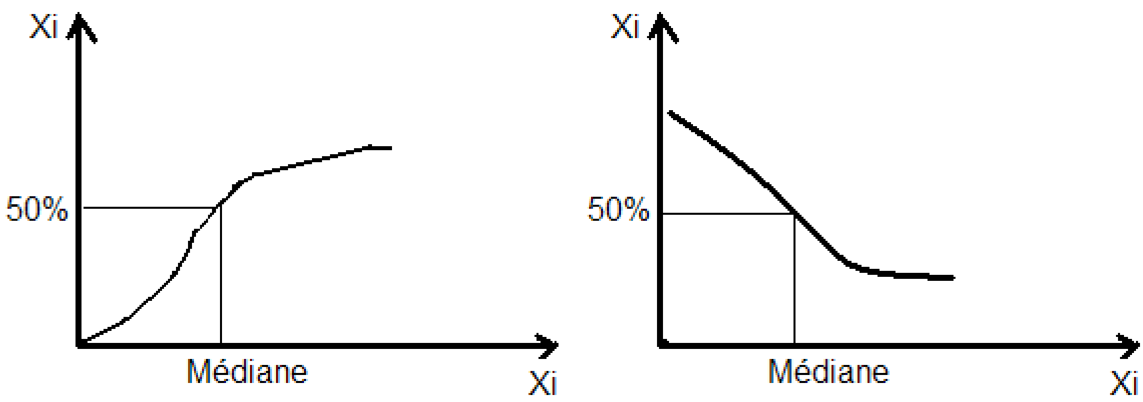
On connaît les salaires des 34 individus 20
On connaît les salaires des 66 individus 25

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

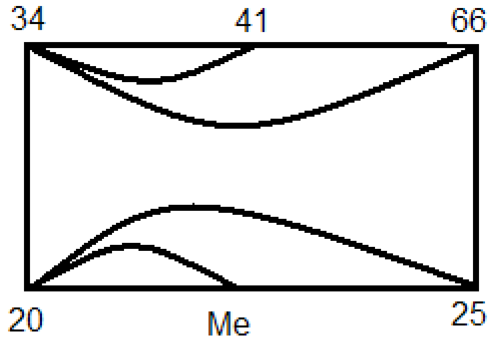


b-4-Remarque :

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

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

Mo =14

Série

Uni modal

Mo = 17

Mo = 39

Série bimodale

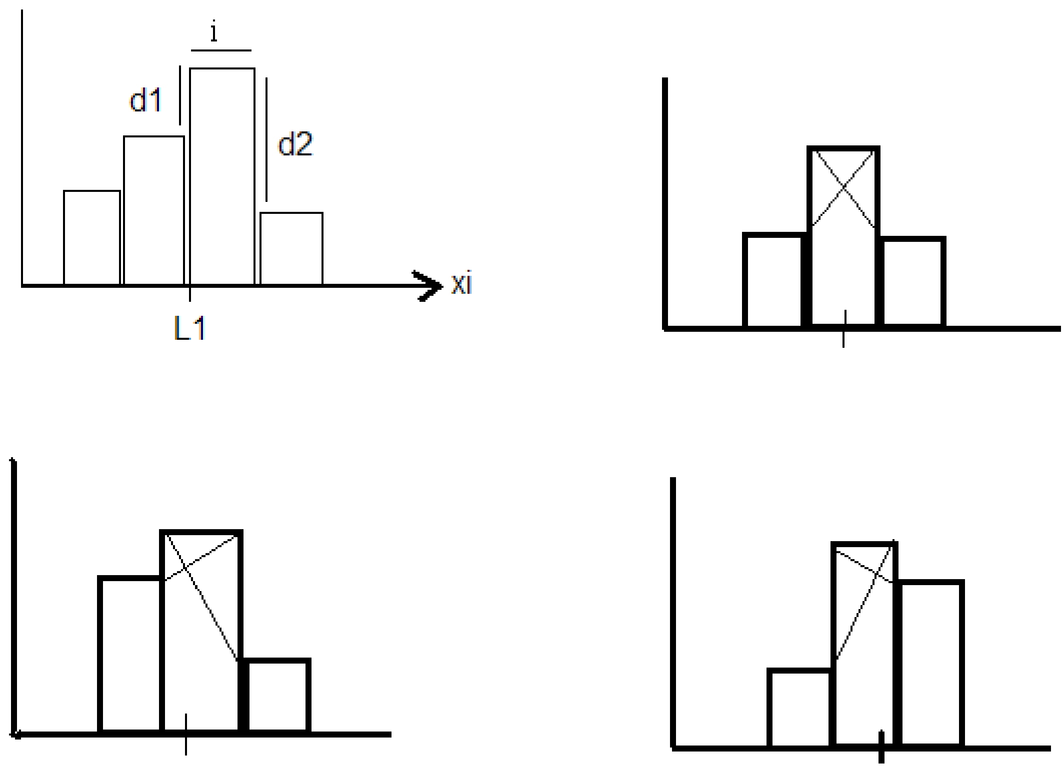
Série plurimodale (série à plusieurs modes)

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 :

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

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

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 : Les Caractéristiques de Dispersion:

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 3 éme 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. $\boxed{25\% < \ ; \ 75\% >}$

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

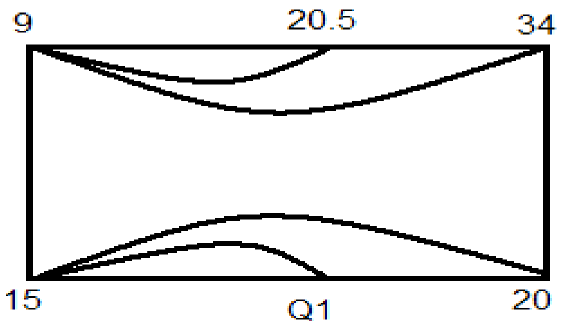
Ni Cum ↗
9
34
66
82

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 :



$$\frac{Q1 - 15}{20 - 15} = \frac{20.5 - 9}{34 - 9}$$

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

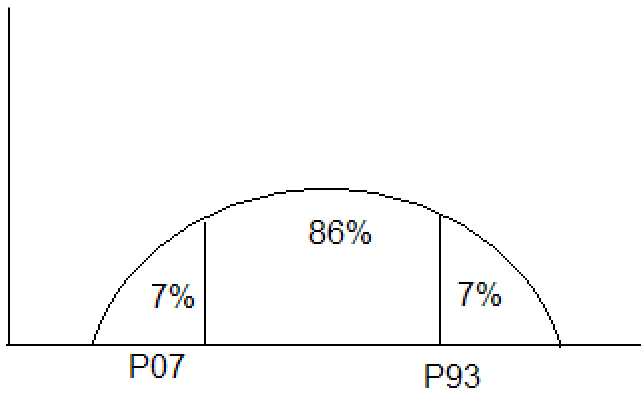
(61,5 – 34) = 27,5 individus Augmentation 5/32 *27,5

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	$\left|xi - \bar{x}\right|$	$ni\left|xi - \bar{x}\right|$
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 $\bar{X}$ = 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

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

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

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 = $\sigma/\overline{X}$
Ex :
$\overline{X}$ =67.75 Kg $\sigma/\overline{X}$=(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

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

Le modèle I est plus faible que le modèle II

Formule développée :

Donc $\sigma = \frac{\sum ni\, xi^2}{\sum ni} - \bar{x}^2$

Poids	ni	xi	xi^2	ni * xi^2
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**

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

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

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 $\dfrac{\Delta m}{\text{Inter variation}}$ 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 $\Bigg\} \Longleftrightarrow$ 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

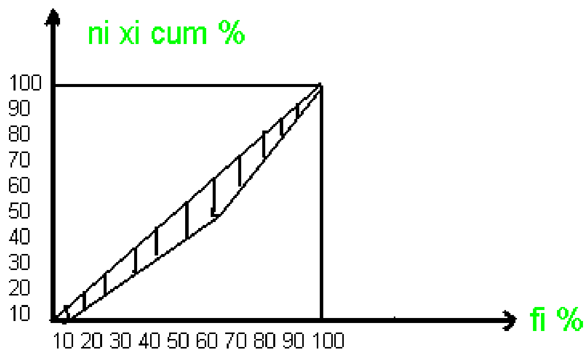
C'est le salaire tel que la moitié de la masse salariale a servi à payer des gens qui gagnent moins que 22.12 dh et l'autre moitié de la masse salariale a servi à payer les gens qui gagnent plus que 22.12 dh

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		



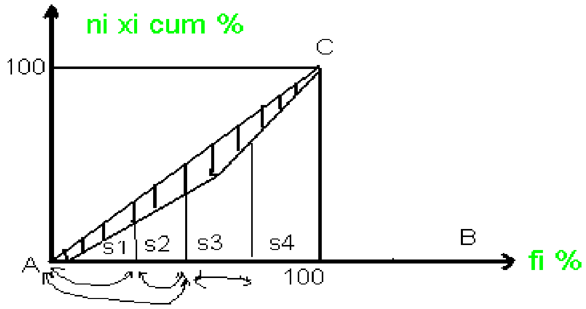
/ : Diagonal de l'égalité
: Aire de concentration

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 :

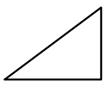
Gini propose de calculer la concentration à l'aide de coefficient suivant :



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

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

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

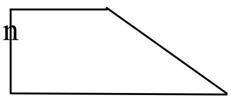


B S=1/2 a*b

A

A

n



S = n/2(a+b)

b

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

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 $\bar{\bar{X}}$

b- la répartition des logements selon la 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 $\overline{\overline{Y}}$

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_2 … X_i … X_k	X_{11} X_{21} … X_{i1} … X_{k1}	X_{12} … X_{i2} … X_{k2}	……… ……… ……… ……… ……… ………	X_{1j} X_{2j} … X_{ij} … X_{kj}	……… ……… ……… ……… ……… ………	X_{1m} X_{2m} … X_{im} … X_{km}	$X_{1.}$ $X_{2.}$ … $X_{i.}$ … $X_{k.}$
total	$x_{.1}$	$x_{.2}$	………	$x_{.j}$	………	$x_{.m}$	$x_{..}$

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

La distribution marginale de X :

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

La distribution marginale de y :

$y(x_i)$	$X_{j.}$
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	$X_{i.}$

III- La régression linéaire

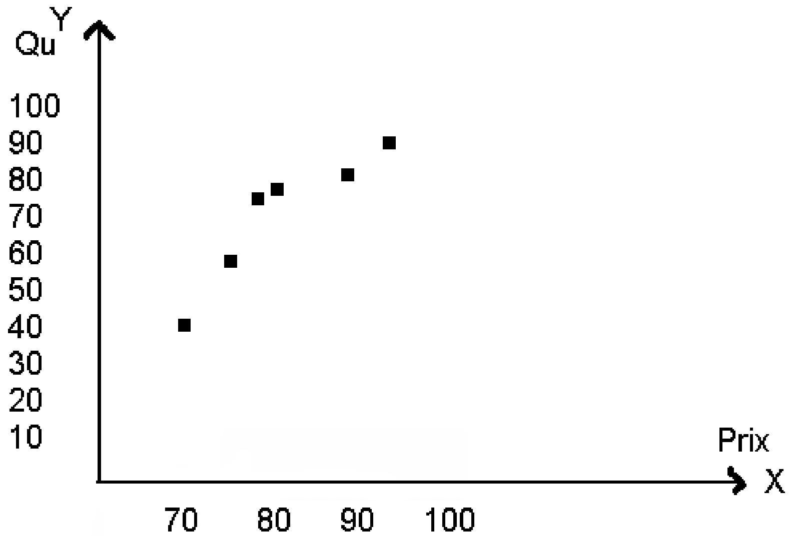
A. Présentation du problème :

Soit le tableau suivant :

qu / Prix	42	51	60	62	74	83	Total
70	1						1
75		1					1
77			1				1
80				1			1
86					1		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.

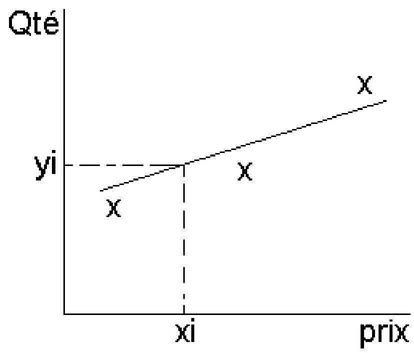
a=pente de droite
b=ordonnée à l'origine

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

<u>Notion de moindres carrés</u> :

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'.

$dZ / db = 2 [\sum (y_i - ax ; -b)] (-1) = 0$ $\Longrightarrow$ $Z = U^2$, $Z' = 2UU'$

$\sum [yi - ax ; -b) = 0$

$\sum yi - a\sum xi - nb = 0$ $U' = (y_i - ax ; -b)$

Divisons par n, on obtient $(\sum yi / n - a\sum xi / n - b = 0$

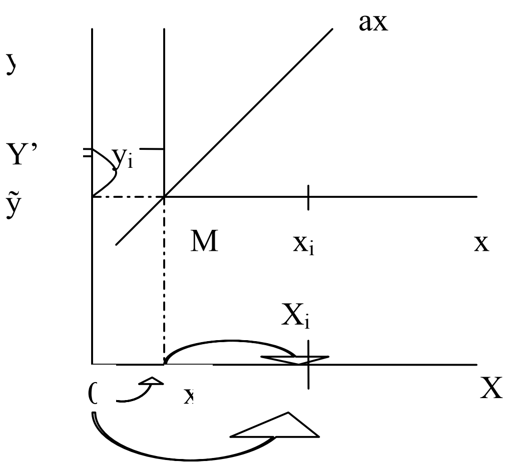
$\boxed{\tilde{y} - a\bar{x} = b}$

Donc :

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

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

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 ($\bar{x}$; $\tilde{y}$).
Procédons un changement d'origine, et prenons comme nouvelle origine le point moyen M(x' ; $\tilde{y}$), les nouvelles cordonnées deviennent :

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

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 $a = \sum x_i y_i / \sum x_i^2 = \sum (x_i - x') (y_i - \tilde{y}) / \sum (x_i - x')^2$

3**- <u>l'équation de la droite de régression</u>** :

Dy(x) =
Y = ax + b

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

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

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

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

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

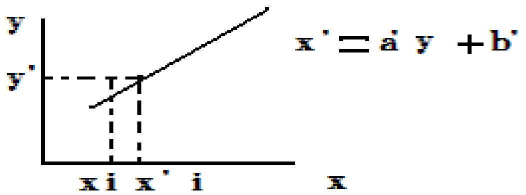
Trouver Dy (x).

xi - $\bar{x}$	yi - ỹ	(xi-$\bar{x}$) (yi - ỹ)	(xi-$\bar{x}$)^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

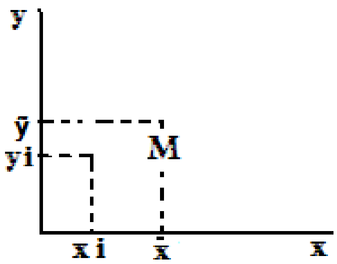
IV- la corrélation linéaire :

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.



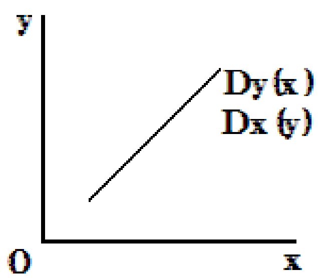
Pour toute yi, nous avons une valeur observée x_i.
Pour toute y_i, nous avons une valeur estimée sur la droite x'_i
Pour toute y_i, nous avons une erreur d'estimation égale à $| x_i – x'_i |$
Dx(y) idéale est tel que : $\sum | x_i – x'_i |$ minimum ou encore $\sum (x_i – x'_i)^2$ minimum
En procédant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

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



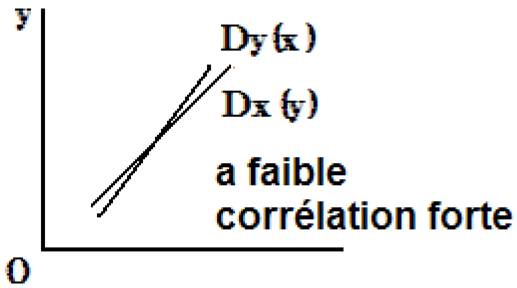
Dans le référentiel XMY nous obtenons 2 droites :
Soit y = ax pour Dy(x)
Soit x = a'y pour Dx(y)
Ou encore y = 1/a' x
4 cas peuvent se produire :

1er cas : les 2 droites sont confondues

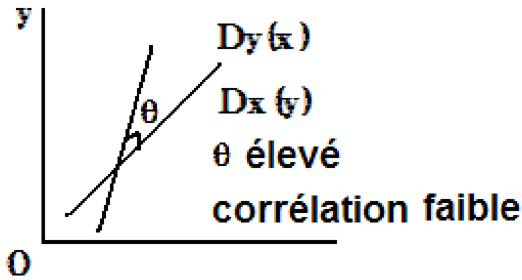


$$\left.\begin{array}{l} Y = ax \\ X = a'y \\ Y = 1/y'x \end{array}\right\} \Leftrightarrow a = 1/a' \Leftrightarrow aa' = 1$$

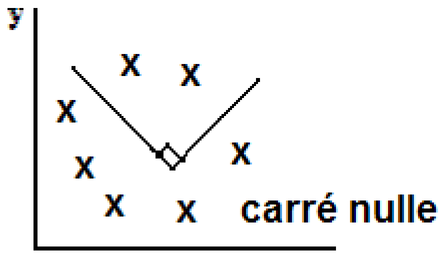
2éme cas : les 2 droite font entre elles un angle très faible :



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 coéff de corrélation la Quantité r tel que : $r^2 = a \cdot a'$, 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 - ỹ	$(x - \bar{x})$ (y_i - ỹ)	$(x - \bar{x})^2$ $(y_i - ỹ)^2$
70 75 77 80 86 33	42 51 60 62 74 83				
			606	339	1110

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

$$\begin{cases} a = \dfrac{\sum x_i y_i}{\sum x^2_i} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \dfrac{606}{339} = 1.79 \\ a' = \dfrac{\sum x_i y_i}{\sum y^2_i} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} = \dfrac{606}{1110} = 0.545 \end{cases}$$

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 :

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

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

$$ỹ = \frac{\sum y_i}{N} \Rightarrow \sum y_i = n\ ỹ$$

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

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

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

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

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

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

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

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

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 = $\dfrac{\text{Cov (xy)}}{бx.бy}$

CHAPITRE IV : ANALYSE DES SERIES CHRONOLOGIQUES.

I – Généralités :

A. Définition :

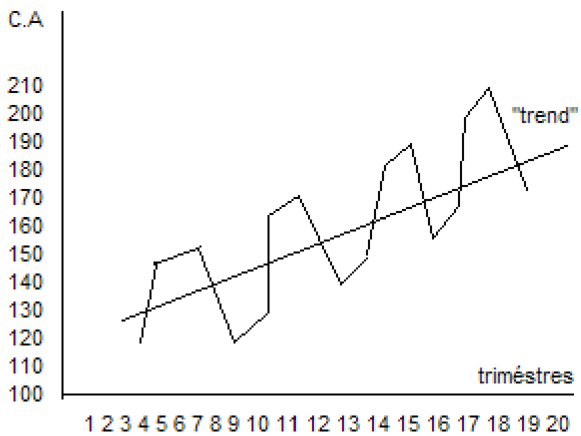
Une série chronologique est une série où les observations de la variable sont faites à des intervalles réguliers de temps.

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 $a_{ii} = 0_{(n\,;p)}$ 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 :

Politique	Sondages, référendums, popularité des partis politiques et de leur chef.
Social	Criminalité, suicide, avortement, racisme, pratiques religieuses, orientations sexuelles, habitudes alimentaires.
Économie	Importations, exportations, prix de vente, taux d'inflation, indice des prix à la consommation (IPC), taux d'intérêt, salaires, taux de chômage, cotes boursières, indices boursiers, déficits gouvernementaux.
Démographie	Taux de mortalité, taux de natalité, population par province, par nationalité.
Culture	Entrées au «box office», cotes d'écoutes.
Études	Résultats scolaires, prêts et bourses, cote R et cote Z.
Sports	Meilleurs compteurs, classement des équipes, salaires des joueurs.

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.

II. Exemples:

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.

III. Étapes d'une enquête statistique :

1. Déterminer la population cible et les caractéristiques de cette population que l'on veut étudier.
2. Déterminer la manière dont l'échantillon va être prélevé.

3. Construire des instruments (questionnaires ou autres).
4. Établir un pré-test ou étude-pilote.
5. Recueillir les données.
6. Compiler les données.
7. Mettre en forme les données.
8. Analyser les données (analyse descriptive ou inférentielle).
9. Interpréter les résultats.
10. Communiquer les résultats.

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 2ème 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 1er et le 9ème 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 3ème 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	**16**
2	6	12	24
3	**3**	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	8	4
1,80-1,90	?	7	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.