

BEMBUS – MALI

Au Service de la Nation

« Le savoir est la plus grande des richesses : on le donne sans réussir à l'épuiser »

« Tabarakalézi Bi-yadihil – Mouloukou WaHouWa Allah-kouli-ché-ine Kadirou »

Bemba dit Bamba TRAORE

Enseignant Chercheur, économètre, gestionnaire de la politique
économique

Tél : (00223) 69 - 50-23-20/ 76 - 43-95-33 mël : bbtraoremali@yahoo.fr

Statistique Descriptive

- 1.- Distribution statistique à une variable x
- 2.- Distribution statistique à deux variables x et y
- 3.- Indices simples et synthétiques

Bamako, novembre 2021.

Chapitre 1 : Distribution statistique à une variable x

1.-/ OBJET DE LA STATISTIQUE

Le but de la statistique est de dégager les significations de données, numériques ou non, obtenues au cours de l'étude d'un phénomène.

Il faut distinguer les **données statistiques** qui sont les résultats d'observations recueillies lors de l'étude d'un phénomène, et la **méthode statistique** qui a pour objet l'étude rationnelle des données.

La méthode statistique comporte plusieurs étapes :

- **La statistique descriptive ou déductive** : c'est l'ensemble des méthodes à partir desquelles on recueille, ordonne, réduit, et condense les données.
- **La statistique mathématique ou inductive** : c'est l'ensemble des méthodes qui permettent de faire des **prévisions**, des interpolations sur une population à partir des résultats recueillis sur un échantillon. Nous utilisons des raisonnements **inductifs** c'est-à-dire des raisonnements de passage du particulier au général. Cette statistique utilise des repères de référence qui sont les **modèles théoriques** (lois de probabilités).

2.- VOCABULAIRE STATISTIQUE

- **La statistique** est la science qui étudie les chiffres.
- **La population statistique** est un ensemble étudié concernant un phénomène donné par la statistique. Exemple: l'ensemble des étudiants de la 1^{ère} SE.
- **Unité statistique** : un élément de l'ensemble étudié s'appelle **unité statistique** ou individu. Exemple: l'ensemble des maliens l'unité statistique un malien. Chaque unité statistique peut être étudiée suivant un ou plusieurs caractères (qualitatif ou quantitatif).

Caractère : c'est un trait déterminé C présent chez tous les individus d'une population sur laquelle on effectue une étude statistique. **Les modalités** Ce sont les différentes situations possibles du caractère.

- **Le Caractère est dit qualitatif**: les modalités du caractère ne sont pas mesurables. Exemple: sexe, ethnie, religion, nationalité, race, couleur des cheveux.

Le caractère qualitatif peut être nominal ou Ordinal

La couleur des vestes des étudiants : Caractère qualitatif nominal

L'appréciation des notes des étudiants : Caractère qualitatif ordinal

- **Caractère est dit quantitatif ou variable statistique**: les modalités du caractère sont mesurable ou repérables exemple: le nombre d'enfant à charge, taille, âge, température ...

- **Variable statistique discrète:** elle est discrète lorsque les modalités sont des nombres entiers. Exemples: nombre d'enfants, nombre de parcelles.
- **Variable statistique continue:** elle est continue lorsque les modalités sont des valeurs comprises entre deux valeurs distinctes (des nombres décimaux) exemple: taille, poids ...

Exemple: Parc - auto on peut s'intéresser.

- ☐ A la couleur des voitures : caractère qualitatif nominal
- ☐ A la puissance des voitures : caractère quantitatif discret
- ☐ A la marque des voitures : caractère qualitatif nominal

Echantillon

C'est un ensemble d'individus prélevés dans une population déterminée

Exemple d'échantillon.

L'échantillon des véhicules automobiles immatriculés dans un département.

3.- Méthode d'observation statistique

Méthode directe : nous nous adressons directement à la population statistique : recensement exhaustif et enquête partielle

Méthode indirecte : on soumet à l'observation statistique les documents qui ont élaborés dans un but autre que statistique : les registres d'Etat civil

4.- Dépouillement et présentations de résultats statistiques

Le dépouillement consiste à regrouper les résultats d'une enquête statistique dans un tableau dans le but de faciliter son exploitation. Ce tableau peut être : de simple entrée ou de double entrée

Tableau : la répartition des employés d'une entité suivant le nombre d'enfants

Nombre d'enfants	0	1	2	3	Total
Employés	12	15	5	8	40

Tableau : la répartition des employés d'une entité suivant le sexe et le nombre d'enfants

Nbre d'enfants	0	1	2	3	Total
Sexes					
Masculin	4	10	1	6	21
Féminin	8	5	4	2	19
Total	12	15	5	8	40

Application : dépouillement et tableau statistique

Une enquête sur les employés d'une entreprise a donné les résultats suivants :

Une enquête sur les employés d'une entreprise a donné les résultats suivants :

Agents	Sexes	Situation matrimoniale	Nombre d'enfants	Salaires	Ages
A	Masculin M	Célibataire	0	175 000F	25
B	Féminin F	Marié	2	420 000 F	30
C	Masculin M	Marié	1	275 000 F	35
D	Féminin F	Célibataire	0	250 000 F	40
E	Masculin M	Marié	1	150 000F	45
F	Masculin M	Célibataire	1	172 000 F	45
G	Féminin F	Célibataire	1	157 000 F	27
H	Féminin F	Marié	2	200 000 F	42
I	Masculin M	Marié	1	205 000 F	37
J	Féminin F	Marié	0	200 000F	26
K	Masculin M	Marié	2	299 000F	34
L	Féminin F	Marié	1	290 000F	49
M	Masculin M	Marié	2	257 000 F	40
N	Masculin M	Célibataire	1	260 000F	31
O	Féminin F	Marié	0	250 000F	28
P	Masculin M	Marié	1	151 000 F	42
Q	Masculin M	Célibataire	2	150 000 F	29
R	Féminin F	Marié	1	406 000F	43
S	Masculin M	Célibataire	0	355 000F	37
T	Masculin M	Marié	2	410 000 F	49

Travail à faire

- Effectuer le dépouillement puis compléter le tableau ;
- Calculer la fréquence ;
- Indiquer la population étudiée;
- Indiquer la nature du caractère ;
- Indiquer la représentation graphique qui convient

a.- La répartition suivant le sexe

b.- La répartition suivant la situation matrimoniale

c.- La répartition suivant le nombre d'enfants

d.- La répartition suivant le salaire

150 000 à 250 000 ; 250 000 à 350 000 ; 350 000 à 450 000.

e.- La répartition suivant le salaire

25 à 30 ; 30 à 35 ; 35 à 40 ; 40 à 45 ; 45 à 50 ; 50 à 55.

f.- La répartition suivant la situation matrimoniale et le sexe

g.- La répartition suivant le nombre d'enfant et le sexe

Solution**a.- La répartition suivant le sexe :**

Caractères (Sexes)	Masculin	Féminin	Total
Effectifs (Employés)	12	8	20
Fréquences (en %)	60%	40%	100%

$$fréquence = \frac{\text{effectif } N_i}{\text{effectif total } N} * 100 \Rightarrow \sum \text{Fréquence} = 100\% \text{ ou } 1$$

La population étudiée : les 20 employés d'une entreprise ;

Le caractère étudié : le sexe → Qualitatif nominal

Représentation graphique qui convient : Diagramme en bandes ou à secteur.

b.- La répartition suivant la situation matrimoniale

Caractères (Matrimoniale)	Célibataire	Marié	Total
Effectifs (Employés)	7	13	20
Fréquence	35%	65%	100%

Le caractère étudié : la situation matrimoniale → Qualitatif nominal

Représentation graphique qui convient : Diagramme en bandes ou à secteur.

c.- La répartition suivant le nombre d'enfants

Caractères Xi (nombre d'enfants)	0	1	2	Total
Effectifs (Employés) Ni	5	9	6	20
Fréquence $fi = \frac{Ni}{N} * 100$	25%	45%	30%	100%
Effectifs cumulés croissants ECC	5	14	20	XXXX
Effectifs cumulés décroissants ECD	20	15	6	XXXX

ECC « on fait la somme des effectifs du 1^{er} chiffre au dernier chiffre »

ECD « on fait la somme des effectifs du dernier chiffre au 1^{er} chiffre »

Le caractère étudié : le nombre d'enfants → Quantitatif discret.

Représentation graphique qui convient : Diagramme en bâtons.

d.- La répartition suivant le salaire

Salaires	[150 000 ; 250 000 [	[250 000 ; 350 000[	[350 000 ; 450 000 [	Total
Effectifs	9	7	4	20
Fréquences	45%	35%	20%	100%
Centre Xi	200 000	300 000	400 000	XXX

Centre de la classe [a , b [est $Xi = \frac{a+b}{2}$

Le caractère étudié : le salaire → Quantitatif continu.

Représentation graphique qui convient : Histogramme.

e.- La répartition suivant l'âge

Caractères (âges)	[25 ; 30 [	[30 ; 35 [	[35 ; 40[	[40 ; 45 [	[45 ; 50 [	Total
Effectifs (Employés)						
Fréquences						
Centre Xi						

5.- Les éléments de calculs de la variable statistique X

Exemple 1 : cas de variable discrète

Soit la répartition des élèves suivant le nombre de frères

Nbre de frères X_i	0	1	2	3	Total
Effectif N_i	5	8	3	4	20

Travail à faire :

- 1.-) Dresse un tableau statistique en y figurant : fréquences ; effectifs cumulés croissants ; effectifs cumulés décroissants.
- 2.-) Trace le diagramme en bâtons.
- 3.-) Détermine la fonction de répartition $F(x)$ puis représente graphiquement.
- 4.-) Calcule la moyenne la variance et l'écart - type

Réponse :

1.-)

La population étudiée : les 20 élèves

Le caractère : nombre de frères : quantitatif discret.

Nbre de frères X_i	0	1	2	3	Total
Effectif N_i	5	8	3	4	20
Fréquence F_i	0,25	0,40	0,15	0,20	1
E C C	5	13	16	20	XXX
E C D	20	15	7	4	XXX

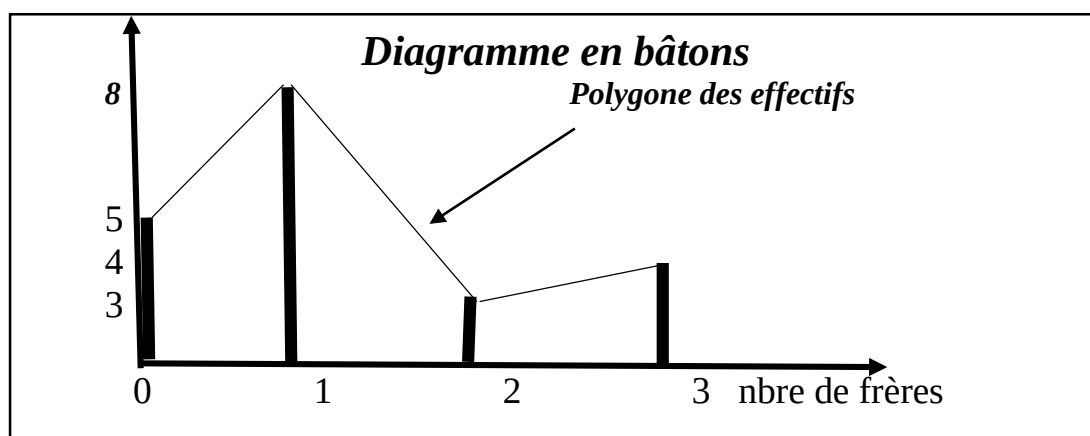
Fréquence $f_i = \frac{\text{effectif } N_i}{\text{effectif total } N} \rightarrow$ elle peut être en % $\rightarrow \sum f_i = 100\%$ ou 1

ECC « on fait la somme des effectifs du 1^{er} chiffre au dernier chiffre »

ECD « on fait la somme des effectifs du dernier chiffre au 1^{er} chiffre »

2.-) **Le diagramme en bâtons.**

On porte en abscisse les caractères X_i et en ordonnée l'effectif (ou la fréquence).



3.-) La fonction de répartition $F(x)$ est comprise entre 0 et 1 :

$$0 \leq F(x) \leq 1$$

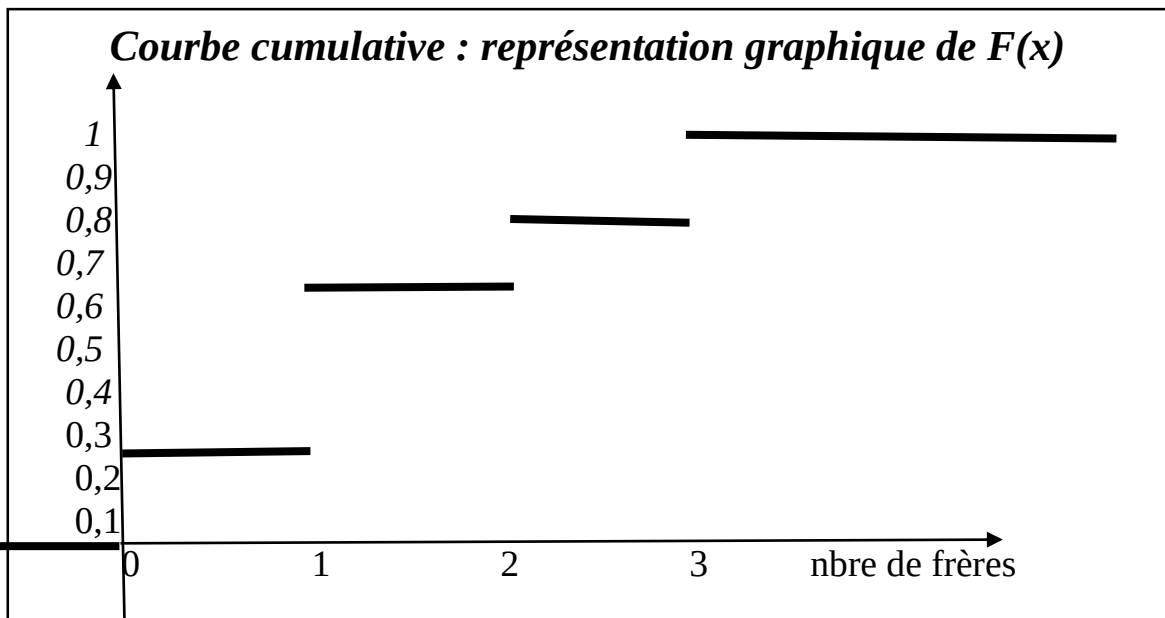
Quand x tend vers $-\infty$ alors $F(x) = 0$

Quand x tend vers $+\infty$ alors $F(x) = 1$

La fonction de répartition est établie en fonction des fréquences cumulées croissantes.

Nbre de frères X_i	0	1	2	3	Total
Fréquence F_i	0,25	0,40	0,15	0,20	1
Fréquences cumulées	0,25	0,65	0,80	1	XXX

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,25 & \text{si } 0 \leq x < 1 \\ 0,65 & \text{si } 1 \leq x < 2 \\ 0,80 & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$



4.-) Calcule la moyenne et la variance.

La moyenne = nombre total d'enfants / effectif total

$$\bar{X} = \frac{\sum N_i \cdot X_i}{N} = \frac{5 \times 0 + 8 \times 1 + 3 \times 2 + 4 \times 3}{20} = \frac{26}{20} = 1,3 \text{ enfants}$$

$$\text{La variance } V(x) = \frac{\sum N_i (X_i - \bar{X})^2}{N} = \frac{\sum N_i \times X_i^2}{N} - (\bar{X})^2$$

$$V(x) = \frac{5 \times 0^2 + 8 \times 1^2 + 3 \times 2^2 + 4 \times 3^2}{20} - (1,3)^2 = 1,11$$

L'écart-type est la racine carrée de la variance $\sigma_x = \sqrt{1,11} = 1,05$ enfants

Exemple 1 : cas de variable continue

Soit la répartition des élèves suivant le poids en kg

Poids kg	[50 ; 55[	[55 ; 60[	[60 ; 65[	[65 ; 70[	[70 ; 75[	Total
Effectif	4	3	6	5	2	20

Travail à faire :

- 1.-) Dresse un tableau statistique en y figurant : fréquences ; effectifs cumulés croissants ; effectifs cumulés décroissants.
- 2.-) Trace l'histogramme (amplitude égale).
- 3.-) Détermine la fonction de répartition $F(x)$ puis représente graphiquement.
- 4.-) Calcule la moyenne la variance et l'écart - type

Réponse :

1.-)

La population étudiée : les 20 élèves

Le caractère : poids : quantitatif continu.

Poids kg	[50 ; 55[	[55 ; 60[	[60 ; 65[	[65 ; 70[	[70 ; 75[	Total
Effectif	4	3	6	5	2	20
Fréquence	0,20	0,15	0,30	0,25	0,10	1
ECC	4	7	13	18	20	XXXX
ECD	20	16	13	7	2	XXX

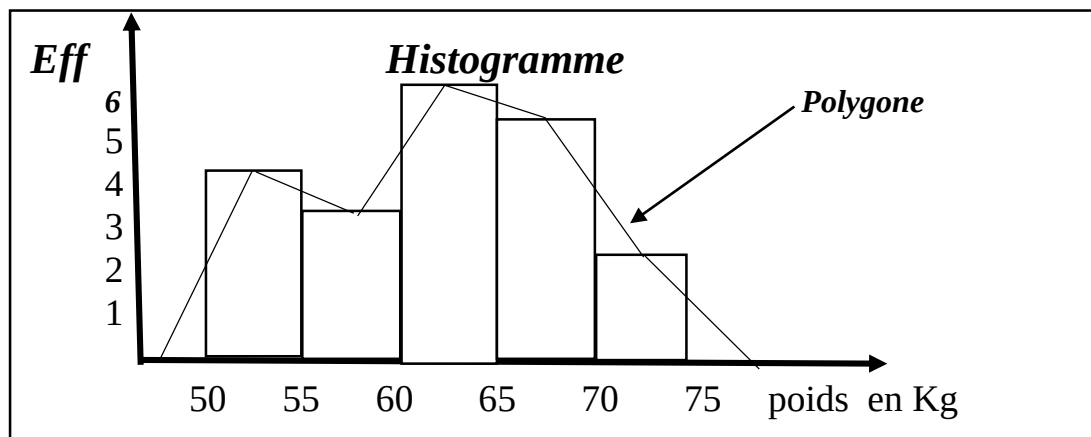
Fréquence $f_i = \frac{\text{effectif } N_i}{\text{effectif total } N} \rightarrow$ elle peut être en % $\rightarrow \sum f_i = 100\%$ ou 1

ECC « on fait la somme des effectifs du 1^{er} chiffre au dernier chiffre »

ECD « on fait la somme des effectifs du dernier chiffre au 1^{er} chiffre »

2.-) L'histogramme

On porte en abscisse les classes et en ordonnée les effectifs (fréquences)



Note : si les amplitudes ne sont pas égales, on corrige les effectifs (fréquences) en fonction de la plus petite amplitude.

$$\text{effectif rectifié} = \frac{\text{effectif de la classe}}{\text{amplitude de la classe}} \times \text{plus petite amplitude}$$

3.-) La fonction de répartition $F(x)$ est comprise entre 0 et 1 :

$$0 \leq F(x) \leq 1$$

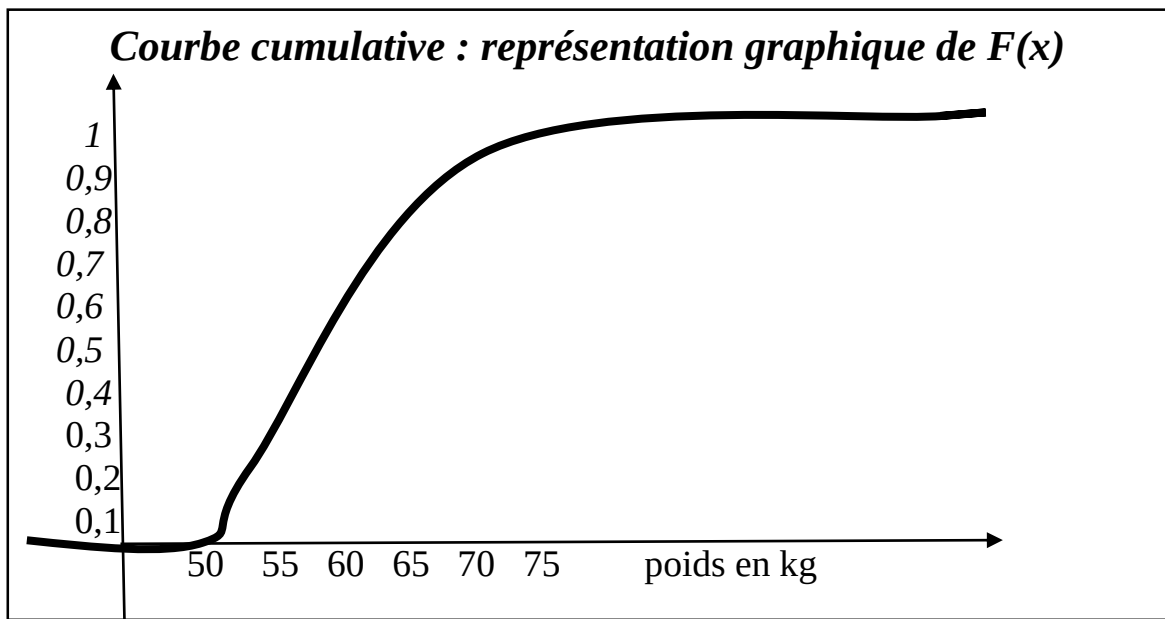
Quand x tend vers $-\infty$ alors $F(x) = 0$

Quand x tend vers $+\infty$ alors $F(x) = 1$

La fonction de répartition est établie en fonction des fréquences cumulées croissantes.

Poids kg	[50 ; 55[	[55 ; 60[	[60 ; 65[	[65 ; 70[	[70 ; 75[	Total
Fréquence	0,20	0,15	0,30	0,25	0,10	1
FCC	0,20	0,35	0,65	0,90	1	XXXX

$$F(x) = \begin{cases} 0 & \text{si } x < 50 \\ 0,20 & \text{si } 50 \leq x < 55 \\ 0,35 & \text{si } 55 \leq x < 60 \\ 0,65 & \text{si } 60 \leq x < 65 \\ 0,90 & \text{si } 65 \leq x < 70 \\ 1 & \text{si } 70 \leq x < 75 \\ 1 & \text{si } x \geq 75 \end{cases}$$



4.-) Calcule la moyenne et la variance.

Poids kg	[50 ; 55[	[55 ; 60[	[60 ; 65[	[65 ; 70[	[70 ; 75[	Total
Centre X_i	52,5	57,5	62,5	67,5	72,5	XXX
Effectif N_i	4	3	6	5	2	20

La moyenne = nombre total des poids / effectif total

$$\bar{X} = \frac{\sum N_i \cdot X_i}{N} = \dots\dots\dots$$

$$\text{La variance } V(x) = \frac{\sum N_i \cdot x \cdot X_i^2}{N} - (\bar{X})^2 = \dots\dots\dots$$

6.- Paramètres caractéristiques.

Il existe 3 types de paramètres :

- paramètres de position (ou de tendance centrale)
- paramètres de dispersion
- paramètres de forme (asymétrie, aplatissement, concentration)

6.1.- Paramètres de position (mode, médiane et moyenne)

Les paramètres de position (mode, médiane, moyenne) permettent de savoir autour de quelles valeurs se situent les valeurs d'une variable statistique.

6.1.1.- Le mode

Le mode, noté Mo , est la modalité qui admet **la plus grande fréquence**.

6.1.2.- La médiane

La médiane est la valeur du caractère qui correspond à l'unité statistique placée au milieu de la population statistique.

Cas des variables statistiques discrètes :

- Si l'effectif total N est impair, $N = 2K + 1$, la médiane est la valeur du $(K+1)^{ième}$ individu (à partir des effectifs cumulés croissants).
- Si l'effectif total est pair, $N = 2K$ la médiane n'est pas une valeur du caractère mais elle vaut $M_e = \frac{x_k + x_{k+1}}{2}$

Cas des variables statistiques continues :

- L'effectif total $N \rightarrow N / 2$
- L'effectif cumulé croissant $ECC \rightarrow ECC1 < \frac{N}{2} < ECC2$
 $\rightarrow$ classe médiane $[a, b[$

$$Mé = a + \frac{(b-a) \left(\frac{N}{2} - ECC1 \right)}{ECC2 - ECC1} \quad \text{ou} \quad Mé = a + \frac{(b-a) \left(\frac{1}{2} - FCC1 \right)}{FCC2 - FCC1}$$

ECC « effectif cumulé croissant » FCC « Fréquence cumulée croissante »

6.1.3.- La moyenne

Il caractérise un individu moyen représentatif du nuage de données. Il existe 4 moyennes : arithmétique ; géométrique ; harmonique et quadratique.

Soient les caractères $x_1, x_2, x_3, \dots, x_n$ et les effectifs $n_1 ; n_2 ; n_3 ; \dots ; n_n$

N.B.: Dans le cas des variables continues de classe $[a, b[$ Centre $x_i = \frac{a+b}{2}$

La moyenne arithmétique $\bar{X}$

$$\bar{X} = \frac{\sum n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \dots + n_k x_k}{N}$$

$$\bar{X} = \frac{\sum x_i}{N} = \frac{x_1 + x_2 + x_3 + \dots + x_k}{N} \quad \text{si } n_1 = n_2 = n_3 = \dots = n_k = 1$$

Propriétés des moyennes

Si $Z = aX \rightarrow \bar{Z} = a\bar{X}$

Si $Z = \frac{X-a}{b} \rightarrow \bar{Z} = \frac{\bar{X}-a}{b}$

Si $Z = aX + bY \rightarrow \bar{Z} = a\bar{X} + b\bar{Y}$

La moyenne géométrique $\bar{X}_G$

$$\bar{X}_G = \sqrt[n]{x_1^{n_1} * x_2^{n_2} * x_3^{n_3} * \dots * x_k^{n_k}} \rightarrow \ln \bar{X}_G = \frac{\sum n_i \ln x_i}{N} = a \rightarrow \bar{X}_G = e^a$$

$$\bar{X}_G = \sqrt[n]{x_1 * x_2 * x_3 \dots * x_k} \text{ si } n_1 = n_2 = n_3 = \dots = n_k = 1$$

Propriété :

Si $Z = aX$ la moyenne géométrique de $\bar{Z}_G = a\bar{X}_G$

Si $Z = aX^\alpha$ la moyenne géométrique de $\bar{Z}_G = a\bar{X}_G^\alpha$

Moyenne harmonique $\bar{X}_H$

$$\bar{X}_H = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \dots + \frac{n_k}{x_k}} = \frac{N}{\sum \frac{n_i}{x_i}}$$

$$\bar{X}_H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_k}} \text{ si } n_1 = n_2 = n_3 = \dots = n_k = 1$$

Retenons :

Si $Z = \frac{1}{X}$ la moyenne harmonique de $X : \bar{X}_H = \frac{1}{\bar{Z}}$

Moyenne quadratique $\bar{X}_q$

$$\bar{X}_q = \sqrt{\frac{n_1 x_1^2 + n_2 x_2^2 + n_3 x_3^2 + \dots + n_k x_k^2}{N}} = \sqrt{\frac{\sum n_i x_i^2}{N}}$$

$$\bar{X}_q = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \dots + x_k^2}{N}} = \sqrt{\frac{\sum x_i^2}{N}} \text{ si } n_1 = n_2 = n_3 = \dots = n_k = 1$$

Exemple 1 : Soient trois réels a, b, c . Calculer les 4 moyennes

Moyenne arithmétique $\bar{X} = \frac{a+b+c}{3}$

Moyenne géométrique $\bar{X}_G = \sqrt[3]{a.b.c}$

Moyenne harmonique $\bar{X}_H = \frac{3}{\frac{1}{a} + \frac{1}{b} + \frac{1}{c}}$

Moyenne quadratique $\bar{X}_Q = \sqrt{\frac{a^2+b^2+c^2}{3}}$ 5, 7, 9

Exemple 2 : soit la série statistique

Caractère xi	1	2	3	4	Total
Effectifs ni	6	4	8	2	20

Calculer la fréquence, l'effectif cumulé croissant puis décroissant

Calculer le Mode, la médiane et les 4 moyennes

Solution

Caractère xi	1	2	3	4	Total
Effectifs ni	6	4	8	2	20
Fréquence $f_i = \frac{n_i}{N}$	0,30	0,20	0,40	0,10	1,00
Effectif cumulé croissant ECC	6	10	18	20	XXX
Effectif cumulé décroissant ECD	20	14	10	2	XXX

Fréquence $f_i = \frac{\text{Effectif}}{\text{Effectif total}} = \frac{n_i}{N} \rightarrow$ somme des fréquences = 1

ECC « on fait la somme des effectifs du 1^{er} chiffre au dernier »

ECD « on fait la somme des effectifs du dernier chiffre au 01^{er} chiffre »

Mode : l'effectif le plus élevé 8 alors la médiane Mé = Xi = 3

Médiane : elle correspond à xi (Mé = xi) qui divise la série en deux parties égales.

- Effectif total N = 20 $\rightarrow$ la moitié $\frac{N}{2} = \frac{20}{2} = 10 \rightarrow$ la Mé est entre le 10^{ème} et 11^{ème} effectif.
- A partir de l'effectif cumulé croissant 10^{ème} : X₁ = 2 et 11^{ème} : X₂ = 3
 $\rightarrow$ la médiane $\frac{X_1 + X_2}{2} = \frac{2+3}{2} = 2,5$

Moyenne arithmétique $\bar{X} = \frac{\sum Ni \cdot Xi}{N} = \frac{6 \cdot 1 + 4 \cdot 2 + 8 \cdot 3 + 2 \cdot 4}{20} = \frac{46}{20} = 2,3$

Moyenne géométrique $\bar{X}_G = \sqrt[n]{\prod X_i^{n_i}} = \sqrt[20]{1^6 \times 2^4 \times 3^8 \times 4^2} = 2,05$

Moyenne harmonique $\bar{X}_H = \frac{n}{\sum \frac{n_i}{x_i}} = \frac{20}{\frac{6}{1} + \frac{4}{2} + \frac{8}{3} + \frac{2}{4}} = 1,79$

Moyenne quadratique $\bar{X}_q = \sqrt{\frac{\sum n_i \cdot x_i^2}{N}} = \sqrt{\frac{6 \cdot 1^2 + 4 \cdot 2^2 + 8 \cdot 3^2 + 2 \cdot 4^2}{20}} = 2,51$

Exemple 3 : soit la série statistique

Caractère xi	1	2	3	4	Total
Effectifs ni	10	5	7	13	35

Calculer la fréquence, l'effectif cumulé croissant puis décroissant

Calculer le Mode, la médiane et les 4 moyennes

Solution

Caractère xi	1	2	3	4	Total
Effectifs ni	10	5	7	13	35
Fréquences Fi	0,29	0,14	0,20	0,37	1
Effectif cumulé croissant ECC	10	15	22	35	XXX
Effectif cumulé décroissant ECD	35	25	20	13	

Mode : l'effectif le plus élevé 13 alors la médiane Mé = Xi = 4

Médiane : elle correspond à xi (Mé = xi) qui divise la série en deux parties égales.

- Effectif total N = 35 → la moitié $\frac{N}{2} = \frac{35}{2} = 17,5 \sim 18$ → la Mé est la valeur du 18^{ème} effectif.

A partir de l'effectif cumulé croissant 18 se trouve pour la 1^{ère} fois dans 22 alors la médiane est 3 → Mé = 3

Calcul des moyennes à partir du tableau

Xi	1	2	3	4	Total	Moyennes
Effec ni	10	5	7	13	35	
Ni* Xi	10	10	21	52	93	$\bar{X} = \frac{\sum Ni * Xi}{N} = \frac{93}{35} = 2,65$
Ni * Xi ²	10	20	63	208	301	$\bar{X}q = \sqrt{\frac{\sum ni * x_i^2}{N}} = \sqrt{\frac{301}{35}} = 2,93$
Ni* ln Xi	0	3,466	7,690	18,02	29,176	$\ln \bar{X}G = \frac{\sum Ni \ln Xi}{N} = \frac{29,176}{35}$ $\ln \bar{X}G = 0,8336 \rightarrow \bar{X}G = 2,30$
$\frac{Ni}{Xi}$	10	2,5	2,333	3,250	18,083	$\bar{X}H = \frac{n}{\sum \frac{ni}{xi}} = \frac{35}{18,083} = 1,94$

Exemple 4 : soit la série statistique

Classe	[10 ; 20[	[20 ; 30[	[30 ; 40[	[40 ; 50[	Total
Centre Xi	15	25	35	45	

Calculer l'effectif cumulé croissant

Calculer le Mode, et la médiane

Calculer le centre Xi et les 4 moyennes

Solution

Classe	[10 ; 20[	[20 ; 30[	[30 ; 40[	[40 ; 50[	Total
Centre Xi	15	25	35	45	
Effectifs ni	1	6	9	5	21
ECC	1	7	16	21	XXXXX

ECC « on fait la somme des effectifs du 1^{er} chiffre au dernier »

Centre de classe [10 ; 20[→ $x = (10 + 20) / 2 = 15$

Mode : l'effectif le plus élevé 9 → classe modale est [30 ; 40[et $Mo = Xi = 35$

Médiane : elle correspond à xi ($Mé = xi$) qui divise la série en deux parties égales.

- Effectif total $N = 21$ → la moitié $\frac{N}{2} = \frac{21}{2} = 10,5$

A partir de l'ECC 10,5 est dans 16 alors la classe médiane [30 ; 40[:

$30 < Mé < 40$ et $7 < 10,5 < 16$ par extrapolation

$$\frac{Mé-30}{40-30} = \frac{10,5-7}{16-7} \Rightarrow Mé = 33,88$$

Rappelons la formule $X < Y < Z$ et $A < B < C$ alors $\frac{Y-X}{Z-X} = \frac{B-A}{C-A}$

La formule de la médiane

$$a < Mé < b \text{ et } ECC1 < \frac{N}{2} < ECC2 \Rightarrow \frac{Mé-a}{b-a} = \frac{\frac{N}{2}-ECC1}{ECC2-ECC1}$$

$$Mé = a + \frac{(b-a) \left(\frac{N}{2} - ECC1 \right)}{ECC2 - ECC1}$$

Calcul des moyennes à partir du tableau

Xi	15	25	35	45	Total	Moyennes
Effec ni	1	6	9	5	21	
Ni* Xi	15	150	315	225	705	$\bar{X} = \frac{\sum Ni \cdot Xi}{N} = \frac{705}{21} = 33,57$
Ni * Xi ²	225	3750	11025	10125	25125	$\bar{X}q = \sqrt{\frac{\sum n_i \cdot x_i^2}{N}} = \sqrt{\frac{25125}{21}} = 34,58$
Ni* ln Xi	2,708	19,313	31,998	19,033	73,052	$\ln \bar{X}G = \frac{\sum Ni \ln Xi}{N} = \frac{73,052}{21}$ $\ln \bar{X}G = 3,479 \Rightarrow \bar{X}G = 32,41$

Exemple 5 : On donne la répartition des élèves suivants, le nombre de frères

Nombre de frères X_i	1	2	3	4	5	Total
Effectifs N_i	10	5	15	7	8	45
Effectifs cumulés ECC	10	15	30	37	45	XXXXX

TAF:

- 1.-) Quelle est la nature de la variable statistique?
- 2.-) Déterminer le mode et la médiane
- 3.-) Calculer la moyenne arithmétique et l'écart type.

Réponse :

La population étudiée : les élèves

Le caractère étudié : le nombre de frère, quantitatif discret (variable statistique discrète)

Mode : l'effectif le plus élevé est 15 alors le mode = 3.

Médiane : effectif total $N = 45 \rightarrow$ moitié $N/2 = 22,5$ soit 23 $\rightarrow$ à partir de l'effectif cumulé croissant 23 appartient à 30 alors la médiane $Mé = 3$.

Calcul de moyenne et l'écart type

Nombre de frères X_i	1	2	3	4	5	Total
Effectifs N_i	10	5	15	7	8	45
$N_i \times X_i$	10	10	45	28	40	133
$N_i \times X_i^2$	10	20	135	112	200	477

$$\bar{X} = \frac{\sum N_i \cdot X_i}{N} = \frac{133}{45} = 2,95 \quad \text{La variation} \quad V(X) = \frac{\sum N_i \cdot X_i^2}{N} - (\bar{X})^2 = 1,8975$$

$$\text{L'écart type} \quad \sigma_X = \sqrt{V(X)} = \sqrt{1,8975} = 1,38$$

Exemple 6: On donne la répartition des élèves suivant la taille en cm

Taille en cm	150-155	155-160	160-165	165-175	175-190
Effectif	17	10	18	20	15

TAF:

- 1.-) Quelle est la nature de la variable statistique?
- 2.-) Déterminer le mode et la médiane
- 3.-) Calculer la moyenne arithmétique et l'écart type.

Solution

Taille en cm	150-155	155-160	160-165	165-175	175-190	Total
Effectif	17	10	18	20	15	80
Fréquences	21,25%	12,5%	22,5%	25%	18,75%	100%
Fréquences cumulées	21,25%	33,75%	56,25%	81,25%	100%	XXX
Effectif rectifiés	$\frac{17}{1} = 17$	$\frac{10}{1} = 10$	$\frac{18}{1} = 18$	$\frac{20}{2} = 10$	$\frac{15}{3} = 5$	XXX

Comment avoir l'effectif rectifié ?

L'amplitude de la classe $[a, b[$ est $b - a$.

La plus petite amplitude est 5. Donc les effectifs qui ne correspondent pas à la plus petite amplitude doivent être corrigés :

$[165 - 175[$ amplitude est égale à 10 = 2 fois la plus petite amplitude alors l'effectif doit être divisé par 2.

$[175 - 190[$ amplitude est égale à 15 = 3 fois la plus petite amplitude alors l'effectif doit être divisé par 3.

Le mode: la classe modale $[160, 165[$, correspond à l'effectif rectifié le plus élevé.

Classe modale $[160, 165[$, le mode = $\frac{160+165}{2} = 162,5$ cm

$$M\acute{e} = a + \frac{(b-a) \left(\frac{1}{2} - FCC1 \right)}{FCC2 - FCC1} = 160 + \frac{(165-160)(50\% - 33,75\%)}{56,25\% - 33,75\%} = 163,61 \text{ cm}$$

6.2.- Paramètres de dispersion

Les paramètres de dispersion (étendue, intervalle interquartile,) sont calculés pour les variables statistiques quantitatives.

Ils ne donnent pas une information complète sur une variable statistique X : en effet, deux variables qui ont la même moyenne peuvent se présenter avec des dispersions très différentes.

L'histogramme, ou le diagramme, des fréquences donnent déjà une idée qualitative de la dispersion.

Etendue

Soit X une variable statistique réelle discrète.

L'étendue ω de X est la différence entre la plus grande valeur de X et la plus petite valeur de X .

Quartiles et déciles.

Pour une variable statistique quantitative réelle continue X , on appelle **quartiles** les nombres réels Q_1 , Q_2 , Q_3 , pour lesquels les fréquences cumulées de X sont respectivement 0,25, 0,50, 0,75.

$$\begin{aligned} \text{1}^{\text{er}} \text{ quartile } Q_1 &= a + \frac{(b-a) \left(\frac{N}{4} - ECC1 \right)}{ECC2 - ECC1} \quad \text{ou} \quad Q_1 = a + \frac{(b-a) (0,25 - FCC1)}{FCC2 - FCC1} \\ \text{2}^{\text{ème}} \text{ quartile } Q_2 &= a + \frac{(b-a) \left(\frac{N}{2} - ECC1 \right)}{ECC2 - ECC1} \quad \text{ou} \quad Q_2 = a + \frac{(b-a) (0,50 - FCC1)}{FCC2 - FCC1} \\ \text{3}^{\text{ème}} \text{ quartile } Q_3 &= a + \frac{(b-a) \left(\frac{3N}{4} - ECC1 \right)}{ECC2 - ECC1} \quad \text{ou} \quad Q_3 = a + \frac{(b-a) (0,75 - FCC1)}{FCC2 - FCC1} \end{aligned}$$

Ecart absolu moyen.

On appelle **écart absolu moyen** de X la moyenne arithmétique des valeurs absolues des écarts de X à sa moyenne. $E = \frac{\sum (X_i - \bar{X})}{N}$

Variance et écart-type.

On appelle **variance** de X , la moyenne arithmétique des carrés des écarts de X à sa moyenne. On appelle **écart-type** de X la racine carrée σ_X de la variance de X .

$$V(X) = \frac{\sum n_i (x_i - \bar{x})^2}{N} \rightarrow V(X) = \frac{\sum N i \cdot x_i^2}{N} - (\bar{X})^2 \rightarrow \text{L'écart type } \sigma_X = \sqrt{V(X)}$$

$$\text{La moyenne arithmétique } \bar{X} = \frac{\sum N i \cdot x_i}{N}$$

$$\text{La variation } V(X) = \frac{\sum N i \cdot x_i^2}{N} - (\bar{X})^2 \rightarrow \text{L'écart type } \sigma_X = \sqrt{V(X)}$$

$V(X)$ sensiblement égale à 0 $\rightarrow$ les valeurs sont proches les unes des autres

$V(X)$ élevée $\rightarrow$ les valeurs sont très écartées

$V(X)$ égale à 0 $\rightarrow$ les valeurs sont identiques

Coefficient de variation.

Pour une variable statistique réelle X , on appelle **coefficient de variation** le rapport $c = \frac{\sigma_X}{\bar{X}}$.

Le coefficient de variation est un nombre sans dimension qui permet de comparer deux variables statistiques de natures différentes.

Moments.

Soit X une variable statistique quantitative réelle.

On appelle **moment centré d'ordre r** de X , la quantité : $\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N}$

On appelle **moment non centré d'ordre r** de X , la quantité : $m_r = \frac{\sum n_i (x_i)^r}{N}$

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N} \rightarrow \mu_2 = \frac{\sum n_i (x_i - \bar{x})^2}{N} = V(X)$$

$$m_r = \frac{\sum n_i (x_i)^r}{N} \rightarrow m_1 = \frac{\sum n_i (x_i)^1}{N} = \bar{X}$$

Exemple :

Considérons la variable statistique X de distribution :

Xi	-1	4	Total
Ni	4	1	5

Calculer μ_2 et μ_3

Solution : la moyenne $\bar{X} = \frac{(-1)(4)+4*1}{5} = 0$

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N} \rightarrow \mu_2 = \frac{\sum n_i (x_i - \bar{x})^2}{N} = \frac{4(-1-0)^2 + 1(4-0)^2}{5} = 4$$

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N} \rightarrow \mu_3 = \frac{\sum n_i (x_i - \bar{x})^3}{N} = \frac{4(-1-0)^3 + 1(4-0)^3}{5} = 12$$

Exemple :

Considérons la variable statistique X de distribution :

Xi	-4	1	Total
Ni	1	4	5

Calculer μ_2 et μ_3

Solution : la moyenne $\bar{X} = \frac{(1)(-4)+4*1}{5} = 0$

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N} \rightarrow \mu_2 = \frac{\sum n_i (x_i - \bar{x})^2}{N} = \frac{1(-4-0)^2 + 4(1-0)^2}{5} = 4$$

$$\mu_r = \frac{\sum n_i (x_i - \bar{x})^r}{N} \rightarrow \mu_3 = \frac{\sum n_i (x_i - \bar{x})^3}{N} = \frac{1(-4-0)^3 + 4(1-0)^3}{5} = 13,6$$

6.3.- Paramètres de forme

Nous définissons les paramètres de forme pour une variable statistique quantitative, discrète ou continue, à valeurs réelles.

Coefficient d'asymétrie.

Il existe plusieurs coefficients d'asymétrie. Les principaux sont les suivants.

Le coefficient d'asymétrie de **Pearson** fait intervenir le mode M_0 : quand il existe, il est défini par : $P = \frac{\bar{X} - M_0}{\sigma_x}$

Le coefficient d'asymétrie de **Yule** fait intervenir la médiane et les quartiles, il est défini par

$$Y = \frac{Q_1 + Q_3 - 2 M_e}{2 (Q_3 - Q_1)}$$

Le coefficient d'asymétrie de **Fisher** fait intervenir les moments centrés, il est défini par

$$F = \frac{\mu_3}{\sigma^3}$$

Lorsque le coefficient d'asymétrie est positif, la distribution est plus étalée à droite : on dit qu'il y a **oblicité à gauche**.

Lorsque le coefficient d'asymétrie est négatif, la distribution est plus étalée à gauche : on dit qu'il y a **oblicité à droite**.

On utilise souvent un coefficient d'asymétrie de **Pearson basé sur les moments**

centrés : $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

Ce coefficient d'asymétrie est toujours positif.

Il est nul pour une distribution à densité de fréquence symétrique, telle la loi de Gauss.

Exemple : Considérons la variable statistique X de distribution :

Xi	-1	4	Total
Ni	4	1	5

Calculer M_0 ; μ_2 ; μ_3 ; P ; F et β_1

Solution : $\bar{X} = 0$; $\sigma_x = 1,84 \Rightarrow \sigma_x^3 = 6,27$

$$\mu_2 = 4 ; \quad M_0 = -1 ; \quad \mu_3 = 12 ; \quad \mu_2^3 = 64 ; \quad \mu_3^2 = 144$$

$$P = \frac{\bar{X} - M_0}{\sigma_x} = \frac{0 - (-1)}{1,84} = 0,54 \quad F = \frac{\mu_3}{\sigma^3} = \frac{12}{6,27} = 1,97$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{144}{64} = 2,25$$

Coefficient d'aplatissement.

Là encore plusieurs définitions sont possibles.

Le coefficient d'aplatissement de **Pearson** est $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Le coefficient d'aplatissement de **Yule** est $F_2 = \frac{\mu_4}{\mu_2^2} - 3$.

On peut se demander pourquoi -3 ?

C'est parce que, en Probabilités, on peut démontrer que le coefficient d'aplatissement de Pearson pour une variable aléatoire réelle qui suit une loi de Gauss, est égal à 3.

Il est alors naturel, pour comparer l'aplatissement d'une distribution statistique à l'aplatissement d'une variable de Gauss, d'introduire le coefficient $F_2 = \beta_2 - 3$.

Si F_2 est égal à 0, le polygone statistique de la variable réduite a le même aplatissement qu'une courbe en cloche, on dit que la variable est **mésokurtique**.

Si F_2 est > 0 , le polygone statistique de la variable réduite est moins aplati qu'une courbe en cloche, on dit que la variable est **leptokurtique**.

Si F_2 est < 0 , le polygone statistique de la variable réduite est plus aplati qu'une courbe en cloche, on dit que la variable est **platykurtique**.

Indice de concentration de Gini.

L'indice de Gini permet de mesurer les inégalités scolaires, les inégalités de statut, les inégalités de salaires, etc.

Indice de GINI

$$g = 2S = 1 - \sum_{i=1}^{i=n-1} (P_{i+1} - P_i)(q_{i+1} + q_i) = 1 - \sum_{i=1}^{i=n-1} f_{i+1}(q_{i+1} + q_i)$$

On note p_i la fréquence cumulée de $e_i + 1$: c'est la proportion de salariés dont le salaire est strictement plus petit que $e_i + 1$.

On note q_i la proportion de masse salariale représentée par les salariés dont le salaire est strictement $q_i = \sum_{k=1}^{i=n-1} f_k \frac{x_k}{\bar{X}}$

Dire que $g = 0$, c'est dire que la courbe de Lorenz coïncide avec la diagonale du carré (égalité absolue).

Dire que $g = 1$, c'est dire que la courbe de Lorenz longe d'abord l'axe des p , puis la droite $p=1$ (inégalité maximale).

Médiale

La **médiale** d'une variable statistique X est la valeur de X qui partage la masse globale en deux parties égales.

Chapitre 2 : Distribution statistique à deux variables x et y

Considérons une population finie sur laquelle nous étudions deux caractères (qualitatifs ou quantitatifs réels) A et B .

Le couple (X, Y) est une **variable conjointe** : c'est une variable statistique à deux dimensions si l'on en élimine les modalités conjointes (A_i, B_j) dont l'effectif est nul.

Une telle variable statistique à deux dimensions peut se représenter par un tableau à double entrée appelé **tableau de contingence**.

1.- Variables marginales. Variables conditionnelles.

Admettons le tableau à double entrée des caractères X et Y

Y_j				
X_i	Y_1	Y_2	Y_3	Total marginal de X
X_1	N_{11}	N_{12}	N_{13}	$N_{1\bullet}$
X_2	N_{21}	N_{22}	N_{23}	$N_{2\bullet}$
Total marginal de Y	$N_{\bullet 1}$	$N_{\bullet 2}$	$N_{\bullet 3}$	N

N_{ij} « effectif de X_i et Y_j » N « effectif total »

Il y a deux distributions marginales : X et Y

Il y a cinq distributions conditionnelles car on a trois colonnes et deux lignes : $3 + 2 = 5$

1°) Les deux distributions marginales sont

Distribution marginale de x	
X_i	$N_{i\bullet}$
X_1	$N_{1\bullet}$
X_2	$N_{2\bullet}$
Total	N

Distribution marginale de y			
Y_j	$N_{\bullet j}$		
Y_1	$N_{\bullet 1}$		
Y_2	$N_{\bullet 2}$		
Y_3	$N_{\bullet 3}$		
Total	N		

2°) Les Cinq distributions conditionnelles :

* Distribution conditionnelle de X par Y_1

X_i	X_1	X_2	Total
Effectif	N_{11}	N_{21}	

* Distribution conditionnelle de X par Y_2

X_i	X_1	X_2	Total
Effectifs	N_{12}	N_{22}	

* Distribution conditionnelle de X par Y_3

X_i	X_1	X_2	Total
Effectif	N_{13}	N_{23}	

* Distribution conditionnelle de Y par X_1

Y_j	Y_1	Y_2	Y_3	Total
Effectif	N_{11}	N_{12}	N_{13}	

* Distribution conditionnelle de Y par X_2

Y_j	Y_1	Y_2	Y_3	Total
Effectif	N_{21}	N_{22}	N_{23}	

Exemple 1 : Soit Z une variable statistique à deux dimensions X, Y :

Y	$Y_1 = 2$	$Y_2 = 3$	Total de distribution marginale de X
X			
$X_1 = 0$	7	2	9
$X_2 = 1$	6	5	11
Total de distribution de Y	13	7	20

Variables marginales.

Les variables statistiques X et Y ainsi définies sont appelées les **variables marginales** de Z.

Variables marginales de Y

Y _j	Effectif
2	13
3	7
Total	20

Variables marginales de X

X _i	Effectif
0	9
1	11
Total	20

* Distribution conditionnelle de X par Y₁

X _i	0	1	Total
Effectif	7	6	13

* Distribution conditionnelle de X par Y₂

X _i	0	1	Total
Effectifs	2	5	7

* Distribution conditionnelle de Y par X₁

Y _j	2	3	Total
Effectif	7	2	9

* Distribution conditionnelle de Y par X₂

Y _j	2	3	Total
Effectif	6	5	11

2.- Calcul de la covariance Cov (x,y) : $Cov (X, Y) = \frac{\sum XY}{n} - \bar{X} * \bar{Y}$

Pour une variable statistique quantitative Z à deux dimensions, de variables marginales X et Y, on définit la covariance de X et Y par l'expression :

$$Cov (X, Y) = \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y})$$

$$Cov (X, Y) = \frac{\sum XY}{n} - \bar{X} * \bar{Y}$$

Si la Cov (X,Y) = 0 alors X et Y sont indépendantes ;

Si la Cov (X, Y) > 0 alors X et Y varient dans le même sens.

Si la Cov (X, Y) < 0 alors X et Y ne varient pas dans le même sens.

On pose $Z = aX + bY \rightarrow$ la variance de Z : $V(Z) = a^2V(X) + b^2V(Y) + 2abCov (X, Y)$

3.- Le Coefficient de corrélation linéaire :

Le coefficient de corrélation mesure l'intensité d'une relation linéaire entre deux variables quantitatives.

$$R_{xy} = \frac{Cov(X,Y)}{\sigma_x * \sigma_y}$$

Le Coefficient de corrélation est compris entre -1 et 1 : $-1 \leq R_{xy} \leq 1$

Si $r_{xy} = 1$ alors X et Y sont fortement corrélées ;

Si $r_{xy} = 0$ alors X et Y sont indépendantes.

4.- Le Coefficient de détermination R^2

Le coefficient de corrélation linéaire $r_{xy} = \frac{Cov(x;y)}{\sigma_x * \sigma_y}$

Le coefficient de détermination $R^2 = \frac{[Cov(x;y)]^2}{V(x)*V(y)} = r_{xy}^2$

Soit le modèle Y en fonction de X : $Y = aX + b$

Le coefficient de détermination $R^2 : 0 \leq R^2 \leq 1$

R^2 est un indicateur de qualité de l'ajustement linéaire entre les variables explicative et expliquée.

Si $R^2 = 1$, on dira que les variations de la variable indépendante sont parfaitement expliquées par la ou les variables indépendantes.

Si $R^2 = 0$, on dira que les variations de la variable dépendante ne sont pas expliquées par la variable indépendante.

Interprétation du coefficient de détermination

Autrement dit le coefficient de détermination r^2 représente la fraction de la variance de y « expliquée » par la corrélation de y avec x.

Par exemple un coefficient de corrélation $r = 0,9$ correspond à un coefficient de détermination $r^2 = 0,81$. Cela signifie que 81% de la variance de y est expliquée par la corrélation : **la corrélation est bonne.**

Mais un coefficient de corrélation $r = 0,5$ correspond à un coefficient de détermination $r^2 = 0,25$.

Cela signifie que seulement 25% de la variance de y est expliquée par la corrélation : **la corrélation est mauvaise.**

4.- Droite de régression

L'équation de la droite de régression de Y en fonction de X

$$\rightarrow y = ax + b \text{ avec } a = \frac{\text{Cov}(X,Y)}{V(X)} = \frac{\sum XY - n \bar{X} \bar{Y}}{\sum X^2 - n(\bar{X})} \text{ et } b = \bar{y} - a \bar{x}$$

La connaissance d'une modalité de la variable X apporte-t-elle une information supplémentaire sur les modalités de la variable Y ?

La réponse à cette question est du domaine de la **régression** : dans un tel cas, on dit que X est la variable explicative et Y la variable expliquée.

Exercice 1 : soit le tableau suivant

X_i	1	0	1	2	1	2	0	1	0	1
Y_i	0	1	2	0	1	1	0	2	1	0

Calculer $\bar{X}$; $\bar{Y}$; $V(x)$; $V(y)$; $\text{Cov}(X,Y)$ et r_{xy} .

Solution 1 :

											Total
X_i	1	0	1	2	1	2	0	1	0	1	9
Y_i	0	1	2	0	1	1	0	2	1	0	8
X^2	1	0	1	4	1	4	0	1	0	1	13
Y^2	0	1	4	0	1	1	0	4	1	0	12
XY	0	0	2	0	1	2	0	2	0	0	7

$$\text{Moyenne } \bar{X} = \frac{\sum X}{N} = \frac{9}{10} = 0,9 \quad \bar{Y} = \frac{\sum Y}{N} = \frac{8}{10} = 0,8$$

$$V(X) = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2 = \frac{13}{10} - \left(\frac{9}{10}\right)^2 = 0,49 \rightarrow \sigma_X = \sqrt{0,49} = 0,7$$

$$V(Y) = \frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2 = \frac{12}{10} - \left(\frac{8}{10}\right)^2 = 0,56 \rightarrow \sigma_Y = \sqrt{0,56} = 0,75$$

$$\text{Cov}(X,Y) = \frac{\sum XY}{n} - \bar{X} * \bar{Y} = \frac{7}{10} - 0,9 * 0,8 = -0,02$$

$$R_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X * \sigma_Y} = -\frac{0,02}{0,7 * 0,75} = -0,04$$

Exercice 2 : soit le tableau suivant

Yj	1	2	Total
Xi			
1	3	2	5
2	1	0	1
3	1	1	2
Total	5	3	8

1°) Calculer $\bar{X}$; $\bar{Y}$; $V(x)$ et $V(y)$

2°) Calculer Cov (x,y) et coefficient de corrélation R x y

Solution 2 :

Distribution marginale de X

X	Ni	Xi * Ni	Ni * Xi²	$\bar{X} = \frac{\sum Ni * Xi}{N} = \frac{13}{8} = 1,625$ $V(X) = \frac{27}{8} - (1,625)^2 = 0,73$
1	5	5	5	
2	1	2	4	
3	2	6	18	
Total	8	13	27	

Distribution marginale de Y

Yj	Ni	Yj * Ni	Ni * Yj²	$\bar{Y} = \frac{\sum Ni * Yj}{N} = \frac{11}{8} = 1,375$ $V(Y) = \frac{17}{8} - (1,375)^2 = 0,23$
1	5	5	5	
2	3	6	12	
Total	8	11	17	

Calcul de la covariance

Yj	1	2	Total	$\sum N_{ij} X_i Y_j = 1 \times 3 \times 1 + 1 \times 2 \times 2 +$ $2 \times 1 \times 1 + 2 \times 0 \times 2 +$ $3 \times 1 \times 1 + 3 \times 1 \times 2 = 18$ $Cov(X, Y) = \frac{\sum N_{ij} X_i Y_j}{n} - \bar{X} \bar{Y}$ $Cov(X, Y) = -0,109375$ $R_{xy} = -0,2636$
1	3	2	5	
2	1	0	1	
3	1	1	2	
Total	5	3	N= 8	

Module 3: Indices

- ✓ Un **indice simple** n'étudie que les variations d'un seul produit,
- ✓ Un **indice composé** prend en compte l'étude de plusieurs produits.

I.- Indice élémentaire

1.1.- Définition de l'indice élémentaire

L'indice élémentaire d'une grandeur G_t par rapport à une grandeur G_k est notée

$I_{t/k} = \frac{G_t}{G_k} \times 100$. (base 100) ou $I_{t/k} = \frac{G_t}{G_k}$ (base 1). Cet indice est dit simple ou élémentaire, car il ne porte que sur une seule grandeur.

1.2.- Calcul de l'indice élémentaire

Indice élémentaire année t par rapport à l'année k : $I_{t/k} = \frac{G_t}{G_k} * 100$

Par exemple : Le prix du mil passe de 250F à 300F de 2000 à 2001. L'indice prix de 2001 base 2000 est $= \frac{300}{250} * 100 = 120$ veut dire « le prix du mil a augmenté de 20% de 2000 à 2001 ».

1.3- Distinction des propriétés de l'indice élémentaire

PP1 : l'indice n'est pas exprimé en pourcentage ;

PP2 : l'indice peut être calculé sur la base 100 ou sur la base 1 ;

Le prix d'un bien est de 140 € en 2010. Le prix passe à 150 € à 2011. Calculer l'indice 2011 base 2010 ?

Indice base 100 : $I_{11/10} = \frac{150}{140} * 100 = 107$

Indice base 1 : $I_{11/10} = \frac{150}{140} * 1 = 1,07$

PP3 : L'indice supérieur à 100 veut dire que la grandeur augmente par rapport à la période base ;

Exemple :

Le prix augmente de 20% de 2005 à 2006, calculer l'indice $I_{06/05}$? $I_{06/05} = 100 + 20 = 120$

Le prix augmente de 5% de 2011 à 2012, calculer l'indice $I_{12/11}$? $I_{12/11} = 100 + 5 = 105$

PP4 : L'indice inférieur à 100 veut dire que la grandeur diminue par rapport à la période base ;

Exemple :

Le prix diminue de 20% de 2005 à 2006, calculer l'indice $I_{06/05}$? $I_{06/05} = 100 - 20 = 80$

Le prix diminue de 5% de 2011 à 2012, calculer l'indice $I_{12/11}$? $I_{12/11} = 100 - 5 = 95$

PP5 : le taux de variation = indice -100; ce taux est exprimé en %.

Exemple :

Indice = 140 veut dire que le taux de variation est de $(140 - 100) : 40\%$

Indice = 90 veut dire que le taux de variation est de $(90 - 100) : -10\%$

PP6 : la réversibilité d'un indice : $I_{t/k} = \frac{100^2}{I_{k/t}}$

Si une grandeur numérique V prend les valeurs Vo et Vt aux instants o, t

alors : $I_{o/t} = \frac{100^2}{I_{t/o}}$

Exemple : Le prix du pain augmente de 40% de 2008 à 2009. Quelle est l'évolution du prix de 2009 à 2008 ?

Réponse : indice de 2008 à 2009 $I_{09/08} = 100 + 40 = 140$

$I_{08/09} = \frac{100}{I_{09/08}} = \frac{100^2}{140} = 71$ soit une diminution de $(71-100) : -29\%$.

Exemple : Le prix du pain diminue de 10% de 2008 à 2009. Quelle est l'évolution du prix de 2009 à 2008 ?

Réponse : indice de 2008 à 2009 $I_{09/08} = 100 - 10 = 90$

$I_{08/09} = \frac{100}{I_{09/08}} = \frac{100^2}{90} = 111$ soit une augmentation de $(111-100) : 11\%$.

PP7 : Transférabilité d'un indice : $I_{p/k} \times I_{t/p} \times \frac{1}{100} = I_{t/p}$

Si une grandeur numérique V prend les valeurs Vo, Vt et Vt' aux instants o, t, t' alors : $I_{t/o} = I_{t/t'} \times I_{t'/t} \times \frac{1}{100}$

Exemple :

Le prix augmente de 30% de 2010 à 2011 et diminue de 10% de 2011 à 2012. Quelle est l'évolution du prix de 2010 à 2012 ?

Le prix augmente de 30% de 2010 à 2011 → indice $I_{11/10} = 130$

Le prix diminue de 10% de 2011 à 2012 → indice $I_{12/11} = 90$

$I_{11/10} \times I_{12/11} \times \frac{1}{100} = I_{12/10} \Rightarrow 130 \times 90 \times \frac{1}{100} = 117 = I_{12/10}$ soit une augmentation de 17%.

Exemple :

Le prix augmente de 18% de 2010 à 2011 et augmente de 5% de 2011 à 2012. Quelle est l'évolution du prix de 2010 à 2012 ?

Le prix augmente de 18% de 2010 à 2011 → indice $I_{11/10} = 118$

Le prix augmente de 5% de 2011 à 2012 → indice $I_{12/11} = 105$

$I_{11/10} \times I_{12/11} \times \frac{1}{100} = I_{12/10} \Rightarrow 118 \times 105 \times \frac{1}{100} = 123,9 = I_{12/10}$ soit une augmentation de 23,9%.

II.- Les indices synthétiques

L'indice est dit synthétique lorsqu'il porte sur l'évolution simultanée de plusieurs grandeurs.

Il est utilisé pour déterminer l'influence des prix ou des quantités d'un panier de produits.

Un indice composé étudie les variations d'un ensemble de produits entre deux époques 0 et 1.

Les formules sont :

	Indice prix	Indice quantités
Laspeyres	$IL.P_{1/0} = \frac{\sum P_1 * Q_0}{\sum P_0 * Q_0}$	$IL.Q_{1/0} = \frac{\sum P_0 * Q_1}{\sum P_0 * Q_0}$
Paasches	$IP.P_{1/0} = \frac{\sum P_1 * Q_1}{\sum P_0 * Q_1}$	$IP.Q_{1/0} = \frac{\sum P_1 * Q_1}{\sum P_1 * Q_0}$

Exemple : En utilisant les renseignements consignés dans le tableau suivant:

Produits	Année 2000		Année 2001	
	Prix unitaire	Quantités	Prix unitaire	Quantités
A	10	30	14	25
B	20	25	26	25
C	50	15	60	30

TAF1 : Sur la base de 2000, calculez l'indice prix de 2001 (pondération Laspeyres) ;

L'indice synthétique des prix (pondération Laspeyres) : de 2001, base 100 (2000)

$$IL.P_{01/00} = \frac{\sum P_{01} * Q_{00}}{\sum P_{00} * Q_{00}} \times 100 = \frac{14 \times 30 + 26 \times 25 + 60 \times 15}{10 \times 30 + 20 \times 25 + 50 \times 15} \times 100 = 127$$

TAF2 : Sur la base de 2000, calculez l'indice quantité de 2001 (pondération Laspeyres) ;

L'indice synthétique des quantités (pondération Laspeyres) : de 2001, base 100 (2000)

$$IL.P_{01/00} = \frac{\sum P_{00} * Q_{01}}{\sum P_{00} * Q_{00}} \times 100 = \frac{10 \times 25 + 20 \times 25 + 50 \times 30}{10 \times 30 + 20 \times 25 + 50 \times 15} \times 100 = 145$$

TAF3 : Sur la base de 2000, calculez l'indice prix de 2001 (pondération Paasches) ;

L'indice synthétique des prix (pondération Paasche) : de 2001, base 100 (2000)

$$IP.P_{01/00} = \frac{\sum P_{01} * Q_{01}}{\sum P_{00} * Q_{01}} \times 100 = \frac{14 \times 25 + 26 \times 25 + 60 \times 30}{10 \times 25 + 20 \times 25 + 50 \times 30} \times 100 = 124$$

Travaux Dirigés de statistique

Remarque

Lorsqu'on multiplie les modalités X_i par m , alors :

Les moyennes $\bar{X}$; $\bar{XG}$; $\bar{XH}$; $\bar{XQ}$ seront multipliées par m .

Le Mode et la médiane seront multipliés par m .

La variance sera multipliée par m^2 .

Lorsqu'on multiplie les effectifs N_i par m , alors :

Les moyennes $\bar{X}$; $\bar{XG}$; $\bar{XH}$; $\bar{XQ}$ resterons invariables.

Le mode et la médiane resteront invariants.

La variance restera invariante.

Exercice 1 : on pose $\bar{X} = 5$

Calculer la moyenne de Z a.-) $Z = 3X$ b.-) $Z = \frac{X}{2}$ c.-) $Z = \frac{X-2}{5}$

Réponse :

a.-) $Z = 3X \rightarrow \bar{Z} = 3\bar{X} = 3 * 5 = 15$

b.-) $Z = \frac{X}{2} \rightarrow \bar{Z} = \frac{\bar{X}}{2} = \frac{5}{2} = 2,5$

c.-) $Z = \frac{X-2}{5} \rightarrow \bar{Z} = \frac{\bar{X}-2}{5} = \frac{5-2}{5} = 0,6$

Exercice 2 : on pose $V(X) = 7$

Calculer la variance de Z a.-) $Z = 3X$ b.-) $Z = \frac{X}{2}$ c.-) $Z = \frac{X-2}{5}$

Réponse : $V(aX + b) = a^2V(X)$

a.-) $V(Z) = 3^2V(X) = 9 * 7 = 63$

b.-) $V(Z) = \frac{1}{2^2} V(X) = \frac{1}{4} * 7 = 1,75$

c.-) $V(Z) = \frac{1}{5^2} * V(X) = \frac{1}{25} * 7 = 0,28$

Exercice 3 :

Soit la série : 12, 17, 10, 17, 15, 17, 16, 2, 18, 15, 13, 14, 18, 17, 16

1.-) Déterminer le mode, la médiane, la moyenne arithmétique de cette série

2.-) Déterminer l'écart-type et l'intervalle interquartile de cette série.

3.-) Construire le diagramme en bâtons, le polygone des fréquences, et la fonction.

Solution exercice 3

Soit la série : 12, 17, 10, 17, 15, 17, 16, 2, 18, 15, 13, 14, 18, 17, 16

Construire un tableau statistique en classant les X_i dans l'ordre :

2 ; 10 ; 12 ; 13 ; 14 ; 15 ; 15 ; 16 ; 16 ; 17 ; 17 ; 17 ; 17 ; 18 ; 18.

X_i	2	10	12	13	14	15	16	17	18	Total
N_i	1	1	1	1	1	2	2	4	2	15
ECC	1	2	3	4	5	7	9	13	15	XXX
$N_i * X_i$	2	10	12	13	14	30	32	68	36	217
$N_i * X_i^2$	4	100	144	169	196	450	512	1156	648	3379
Fréquence										

La variance $V(X) = \frac{3379}{15} - (14,46)^2 = 16,175 \rightarrow$ l'écart-type = 4,02

Le mode : 17 ; la médiane : 16 ; moyenne arithmétique = 14,46

Moy arith = $\frac{2+10+12+13+14+15+15+16+16+17+17+17+17+18+18}{15} = \frac{217}{15} = 14,46$

La variance $V(X) = \frac{3379}{15} - (14,46)^2 = 16,175 \rightarrow$ l'écart-type = 4,02

Médiane : effectif total 15 $\rightarrow 15 * 0,50\% = 7,5 \rightarrow$ suivant l'ECC Mé = 16

L'intervalle interquartile : $Q_3 - Q_1 = 12 - 10 = 2$

Q_1 : effectif total 15 $\rightarrow 15 * 25\% = 3,75 \rightarrow$ suivant l'ECC $Q_1 = 10$

Q_3 : effectif total 15 $\rightarrow 15 * 75\% = 11,25 \rightarrow$ suivant l'ECC $Q_3 = 12$

Exercice 4 : Indice

Vous disposez de l'évolution du prix de la viande

Années	2000	2001	2002	2003
Prix	1020F	1050F	1080F	1200F

Question 1 : Sur la base de 2000, calculez les indices de 2001 ; 2002 ; 2003.

Réponse :

Indice 2001 base 2000 = $\frac{\text{prix 2001}}{\text{prix 2000}} = \frac{1050}{1020} * 100 = 103$ soit une augmentation de 3%

Indice 2002 base 2000 = $\frac{\text{prix 2002}}{\text{prix 2000}} = \frac{1080}{1020} * 100 = 106$ soit une augmentation de 6%

Indice 2003 base 2000 = $\frac{\text{prix 2003}}{\text{prix 2000}} = \frac{1200}{1020} * 100 = 118$ soit une augmentation de 18%

Question 2 : Sur la base de 2003, calculez les indices de 2000 ; 2001 ; 2002.

Réponse :

Indice 2000 base 2003 = $\frac{\text{prix 2000}}{\text{prix 2003}} = \frac{1020}{1200} * 100 = 85$ soit une diminution de 15%

Indice 2001 base 2003 = $\frac{\text{prix 2001}}{\text{prix 2003}} = \frac{1050}{1200} * 100 = 87,5$ soit une diminution de 12,5%

Indice 2002 base 2003 = $\frac{\text{prix 2002}}{\text{prix 2003}} = \frac{1080}{1200} * 100 = 90$ soit une diminution de 10%

Exercice 5 : indice

Quel est l'indice du prix base 100 en 2010 d'un produit valant 250F en 2010 et 300F en 2012 ?

Réponse : **Circularité ou transférabilité : si une grandeur numérique V prend les valeurs Vo, Vt et Vt' aux instants o, t, t'**

alors : $I_{t/o} = I_{t/t'} * I_{t'/t} * \frac{1}{100}$

Le chiffre d'affaire d'une entreprise a augmenté de 30% de 2010 à 2011 et a diminué de 25% de 2011 à 2012. Le CA a-t-il augmenté ou diminué de 2010 à 2012 ?

Le CA a augmenté de 30% de 2010 à 2011 alors l'indice = $100 + 30 = 130$

→ $I_{11/10} = 130$

Le CA a diminué de 25% de 2011 à 2012 alors l'indice = $100 - 25 = 75$ → $I_{12/11} = 75$

$I_{11/10} = 130$ $I_{12/11} = 75$ calculer $I_{12/10} = ?$

On sait que : $I_{t/k} * I_{k/p} * \frac{1}{100} = I_{t/p}$ → $I_{12/11} * I_{11/10} * \frac{1}{100} = I_{12/10}$

→ $I_{12/10} = 75 * 130 * \frac{1}{100} = 97,50$ soit une diminution de $97,50 - 100 = -2,5\%$ soit une diminution de 2,5%.

Exercice 6 : indice

Si un prix augmente de 20% de 2010 à 2012 que dire de son évolution de 2012 à 2010 ?

Réponses : Connaissant l'évolution de 0 à t d'une certaine quantité, peut-on connaître l'évolution de t à 0 ?

Si une grandeur numérique V prend les valeurs Vo et Vt aux instants o, t

alors : $I_{o/t} = \frac{100^2}{I_{t/o}}$. $I_{12/10} = 120$ → $I_{10/12} = \frac{100}{120} * 100 = 83,33$

soit une diminution de $83,33 - 100 = -16,67\%$.

Exercice 7 : indice

Soit P et Q les prix et quantités d'un produit vendu par une entreprise. Si le prix de ce produit augmente de 60% de 2000 à 2010 et si les quantités vendues ont diminué de 50% de 2000 à 2010, quelle est l'évolution des recettes de 2000 à 2010 ?

Réponse :

Recette = Prix * quantité → indice recette = Indice Prix * indice quantité * $\frac{1}{100}$

Le prix de ce produit augmente de 60% de 2000 à 2010 → $IP_{10/00} = 160$

La quantité de ce produit diminué de 50% de 2000 à 2010 → $IQ_{10/00} = 50$

→ indice recette = $160 * 50 * \frac{1}{100} = 80$ soit une diminution de 20%.

Exercice 8 :

Le tableau suivant donne la répartition des salaires dans une entreprise.

Salaire en millier de F	20 à 40	40 à 50	50 à 70	70 à 110	110 à 150
Effectifs	8	12	20	6	4
Effectifs cumulés	8	20	40	46	50
Fréquences cumulées	16%	40%	80%	92%	100%

1/ La moitié des employés ont un salaire inférieur ou égal à un montant S ; trouver S.

2/ Quelle est la proportion des individus qui ont un salaire inférieur ou égal à 60 000F ? Comment appelle-t-on ce nombre ?

3/ Quelle est la proportion d'individus ayant un salaire supérieur ou égal à 50 000F ? Comment appelle-t-on ce nombre ?

Solution :

1/ La moitié des employés = $50/2 = 25$ suivant les effectifs cumulés $20 < 25 < 40$ et $50 < S < 70$.

$$S = 50\,000 + (70\,000 - 50\,000) \cdot x$$

$$S = 50.000 + (70.000 - 50.000) \cdot \frac{25 - 20}{40 - 20} = 55.000 \text{ FCFA}$$

2/ soit x cette proportion : 60 000 F est dans la classe 50 000 et 70 000F :

$$50\,000 < 60\,000 < 70\,000 \text{ et } 40\% < x < 80\%$$

$$50.000 + (70.000 - 50.000) \cdot \frac{x - 0,40}{0,80 - 0,40} = 60.000 \text{ On trouve } x = 0,60 = 60\%$$

3/

Supérieur ou égal à 50 000F : $20 + 6 + 4 = 30$ sur 50 soit 60%

4/ $40 < Q_1 < 50$ et $0,16 < 0,25 < 0,40$

$$Q_1 = 40.000 + (50.000 - 40.000) \cdot \frac{0,25 - 0,16}{0,40 - 0,16} = 43.750 \text{ FCFA}$$

$$Q_2 = 55\,000 \text{ F}$$

$50 < Q_3 < 70$ et $0,40 < 0,75 < 0,80$

$$Q_3 = 50.000 + (70.000 - 50.000) \cdot \frac{0,75 - 0,40}{0,80 - 0,40} = 67.500 \text{ FCFA}$$

Exercice 9 :

Soit X une variable statistique, b étant un réel positif. On pose $Z = bx$. Exprimer la moyenne géométrique de Z en fonction de celle de X.

Réponse :

Par définition la moyenne géométrique de X est : $\ln \bar{X}_G = \frac{\sum n_i \ln x_i}{N}$

Alors la moyenne géométrique de Z :

$$\ln \bar{Z}_G = \frac{\sum n_i \ln Z_i}{N} = \frac{\sum n_i \ln b X_i}{N} = \frac{\sum n_i \ln b}{N} + \frac{\sum n_i \ln X_i}{N} = \ln b + \ln \bar{X}_G = \ln b \bar{X}_G$$

$$\text{Alors } \bar{Z}_G = b \bar{X}_G.$$

Exercice 10 : Soit X une variable statistique positive de moyenne arithmétique 12,100 et d'écart type 4,783

1°) calculer la moyenne quadratique $\overline{X}_q$

2°) On pose $Z = \frac{X-2\bar{X}}{\sigma^2}$ Exprimer $\bar{Z}$ et $V(Z)$ en fonction de $\bar{X}$ et $V(X)$ calculer $\bar{Z}$ et $V(Z)$

3°) On pose $= \frac{1}{\bar{X}}$. Quelle est la valeur de $\bar{X}H$?

Réponse :

1/ $V(X) = \overline{X}_q^2 - \bar{X}^2 = \sigma^2 \Rightarrow \overline{X}_q = \sqrt{12,100 + 4,783} = 4,109$

2/ par définition la moyenne arithmétique $\bar{X} = \frac{\sum n_i * X_i}{N}$

alors $\bar{Z} = \frac{\sum n_i * Z_i}{N} = \frac{\sum n_i * \left(\frac{X-2\bar{X}}{\sigma^2}\right)}{N} = \frac{\bar{X}-2\bar{X}}{\sigma^2} = \frac{-\bar{X}}{\sigma^2}$

la variance $V(Z) = V\left(\frac{X-2\bar{X}}{\sigma^2}\right) = \frac{1}{\sigma^4} V(X)$

3/ la moyenne harmonique $\bar{X}H = \frac{n}{\sum \frac{1}{x_i}} = \frac{n}{\sum y_i} = \frac{1}{\bar{y}}$

Exercice 11 : Démontrer que la moyenne arithmétique d'une variable statistique est inférieure ou égale à sa moyenne quadratique

Réponse : Nous devons comprendre que la variance $V(X) \geq 0$ positif ou nulle et que $V(X) = (\overline{X}_q)^2 - (\bar{X})^2$ alors $(\overline{X}_q)^2 - (\bar{X})^2 \geq 0 \Rightarrow (\overline{X}_q)^2 \geq (\bar{X})^2$ ainsi $\overline{X}_q \leq \bar{X}$.

Exercice 12 : Dans une société de gardiennage, le salaire mensuel moyen est de 18.000 F. Les salaires mensuels moyens des employés d'hommes et des employés femmes étant respectivement 20.000 F et 12.000 F. Calculer le % d'hommes et de femmes de société.

Si cette société emploie 20 personnes, déterminer le nombre d'homme et de femmes.

Réponses : f_1 la fréquence d'homme, f_2 la fréquence des femmes

$$\begin{cases} f_1 + f_2 = 1 \\ \overline{X}_1 f_1 + \overline{X}_2 f_2 = \bar{X} \end{cases} \quad \begin{cases} f_1 + f_2 = 1 \\ 20.000 f_1 + 12.000 f_2 = 18.000 \end{cases} \quad \begin{cases} f_1 + f_2 = 1 \\ 5 f_1 + 3 f_2 = 4,5 \end{cases}$$

On trouve $f_2 = 0,25 = 25\%$ et $f_1 = 0,75 = 75\%$

Le nombre d'homme $20 \times 0,75 = 15$

Le nombre de femme $20 \times 0,25 = 5$

Exercice 13 : Un cycliste parcourt une certaine distance D en trois étapes D_1 , D_2 , D_3 , avec les vitesses respectives V_1 , V_2 et V_3 . Calculer la vitesse moyenne de ce cycliste sur le trajet D.

Quelle est la nature de cette moyenne. Que devient la formule si $D_1=D_2=D_3 = D/3$

Réponse : $\bar{V} = \frac{D_1 + D_2 + D_3}{t_1 + t_2 + t_3} = \frac{\sum D_i}{\sum \frac{D_i}{V_i}}$, c'est la moyenne harmonique.

$$D_1 = D_2 = D_3 = D/3 \rightarrow \bar{V} = \frac{D_1 + D_2 + D_3}{t_1 + t_2 + t_3} = \frac{\sum D_i}{\sum \frac{D_i}{V_i}} = \frac{D/3 + D/3 + D/3}{\frac{D/3}{V_1} + \frac{D/3}{V_2} + \frac{D/3}{V_3}} = \frac{D}{\frac{1}{V_1} + \frac{1}{V_2} + \frac{1}{V_3}} = \frac{D}{\sum \frac{1}{V_i}}$$

Exercice 14 : Soit X une variable statistique positive dont la distribution est la suivante

X	X ₁	X ₂	16	Total
EFF.	2	2	1	5

Sachant que les moyennes géométriques et arithmétiques sont respectivement 4 et 5,6. Calculer X₁ et X₂. En déduire $\bar{X}_H$

Réponse : on trouve X₁ = 2 ; X₂ = 4

Exercice 15 : On dispose du tableau suivant :

X _i	3	5	10	15	20
n _i	6	a	9	b	10

Calculer les effectifs manquants.

On précise que $\bar{X} = 12,26$ $\bar{X}_H = 25/3$

Exercice 16 : A partir de la distribution suivant

X _i	X ₁	5	X ₃	15	20
n _i	3	1	6	12	8

Et sachant que $\bar{X} = 13,8$ $\bar{X}_H = 10$ calculer X₁ et X₃.

Exercice 17: On se trouve face du tableau suivant

Variable x _i	2 à 4	4 à 6	6 à 10	10 à 16	16 à 30
Fréquences	-----	0,25	0,30	-----	-----

Sachant que la moyenne arithmétique de cette distribution est égale à 9 et que sa variance égale 32,5 calculer les fréquences manquantes f₁, f₄, f₅.

Sujet n°1 de Statistique descriptive / L1 – 2015 – 2016 / 1 heure**Exercice 1 :**

1.- Répondez par vrai (V) ou faux (F) les expressions suivantes : le coefficient de corrélation mesure.

a.-) L'intensité d'une relation linéaire entre deux variables quantitatives.

b.-) L'intensité d'une relation quelconque entre deux variables qualitatives.

b.-) Le degré de dépendance entre une variable qualitative et variable quantitative.

Réponse : a

2.- Cochez la bonne réponse selon la nature des indicateurs statistiques suivants

Indicateurs	Moyenne	Ecart-type	Mode	Ecart-interquartile	Médiane
Tendance centrale					
Dispersion					

Réponse :

Paramètres de position (ou de tendance centrale) sont : mode, médiane et moyenne

Paramètres de dispersion sont : étendue ; écart - interquartile ; quartiles, déciles , variance, moments, coefficient de variation, écart absolu moyen.

3.- On dispose du tableau suivant

Modalités xi	$X_1=a$	$X_2=b$	16	Total
Effectifs ni	2	1	1	4

Sachant que la moyenne arithmétique $M = 4$ et la moyenne géométrique $G = 4$, déterminer x_1 et x_2 où ($x_1 < x_2$).

Réponse : $\bar{X} = \frac{2a+1b+1*16}{4} = 4 \Rightarrow 2a + b = 0$

$\bar{X}G = \sqrt[4]{a^2 x b^1 x 16^1} = 4 \Rightarrow a^2 x b^1 x 16^1 = 4^4 \Rightarrow a^2 x b^1 = 16$

$\begin{cases} 2a + b = 0 \\ a^2 b = 16 \end{cases} \text{ alors } \begin{cases} b = -2a \\ a^2 b = 16 \end{cases} \Rightarrow \begin{cases} b = -2a \\ a^2(-2a) = 16 \end{cases} \Rightarrow \begin{cases} b = -2a \\ a^3 = -8 \end{cases}$

$\Rightarrow \begin{cases} b = -2a \\ a = \sqrt[3]{-8} = -2 \end{cases} \Rightarrow \begin{cases} b = 4 \\ a = -2 \end{cases}$

Exercice 2 :

A.-) Une usine fabrique quatre produits : A, B, C, D, dont les prix unitaires et les chiffres de production 2010 et 2011 sont indiqués ci-dessous :

Produits	Prix unitaires (francs)		Quantités (nombres d'unités)	
	2010	2011	2010	2011
A	3	3,20	2000	3 000
B	10	11	500	800
C	8	7,50	1000	500
D	4	4,20	4000	3 000

Calculer l'indice de PAASCHE, pondérées par les prix, en 2011 base 100 en 2010.

Réponse : l'indice de PAASCHE : $IPP_{2011/2010} = \frac{\sum P_{2011} \times Q_{2010}}{\sum P_{2010} \times Q_{2011}}$

$$IPP_{2011/2010} = \frac{\sum P_{2011} \times Q_{2011}}{\sum P_{2010} \times Q_{2011}} = \frac{3,20 \times 3000 + 11 \times 800 + 7,50 \times 500 + 4,20 \times 3000}{3 \times 3000 + 10 \times 800 + 8 \times 500 + 4 \times 3000} = \dots\dots$$

B.-) Une population statistique se présente comme suit :

Valeur de la variable	[0 ; 4 [	[4 ; 10 [	[10 ; 20 [	[20 ; 40 [
Effectifs	4	20	14	2

1.- Calculer la moyenne et la variance de cette série.

2.- Chacune des classes de la distribution précédente est divisée en deux classes de même amplitude, auxquelles on fait correspondre un effectif moitié de l'effectif initial de la classe qui a été divisée. Faire un nouveau tableau. Calculer à nouveau la moyenne et la variance.

Réponse :

Valeur de la variable	[0 ; 4 [	[4 ; 10 [	[10 ; 20 [	[20 ; 40 [	Total
Centre X_i	2	7	15	30	XXX
Effectifs N_i	4	20	14	2	40
$N_i \times X_i$	8	140	210	60	418
$N_i \times X_i^2$	16	980	3150	1800	5946

La moyenne $\bar{X} = \frac{\sum N_i \times X_i}{N} = \frac{418}{40} = 10,45$

La variance $V(X) = \frac{\sum N_i \times X_i^2}{N} - (\bar{X})^2 = \frac{5946}{40} - 10,45^2 = 39,4475$

2.-)

Valeur	[0 ; 2 [	[2 ; 4 [	[4 ; 7 [	[7 ; 10 [	[10 ; 15 [	[15 ; 20 [	[20 ; 30 [	[30 ; 40 [
Eff N_i	2	2	10	10	7	7	1	1
X_i	1	3	5,5	8,5	12,5	17,5	25	35
$N_i \times X_i$								
$N_i \times X_i^2$								

La moyenne $\bar{X} = \frac{\sum N_i \times X_i}{N} = \dots\dots\dots$

La variance $V(X) = \frac{\sum N_i \times X_i^2}{N} - (\bar{X})^2 = \dots\dots\dots$

Sujet n°2 de Statistique descriptive / L1 – 2015 – 2016 / 1,5 heures**Exercice 1 :**

Données reparties en classe de l'âge des enfants de 0 à 25 ans :

Ages	[0, 5[	[5, 10[	[10, 15[	[15, 20[	[20, 22[	[22, 25[	Total
Effectif	2	5	14	2	1	1	25

- 1.-) Quelle est la population étudiée ? quel est son effectif ?
- 2.-) Quelle est la variable étudiée ? Préciser sa nature ?
- 3.-) Quel graphique peut-on utiliser pour représenter cette distribution statistique ?
- 4.-) Déterminer une valeur approchée d'âge modal.
- 5.-) Calculer à l'aide d'une interpolation linéaire une valeur approchée de la médiane et des quartiles Q1 et Q3. En déduire l'intervalle interquartile.
- 6.-) Déterminer les pourcentages des enfants ayant un âge inférieur à 20 ans et ceux ayant au moins 15 ?
- 7.-) Déterminer l'âge moyen et l'écart-type. En déduire le coefficient de variation.

Exercice 2 :

On se propose d'étudier les deux variables statistiques suivantes : X (taux de postes de télévisions déclarés) et Y (taux hebdomadaire des spectateurs de cinéma). Les valeurs données dans le tableau suivant :

X	5,6	6,0	5,4	4,9	5,5	3,6	4,1	3,9	3,6	2,7
Y	8,2	7,3	6,2	5,6	5,5	5,1	4,9	4,6	3,4	3,2

- 1.-) En portant le nombre de postes de télévisions déclarés aux abscisses et taux hebdomadaire des spectateurs de cinéma ordonnées, établir une représentation graphique du tableau.
- 2.-) Déterminer l'équation de la droite de régression de Y en fonction de X.
- 3.-) Donner l'expression et la valeur numérique du coefficient de corrélation linéaire.
- 4.-) Quelle est la valeur du coefficient de détermination ? puis interpréter le résultat.

Solution / Sujet n°2 / L1 – 2015 – 2016 / 1,5 heures**Exercice 1 :**

- 1.-) La population étudiée : les enfants de 0 à 25 ans → l'effectif est = 25
- 2.-) La variable étudiée : âge → Sa nature : quantitative discrète
- 3.-) Graphique pour représenter cette distribution statistique : histogramme
- 4.-) Le mode ?

Les amplitudes ne sont égales, dans ce cas le mode est la valeur du caractère qui a l'effectif rectifié le plus élevé.

Données reparties en classe de l'âge des enfants de 0 à 25 ans :

Ages	[0, 5[	[5, 10[	[10, 15[	[15, 20[	[20, 22[	[22, 25[	Total
Ampl	5	5	5	5	2	3	
Effectif	2	5	14	2	1	1	25
Eff corr	$\frac{2}{5} * 2$	$\frac{5}{5} * 2$	$\frac{14}{5} * 2$	$\frac{2}{5} * 2$	$\frac{1}{2} * 2$	$\frac{1}{3} * 2$	

$$\text{Effectif corrigé} = \frac{\text{effectif de la classe}}{\text{amplitude de la classe}} * \text{plus petite amplitude}$$

L'effectif rectifié le plus élevé est $\frac{14}{5} * 2 \rightarrow$ classe modale [10, 15[→ Mode : 12,5 ans.

5.-) la médiane

Ages	[0, 5[	[5, 10[	[10, 15[	[15, 20[	[20, 22[	[22, 25[	Total
Effectif	2	5	14	2	1	1	25
ECC	2	7	21	23	24	25	XXX

Effectif total 25 → $\frac{N}{2} = \frac{25}{2} = 12,5 \rightarrow$ suivant l'effectif cumulé croissant 12,5 est entre 7 et 21 et la classe médiane est [10, 15[

$$7 < 12,5 < 21 \text{ et } 10 < \text{Mé} < 15 \rightarrow \frac{\text{Mé}-10}{15-10} = \frac{12,5-7}{21-7} \rightarrow \text{Mé} = 11,96 \text{ ans}$$

Les quartiles Q1 et Q3

1^{er} quartile 25% de N Q1 ? Effectif total 25 → $\frac{N}{4} = \frac{25}{4} = 6,25 \rightarrow$ suivant l'effectif cumulé croissant 6,25 est entre 2 et 7 et la classe de Q1 est [5 ; 10[

$$2 < 6,25 < 7 \text{ et } 5 < Q1 < 10 \rightarrow \frac{Q1-5}{10-5} = \frac{6,25-2}{7-2} \rightarrow Q1 = 9,25 \text{ ans}$$

3^{ème} quartile 75% de N Q3 ? Effectif total 25 → $\frac{3N}{4} = \frac{3*25}{4} = 18,75 \rightarrow$ suivant l'effectif cumulé croissant 18,75 est entre 7 et 21 et la classe de Q3 est [10 ; 15[

$$7 < 18,75 < 21 \text{ et } 10 < Q3 < 15 \rightarrow \frac{Q3-10}{15-10} = \frac{18,75-7}{21-7} \rightarrow Q3 = 14,196 \text{ ans.}$$

$$\text{Ecart interquartile} = Q3 - Q1 = 14,196 - 9,25 = 4,95 \text{ ans.}$$

6.-)

Le pourcentage des enfants ayant un âge inférieur à 20 ans

$$(2 + 5 + 14+2) = 23 \text{ enfants sur 25 soit } \frac{23}{25} * 100 = 92\%$$

Le pourcentage des enfants ayant un âge au moins 15 ans

$(2 + 1 + 1) = 4$ enfants soit $\frac{4}{25} * 100 = 16\%$

7.-) L'âge moyen et l'écart-type. En déduire le coefficient de variation.

Ages	[0, 5[	[5, 10[	[10, 15[	[15, 20[	[20, 22[	[22, 25[	Total
Xi	2,5	7,5	12,5	17,5	21	23,5	
Ni	2	5	14	2	1	1	25
Ni x Xi	5	37,5	175	35	21	23,5	297
Ni x Xi ²	12,5	281,25	2187,5	612,5	441	552,25	4087

La moyenne $\bar{X} = \frac{\sum N_i x X_i}{N} = \frac{297}{25} = 11,88$

La variance $V(X) = \frac{\sum N_i x X_i^2}{N} - (\bar{X})^2 = \frac{4087}{25} - (11,88)^2 = 22,3456$

L'écart type $= \sqrt{22,3456} = 4,73$

Le coefficient de variation $c = \frac{\sigma_x}{\bar{X}} = \frac{4,73}{11,88} = 0,3979$.

Solution / exercice 2 :

X	5,6	6,0	5,4	4,9	5,5	3,6	4,1	3,9	3,6	2,7
Y	8,2	7,3	6,2	5,6	5,5	5,1	4,9	4,6	3,4	3,2

$\sum X = 45,3 \rightarrow \bar{X} = \frac{45,3}{10} = 4,53$

$\sum Y = 54 \rightarrow \bar{Y} = \frac{54}{10} = 5,4$

$\sum X^2 = 216,01 \rightarrow V(X) = \frac{216,01}{10} - (4,53)^2 = 1,0801$

$\sum Y^2 = 313,56 \rightarrow V(Y) = \frac{313,56}{10} - (5,4)^2 = 2,196$

$\sum X * Y = 258,16 \rightarrow Cov(X, Y) = \frac{258,16}{10} - 4,53 * 5,4 = 1,354$

2.-) L'équation de la droite de régression de Y en fonction de X

$\rightarrow y = ax + b$ avec $a = \frac{Cov(X,Y)}{V(X)} = \frac{\sum XY - n \bar{X} \bar{Y}}{\sum X^2 - n(\bar{X})}$ et $b = \bar{y} - a \bar{x}$

$a = \frac{Cov(X,Y)}{V(X)} = \frac{1,354}{1,0801} = 1,25$ et $b = 5,4 - 1,25 * 4,53 = -0,2625$

3.-) l'expression et la valeur numérique du coefficient de corrélation linéaire.

$R_{xy} = \frac{Cov(X,Y)}{\sigma_x * \sigma_y} = \frac{1,354}{\sqrt{1,0801 * 2,196}} = 0,879$

4.-) la valeur du coefficient de détermination

$R_{xy}^2 = (0,879)^2 = 0,77$ Y est expliqué par X à 77%.

Sujet n°3 de Statistique descriptive / L1 – 2015 – 2016 / 1 heure**Exercice 1 :**

Répondre par vrai ou faux

1.-) Dans une série statistique si les effectifs sont multipliés par 3 alors

a.-) La moyenne arithmétique reste invariante

b.-) Le mode est multiplié par 3

c.-) La médiane est multipliée par 3

Réponse : a2.-) Dans une série statistique si toutes les modalités sont multipliées par un nombre réel m alors :a.-) La moyenne arithmétique est multipliée par m ;b.-) Le mode est multiplié par m ;c.-) La médiane est multipliée par $m/2$.d.-) L'écart-type est multiplié par m .**Réponse : a et b**

3.-) La caractéristique de dispersion usuelle qui est sans unité est :

a.-) La variance ;

b.-) L'écart absolu moyen ;

c.-) L'écart-type ;

d.-) Le coefficient de variation.

Réponse : d**Exercice 2 :**

Voici la répartition des notes obtenues sur 20 par 20 candidats :

Notes	[0 ; 7[	[7 ; 9[	[9 ; 10[	[10 ; 12[	[12 ; 14[	[14 ; 16[	[16 ; 18[	[18 ; 20[	Total
Ni	2	3	5	4	2	1	2	1	
ECC	2	5	10	14	16	17	19	20	
ECD	20	18	15	10	6	4	3	1	

1.-) Quelle est la population étudiée ? Et quel est son effectif ?

2.-) Quelle est la variable étudiée ? Préciser sa nature.

3.-) Quel graphique peut-on utiliser pour représenter cette distribution statistique ?

4.-) Déterminer une valeur approchée de la note modale.

5.-) Calculer à l'aide d'une interpolation linéaire la valeur approchée de la médiane.

6.-) Quel est le nombre de candidats dont la note est inférieure à 10 ?

7.-) Quel est le % des candidats percevant au moins 9 ?

Réponse : La population : candidats . Effectif = 20

Variable : les notes. Nature : Quantitatif continu

3.-) Graphique : histogramme

4.-) La classe modale : [9 ; 10[et le mode = 9,5

5.-) La médiane $\rightarrow$ effectif total 20 $\rightarrow$ Moitié : $20/2 = 10$

Suivant l'ECC la médiane = 10.

6.-) Le nombre de candidats dont la note est inférieure à 10 est = 10

7.-) Le % des candidats percevant au moins 9 : $\frac{15}{20} \times 100 = 75\%$.

Exercice 3 : Examen de passage 2008**1/ Ecrire le numéro de la bonne réponse (unique).**

Le tableau suivant donne la répartition des ménages d'une population selon le nombre d'automobiles qu'ils possèdent.

Nombre d'autos	0	1	2	3	4
Nombre de ménages	500	2421	900	156	23

La représentation graphique qui convient est :

- a.-) Un diagramme en bâtons b.-) un histogramme

2/ Pour la distribution ci-dessus, un graphique en secteurs circulaires convient-il ? Si oui quelle est la mesure du secteur angulaire qui représente la modalité 2 sur un demi-cercle ?

Réponse :

1/ La représentation graphique qui convient, pour les variables discrète, est le diagramme en bâtons.

2/ Pour la distribution ci-dessus, un graphique en secteurs circulaires convient. La mesure du secteur angulaire qui représente la modalité 2 sur un demi-cercle est = $180^\circ \times \text{fréquence} = 180^\circ \times (900 / 4\ 000) = 40,5^\circ$.

Exercice 4 : Examen de passage 2008

Soit la distribution des lettres par une entreprise à ses clients selon le temps d'acheminement (en jours) par la poste :

Temps d'acheminement	1	2	3	4	5
Nombre de lettres	120	50	20	6	4

1/ Quelle est la population étudiée ? Quelle est l'unité statistique ? Quel est le caractère ?

2/ (a) Quelle est la proportion de lettres dont la durée d'acheminement est inférieure ou égale à 3 jours ? Comment appelle-t-on ce nombre ?

(b) Quelle est la proportion de lettres dont la durée d'acheminement est supérieure ou égale à 4 jours ? Comment appelle-t-on ce nombre ?

Réponse :

1/ La population étudiée est : lettres

L'unité statistique est une lettre.

Le caractère : Temps d'acheminement (variable statistique discrète).

2/ (a)

La proportion de lettres dont la durée d'acheminement est inférieure ou égale à 3 jours est : $120 + 50 + 20 = 190$ lettres sur 200 soit 95%

Ce nombre est 190.

(b) La proportion de lettres dont la durée d'acheminement est supérieure ou égale à 4 jours : $6 + 4 = 10$ lettres sur 200 soit 5%. Ce nombre est 10 lettres.