La Compagne Etudiante pour Résumer les Cours et Organiser les Polycopies

Option : Science Economique et Gestion

Module : METHODES QUANTITATIVES I

Matière : Statistique Descriptive I

Semestre : 1

Type de document : Résumés des Chapitres

Remarque :

- Ce document présent le résumé de cour, il concerne des définitions, des exemples, des interprétations, etc. ;
- Les interprétations doit être présenter au feuille d'examen ;
- Les formules et les unités statistiques doit être aussi préciser au feuille d'examen,
- Les titres des graphiques doit être préciser aussi.

Année universitaire 2012-2013

Chapitre I - Introduction à la méthode statistique :

Section I - Concept de base en Statistique Descriptive :

I-1 Population - Unité Statistique:

- Une population statistique est l'ensemble de références pour une étude statistique. Chaque élément de cet ensemble est appelé unité (ou individu) statistique.

Une population statistique peut être "fini" ou "infinie".

Exemple des unités statistiques :
- Être humain : ouvrier, cadre, étudiant...
- Un objet concert : voiture, appartement, ...
- Un objet abstrait : note d'examen, ...

Échantillon : c'est la partie de la population qui est observée.
Recensement : c'est le processus consistant à observer chaque individu de la population.
Sondage : c'est le processus consistant à observer une partie de la population.

I-2 Caractères :

Pour décrire une population, en classe les individus selon une ou plusieurs caractéristiques appelées caractères.
Un caractère est le trait commun à toutes les unités statistiques d'une population que l'on désire étudier. C.-à-d. c'est ce qui est observé ou mesuré sur les individus d'une population statistique

Exemples:

- La population d'un pays ou les salariés d'une entreprise pourront être décrits par des caractères tel que : sexe, âge, nationalité, nombre d'enfant, calcification...

- La production automobile pourra être repérée par les caractères : couleurs, type, puissance du moteur, etc.

- Un caractère peut être qualitatif lorsqu'il n'est mesurable.

Exemple : nationalité, couleur, situation matrimonial, ...

- Un caractère peut être quantitatif lorsqu'il est mesurable ou exprimé par un nombre.

Exemple : Le nombre d'enfant par famille, l'âge, la taille, le poids ou la longueur d'un individu ...

Exemple des caractères qualitatifs et caractères quantitatifs :

Caractère Qualitatif	Variables Quantitatif	
	Continue	Discrètes
Sexe : masculin ou féminin Catégorie socio-professionnelle: - chômeurs - ouvriers Etat matrimonial : célibataire. etc.	Age Salaires Notes Chiffre d'affaires Durée d'étude Poids, etc.	Nombre des produits Nombre des salariés Nombre d'entreprise Population d'un territoire Nombre d'enfants à charge Etc.

Une variable statistique : c'est le nombre affecté à chaque unité statistique est une mesure du caractère. Elle peut distinguer entre variables statistiques discrètes et variables statistiques continues.

- une variable statistique discrète ou discontinue lorsqu'elle on peut prendre des nombres entiers 0, 1, 2, 3, ... appartient à l'ensemble IN.

Exemple : nombre d'étudiants dans la faculté, nombre des entreprises dans une région, ...

- Une variable statistique est continue lorsqu'elle peut prendre les valeurs numériques de l'ensemble IR.

Exemple: la taille, l'âge, le poids ...

I-3 Modalités (x_i):

Les modalités x_i d'un caractère sont les différentes caractéristiques que peuvent présenter les unités statistiques d'une certaine population. Chaque individu présente une et une seule modalité à la fois (exhaustivité et disjonctive)

Section II - Présentation des Résultats:

II-1 Tableaux statistiques :

Les résultats d'une enquête sont présentés dans deux tableaux :

II.1.1 - Tableau statistique à simple entrée :

C'est un tableau utilisé lorsqu'on étudie une population à un seul caractère. Il composé de deux colonnes :

- Le premier composé des modalités prises par le caractère. Ces modalités sont distinguées par x_i
- La seconde est réservé pour les effectifs correspondants aux différents modalités et qui sont distinguées par n_i.

La forme générale d'un tableau statistique à simple entrée est la suivante :

Modalités en valeurs	Effectifs
X_i X_1 X_2 . . X_i . . X_k	n_i n_1 n_2 . . n_i . . n_k
Total	N ou n

N ou n : effectif total de la population ($n = n_1 + n_2 + n_3 + ... + n_i + … + n_k$)

II.1.2. Tableau statistique à double entrée :

Ce tableau est utilisé lorsque l'étude statistique porte sur 2 caractères (X et Y) de chacun des éléments de la population ou de l'échantillon (**Exemple** : taille et poids, niveau d'étude et salaire …).
X peut prendre p modalités (x_1, x_2 .... x_p) et Y peut prendre q modalités (y_1, y_2 ...... y_q).
Soit n_{ij} = le nombre d'individus ayant à la fois les modalités x_i de X et y_i de Y.
i = 1, ..., p et j= 1, ..., q
Les différentes modalités peuvent être des entiers naturels ou des classes selon que les variables étudiées discrètes ou continues. Les différentes classes doivent être exhaustives et incompatibles.
Le tableau bien présenté est un tableau qui concerne :

- Un titre complet qui résume l'information nécessaire.
- Les unités utilisés présenté clairement (**Exemple :** Dirhams, millions dirhams, m^2, kg …)
- Les origines et les ressources : c'est une référence à laquelle on se réfère. (**Exemple :** Haut-commissariat du plan, FMI, annuaire statistique…).

Section III - Sommation :

III-1 Définition :

Soit n_i = le nombre d'individus ayant le caractère xi dans l'échantillon étudié.
Supposant que le caractère peut prendre 10 modalités en on cherche à calculer le nombre total des individus dans l'échantillon. Ce nombre est donné par : $n_1 + n_2 + \dots + n_{10} = n$
n est donc la somme pour i variant de 1 à 10 de n_i.
On écrit : $\sum_i^{10} n_i = n$
On lit la somme pour i = 10 des valeurs des n_i.
$n_1 + n_2 + \dots + n_k$ s'écrit sous la forme :

III-2 Propriétés des Sommations:

- $P_1 = \sum_{i=1}^{n} X_i = x_1 + x_2 + \cdots + x_n$
- $P_2 = \sum_{i=1}^{n} (X_i + a) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} a = \sum_{i=1}^{n} X_i + na$
- $P_3 = \sum_{i=p}^{n} (X_i + a) = \sum_{i=p}^{n} X_i + \sum_{i=p}^{n} b = \sum_{i=p}^{n} X_i + (n-p+1)b$
- $P_4 = \sum_{i=1}^{n} a X_i = a \sum_{i=1}^{n} X_i$

- $P_5 = \sum_{i=P}^{n}(aX_i + b) = a\sum_{i=P}^{n} X_i + (n-p+1)b$
- $P_6 = \sum_{i=1}^{n}(X_i + a)^2 = \sum_{i=1}^{n} X_i^2 + 2aX_i + a^2 = \sum_{i=1}^{n} X_i^2 + 2a = \sum_{i=1}^{n} X_i + na^2$
- $P_7 = \sum_{i=1}^{n}(aX_i + b)^2 = a^2\sum_{i=1}^{n} X_i^2 + 2ab\sum_{i=1}^{n} X_i + nb^2$

III-3 Quelques sommes classiques :

- Somme de n premiers entiers naturels :

$1 + 2 + 3 + ... + n = \frac{n(n+1)}{2}$

- Somme des carrés des n premiers entiers naturels:

$1^2 + 2^2 + 3^2 + ... + n^2 = \frac{n(n+1)(2n+1)}{6}$

- Sommes des n termes d'une progression arithmétique :

Soit une progression arithmétique de raison r et constituées des termes U_1 , U_2 tel que

$U_2 = U_1 + r$
$U_3 = U_1 + 2r$
.
.
$U_n = U_1 + (n-1)\,r$

$$\sum_{i=1}^{n} U_i = \frac{U_1 + U_n}{2} n = \frac{2U_1 + (n-1)r}{2}$$

- Sommes des n termes d'une progression géométrique :

Soit une progression arithmétique de raison q

$U_n = U_1 \cdot q$ avec $n=1,\ldots\ldots\ldots\ldots n$

$$\sum_{i=1}^{n} U_1 = U_1 \frac{1-q^n}{1-q}$$

Section IV - Fréquences et effectifs cumulés :

IV.1- Fréquence relative F(x):

On appelle fréquence relative le rapport $\frac{\textit{effectif de la valeur}}{\textit{effectif total}}$, On note $fi = \frac{n_i}{\sum_{i=1}^{k} n_i} = \frac{n_i}{n}$

Remarque :

- Ne pas confondre entre fréquence absolue et fréquence relative.
- La fréquence absolue est définie comme l'effectif de chaque classe et noté ni ; alors que la fréquence relative est le rapport : $\frac{n_i}{n}$
- La somme de fréquences relatives est égale à 1 ==> $\sum_{i=1}^{k} f_i = \sum_{i=1}^{k} \frac{n_i}{n} = \frac{n}{n} = 1$
- En multipliant une fréquence fi par 100, on obtient le taux de pourcentage de la population total que représente n_i.

IV.2- Effectifs cumulés :

Les effectifs cumulés (fréquence relatives cumulées) s'obtiennent en totalisant l'effectif (fréquence) d'une classe déterminée avec les effectifs (fréquences) des classes antérieurs ou des classes antérieurs ou des classes postérieures. Dans le premier vas ; on parle des effectifs cumulés croissants. Dans le deuxièmes cas, ou parle des effectifs (fréquences) cumulé(e)s décroissant(e)s.

Représentation d'un tableau relatif à la répartition de la population étudiée selon les effectifs ou fréquences simples, les effectifs ou les fréquences cumulées.

Modalité des Caractères	Fréquences absolues	Fréquences absolues cumulés croissantes	Fréquences absolues cumulés décroissantes	Fréquences Relatives
X_1	n_1	n_1	$n_k + \ldots + n_i + \ldots + n_2 + n_1 = n$	f_1
X_2	n_2	$n_1 + n_2$	$n_k + \ldots + n_i + \ldots + n_2$	f_2
.	.	.	.	.
.	.	.	.	.
X_i	n_i	$n_1 + n_2 + \ldots + n_i$	$n_k + \ldots + n_i$	f_i
.	.	.	.	.
.	.	.	.	.
X_k	n_k	$n_1 + n_2 + \ldots + n_i + \ldots + n_k$	n_k	f_k
Total	n			1

L'effectif cumulé croissante N(x) (ou la fréquence relative cumulé croissante) est la somme (additionne) de haut en bas, et elle correspond à la notion de "**Au plus** (≥)" ou bien "**Moins de** (<)".

Dans le cas d'une variable continue, le cumul s'interprète par rapport à l'extrémité <u>supérieure</u> des classes.

L'effectif cumulé décroissante D(x) (ou la fréquence relative cumulé décroissante) est la somme de bas en haut, elle correspond à la notion de "**Au moins** (≤)" ou bien "**Plus de** (>)".

Dans le cas d'une variable continue, le cumul s'interprète par rapport à l'extrémité inférieure des classes.

Exemple : On considère 30 entreprises (n_i) classées en fonction du nombre des employeurs (X_i). Les résultats sont récapitulés dans le tableau suivant :

Modalités du caractère (X_i)	Effectif (n_i) ou fréquences absolues	f_i Fréquences relatives
6	6	0,2
7	5	0,167
8	5	0,167
9	4	0,133
10	4	0,133
12	4	0,133
15	2	0,067
Total	30	1

Le calcul les effectifs cumulés croissantes et décroissantes nous donne :

X_i	n_i	f_i Fréquences relatives	Effectifs cumulés croissantes	Effectifs cumulés décroissantes
6	6	0,2	6	30
7	5	0,167	11	24
8	5	0,167	16	19
9	4	0,133	20	14
10	4	0,133	24	10
12	4	0,133	28	6
15	2	0,067	30	2
Total	30	1		

Exemple : Déterminer l'effectif et la fréquence :

a) Des entreprises ayant moins de 9 employeurs.
b) Des entreprises ayant au moins 9 employeurs.
c) Des entreprises ayant plus de 9 employeurs.
d) Des entreprises ayant au plus 9 employeurs.

Nombre des personnes (X_i)	Nombres d'entreprises (n_i)	Fréquences relatives	Effectifs cumulés croissantes	Effectifs cumulés décroissantes
6	6	0,2	6	30
7	5	0,167	11	24
8	5	0,167	16	19
9	4	0,133	20	14
10	4	0,133	24	10
12	4	0,133	28	6
15	2	0,067	30	2
Total	30	1		

a) 16 entreprises ayant moins de 9 employeurs soit 53,4 % de l'effectif total (lecture de l'effectif cumulé **croissante**)

b) 14 entreprises ayant au moins 9 employeurs soit 46,6% de l'effectif total (lecture de l'effectif cumulé **décroissante**)

c) 10 entreprises ayant plus de 9 employeurs soit 33,3 % de l'effectif total (lecture de l'effectif cumulé **décroissante**)

d) 20 entreprises ayant au plus 9 employeurs soit 66,7% de l'effectif total (lecture de l'effectif cumulé **croissante**)

- Le cas de (a) et (b) :

Nombre des personnes (X_i)	Nombres d'entreprises (n_i)	Effectifs cumulés croissantes	Effectifs cumulés décroissantes
6	6	6	30
7	5	11	24
8	5	16 a)	19
9	4	20	14 b)
10	4	24	10
12	4	28	6
15	2	30	2
Total	30		

- Le cas de (c) et (d) :

Nombre des personnes (X_i)	Nombres d'entreprises (n_i)	Effectifs cumulés croissantes	Effectifs cumulés décroissantes
6	6	6	30
7	5	11	24
8	5	16	19
9	4	20 c)	14
10	4	24	10 d)
12	4	28	6
15	2	30	2
Total	30		

Chapitre II Étude des distributions statistiques à une dimension

Section I - Représentation graphique :

I.1- Représentation graphique des distributions à caractère qualitatif:

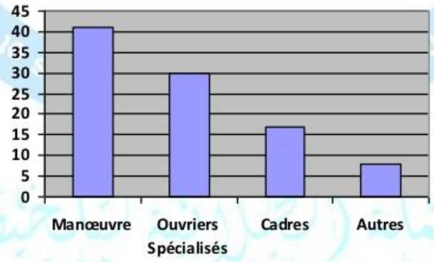
Les formes des graphiques les plus couramment utilisés en cas d'une distribution à caractère qualitatif sont :

I.1.1- les diagrammes en barres (ou les tuyaux d'orgue) :

Il consiste à représenter chaque modalité du caractère qualitatif étudié, par un rectangle dont la base est constante et dont la hauteur est proportionnelle à l'effectif.

Exemple :

Modalités du caractère (X_i)	Effectif (n_i) ou fréquences absolues	f_i Fréquences relatives
Manœuvre	41	0,43
Ouvriers Spécialisés	30	0,31
Cadres	17	0,18
Autres	8	0,08
Total	96	1



I.1.2- La représentation par secteurs (le diagramme séculaire) :

L'effectif total est représenté par la surface d'un cercle, chaque effectif de modalité est représenté par secteur.

L'angle θ_i d'un secteur relative à une modalité est déterminé par : $\theta_i = 360° \cdot \frac{n_i}{n} = 360° \cdot f_i$

Exemple de calculer l'angle θ_i d'un secteur :

$$\begin{aligned}\sum_{i=1}^{k}\theta_i &= \sum_{i=1}^{k}360° \\ &= 360°\sum_{i=1}^{k}f_i \\ &= 360°\end{aligned}$$

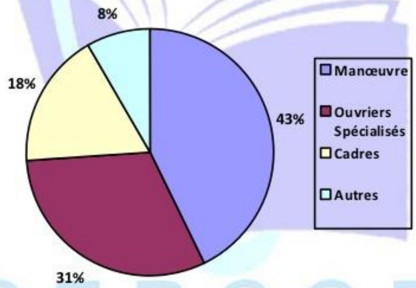
$\theta_1 = 360° \times 0{,}43 = 155°$

$\theta_2 = 360° \times 0{,}31 = 111°$

$\theta_i = 360° \times 0{,}18 = 65°$

$\theta_i = 360° \times 0{,}08 = 29°$

$\sum_{i=1}^{k} \theta_i = 155° + 111° + 65° + 29°$



Remarque :

On peut utiliser le diagramme semi-circulaire. A ce moment l'angle θ au centre relatif à une modalité i est égal à : $\theta_i = 180° \cdot f_i$

<u>I.2- Représentation graphique des distributions à caractère quantitatif:</u>

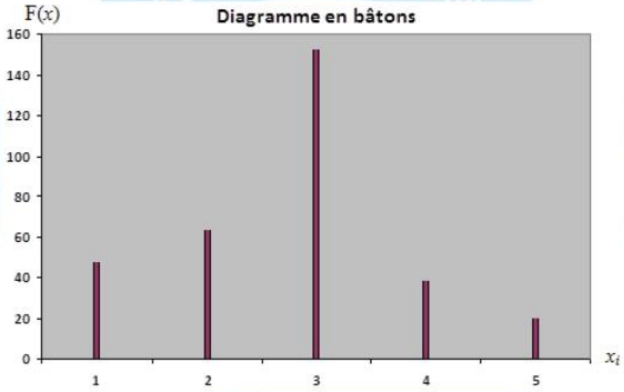
Le type de représentation graphique dépend de la nature discrète ou continue de la variable statistique.

I.2.1. Représentation graphique des variables discrètes :

Trois modes de représentation différents sont utilisés pour illustrer une distribution d'effectifs ou une distribution de fréquences relatives (non cumulée ou cumulée).

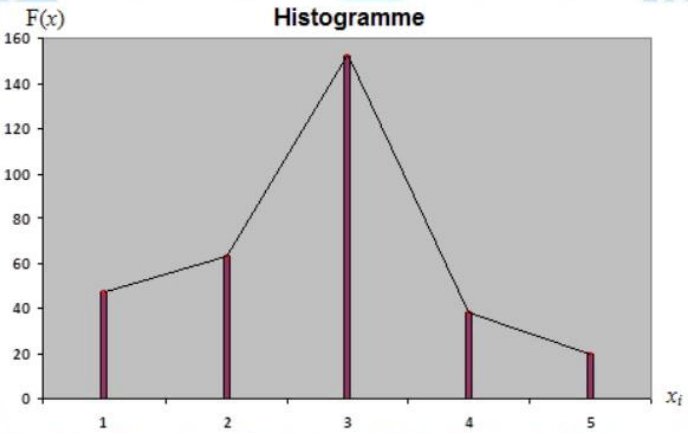
a) Diagramme en bâtons :

On place en abscisse les modalités différentes valeurs x_i de la valeur statistique. En chacun de ces points ; on trace parallèlement à l'axe des ordonnées un segment vertical (bâton) de longueur



b) Histogramme :

Il s'agit de relier entre les sommets des bâtons. F(*x*) x_i

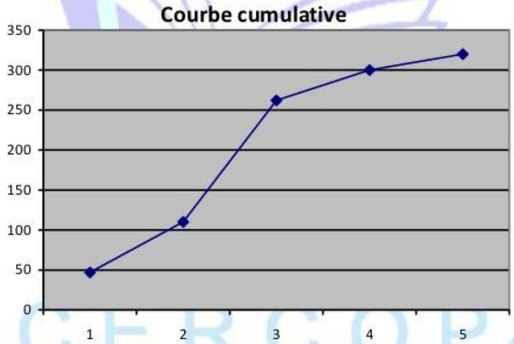


c) Courbe cumulative ou courbe escalier :

- **La courbe cumulative :**

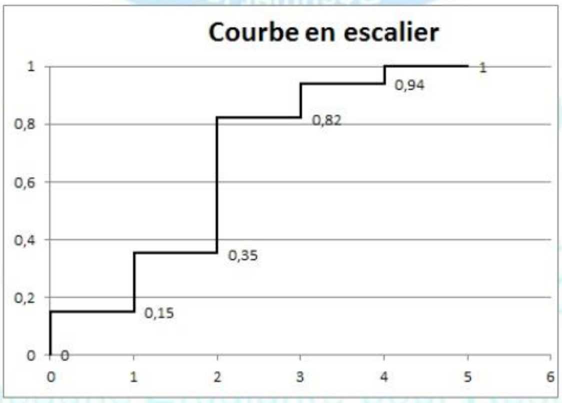
La courbe cumulative est une fonction de répartition de la forme : $F(xi)= f(X<x_i)$ pour une variable statistique X se définie par la correspondance entre xi et la propension d'élément dans l'ensemble ayant une valeur strictement inférieur à x_i.

X_i	f_i	F(x)
1	0,15	0,15
2	0,20	0,35
3	0,47	0,82
4	0,12	0,94
5	0,06	1
Total	1	



- **La courbe cumulative en escalier :**

La fonction de répartition est représentée par une courbe en escaliers dont des paliers horizontaux ont pour ordonnées les f_i ou n_i cumulées, les marches de l'escalier correspondant aux valeurs possibles de xi et ont des hauteurs proportionnelles aux fréquences cumulées f_i ou n_i cumulés.



I.2.2 Représentation graphique des valeurs continues :

Les types de représentations graphiques des séries continues sont :

- L'histogramme ou diagramme différentiel des fréquences ou les effectifs.
- La courbe cumulative ou diagramme intégral de fréquences cumulées ou les effectifs cumulés.

1.2.2.1- L'histogramme ou diagramme différentiel des fréquences ou les effectifs.

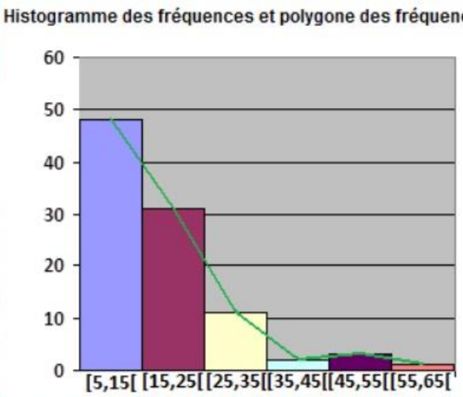
En abscisse, les limites des classes d'égale amplitude créée se forme d'un segment qui correspond un rectangle dont la hauteur est égale à l'effectif ou la fréquence de la classe.

Exemple :

Classe d'âge	Effectifs	Effectifs cumulés croissants
[5,15[	48	48
[15,25[	31	79
[25,35[	11	90
[35,45[	2	92
[45,55[	3	95
[55,65[	1	96
Total	**96**	

a) ***Cas d'amplitude égale :***

Dans ce cas, les hauteurs des rectangles de l'histogramme correspondent aux effectifs ou aux fréquences observées pour chaque classe.



b) ***Cas d'amplitude inégale :***

Dans ce cas, il faut corriger les effectifs. L'effectif corrigé est obtenu par une choisir d'une amplitude de référence, pour chaque classe donnée, si l'amplitude est le double, le triple…… l'amplitude de référence, il faut diviser par 2, 3, ... l'effectif observé. Si l'amplitude de la classe étudier est égale à la moitié, au tiers…… de l'amplitude de référence, il faut multiplier par 2, par 3, … l'effectif de la classe observé.

$$a_i = BS - BI$$

BS : borne supérieur

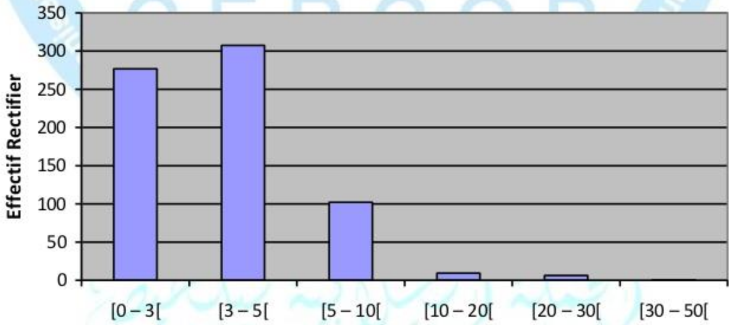
BI : borne inférieur
a_i : amplitude unité

En abscisse, les limites des classes d'égale amplitude créée se forment d'un segment qui correspond un rectangle dont la hauteur des rectangles, égale à l'effectif rectifié (corrigé) de la classe.

Formule de l'effectif rectifié ou l'hauteur corrigée (h_i) : $h_i = \frac{n_i}{a_i}$

Exemple :

Classes	Effectifs n_i	Amplitude a_i	Effectifs rectifiés $h_i = n_i / a_i$
[0 – 3[	830	3	276,7
[3 – 5[	615	2	307,5
[5 – 10[	510	5	102,0
[10 – 20[	92	10	9,2
[20 – 30[	63	10	6,3
[30 – 50[	15	20	0,75



1.2.2.2- Le polygone des effectifs ou des fréquences :

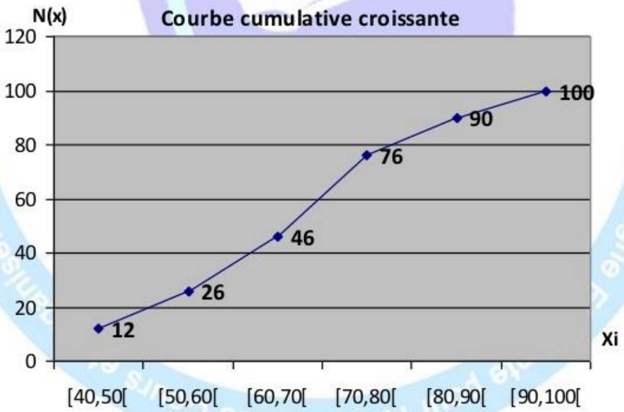
C'est la courbe obtenue en joignant par des segments de droite les milieux des bases supérieures des rectangles de l'histogramme des effectifs, ainsi que ceux de deux classes fictives d'effectifs nuls situées de part et d'autre de l'histogramme. La surface du domaine compris entre l'axe des abscisses et le polygone des effectifs doit être égale à la surface des rectangles de l'histogramme.

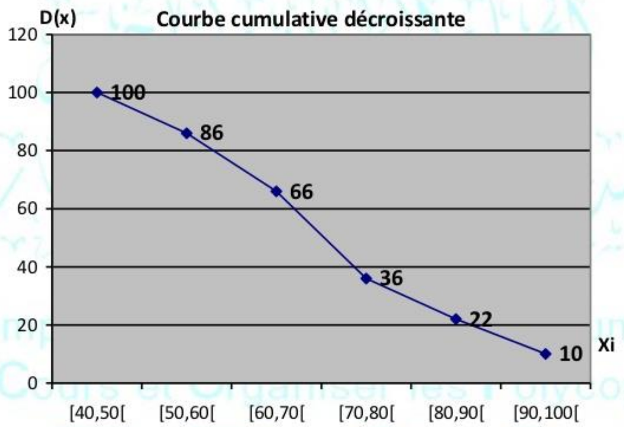
1.2.2.3- Courbe cumulative (ou : Polygone des fréquences relatives cumulés):

C'est la représentation graphique de la fonction de répartition F(x). Il s'obtient en joignant par des segments de droite les extrémités supérieures droites des rectangles de l'histogramme des effectifs cumulés croissants.

Exemple :

Salaire (x_i)	Effectifs n_i	N(x)	D(x)
[40,50[	12	12	100
[50,60[	14	26	86
[60,70[	20	46	66
[70,80[	30	76	36
[80,90[	14	90	22
[90,100[	10	100	12
Total	**100**		





Section II – Caractéristiques de Tendance Centrale :

II.1- La moyenne arithmétique :

La moyenne arithmétique d'une série est égale à la somme des produits de chaque variable x_i par le nombre de fois où $\overline{X}$ elle est répétée (pondérer) sur l'effectif total.

La moyenne arithmétique de X notée est définie par :

$$\overline{X} = \sum_{i=1}^{k} \frac{x_i n_i}{n} \text{ ou } \overline{X} = \sum_{i=1}^{k} f_i x_i$$

Le cas d'une variable statistique continue : $\overline{X} = \sum_{i=1}^{k} \frac{C_i n_i}{n}$

C_i : le centre des classes

Interprétation :

$\overline{X}$ Est le xi moyen obtenu par l'ensemble des n.
Le x_i moyen est donnée par : $\overline{X}$ =……

II.2- Autres types de moyennes :

II.2.1- La moyenne géométrique :

La moyenne géométrique de X est notée G(X) ; C'est la racine carré nième

$$G = \sqrt[n]{X_1 \cdot X_2 ... X_n} \text{ Ou } G = \left[\prod_{i=1}^{n} X_i\right]^{\frac{1}{n}} = \sqrt[n]{\prod X_1}$$

Le calcul se fait par les logarithmes : $\log G = \frac{1}{n}\sum_{i=1}^{k} n_i \log X_i = \sum_{i=1}^{k} f_i \log X_i$

Log G est une moyenne arithmétique simple des logarithmes de la variable X_i.
L'utilisation de la moyenne géométrique est recommandée dans certains problèmes économiques ; notamment pour l'analyse des phénomènes de croissances et de mesure du taux de croissance moyen.

II.2.2- La moyenne harmonique :

Soit X = (x_1, x_2,…, x_k) avec les effectifs (n_1, n_2,…, n_k) tel que $n = \sum_{i=1}^{k} n_i$

La moyenne harmonique est l'inverse de la moyenne arithmétique des inverses des valeurs de X.

$$H = \frac{n}{\sum_{i=1}^{n} \left(\frac{1}{X_i}\right)}$$

- La moyenne arithmétique des inverses des valeurs de X c'est $\frac{1}{H} = \frac{\sum_{i=1}^{n} \frac{n_i}{x_i}}{n}$

L'utilisation de la moyenne harmonique est recommandée lorsque l'inverse des valeurs a un sens. On l'emploi dans le calcul des moyennes de pourcentages et de rapports ; notamment dans celui des durées moyennes et de vitesses moyennes.

II.2.3- La moyenne quadratique :

Soit X = (x_1, x_2,…, x_k) avec les effectifs (n_1, n_2,…, n_k) tel que $n = \sum_{i=1}^{k} n_i$

La moyenne quadratique est la racine carrée de la moyenne arithmétique des carrés des x_i.

$$Q = \sqrt{\frac{\sum_{i=1}^{k} n_i x_i^2}{n}} = \left(\frac{\sum_{i=1}^{k} n_i x_i^2}{n}\right)^{\frac{1}{2}}$$

II.2.4- Moyenne d'ordre « r », ou moyenne généralisé :

$$M_r = \left[\frac{1}{2}\sum_{i=1}^{K} n_i x_i^r\right]^{\frac{1}{r}} \text{ avec } n = \sum_{i=1}^{k} n_i$$

Si $r = 1 \Leftrightarrow M_r = \overline{X}$ (Moyennearithmétique)

Si $r = 2 \Leftrightarrow M_r = Q$(Moyennequadratique)

Si $r = -1 \Leftrightarrow M_r = H$ (Moyenneharmonique)

Si $r = \varepsilon \Leftrightarrow M_r = G$ (Moyennegéométrique)

Avec ε nombre positif trés faible.

<u>II.3- La médiane (M_e) :</u>

a- <u>Définition :</u>

La médiane M_e d'une variable statistique est la valeur numérique qui partage la série préalablement rangée par ordre croissant ou décroissant en deux parties égales.

b- <u>Calcul de la médiane :</u>

- <u>Le cas des effectifs impairs :</u>

Exemple :

Lors d'un contrôle, dans une classe de 11 élèves, les notes obtenues ont été les suivantes :
04 – 16 – 15 – 10 – 09 – 11 – 06 – 12 – 11 – 15 – 14 – 13 – 08

Puisqu'il y a 11 élèves, la 7ème note partagera la classe en deux groupes de même effectif (6 élèves auront une note inférieure ou égale à cette 7ème note et 6 élèves auront une note supérieure ou égale à cette 7ème note).

En classant les 11 notes (par ordre croissant), nous avons :

$$\underbrace{04 - 06 - 08 - 09 - 10 - 11}_{\text{7 notes inférieure ou égale à 11}} - 11 - \underbrace{12 - 13 - 14 - 15 - 15 - 16}_{\text{7 notes supérieure ou égale à 11}}$$

11 est la valeur médiane

- Le cas des effectifs pairs :

Exemple :

Considérons les notes d'un devoir (effectif de la classe : 14), ordonnons ces notes par ordre croissant.

04 - 07 - 07 - 08 - 10 - 10 - 11 - 12 - 14 - 15 - 15 - 18 - 18 - 20

Partageons cette série statistique en deux groupes de même effectif. Comme l'effectif total est 14, nous pouvons créer deux groupes de 7 notes.

$$\underbrace{04 - 07 - 07 - 08 - 10 - 10 - 11}_{\text{7 notes dans le 1}^{\text{er}}\text{ groupe}} - \underbrace{12 - 14 - 15 - 15 - 18 - 18 - 20}_{\text{7 notes dans le 2}^{\text{éme}}\text{ groupe}}$$

Il n'y a pas de note correspondant à la valeur centrale. N_i 11 (dernière note du premier groupe), N_i 12 (première note du second groupe) ne partagent la série en deux groupes de même effectif.

Dans ce cas, la valeur centrale est le "milieu", c'est à dire la moyenne des valeurs centrales. La médiane est donc égale à :

$$\frac{11+12}{2} = 11{,}5$$

La médiane de cette série statistique est 11,5

- Dans le cas d'une variable statistique continue, la médiane existe toujours ;

La médiane divise la série de xi éléments en deux sous-ensembles égaux. Cette valeur (?) se trouve dans la classe [...-...[qui est la classe médiane.

$$\frac{M_e - BI}{BS - BI} = \frac{\frac{n}{2} - F(i-1)}{F(i) - F(i-1)} \Rightarrow M_e = BI + \left(\frac{\frac{n}{2} - F(i-1)}{F(i) - F(i-1)} \right)(BS - BI)$$

BI = borne inferieur de la classe médiane
BS = borne supérieur de la classe médiane
$F(i)$ = fréquence relative cumulée de la classe i
$F(i\text{-}1)$ = fréquence relative cumulée de la classe i – 1

Interprétation :

- Il y a $\frac{n}{2}$ de n_i qui ont un x_i inférieur à M_e et $\frac{n}{2}$ des autres qui ont un x_i supérieur à M_e.
- ($\frac{n}{2}$ x100) % des n_i qui ont un x_i inférieur à M_e et ($\frac{n}{2}$ x100) % ont un x_i supérieur à M_e. ($\frac{n}{2}$ x 100 = pourcentage)

- Dans le cas d'une variable statistique discrète, la médiane peut ne pas exister ;

Le calcul de la médiane passe par la construction d'une colonne F(x) ou N(x) de façon à déterminer la médiane telle que F(M_e)=1/2=50% ou N(M_e) = n/2

La médiane est la valeur x_i qui correspond à la ligne la plus basse des deux.

Exemple : soit le tableau suivant qui représente 320 appartements classés par le nombre des chambres :

x_i	n_i	N(x)	f_i	F(x)
1	47	47	0,15	0,15
2	63	110	0,20	0,35
3	152	262	0,47	0,82
4	38	300	0,12	0,94
5	20	320	0,06	1
Total	320			

La médiane est la valeur centrale de x_i : M_e = 3 F(x)=1/2=50% ou N(x) = 320/2 = 160

Interprétation :

M_e = 3, cela veut dire qu'il y a 50% (ou 160) d'appartements ayant moins de 3 chambres et 50% (ou 160) d'appartements ayants plus de 3 chambres.

Interprétation :

M_e = ?, cela veut dire qu'il y a 50% (ou n/2) des n_i ayant moins de M_e des x_i et 50% (ou n/2) des n_i ayant plus de M_e des x_i.

- Détermination graphique :

A partir de ce diagramme en bâtons, nous réalisons un diagramme représentant les effectifs cumulés en fonction du nombre de x_i. Le bâton qui situé au centre des autres bâtons c'est le bâton qui présente la valeur médiane.

Exemple :

Note / 20 x_i	Nombre des élèves n_i	Effectifs cumulés croissantes
6	2	2
8	6	8
10	9	17
12	9	26
15	4	30
Total	30	

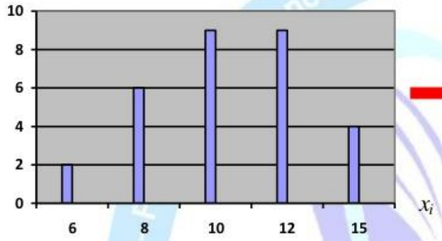


Diagramme des effectifs

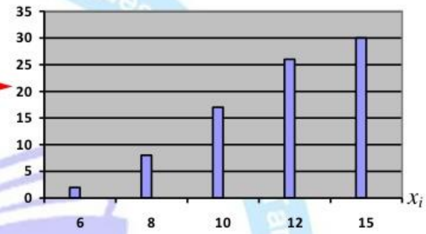


Diagramme des effectifs cumulés

La valeur médiane d'après la détermination graphique c'est 10.

II.4- Le mode (M_o) :

Le mode M_o est la valeur maximale de la variable où s'effectif le plus grand.

- Si la variable est discrète :

Le mode est bien défini : il correspond à la valeur X_i la plus fréquente dans un tableau ou un effectif plus grand

- Si la variable est continue :

Le mode est défini par la classe modale qui correspond l'effectif plus grand.

$$M_O = L_1 + \frac{d_1}{d_1 + d_2}(L_2 - L_1)$$

L_1 = borne <u>inférieur</u> de la classe modale
L_2 = borne <u>supérieur</u> de la classe modale
d_1 = différence entre l'effectif de la classe modale et l'effectif de la classe <u>inférieur</u> à la classe modale
d_2 = différence entre l'effectif de la classe modale et l'effectif de la classe <u>supérieur</u> à la classe modale

Interprétation :

La classe modale est …: c'est la classe à laquelle corresponde le plus grand effectif corrigé.

II.5- La médiale (M_l) :

a- *Définition :*

La médiale M_l est la valeur qui partage la masse $x_i.n_i$ en deux sous-ensembles égaux. Le calcul de la médiale passe par la formule de l'interpolation linéaire en utilisant la colonne de fréquences relatives cumulées croissantes F(x).

Remarque pour l'interprétation : Pour une distribution statistique donnée, la médiale est toujours : $M_l \geq M_e$

$$M_I = BI + \left(\frac{0,5 - F(i-1)}{F(i) - F(i-1)}\right)(BS - BI)$$

- *Les quantiles :*

Ce sont les valeurs du caractère x_i qui partagent la série en quatre sous-ensembles égaux. Ils sont donc au nombre de 9 : D_1 D_2 D_3. Ils contiennent des 10% des observations, soit un dixième de l'effectif (n/10)

- *Les quartiles :*

Ce sont les valeurs du caractère x_i qui partagent la série en quatre sous-ensembles égaux. Ils sont donc au nombre de trois : Q_1 Q_2 Q_3. Ils contiennent des 25% des observations, soit un quart de l'effectif (n/4)

- *Les déciles :*

Ce sont les valeurs du caractère x_i qui partagent la série en dix sous-ensembles égaux. Ils sont donc au nombre de 9 : D_1 D_2 D_3. Ils contiennent des 10% des observations, soit un dixième de l'effectif (n/10)

- *Les Centiles :*

Ce sont les valeurs du caractère x_i qui partagent la série en cent sous-ensembles égaux. Ils sont donc au nombre de 99 : P_1 P_2 P_3. Ils contiennent des 10% des observations.

b- *Exemple de la médiale :*

Soit la série suivant représentant un échantillon de 43 salariés selon leurs classes de salaire net exprimé en milliers de dirham :

Salaires en milliers DHs x_i	Effectifs n_i	Centre C_i	f_i	F(X)
[4 - 6[	5	5	0,11	0,11
[6 - 8[	8	7	0,19	0,30
[8 - 10[	12	9	0,28	0,58
[10 - 12[	10	11	0,23	0,80
[12 - 14[	8	13	0,19	1,00
Total	43		1	

1) Détermination de la médiane :

- La classe médiane : [8 - 10[
- Interpolation linière :

$$M_e = BI + \left(\frac{0,5 - F(i-1)}{F(i) - F(i-1)}\right)(BS - BI)$$

$$M_e = 8000 + \left(\frac{0,5 - 0,30}{0,58 - 0,30}\right)(10000 - 8000)$$

$$M_e = 9428 Dhs$$

2) Détermination de la médiale :
- La classe médiale : [10 - 12[

$$M_l = BI + \left(\frac{0,5 - F(i-1)}{F(i) - F(i-1)}\right)(BS - BI)$$

$$M_l = 10000 + \left(\frac{0,5 - 0,47}{0,74 - 0,47}\right)(12000 - 10000)$$

$$M_l = 10222 Dhs$$

Section III – Les caractéristiques de dispersion :

III.1- Étendue :

L'étendue est la différence entre la plus grand et la plus petite des valeurs possibles de la série. On écrit :

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

III.2- Intervalle interquartile (I) :

C'est la différence entre le troisième quartile et le premier quartile. Il contient 50% des observations.

$$I = Q_3 - Q_1$$

1^er quartile (Q_1) : 0,25
2^éme quartile (Q_2) : 0,50
3^éme quartile (Q_3) : 0,75

Pour calculer les quartiles, on a utilisé l'interpolation linéaire :

$$\Rightarrow \frac{Q_1 - L_1}{L_2 - L_1} - \frac{0,25 - f_i(L_1)}{f_i(L_2) - f_i(L_1)}$$

$$Q_1 = L_1 + \frac{0,25 - f_i(L_1)}{f_i(L_2) - f_i(L_1)}(L_2 - L_1)$$

Remarque : On peut utiliser pour calculer Q_3 : 0,75.

Interprétation :

50% des n_i ont un x_i compris entre Q_3 et Q_1 ; 25% des n_i ont un x_i inférieur à Q_1 et 50% des n_i ont un x_i supérieur à Q_3.

III.3- Écart absolue moyenne par rapport à la moyenne (e) :

C'est la moyenne arithmétique des écarts (en valeurs absolues) entre chacune des valeurs possibles de la variable x et la moyenne arithmétique x. on note :

$$e = \frac{\sum_{i=1}^{k} n_i \left| x_i - \overline{X} \right|}{n} = \sum_{i=1}^{k} f_i \left| x_i - \overline{X} \right|$$

Interprétation :

En moyenne : Les x_i des n_i s'écartent d'environ ***e*** de la moyenne arithmétique des x_i

Remarque :

Nous pouvons définir aussi l'écart absolu par à la médiane qui s'écrit :

$$e = \frac{\sum_{i=1}^{k} n_i \left| x_i - M_e \right|}{n} = \sum_{i=1}^{k} f_i \left| x_i - M_e \right|$$

III.4- Variance (*V(x)*) :

C'est la moyenne arithmétique des carrés des écarts des valeurs X par rapport à leur moyenne arithmétique.

$$V(x) = \sigma^2 = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{X} \right)^2}{n} \text{ ou } V(x) = \sigma^2 = \sum_{i=1}^{k} f_i \left(x_i - \overline{X} \right)^2$$

III.5- Écart-type (σ) :

C'est la racine carrés positive de la variance. $\sigma = \sqrt{V(x)}$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{X} \right)^2}{n}} \text{ ou } \sigma = \sqrt{\sum_{i=1}^{k} f_i \left(x_i - \overline{X} \right)^2}$$

III.6- Coefficient de variation (C_V) :

Le coefficient de variation à la moyenne d'une distribution est le rapport de l'écart-type à la moyenne arithmétique :

$$C_v = \frac{\sigma}{\overline{X}}$$

Table des matières