Module : Statistiques Descriptives

I- TERMINOLOGIE :

1. **Statistique :** La statistique est une méthode scientifique dont l'objet est de recueillir, d'organiser, de résumer et d'analyser les données d'une enquête, d'une étude o d'une expérience, aussi bien que de tirer les conclusions logiques et de prendre les décisions qui s'imposent à partir des analyses effectuées.
2. **Population :** Ensemble d'individus définis par une propriété commune donnée.

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

3. **Echantillon :** Sous-ensemble de la population.

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

4. **Individu ou unité statistique :** Chaque élément de la population ou de l'échantillon.

Exp : dans l'exemple précédant, chaque ampoule constitue un individu ou une unité statistique.

5. **La taille :** Représente le nombre d'individus d'un échantillon ou d'une population. Elle est symbolisée par « n » dans le cas d'un échantillon et par « N » dans le cas d'une population.
6. **Le caractère :** C'est l'aspect particulier que l'on désire étudier.

Exp : concernant un groupe de personnes, on peut s'intéresser à leur âge, leur sexe leur taille…

7. **Les modalités:** Les différentes manières d'être que peut présenter un caractère.

Exp 1 : le sexe est un caractère qui présente deux modalités : féminin ou masculin
Exp 2 : quant au nombre d'enfants par famille, les modalités de ce caractère peuvent être 0,1 2,3…,20.

8. **Caractère qualitatif :** Ses modalités ne s'expriment pas par un nombre

Exp : la religion, le sexe, l'opinion…

9. **Caractère quantitatif :** Ses modalités sont numériques.

Exp : l'âge, la taille, le poids…

10. **Caractère quantitatif discret :** L'ensemble des valeurs que peut prendre le caractère est fini ou dénombrable. Le plus souvent, ces valeurs sont entières.

Exp : le nombre d'enfant dans une famille, le nombre de téléviseurs par foyer et la pointure des souliers.

11. **Caractère quantitatif continu :** Le caractère peut prendre théoriquement n'importe quelle valeur dans un intervalle donné de nombres réels.

Exp : la taille d'un individu, le poids…

12. **Série statistique :** L'ensemble des différentes données associées à un certain nombre d'individus.

Exp : la série suivante résulte d'une courte enquête auprès de quelques personnes pour connaître leur âge :
18 21 19 19 17 22 27 18 18 17 20 20 23

II- TABLEAUX STATISTIQUES

Le tableau brut se présente sous la forme suivante:

Individu	Variable
1	X_1
2	X_2
.	.
.	.
.	.
n	X_n

Exemple d'application : On une série comme suit :
1.3.5.6.7.9.4.5.2.3.4.5.8.7.9.8.9.10.10.1.2.4.3.5.6.9.7.8.1.5
TAF 1 : présenter cette série dans un tableau.
TAF 2 : présenter dans un tableau les ECC (effectifs cumulé croissant), les ECD (effectifs cumulé décroissant), les Fi (fréquences), les FCC (fréquences cumulé croissant), les FCD (fréquences cumulé décroissant).

Exercice : soient des séries suivants :

1^er^ série : 3.1.5.6.4.3.8.7.5.6.9.10.3.4.5.6.2.3.4.5.6.9.8.6.4.3.2.2.4.5.10.10.10. 2.2.3.4.6.7.5
2^eme^ série: 8.7.5.6.9.10.3.4.5.6.2.3.4.5.6.9.8.6.4.3.2.2.4.5.10.10.10. 2.2.3.4.6.7.5.4.3.2.3.4.5
3^eme^ série: 6.4.3.8.7.5.6.9.10.3.4.5.6.2.3.4.5.6.9.8.6.4.3.2.2.4.5.10.10.10. 2.2.3.4.6.7.5.4.3.2
4^eme^ série: 4.3.2.3.4.5.6.4.3.8.7.5.6.9.10.3.4.5.6.2.3.4.5.6.9.8.6.4.3.2.2.4.5.10.10.10. 2.2.3.4

TAF : présenter dans un tableau les ECC (effectifs cumulé croissant), les ECD (effectifs cumulé décroissant), les Fi (fréquences), les FCC (fréquences cumulé croissant), les FCD (fréquences cumulé décroissant). pour chaque série

Elabore par : Mohamed OUTRGA 06 91547731

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

- Pour une variable ***qualitative***, les modalités ne sont pas mesurables.
- Pour une variable ***quantitative***, les modalités sont mesurables. Ce sont :
  - des valeurs numériques ponctuelles lorsque la variable est *discrète*
  - des intervalles lorsque la variable est *continue*

Modalités	Effectifs
C_1	n_1
C_2	n_2
.	
.	
C_k	n_k

Application :
Nous étudions une population de 1000 entreprises selon le caractère modalité « forme juridique ».
Les modalités retenues : S.A (Société Anonyme), SARL (Société A Responsabilité Limitée), EI (Entreprise Individuelle), SNC (Société en Nom Collectif).
Leurs effectifs respectifs : 200, 400, 340, 60.

T.A.F :
1. Préciser la population étudiée, sa taille, le caractère sa nature, et le nombre de modalités
2. Présentez cette série dans un tableau.

Exercice 1 :

Dans un ensemble résidentiel, on considère 320 appartements classés selon le nombre de pièce, soit le tableau suivant :

Nombre de pièces	*ni*
1	47
2	63
3	152
4	38
5	20
TOTAL	320

Travail à faire :
1-Préciser la population étudiée, sa taille, le caractère sa nature, et le nombre de modalités
2-Calculer les fréquences relatives

Exercice 2 :

Les renseignements fournis par 1000 membres d'une association sortie ont été consignés dans les tableaux suivants :

Tableau N° 01 :

Situation familiale	Effectifs
Célibataire	476
Marié	396
Veuf	51
Divorcé	77
TOTAL	1000

Tableau N° 02 :

Nombre d'enfant à charge	Effectifs
0	498
1	11
2	114
3	129
4	100
5	48
TOTAL	1000

Tableau N° 03 :

Age	Effectifs
Moins de 20	105
20-30	456
30-40	242
40-50	179
50 et plus	18
TOTAL	1000

Travail à faire :
1/ Préciser la population étudiée et sa taille.
2/Préciser le caractère, la nature de caractère, le nombre de modalité dans les tableaux ci-dessus.
3/Interpréter la valeur 51, 114,105.
4/ Calculer les fréquences pour les trois tableaux.

Elabore par : Mohamed OUTRGA 06 91547731

III- REPRÉSENTATIONS GRAPHIQUES

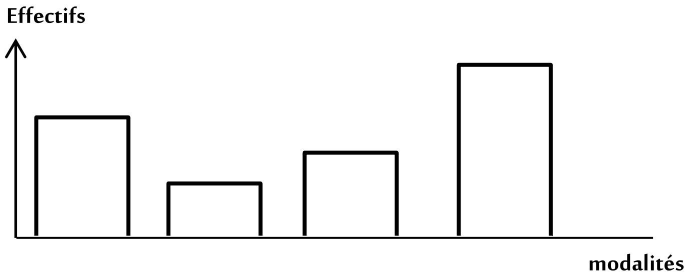
Lorsqu'on observe un caractère sur des individus, on aboutit à un tableau de chiffres peu parlant. L'objectif est de donner une représentation graphique de ce tableau qui permette d'un seul coup d'œil d'avoir une idée de la manière dont se répartissent les individus.

A- Variable qualitative :

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

  - les modalités de la variable sont placées sur une droite horizontale (attention: ne pas orienter cette droite car les modalités ne sont pas mesurables et il n'y a donc pas de relation d'ordre entre elles).

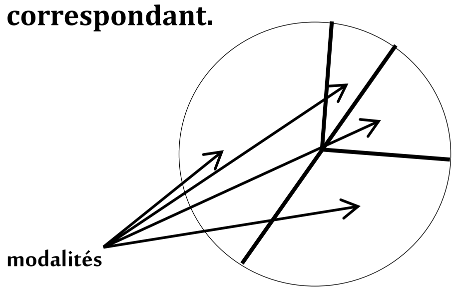
  - les effectifs (ou les fréquences) sont placés sur un axe vertical. La hauteur du tuyau est proportionnelle à l'effectif.



- les diagrammes à secteurs (ou camemberts)

  - L'effectif total est représenté par un disque.

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



Application :

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

Série du Bac xi	Nombre de candidats ni
Sciences économiques	200
Sciences mathématiques	300
Sciences expérimentales	500
Total	1000

TAF:

représentez cette distribution en Tuyaux d'orgues et Diagramme circulaire.

B- Variable quantitative :

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

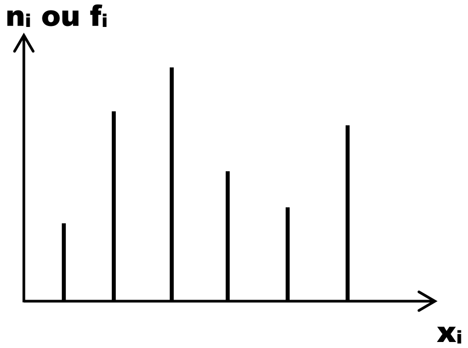
Deux types de graphiques sont intéressants de représenter:

<u>a) les diagrammes différentiels</u> qui mettent en évidence les différences d'effectifs (ou de fréquences) entre les différentes modalités ou classes.

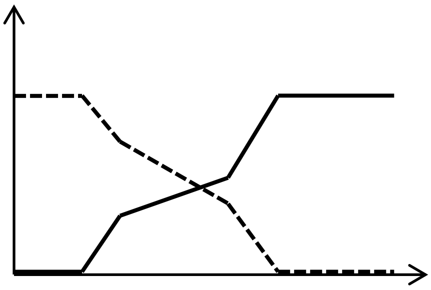
<u>b) les diagrammes cumulatifs</u> qui permettent de répondre aux questions du style "combien d'individus ont pris une valeur inférieure (ou supérieure) à temps?".

1) Variable discrète

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



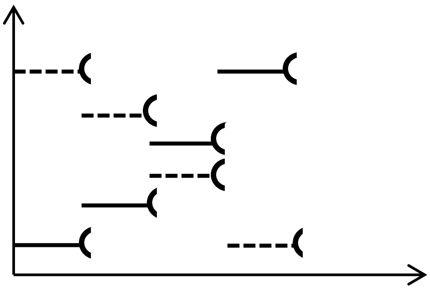
Diagrammes cumulatifs :
ils permettent de visualiser l'évolution des effectifs (fréquences) cumulés croissants ou décroissants.



Se souvenir:

(au plus x) équivalent à (< x) donc utiliser N(x) ou F(x)

(plus que x) équivalent à (> x) donc utiliser N '(x) ou F '(x)



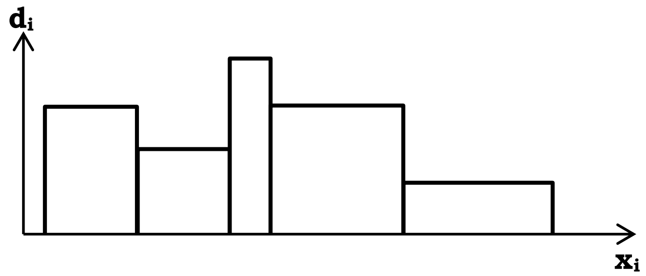
Application :

Représentez graphiquement la distribution des 50 étudiants en fonction du nombre de personnes par ménage suivante :	Nombre de personnes par ménage xi	Nombre d'étudiants ni
	3	5
	4	15
	6	15
	7	10
	8	5
	Total	50

2) Variable classée

Diagramme différentiel (l'histogramme) : C'est un ensemble de rectangles contigus, chaque rectangle associé à chaque classe ayant une surface proportionnelle à l'effectif (fréquence) de cette classe.

Attention: Avant toute construction d'histogramme, il y a lieu de regarder si les classes sont d'amplitudes égales ou inégales.



Le cas des classes d'amplitudes égales ne pose aucune difficulté car il suffit de reporter en ordonnée l'effectif (la fréquence).

Dans le cas d'amplitudes inégales on reporte en ordonnée la densité di (effectif divisé par l'amplitude de la classe)

Application :

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

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

IV- LES PARAMÈTRES DE TENDANCE CENTRALE ET DE POSITION

<table>
<tr><th>1. Moyenne Arithmétique ($\bar{X}$) :</th><th>Formule</th><th>Exemple</th></tr>
<tr><td>Cas 1 : Série statistique</td><td>$\bar{X} = \frac{1}{N}\sum x_i$</td><td>1.2.3.4.5.6.7.8.9
$\bar{X} = \frac{1}{9} \times (1+2+3+4+5+6+7+8+9) = \boxed{5}$</td></tr>
<tr><td>Cas 2 : variable statistique discret</td><td>$\bar{X} = \frac{1}{N}\sum n_i x_i$</td><td>
<table>
<tr><th>x_i</th><th>n_i</th><th>$n_i x_i$</th></tr>
<tr><td>1</td><td>4</td><td>4</td></tr>
<tr><td>2</td><td>6</td><td>12</td></tr>
<tr><td>3</td><td>2</td><td>6</td></tr>
<tr><td>4</td><td>8</td><td>32</td></tr>
<tr><td>Total</td><td>20</td><td>54</td></tr>
</table>
$\bar{X} = \frac{1}{N}\sum n_i x_i = \frac{1}{20} \times 54 = \boxed{2,7}$</td></tr>
<tr><td>Cas 3 : variable statistique contenu</td><td>$\bar{X} = \frac{1}{N}\sum n_i c_i$</td><td>
<table>
<tr><th>x_i</th><th>n_i</th><th>c_i</th><th>$n_i c_i$</th></tr>
<tr><td>1 - 2</td><td>2</td><td>(1+2)/2=1,5</td><td>3</td></tr>
<tr><td>2 - 3</td><td>2</td><td>2,5</td><td>5</td></tr>
<tr><td>3 - 4</td><td>2</td><td>3,5</td><td>7</td></tr>
<tr><td>4 - 5</td><td>2</td><td>4,5</td><td>9</td></tr>
<tr><td>total</td><td>8</td><td>---</td><td>24</td></tr>
</table>
$\bar{X} = \frac{1}{N}\sum n_i c_i = \frac{1}{8} \times 24 = \boxed{3}$</td></tr>
</table>

<table>
<tr><th>2. Mode (Mo) :</th><th>Formule</th><th>Exemple</th></tr>
<tr><td>Cas 1 : Série statistique</td><td rowspan="2">C'est la valeur observée d'effectif maximum</td><td>1.2.3.4.5.6.7.8.9 → n'est pas de Mode
1.2.3.4.3.5.6.3.1 → le Mode est 3
1.2.3.4.2.5.6.4.7 → le Mode est 2 et 4</td></tr>
<tr><td>Cas 2 : variable statistique discret</td><td>
<table>
<tr><th>x_i</th><th>n_i</th></tr>
<tr><td>1</td><td>4</td></tr>
<tr><td>$\boxed{2}$</td><td>8</td></tr>
<tr><td>3</td><td>2</td></tr>
<tr><td>4</td><td>6</td></tr>
<tr><td>Total</td><td>20</td></tr>
</table>
le Mode est 2</td></tr>
<tr><td>Cas 3 : variable statistique contenu</td><td>$Mo = l_0 + a_i\left(\frac{d_1}{d_1 + d_2}\right)$</td><td>
<table>
<tr><th>x_i</th><th>n_i</th></tr>
<tr><td>1 - 2</td><td>10</td></tr>
<tr><td>2 - 3</td><td>20</td></tr>
<tr><td>3 - 4</td><td>30</td></tr>
<tr><td>4 - 5</td><td>10</td></tr>
<tr><td>total</td><td>70</td></tr>
</table>
$Mo = l_0 + a_i\left(\frac{d_1}{d_1 + d_2}\right)$
$= 3 + (4-3) \times \left(\frac{30-20}{(30-20)+(30-10)}\right)$
$= \boxed{3,5}$</td></tr>
</table>

Elabore par : Mohamed OUTRGA 06 91547731

3. Médiane (Me) :

1er cas N est paire

<table>
<tr><th></th><th>Formule</th><th>Exemple</th></tr>
<tr><td>Cas 1 : Série statistique</td><td>$Me = \frac{1}{2}\left[x\left(\frac{N}{2}\right) + x\left(\frac{N}{2}+1\right)\right]$</td><td>1 . 2 . 3 . 4 . 5 . 6 . 7 . 8 N = 8
x_4 x_5
$Me = \frac{1}{2}\left[x\left(\frac{8}{2}\right) + x\left(\frac{8}{2}+1\right)\right] = \frac{1}{2}(x_4 + x_5) = \frac{1}{2}(4+5) = \boxed{4,5}$</td></tr>
<tr><td>Cas 2 : variable statistique discret</td><td>$x_{\frac{n}{2}} \le Me \le x_{\frac{n}{2}+1}$
$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$</td><td><table><tr><th>$x_i$</th><th>$n_i$</th><th>$n_i \nearrow$</th></tr><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td>4</td><td>5</td></tr><tr><td>3</td><td>2</td><td>7</td></tr><tr><td>4</td><td>3</td><td>10</td></tr><tr><td>Total</td><td>10</td><td>---</td></tr></table>N = 10 , $\frac{N}{2} = 5$
$5 \le Me \le 7$
$Me = \frac{x_5 + x_7}{2} = \frac{2+3}{2} = \boxed{2,5}$</td></tr>
<tr><td>Cas 3 : variable statistique contenu</td><td>$Me = l_0 + a_i\left(\frac{\frac{N}{2} - (N_i - 1)}{n_i}\right)$</td><td><table><tr><th>$x_i$</th><th>$n_i$</th><th>$n_i \nearrow$</th></tr><tr><td>1 - 2</td><td>10</td><td>10</td></tr><tr><td>2 - 3</td><td>20</td><td>30</td></tr><tr><td>3 - 4</td><td>30</td><td>60</td></tr><tr><td>4 - 5</td><td>10</td><td>70</td></tr><tr><td>total</td><td>70</td><td>---</td></tr></table>N = 70 , $\frac{N}{2} = 35$
$Me = 3 + 1\left(\frac{\frac{70}{2} - 30}{30}\right) = \boxed{3,17}$</td></tr>
</table>

2éme cas N est impaire

<table>
<tr><th></th><th>Formule</th><th>Exemple</th></tr>
<tr><td>Cas 1 : Série statistique</td><td>$Me = x\left(\frac{N+1}{2}\right)$</td><td>1 . 2 . 3 . 4 . 5 . 6 . 7 . 8 . 9 N = 9
$Me = x\left(\frac{N+1}{2}\right) = x\left(\frac{9+1}{2}\right) = x\frac{10}{2} = x_5 = \boxed{5}$</td></tr>
<tr><td>Cas 2 : variable statistique discret</td><td>La médiane est l'observation de rang $\frac{n+1}{2}$
$Me = x_{\frac{n+1}{2}}$</td><td><table><tr><th>x_i</th><th>n_i</th><th>$n_i \nearrow$</th></tr><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td>4</td><td>5</td></tr><tr><td>3</td><td>1</td><td>6</td></tr><tr><td>4</td><td>3</td><td>9</td></tr><tr><td>Total</td><td>9</td><td>----</td></tr></table>N = 9 , $\frac{N+1}{2} = 5$
$Me = x_{\frac{n+1}{2}} = x_5 = \boxed{2}$</td></tr>
<tr><td>Cas 3 : variable statistique contenu</td><td>$Me = l_0 + a_i\left(\frac{\frac{N+1}{2} - (N_i - 1)}{n_i}\right)$</td><td><table><tr><th>$x_i$</th><th>$n_i$</th><th>$n_i \nearrow$</th></tr><tr><td>1 - 2</td><td>10</td><td>10</td></tr><tr><td>2 - 3</td><td>20</td><td>30</td></tr><tr><td>3 - 4</td><td>30</td><td>60</td></tr><tr><td>4 - 5</td><td>11</td><td>71</td></tr><tr><td>total</td><td>71</td><td>----</td></tr></table>N = 71 , $\frac{N+1}{2} = 36$
$Me = 3 + 1\left(\frac{\frac{71+1}{2} - 30}{30}\right) = \boxed{3,2}$</td></tr>
</table>

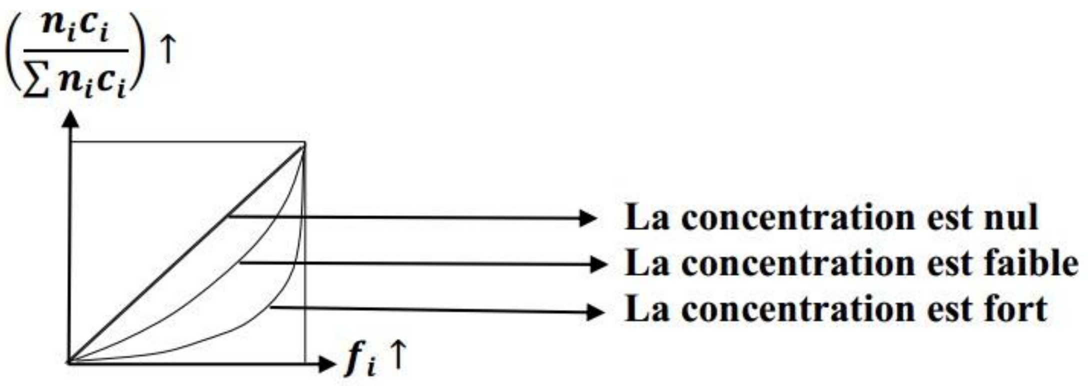
4. Médiale ($M\ell$) : la médiale est partage la masse total $\sum n_i c_i$ en deux parties égaux

V- LES PARAMÈTRES DE DISPERSION

1. Etendue (E) :	$X_{max} - X_{min}$	
2. Variance (σ^2) :	Cas 1 : Série statistique	$\sigma^2 = \frac{1}{N}\sum x_i^2 - \bar{X}^2$
	Cas 2 : variable statistique discret	$\sigma^2 = \frac{1}{N}\sum n_i x_i^2 - \bar{X}^2$
	Cas 3 : variable statistique contenu	$\sigma^2 = \frac{1}{N}\sum n_i c_i^2 - \bar{X}^2$
3. L'écart type (σ) :	$\sigma = \sqrt{\sigma^2}$	
4. Coefficient de variation (CV) :	$CV = \frac{\sigma}{\bar{X}} \times 100$	
5. Les quartiles	$Q_3 = l_0 + a_i\left(\frac{75\%N - (75\%N - 1)}{n_i}\right)$	I i q = $[Q_1; Q_3]$
	$Q_1 = l_0 + a_i\left(\frac{25\%N - (25\%N - 1)}{n_i}\right)$	E i q = $Q_3 - Q_1$
6. Les déciles	$d_1 = l_0 + a_i\left(\frac{10\%N - (10\%N - 1)}{n_i}\right)$	I i d = $[d_1; d_9]$
	$d_9 = l_0 + a_i\left(\frac{90\%N - (90\%N - 1)}{n_i}\right)$	E i d = $d_9 - d_1$

VI- LES PARAMÈTRES DE CONCENTRATION

a) La courbe de concentration de Lorentz :



b) L'indice de concentration de GINI

$$IC = \frac{\text{Médiale} - \text{Médiane}}{\text{Etendue}} \times 100$$

VII- EXERCICES :

Exercice 1 :

Soit le tableau suivant relatif au nombre d'enfants par famille:

Nombre d'enfants	ni
0	1
1	8
2	9
3	5
4	20
5	3
Total	46

1/- Calculez la moyenne arithmétique ? Et interpréter le résultat.
2/- Déterminez la valeur de du mode.
3/- Calculer la Médiane.
4/-Calculer la Variance ? Et déduire l'écart-type.
5/- Déterminez l'Etendue
6/ -Coefficient de variation

Exercice 2 :

On considère 75 ateliers d'artisans classés en fonction du nombre des heures travaillées :

Classes	ni
[50-70[	6
[70-100[	9
[100-130[	15
[130-150[	23
[150-180[	17
[180-200[	5
Total	75

1/- Calculez la moyenne arithmétique, et interpréter le résultat.
2/- Calculez le mode, et interpréter le résultat.
3/- Calculez la médiane, et donner leur signification.
4/- Calculez la Variance ? Et déduire l'écart-type.
5/- Calculez Coefficient de variation
6/- Calculez l'intervalle interquartile et interdécile
7/- Calculer L'indice de Gini ?
8/- Calculer la médiale ?

Exercice 3 :

On analyse les dépenses annuelles pour les loisirs des ménages, les résultats sont résumés dans le tableau ci-après.

Dépenses en (1000dh)	Nbre de ménage
0-10	10
10-20	30
20-30	40
30-40	70
40-50	50
Total	200

1/- Calculez la moyenne arithmétique, et interpréter le résultat.
2/- Calculez le mode, et interpréter le résultat.
3/- Calculez la médiane, et donner leur signification.
4/- Calculez la Variance ? Et déduire l'écart-type.
5/- Calculez Coefficient de variation
6/- Calculez l'intervalle interquartile et interdécile
7/- Calculer L'indice de Gini ?
8/- Calculer la médiale ?

Elabore par : Mohamed OUTRGA 06 91547731