

Chapitre I: Statistique descriptive

terminologie - tableau statistique - représentation graphique

I - Méthodologie

I. 1) Notion de 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 ou d'une expérience. Elle vise à tirer des conclusions logiques et de prendre les décisions qui s'imposent à partir des analyses effectuées. Une étude statistique porte toujours sur une population. Une population stat. est l'ens. des individus sur lesquelles portent l'étude stat. Si l'étude porte sur une partie de la population alors le référentiel sera appelé échantillon. Le nombre d'individus qui constituent la population représente la taille de la population. Une population stat. est étudiée au lieu d'un ou de plusieurs caractères. Un caractère stat. est une propriété qui est commune à l'ens. des individus qui constituent la population. Un caractère stat. peut se trouver dans une situation appelée modalité du caractère. Un caractère est dit qualitatif lorsque ces modalités ne peuvent pas être exprimées par des valeurs numériques. Un caractère est quantitatif lorsque ces modalités peuvent être exprimées par des valeurs numériques. Un caractère quantitatif est dit discontinu ou discret si ces modalités peuvent prendre des valeurs finies ou dénombrables. Le plus souvent, ce sont des valeurs entières. Un caract. quantitatif est dit continu si ces modalités peuvent

appartenir à un interval fixé à priori, les modalités sont exprimés par des intervalles ou classes

I. 2) Les Tableaux Statistiques

1- Cas d'une seule variable : tableau statistique à simple entrée

Le tableau brut se présente sous la forme suivante :

FR
à l'échelle
relative : pourcentage
de la somme des effectifs (N)

Indiv	C_i
1	O_1
2	O_2
3	O_3
...	...
n	...

Caractères	Effectif
C_1	n_1
C_2	n_2
...	...
C_k	n_k
Total	n

$$n = \sum_{j=1}^k n_j$$

definit le caractère quantitatif continu

FR
↑
%

Modalité	Eff	%	ECC
0	6	12	6
1	12	24	18
2	20	40	38
3	8	16	46
4	2	4	50
5	2	4	50
Total	50	100	

Ex

Revenue!	n_i	d_i	U_i	n_{ir}	LCC	ECC
100-200	15	0,15	2	15	15	100
200-300	10	0,10	2	10	25	85,00
300-400	30	0,30	2	30	55	45
400-600	60	0,20	2	20	85	45
600-700	3	0,03	1	3	88	5
700-900	1	0,005	2	0,5	89	2
900-1000	1	0,01	2	2	100	1
T	100					

amplitude $a_i = e_{i+1} - e_i$ extrémité

centre $C_i = \frac{e_i + e_{i+1}}{2} = m_i$

densité $d_i = \frac{n_i}{a_i}$

effectif rectifié (n_{ir}) : $n_{ir} = \frac{n_i}{U_i}$ nombre d'unité d'amplitude

$U_a =$ unité d'amplitude

$U_a = 100$

$U_i = \frac{a_i}{U_a}$ ou

$n_{ir} = d_i \times U_a$ / $n_{ir} = \frac{n_i}{\frac{a_i}{U_a}} = \frac{n_i \times U_a}{a_i}$

$\frac{\text{bas} - \text{haut}}{a_i}$

il ne faut jms calculer l'effectif cumulé rectifié.

2- Cas de 2 variables : tableau à double entrée :

Elle se présente sous cette forme :

X \ Y	y ₁	y ₂	y ₃	y _j	y _l	Total
x ₁	n _{1,1}	n _{1,2}	n _{1,3}	n _{1,j}	n _{1,l}	n _{1..}
x ₂	n _{2,1}	n _{2,2}	n _{2,3}	n _{2,j}	n _{2,l}	n _{2..}
x _i	n _{i,1}	n _{i,2}	n _{i,3}	n _{i,j}	n _{i,l}	n _{i..}
x _k	n _{k,1}	n _{k,2}	n _{k,3}	n _{k,j}	n _{k,l}	n _{k..}
Total	n _{.1}	n _{.2}	n _{.3}	n _{.j}	n _{.l}	n _{..}

distrib³
fondamentale

distrib² marginale dans une populat²

donne un caractère (3)
la répartit² d² seul caractère

X	eff
x ₁	n _{1..}
⋮	⋮
x _k	n _{k..}
T	n _{..}

Y	eff
y ₁	n _{.1}
y ₂	n _{.2}
⋮	⋮
y _l	n _{.l}
T	n _{..}

I. 3) Les représentations graphiques

08/12/2018

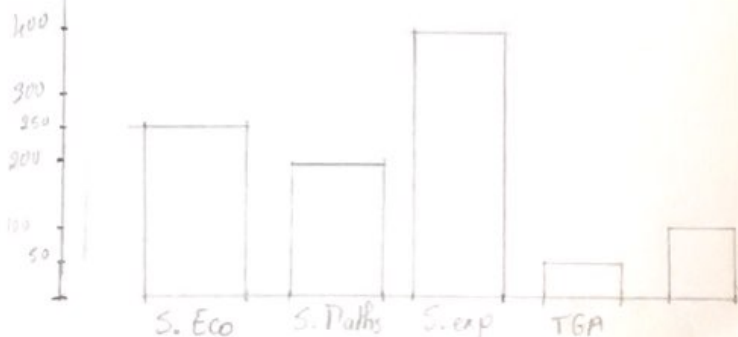
Ex. I. 3) 1- Représentation graphique d'un caract. qualitatif
Dans le tableau statistique donnant la répartition qualitative chaque modalité x_i est associé à un effectif n_i la seule représentat² qui ns intéresse est celle des effectifs n_i ou des

Suivant la variable observée, de nombreuses représentat² + ou - informatives peuvent être utiliser. Cependant les 2 les plus classiques sont : • les diagrammes en tuyaux d'orgues ou en barres ou diagramme en bande. Les modalités de la variable sont placés sur une droite horizontale (attention ne pas orientée cette droite car les modalités ne sont pas mesurables et il n'y a pas donc de relations d'ordres entre elles.

Les effectifs ou les fréquences sont placés sur un axe vertical, la hauteur du tuyau ou de la barre est proportionnelle à l'effectif.

Ex: la répartition des candidats convoqués pour participer à un test d'admissibilité à la formation en management est la suivante.

Serie	Effectif
S. Eco	250 $\times 90$
S. Maths	200 $\times 72$
S. exp	400 $\times 144$
TGA	50 $\times 18$
TGC	100 $\times 36$
TOTAL	1000



permet de représenter l'importance absolue de chaque modalité

• Les diagrammes à secteurs ou camembert
 L'effectif total est représenté par un disque. Chaque modalité est représentée par un secteur circulaire dont la surface (l'angle au centre) est proportionnel à l'effectif correspondant.

$360^\circ \rightarrow 1000$
 $\Rightarrow \frac{250 \times 360}{1000} = 90^\circ$
 $\frac{200 \times 360}{1000} = 72^\circ$
 $\frac{400 \times 360}{1000} = 144^\circ$
 $\frac{50 \times 360}{1000} = 18^\circ$

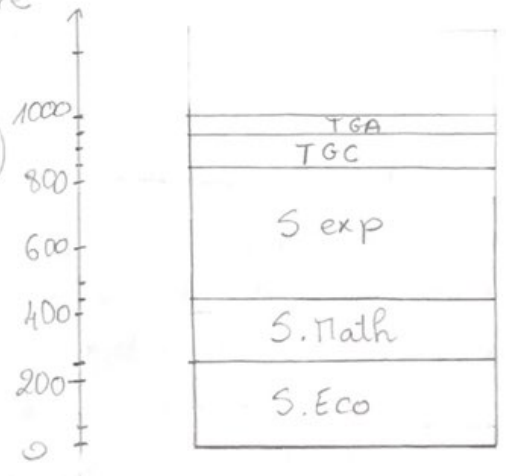


on calcule les pourcentages

- S. Eco = 25%
- S. Math = 20%
- S. exp = 40%
- TGC = 10%
- TGA = 5%

• ou en barre

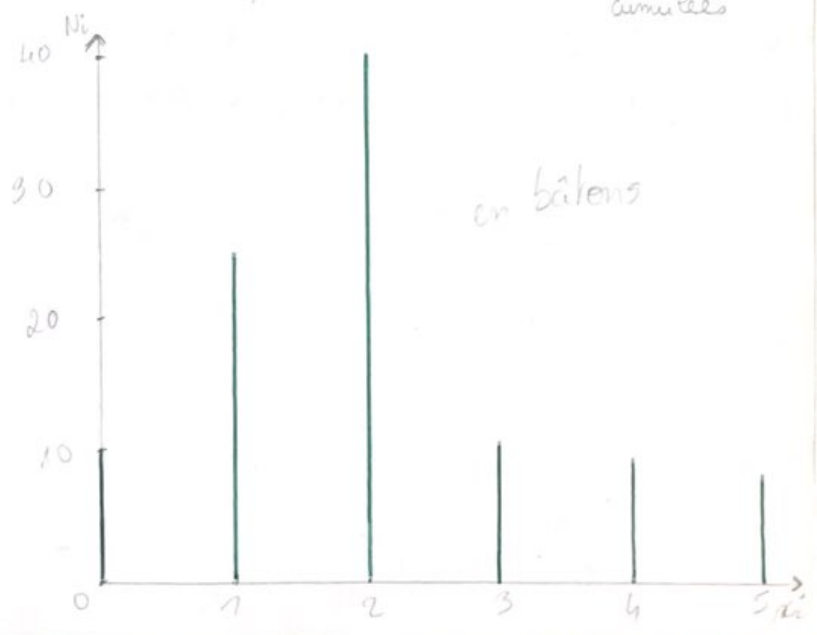
(doit être effacé ensuite)



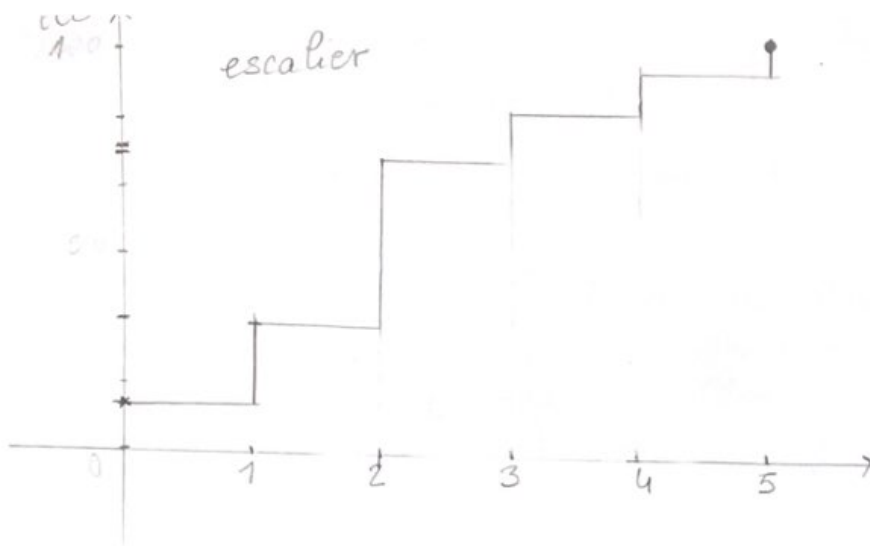
I.3) 2- Représentation graphique d'un caractère quantitatif
 Caract. q. continu \rightarrow diagramme différentiel (1) \rightarrow diagramme en bâtons
 Caract. q. discontinu \rightarrow diagramme intégral (ECC) (2) \rightarrow histogramme
 \rightarrow diagramme en escalier
 \rightarrow courbes de fréquences cumulée

Ex

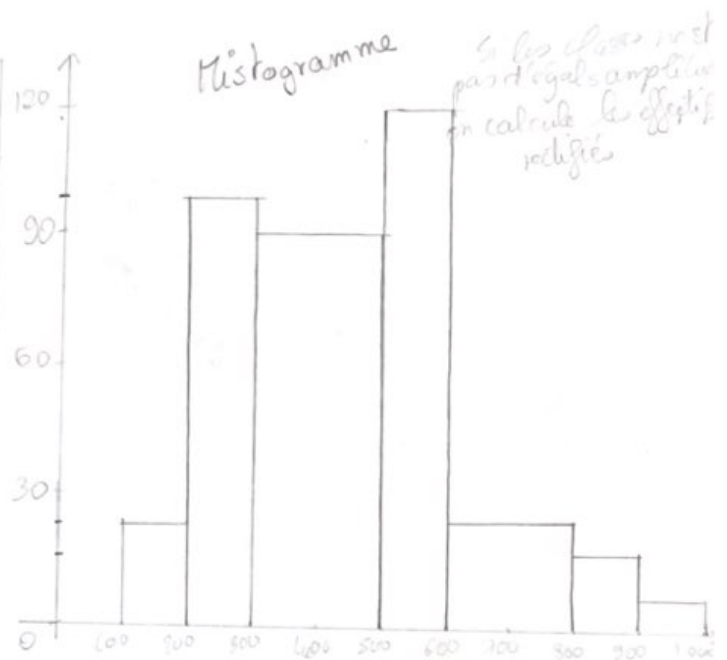
Nb d'enf.	Nb de ménages	ECC
0	10	10
1	25	35
2	40	75
3	10	85
4	8	93
5	7	100
Total	100	



en bâtons



Sal. men (um)	ni	fid	hi	hir	icc
100		500			0
200	25	475	1	25	25
300	100	375	1	100	125
500	180	195	2	90	305
600	120	75	1	120	425
800	50	25	2	25	475
900	15	10	1	15	490
1000	10	0	1	10	500
T	500				



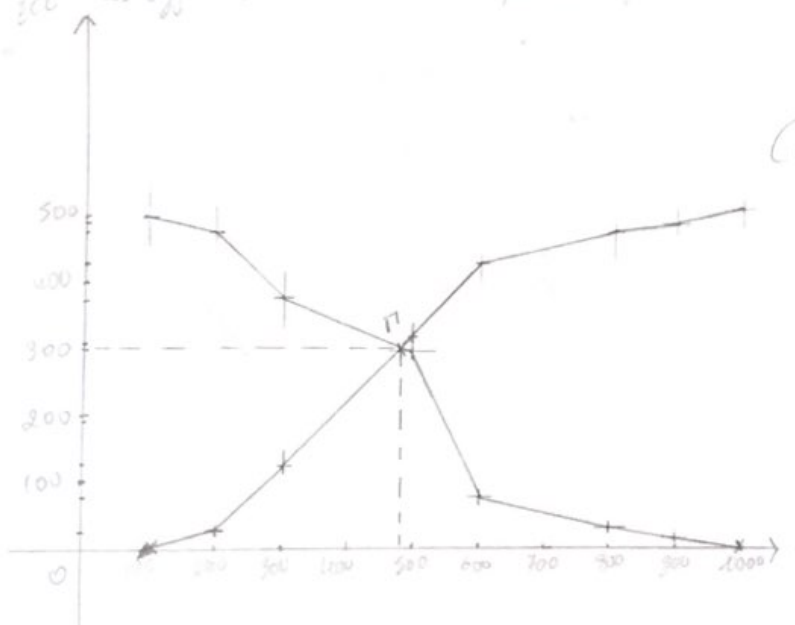
Les eff. relat. utilisés uniquement pour l'histogramme

on relie par des segments de droites;

Courbes des fréquences cumulées croissantes

Le point d'intersection

La médiane modale du caractère permettant de le diviser en 2



Chapitre II - Analyse numérique des séries statistiques

20/12/2018 (B)

I - Les paramètres de tendance centrale et de position

I-1) Le mode

C'est une modalité du caractère qui est le plus représentée.
Modalité qui possède l'effectif le plus représenté.
C'est la valeur observée d'effectif maximum. Cas d'un caractère quantitatif discontinu :

ECC	x_i	n_i	$\frac{n_i}{x_i}$
5	0	5	0
15	1	10	10
40	2	25	50
45	3	5	15
48	4	3	12
50	5	2	10
Σ		50	97

$M_0 = 2$

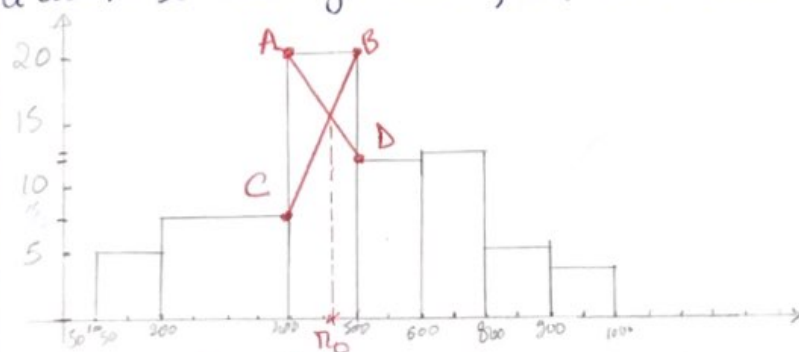
Classe modale: $[400; 500[$

$n_0 \approx 450$

Attention !

Classes	n_i	x_i	$n_i x_i$	ECC	$n_i x_i$
100	5	150	5	0	750
200	15	300	15/2	5	4500
400	20	450	20	20	9000
500	12	550	12	40	6600
600	25	700	25/2	52	17500
800	5	950	5	77	4250
900	3	950	3	82	2850
1000				85	
Total	85				45450

Pour déterminer le mode par le calcul, on représente d'abord l'histogramme ; en calculant l'abscisse du point d'intersection des deux droites passant par A, B, C, D



I-2) La médiane

Cas d'une série d'observation: $x_i: 1 \ 3 \ 4 \ 10 \ 1 \ 2 \ 7 \ 8$

① ranger les observations par val croissante ou décroissante

$x_n: 1; 1; 2; 3; 4; 7; 8; 10$

② regarder la parité du nombre d'observation

* n paire: $n = 2p$ avec $p = 4$ donc on aura un interval médian et elle sera det. par les observations qui occupent les positions p et $p+1 \Rightarrow [3; 4]$

* Si n impair $\Rightarrow n = 2p + 1$ on aura une médiane Me qui va occuper la position $(p + 1)$. ⑦

* **Cas d'un tableau statistique décrivant la distribution d'un caractère quantitatif discontinu**: La médiane se det. de la m^{re} façon que p_i en a une série statistique.

On calcule d'abord les ECC

n = paire : $n = 2p$ avec $p = 25$

$$Me = [2]$$

* // continu

Lorsque ns avons un tableau statistique qui décrit la distrib^{ti} d'un caract. quantitatif continu, on va considérer que la médiane est une modalité du caractère qui est associé à un $EC = 1/2 ET$ ou à une $FC = 50\%$

Ns voulons det la médiane dans ce tableau

	Mod. du caract	ECC
$Me \rightarrow 85/2 = 42,5$	1) 500	$\rightarrow 40$
\Rightarrow Classe médiane = $[500, 600[$	2) Me	$\rightarrow 85/2 = 42,5$
	3) 600	$\rightarrow 52$

On utilise la technique de l'interpolation linéaire consiste à faire une approximation: $\frac{(2) - (1)}{(3) - (1)} = \frac{Me - 500}{600 - 500} = \frac{42,5 - 40}{52 - 40} = \frac{2,5}{12}$

$$\frac{Me - 500}{600 - 500} = \frac{2,5}{12} \Rightarrow \frac{Me - 500}{100} = \frac{2,5}{12} \Rightarrow \frac{2,5 \times 100}{12} = \frac{1}{12} Me - \frac{1}{12} \times 500$$

$$\Rightarrow Me = 500 + \frac{2,5}{12} \times 100$$

I-3) La moyenne arithmétique

* **Cas d'une série d'observation**: Supposons qu'une population statistique soit étudiée au vue du caract. revenu. La moyenne arithmétique de ces observations notées $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$. Et cette moy. arithmétique qui par df est une moy. arith. simple représente le revenu que chaque individus aura reçu si la masse globale des revenus était répartie de façon égale

$$x_i : x_1, x_2, \dots, x_n$$

A chaque fois qu'on a des observations individualisées, il faut calculer la moyenne arith. simple

* **Cas d'un tableau statistique**

Les observations sont affectées d'un coeff de pondératⁿ représentée par

l'effectif ou par la fréquence relative. Dans ce cas on va calculer une moyenne arithmétique pondérée notée \bar{x}_p (8)

$$\bar{x}_p = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k n_i x_i}{\sum_{i=1}^k n_i}$$

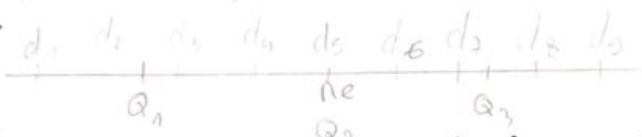
Lorsqu'on a un tableau quantitatif à caractère continu, on calcule d'abord les centres de classes

$$T_1: \bar{x}_p = \frac{97}{50} = 1,94$$

$$T_2: \bar{x}_p = \frac{45450}{85} = 534,7$$

I- 4) Les quartiles

Ce sont des paramètres de positions. On a une médiane qui sépare les observations en 2 groupes d'effectifs égaux. Il y'a 3 quartiles qui sépare les observations en 4 groupes d'effectifs égaux. Il y'a 9 déciles qui séparent les observations en 10 groupes d'effectifs égaux.



Il y'a 99 centiles C_1, C_2, \dots, C_{99} qui sépare les observations en 100 groupes d'effectifs égaux. La détermination de ces caractéristiques est identique à celle de la médiane.

Les quartiles st obtenus lorsqu'on a cumulé 25, 50, 75% de la population. Les déciles st obtenus lorsqu'on a cumulé 10, 20, 30%, de la population. Les centiles st obtenus lorsqu'on a cumulé 1, 2, ..., 99% de la population. La notion de decile et centile n'a de sens que s'il y'a beaucoup d'observations et donc essentiellement pour une variable classée.

24/01/2019

II- Les paramètres de dispersion

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

II- 1) L'étendue ou range

C'est un paramètre étendue de dispersion. C'est la \neq entre la plus grande observation et la plus petite.

(9)

II-2) : Intervall interquartile ou écart interquartile
 C'est le \neq entre le Q_3 et Q_1 ; $E/Q = Q_3 - Q_1$

II-3) l'écart absolue moyenne

Soit une série d'observation :

$$x_i : x_1, x_2, x_3, \dots, x_n \quad \bar{x}$$

$$x_i - \bar{x} : x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$$

may. arithmétique des écarts par rapport à la moyenne :

$$\bar{x} = \frac{\sum x_i}{n} \Rightarrow x_i - \bar{x} = \frac{\sum (x_i - \bar{x})}{n} = \frac{\sum x_i}{n} - \frac{\sum \bar{x}}{n} = \bar{x} - \frac{n \cdot \bar{x}}{n} = \bar{x} - \bar{x} = 0$$

pr éviter le pb de signe, on utilise la valeur absolue :

$$|x_1 - \bar{x}| \quad |x_2 - \bar{x}| \quad \dots \quad |x_n - \bar{x}| \Rightarrow$$

$$EAM = \frac{\sum |x_i - \bar{x}|}{n} ; \quad EAM_p = \frac{\sum n_i |x_i - \bar{x}_p|}{\sum n_i}$$

écart absolue
moyenne

$$\text{ex: } 100 \quad 150 \quad 90 \quad 60 \quad 10 \quad 40$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{450}{6} = 75$$

$$x_i - \bar{x} = 100 - 75 = 25; 75; 15; -15; -65; -35 =$$

Vérifions que $\sum (x_i - \bar{x}) = 0$

$$|x_i - \bar{x}| : 25; 75; 15; 15; 65; 35$$

$$EAM = \frac{\sum |x_i - \bar{x}|}{n} = \frac{230}{6} = 38,3$$

Si on a un tableau statistique : on calcule EAM_p

Supposons qu'on a :

x_i	n_i	$n_i x_i$	$ x_i - \bar{x}_p $	$n_i x_i - \bar{x}_p $
5	10	50	8,5	85
10	20	200	3,5	70
15	10	150	1,5	15
25	5	125	11,5	57,5
30	5	150	16,5	82,5
T	50	675	/ / /	310

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{675}{50} = 13,5$$

$$EAM_p = \frac{310}{50} = 6,2$$

Si on avait un tableau avec caractère quantitatif continu, les centres de classe allaient représenter les x_i .

II-4) la variance et l'écart-type

$$(x_1 - \bar{x})^2 : (x_2 - \bar{x})^2 \quad (x_2 - \bar{x})^2 \quad \dots \quad (x_n - \bar{x})^2$$

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

(10)

La variance est la moyenne arithmétique du carré des écarts par rapport à la moyenne arithmétique

$$\sigma(X) = \sqrt{V(X)}$$

La variance n'est pas un paramètre de dispersion comme l'écart-type

$$V_p(X) = \frac{\sum n_i (x_i - \bar{x}_p)^2}{\sum n_i}$$

La formule pondérée est utilisée

à chaque fois qu'on a un tableau

$$\sigma_p(X) = \sqrt{V_p(X)}$$

ça c'est des formules de définitions ↑

formules développées :

$$V(X) = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} = \frac{\sum x_i^2}{n} - 2\bar{x} \left(\frac{\sum x_i}{n} \right) + \bar{x}^2$$

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

$$2\bar{x}^2 + \bar{x}^2 = -\bar{x}^2$$

$$\Rightarrow \boxed{\begin{aligned} V(X) &= \frac{\sum x_i^2}{n} - \bar{x}^2 \\ V_p(X) &= \frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}_p^2 \end{aligned}}$$

$(x_i - \bar{x})^2$: 625, 5625, 225, 225, 4225, 1225

$$V(X) = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{12150}{6} = 2025$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{2025} = 45$$

pour le tableau

$$V_p(X) = \frac{\sum n_i (x_i - \bar{x}_p)^2}{\sum n_i} = \frac{3012,5}{50} = 60,25$$

$n_i(x_i - \bar{x}_p)^2$: 722,5; 245; 22,5; 661,25; 1361,25
T = 3012,5

$$\sigma_p(X) = 7,76$$

Si on ne donne uniquement n_i et x_i , on calcule d'abord $x_i n_i$, $x_i - \bar{x}_p$, $(x_i - \bar{x}_p)^2$, $n_i(x_i - \bar{x}_p)^2$ / simple

pondéré on fait d'abord : $\alpha_i, n_i, n_i \alpha_i, n_i \alpha_i^2$.

(11)

#détermination de la variance en passant par une variance auxiliaire ou provisoire

Supposons qu'on a :

Revenu	n_i	α_i	$n_i \alpha_i$	$n_i \alpha_i^2$	$\alpha_i' = \frac{\alpha_i - 650}{100}$	$n_i \alpha_i'^2$	$n_i \alpha_i'^2$
100							
100	10	150	1500	22500	-5	-50	250
200	20	250	5000	125000	-4	-80	320
300	35	350	12250	428750	-3	-105	315
400	25	500	12500	625000	-1,5	-37,5	56,25
600	70	650	13000	845000	0	0	0
700	10	750	7500	562500	1	10	10
800	5	850	4250	361250	2	10	20
900	5	950	4750	451250	3	15	45
1000							
T	130		60750	3421250		-237,5	1016,25

$$\bar{\alpha}_p = \frac{\sum n_i \alpha_i}{\sum n_i} = \frac{60750}{130} = 467,30769$$

$$V_p(x) = \frac{\sum n_i \alpha_i^2}{\sum n_i} - \bar{\alpha}_p^2 = \frac{3421250}{130} - (467,30769)^2 = 44796,5$$

$$\Rightarrow G_p(x) = 211,65$$

$$\alpha_i' = \frac{\alpha_i - b}{a}$$

$$\bar{\alpha}_p \quad \bar{\alpha}_p'$$

$$V_p(x) \quad V_p(x')$$

$$\bar{\alpha}' = \frac{\sum \alpha_i'}{n} = \frac{\sum (\frac{\alpha_i - b}{a})}{n}$$

$$= \frac{1}{a} \left(\frac{\sum n_i \alpha_i}{n} - \frac{\sum b}{n} \right)$$

$$= \frac{\bar{\alpha} - b}{a} \Rightarrow \bar{\alpha}_p = a \bar{\alpha}_p' + b$$

$$V(x') = \frac{\sum (\alpha_i' - \bar{\alpha}')^2}{n}$$

$$V(x') = \frac{\sum \left(\frac{\alpha_i - b}{a} - \frac{\bar{\alpha} - b}{a} \right)^2}{n} = \frac{1}{a^2} \frac{\sum (\alpha_i - \bar{\alpha})^2}{n} = \frac{1}{a^2} V(x)$$

$$\bar{\alpha}_p = a \bar{\alpha}_p' + b$$

$$V(x) = a^2 V(x')$$

a et b et des constantes à choisir
d'abord on choisit b parmi les valeurs
médianes de α_i

Posons $b = 650$

Après choix de b , on choisit a de tel sorte qu'elle ait un diviseur commun de n donc $a = 100$

$$\bar{x}'_p = -\frac{237,5}{130}$$

$$V_p(x') = \frac{1016,25}{130} - \left(\frac{237,5}{130}\right)^2 = 4,47965$$

$$\bar{x}_p = 100 \left(-\frac{237,5}{130}\right) + 650 = 467,30769$$

$$V_p(x) = 100^2 \times 4,47965 = 44796,5$$

Arrondir au m²/rang

14/03/2019

Exercice d'application :

Sol. mens	n_i	ECC	\bar{x}'_i	$n_i \bar{x}'_i^2$	$\bar{x}_i = p_i$	$n_i \bar{x}_i$	$n_i \bar{x}_i^2$	$n_i \bar{x}_i^3$
10		10						
20	180	180	-6	6480	15	2700	40500	-1080
30	210	390	-4	3360	25	5250	131250	-840
40	470	860	-2	1880	35	16450	575750	-940
50	108	968	0	/	45	4860	218700	0
60	850	1818	2	3400	55	46750	2571250	1700
70	110	1928	4	1760	65	7150	464750	440
Total	1928	/	-6	16880		83160	4002200	-720

1) Calculer la médiane

2) " les quartiles

3) Calculer la variance directement puis par une variance provisoire

$$T = 1928$$

$$n_{te} = \%50 = 964$$

964 compris entre 40 et 50 \Rightarrow classe médiane

$$40 \rightarrow 860$$

$$n_{te} \rightarrow 964$$

$$50 \rightarrow 968$$

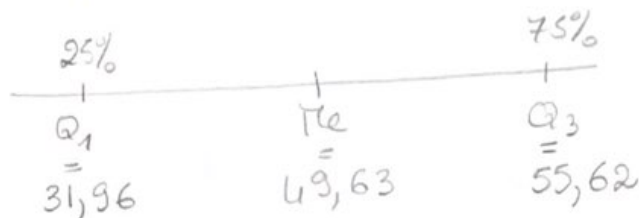
$$\frac{n_{te} - 40}{50 - 40} = \frac{964 - 860}{968 - 860} \Rightarrow$$

$$\frac{176-40}{10} = \frac{104}{108} \Rightarrow 176-40 = \frac{1040}{108} \Rightarrow 176 = 49,63$$

(13)

2°) Calculons les quartiles

$$176 = 49,63$$



3°) Calcul direct de la variance

$$V_p(x) = \frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}_p^2 \left(\frac{\sum n_i x_i}{\sum n_i} = \frac{83160}{1928} = 43,13 \right)$$

$$V_p(x) = \frac{4002200}{1928} - (43,13)^2 = 215,63$$

recalcul de la variance provisoire

$$V_p(x') = \frac{\sum (x'_i - \bar{x}')^2}{n}$$

On choisit $b = 45$

$$a = 5$$

$$\bar{x}'_p = \frac{\sum n_i x'_i}{\sum n_i} = \frac{-1720}{1928} = -0,37$$

$$\text{avec } x'_i = \frac{x_i - b}{a}$$

$$\Rightarrow V_p'(x)^2 = \frac{16880}{1928} - (0,37)^2 = 8,62$$

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

$$\bar{x}_p = 5 \times (-0,37) + 45 = 43,15$$

$$V_p(x) = 5^2 \times V_p'(x) = 215,46$$

21/03/2019

III Les paramètres de concentration

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

signification revenue

en mille

(14)

III 1- Notion de valeur globale

Soit x_i les modalités d'un caractère ou les c_i .

Soit n_i les effectifs correspondants

La valeur globale de la série (x_i, n_i) c'est la valeur $g_i = n_i x_i$

III 2- La médiale

La médiale de la série (x_i, n_i) est la médiane (x_i, g_i)

masse globale de revenus distribués

Ex:

Rev (um)	n_i	x_i	$g_i = n_i x_i$	$g_i \%$	$F_i \uparrow$	$G_i \uparrow$
-1000					0	0
-10000	41	5500	225500	4,11	41	4,11
-50000	37	30000	1.110.000	20,21	78	24,35
-100000	10	75000	750000	13,67	88	38,02
-200000	6	150000	900000	16,41	94	54,43
-500000	4	350000	1.400.000	25,52	98	79,95
-600000	2	550000	1.100.000	20,05	100	100
T	100		5.485.500	100		

h_i = fréquence (on aura la m^{re} chose après calcul

Calcul de la médiane

10000 \rightarrow 41

$n_e \rightarrow$ 50 \Rightarrow

50000 \rightarrow 78

$$n_e - 10000 = \frac{9}{37} \times 40000$$

$$n_e = 19729,73$$

Calcul de la médiale

100000 \rightarrow 38,02

ML \rightarrow 50

200000 \rightarrow 54,43

$$ML - 100000 = \frac{11,98}{16,41} \times 100000$$

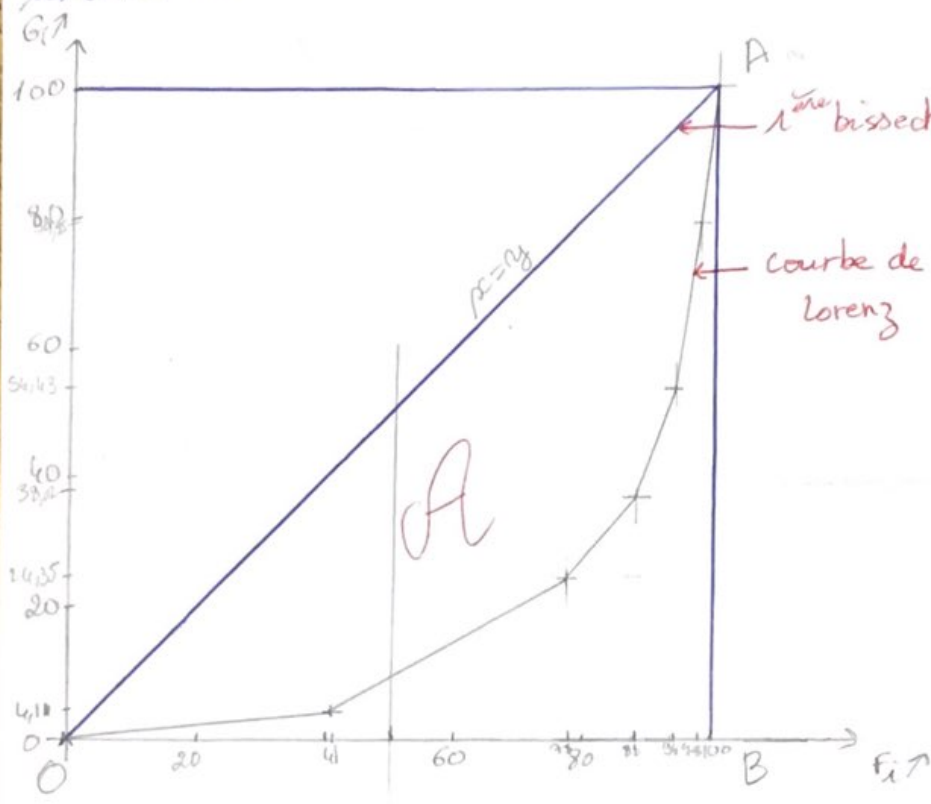
$$ML = 173004,27$$

La médiale est une modalité de la valeur globale tel que

Quand il y'a une grande différence entre la médiale et la médiane, on dit qu'on a une concentration

III 3- Courbe de concentration ou courbe de Lorenz

c'est la courbe obtenue en représentant en abscisse les FCC de la série (x_i, n_i) et en ordonnée, les FCC de la série (x_i, g_i) . L'allure de la courbe permet d'avoir une idée de la concentration



Si la courbe de concentration est confondue avec la 1^{ère} bissectrice, on dit qu'on a pas de concentration. Si c'est proche, on a une faible concentration. Si c'est éloigné, on a une forte concentration. C'est le cas ici.

L'indice de Gini = indice de concentration

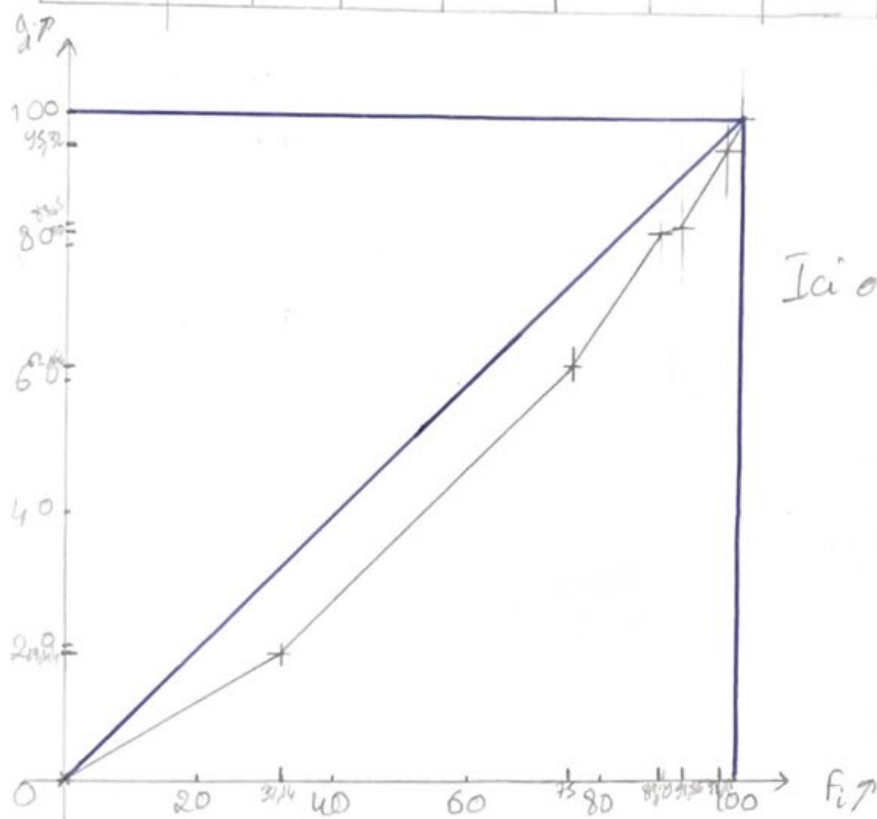
$$I_G = \frac{A}{\text{Surf}(OAB)} = \frac{A}{\frac{5000}{2}} \quad (\text{par intégration par parties})$$

Calculons en portant l'équation de la droite (A voir)

$$y = ax + b \quad \text{ou} \quad ax + by + c = 0$$

caractères pour la masse globale sont des caractères susceptibles d'addition.

Classes	n_i	x_i	g_i	$g_i\%$	$F_i\%$	$g_i\uparrow$	$F_i\uparrow$
100						0	0
200	90	150	13500	19,44	32,14		
300	120	250	30000	43,2	42,86	19,44	32,14
400	40	350	14000	20,16	11,29	62,64	75
500	10	450	450	0,65	3,57	82,8	89,29
600	15	550	8250	11,87	5,36	83,45	92,86
700	5	650	3250	4,68	1,78	95,32	98,22
800						100	100
T	280		69450	100	100		



Ici on a une concentration faible