15/11/2018 (1)

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 tjrs sur une population. Une populat° stat. est l'ens ds individus sur lesquelles portent l'étude stat. Si l'étude porte sur une partie de la populat° alors le référentiel sera appelé échantillon. Le nbre d'individus qui constituent la populat° représente la taille de la populat°. Une populat° stat. est étudiée au Mue d'1 ou de plrs caractères. Un caractère stat. est une propriété qui est commune à l'ens des individus qui constituent la populat°. Un caractère stat. peut se trouver dans ≠ situat° appelé modalité du caractère. Un caractère est dit qualitatif lorsque ces modalités ne peuvent pas être exprimer par des valeurs numériques. Un caractère est quantitatif lorsque ces modalités peuvent être exprimer par des valeurs numériques. Un caractère quantitatif est dit discontinue ou discret si ces modalités peuvent prendre des valeurs finies ou dénombrables. Le plus svt c'es valeurs st entières. Un caract. quantitatif est dit continu si ces ≠ modalités peuvent appartenir à un intervalle fixé à priori, les modalités st ont exprimés par des intervals 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 :

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

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

définit le caractère quantitatif continu

Nb enf	Eff	f_i %	ECC
0	6	12	6
1	12	24	18
2	20	40	38
3	8	16	46
4	2	4	48
5	2	4	50
Total	50	100	

Ex

Revenue	n_i	d_i	U_i	n_{ir}	ECC	ECD
[100-200[	15	0,15	1	15	15	100
[200-300[	10	0,10	1	10	25	85
[300-400[	30	0,30	1	30	55	75
[400-600[	40	0,20	2	20	95	45
[600-700[	3	0,03	1	3	98	5
[700-900[	1	0,005	2	0,5	99	2
[900-1000]	1	0,01	1	1	100	1
T	100					

amplitude $a_i = e_{i+1} - e_i$ extrémité ; centre $c_i = \frac{e_i + e_{i+1}}{2} = x_i$

densité $d_i = \frac{n_i}{a_i}$; effectif rectifié (n_{ir}) : $n_{ir} = \frac{n_i}{U_i}$ → nbre d'unité d'amplitude

U_a = unité d'amplitude $\quad U_a = 100$

$n_{ir} = d_i \times U_a$ / $n_{ir} = \frac{n_i}{\frac{a_i}{U_a}} = \frac{n_i \times U_a}{a_i}$ $\quad U_i = \frac{a_i}{U_a}$ ou $\frac{b_{sup} - b_{inf}}{a_i}$

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	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	$n_{1,j}$	$n_{1,l}$	$n_{1\bullet}$
x_2	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	$n_{2,j}$	$n_{2,l}$	$n_{2\bullet}$
x_i	$n_{i,1}$	$n_{i,2}$	$n_{i,3}$	$n_{i,4}$	$n_{i,5}$	$n_{i\bullet}$
x_k	$n_{k,1}$	$n_{k,2}$	$n_{k,3}$	$n_{k,j}$	$n_{k,l}$	$n_{k\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	$n_{\bullet j}$	$n_{\bullet l}$	$n_{\bullet\bullet}$

distribut° conditionnelle

distribut° marginale donne un caractère dans une populat°

X	eff
x_1	$n_{1\bullet}$
$\vdots$	$\vdots$
x_k	$n_{k\bullet}$
T	$n_{\bullet\bullet}$

, la repartit° d'1 seul caractère

Y	eff
y_1	$n_{\bullet 1}$
y_2	$n_{\bullet 2}$
$\vdots$	
y_l	$n_{\bullet l}$
T	$n_{\bullet\bullet}$

I. 3) Les représentations graphiques

08/12/2018

II. I. 3) 1- Représentation graphique d'un caract. qualitatif

Dans le tableau statistique donnant la répartition qualitatif chaque modalité x_i est associé à un effectif n_i La seule représentat° qui ns interesse est celle des effectifs n_i ~~ou des~~

Suivant la variable observée, de nombreuses représentat° + ou - informatives peuvent être utiliser. Cependt 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.

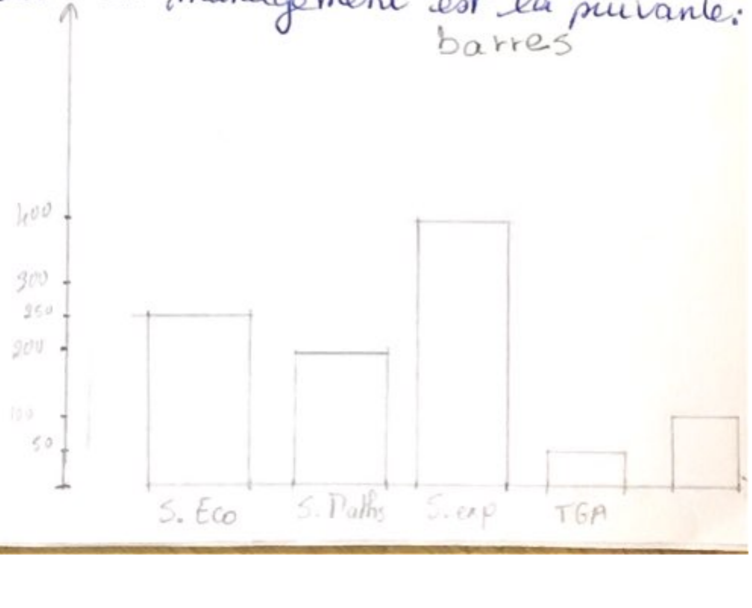
Les effectifs ou les fréquences sont placés sur un axe vertical, la hauteur du tuyau ou de la barre est proportionnelle à l'effectif.

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	×90
S. Maths	200	×72
S. exp	400	×144
TGA	50	×18
TGC	100	×36
TOTAL	1000	

sert à représenter l'importance absolue de chaque modalité

barres



• Les diagrammes à secteurs ou camembert (4)

L'effectif total est représenté par un disque, Chaque modalité est représenté par un secteur circulaire dont la surface (l'angle au centre est proportionnel à l'effectif correspondant.

$360° \to 1000$

$\Rightarrow \frac{250 \times 360}{1000} = 90°$ $\quad \frac{100 \times 360}{1000} = 36°$

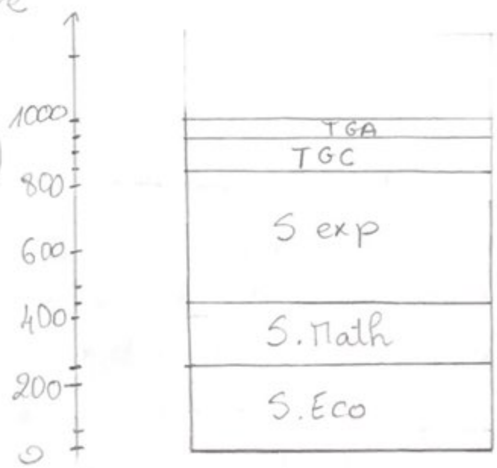
$\frac{200 \times 360}{1000} = 72°$

$\frac{400 \times 360}{1000} = 144°$

$\frac{50 \times 360}{1000} = 18°$

• ou en barre

(doit être effacer ensuite)



on calcule les pourcentages

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

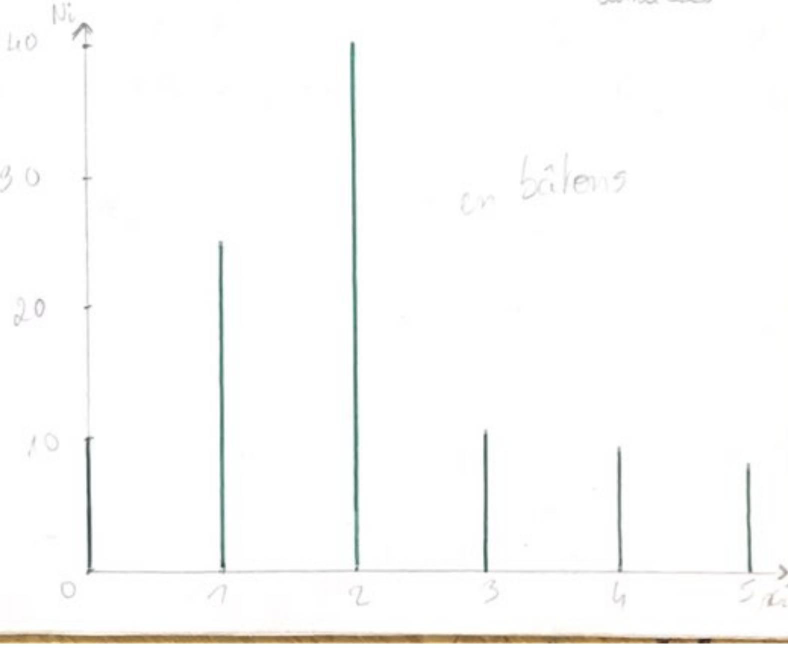
I.3) 2- Représentation graphique d'un caractère quantitatif

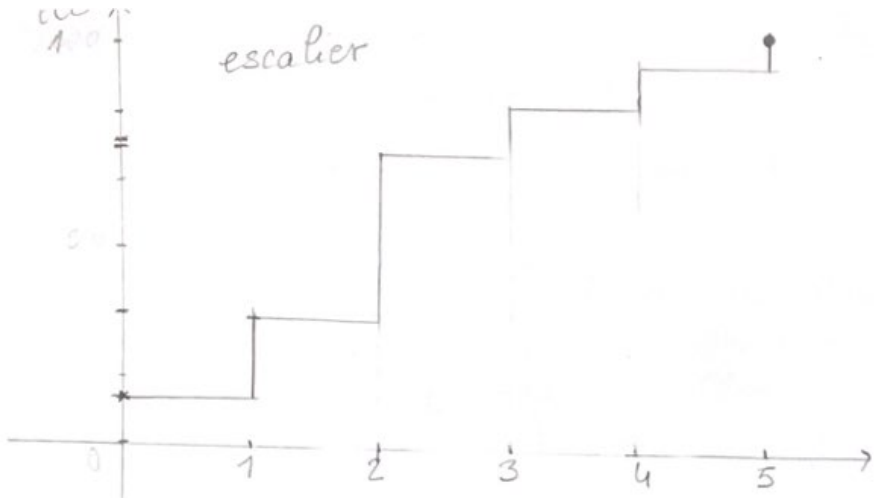
Caract. q. continu → diagramme différentiel (1) disc → diagramme en bâton ; cont → histogramme
→ diagrame intégral (ECC) cumulatif FF(%) (2) disc → diagramme en escalier ; cont → Courbes des fréquences cumulées

Caract. q. discontinu →

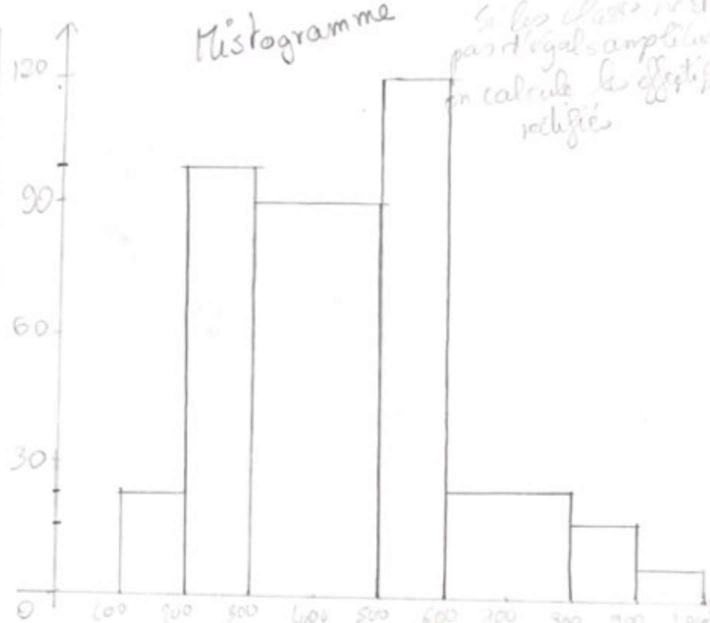
Ex.

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	





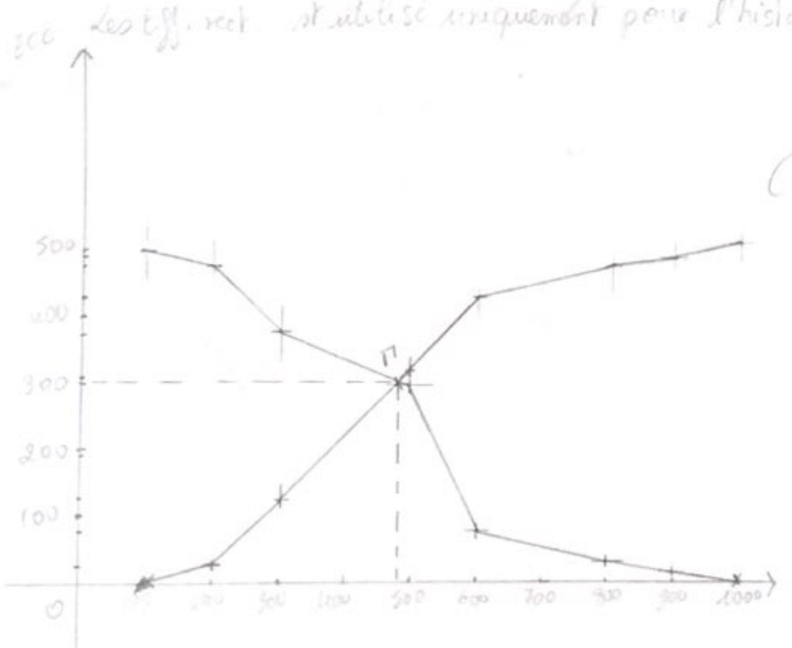
Sal men (um)	ni	ECD	ai	nir	ICC
100		500			0
	25		1	25	
200		475			25
	100		1	100	
300		375			125
	180		2	90	
500		195			305
	120		1	120	
600		75			425
	50		2	25	
800		25			475
	15		1	15	
900		10			490
	10		1	10	
1000		0			500
T	500				



Si les classes ne st pas d'égales amplitude on calcule les effectifs rectifiés

Les eff. rect st utilisé uniquement pour l'histogramme

on relie par des segments de droites;



Courbes des fréquences cumulées croissante

Le point d'intersection

La médiane modalité du caractère permettant de le diviser en 2

Chapitre II - Analyse numérique des séries statistiques

20/12/2018 (6)

I- Les paramètres de tendance centrale et de position

I- 1) Le mode

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

pour représenter uniquement le mode ou l'histogramme

ECC	x_i	n_i	$n_i x_i$
5	0	5	0
15	1	10	10
40	2	25	50
45	3	5	15
48	4	3	12
50	5	2	10
	T	50	97

T_1

$\to M_o = 2$

Classe Modale : $[400;500[$

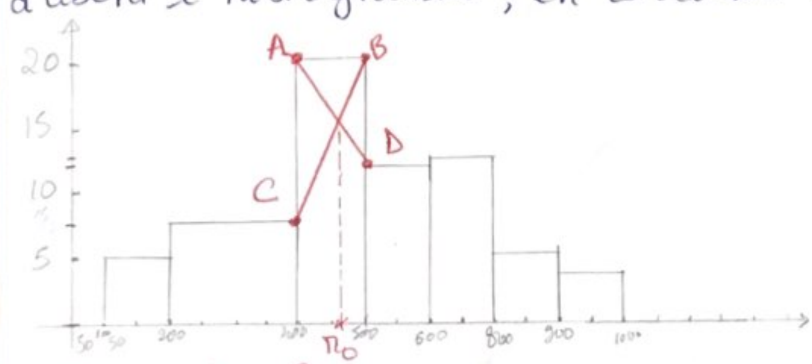
$M_o \simeq 450$

Attention

$T_2 \to$

Classes	n_i	x_i	n_{ir}	ECC	$n_i x_i$
100				0	
	5	150	5		750
200				5	
	15	300	15/2		4500
400				20	
	20	450	20		9000
500				40	
	12	550	12		6600
600				52	
	25	700	25/2		17500
800				77	
	5	850	5		4250
900				82	
	3	950	3		2850
1000				85	
Total	85				45450

Pour déterminer le mode par le calcul, on représente d'abord l'histogramme ; en calculant l'abcisse 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_i : 1;1;2;3;4;7;8;10

② regarder la parité du nbre 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 dét. de la mᵉ façon que si on 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 distribut° 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 dét la médiane dans ce tableau

	Mod. du caract		ECC
1)	500	$\rightarrow$	40
2)	Me	$\rightarrow$	$85/2 = 42{,}5$
3)	600	$\rightarrow$	52

$Me \rightarrow 85/2 = 42{,}5$

$\Rightarrow$ Classe médiane $= [500, 600[$

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

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

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

$Me = 520$

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'1 coeff de pondérat° représentée par l'effectif ou par la fréquence relative. Dans ce cas on va calculer une moyenne arithmétique pondérée notée $\bar{x}_p$

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

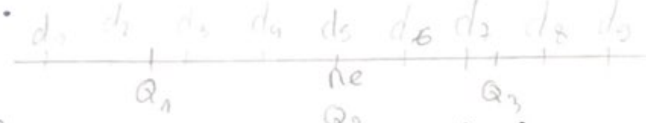
Lorsqu'on a un tableau quantitatif à caractère continu, on calcule d'abord les centres de classes

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 des 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 ... 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 obtenues lorsqu'on a cumulés 25, 50, 75% de la population. Ls déciles st obtenus lorsqu'on a cumulés 10, 20 . 90% de la populat°. Ls centiles st obtenus lorsqu'on a cumulés 1, 2, .. 99 % de la populat°. La notion de décile 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 essayent de synthétiser par une seule valeur numérique, la dispersion de ttes ls valeurs observées

II - 1) L'étendue ou range

C'est un paramètre eltaire de dispersion. C'est la ≠ entre la plus grande observation et la plus petite.

II-2) : Interval interquartile ou écart interquartile

C'est le ≠ entre le Q_3 et Q_1 ; $E/Q = Q_3 - Q_1$

II-3) l'écart absolue moyenne

Soit une série d'observation :

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

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

moy. arithmétiq ds écarts par rapport à la moyenne :

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

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

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

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

écart absolue moyenne

Ex : 100 150 90 60 10 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	$n_i x_i$	$\lvert x_i - \bar{x}_p \rvert$	$n_i \lvert x_i - \bar{x}_p \rvert$
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 allait representer les x_i.

II-4) La variance et l'écart-type

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

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

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

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

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

La formule pondéré est utilisée à chaque fois qu'on a un tableau

ça c'est des formules de définitions

formules développées :

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

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

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

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

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

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

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

pour le tableau

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

$n_i(x_i-\bar{x}_p)^2$ = 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, et qu'on ne demande la variance 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 : x_i, n_i, $n_i x_i$, $n_i x_i^2$.

Détermination de la variance en passant par une variance auxiliaire ou provisoire

Supposons qu'on a :

Revenu	n_i	x_i	$n_i x_i$	$n_i x_i^2$	$x_i' = \frac{x_i - 650}{100}$	$n_i x_i'$	$n_i x_i'^2$	
100 – 200	10	150	1500	225000	−5	−50	250	
200 – 300	20	250	5000	1250.000	−4	−80	320	
300 – 400	35	350	12250	4287500	−3	−105	315	
400 – 600	25	500	12500	6250000	−1,5	−37,5	56,25	
600 – 700	20	650	13000	8450000	0	0	0	
700 – 800	10	750	7500	5625000	1	10	10	
800 – 900	5	850	4250	3612500	2	10	20	
900 – 1000	5	950	4750	4512500	3	15	45	
T	130	/	60750	34212500		−237,5	1016,25	

$$\bar{x}_p = \frac{\sum n_i x_i}{\sum n_i} = \frac{60750}{130} = 467,30769$$

$$V_p(X) = \frac{\sum n_i x_i^2}{\sum n_i} - \bar{x}_p^2 = \frac{34212500}{130} - (467,3076)^2 = 44\,796,5$$

$$\Rightarrow \sigma_p(X) = 211,65$$

$$x' = \frac{x_i - b}{a} \qquad \bar{x}_p \;(\bar{x}'_p) \qquad V_p(X) \;\; V_p(X')$$

$$\bar{x}' = \frac{\sum x_i'}{n} = \frac{\sum\left(\frac{x_i - b}{a}\right)}{n} = \frac{1}{a}\left(\frac{\sum x_i}{n} - \frac{\sum b}{n}\right) = \frac{\bar{x} - b}{a} \Rightarrow \bar{x}_p = a\bar{x}_p' + b$$

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

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

$$\bar{x}_p = a\bar{x}'_p + b$$
$$V(X) = a^2 V(X')$$

a et b sont des constantes à choisir. D'abord on choisit b parmi les valeurs médianes de x_i

Posons b = 650
Après choix de b, on choisit a de tel sorte qu'elle soit un diviseur commun de x_i 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 m^2 rang

14/03/2019

Exercice d'application :

Sal. mens	n_i	ECC	x'_i	$n_i x'^2_i$	$c_i = x_i$	$n_i x_i$	$n_i x_i^2$	$n_i x'_i$
10		0						
	180		−6	6480	15	2700	40500	−1080
20		180						
	210		−4	3360	25	5250	131250	−840
30		390						
	470		−2	1880	35	16450	575750	−940
40		860						
	108		0	/	45	4860	218700	0
50		968						
	850		2	3400	55	46750	2571250	1700
60		1818						
	110		4	1760	65	7150	464750	440
70		1928						
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

Me = %50 = 964

964 compris entre 40 et 50 ⟹ classe médiane

40 → 860

Me → 964

50 → 968

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

$\frac{Me-40}{10} = \frac{104}{108} \Rightarrow Me-40 = \frac{1040}{108}$

$\Rightarrow Me = 49,63$

2°) Calculons les quartiles

$Me = 49,63$

25%		75%
$Q_1 = 31,96$	$Me = 49,63$	$Q_3 = 55,62$

3°) Calcul direct de la variance

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

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

calcul de la variance provisoire

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

On choisit $b = 45$

$a = 5$

$\bar{x}'p = \frac{n_i x'_i}{n_i} = \frac{-720}{1928} = -0,37$

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

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

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

on doit trouver la mê valeur

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

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

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 concrète

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)

Ex:

masse globale de revenue distribuée

Rev (um)	n_i	x_i	$g_i = n_i x_i$	g_i %	$F_i\nearrow$	$G_i\nearrow$
1000					0	0
	41	5500	225500	4,11		
10000					41	4,11
	37	30000	1110000	20,24		
50000					78	24,35
	10	75000	750000	13,67		
100000					88	38,02
	6	150000	900000	16,41		
200000					94	54,43
	4	350000	1400000	25,52		
500000					98	79,95
	2	550000	1100000	20,05		
600000					100	100
T	100		5485500	100		

= fréquence (on aura la m chose après calcu

Calcul de la médiane

$10000 \longrightarrow 41$
$Me \longrightarrow 50 \Rightarrow$
$50000 \longrightarrow 78$

$$Me - 10000 = \frac{9}{37} \times 40000$$
$$Me = 19729{,}73$$

Calcul de la médiale

$100000 \longrightarrow 38{,}02$
$ML \longrightarrow 50$
$200000 \longrightarrow 54{,}43$

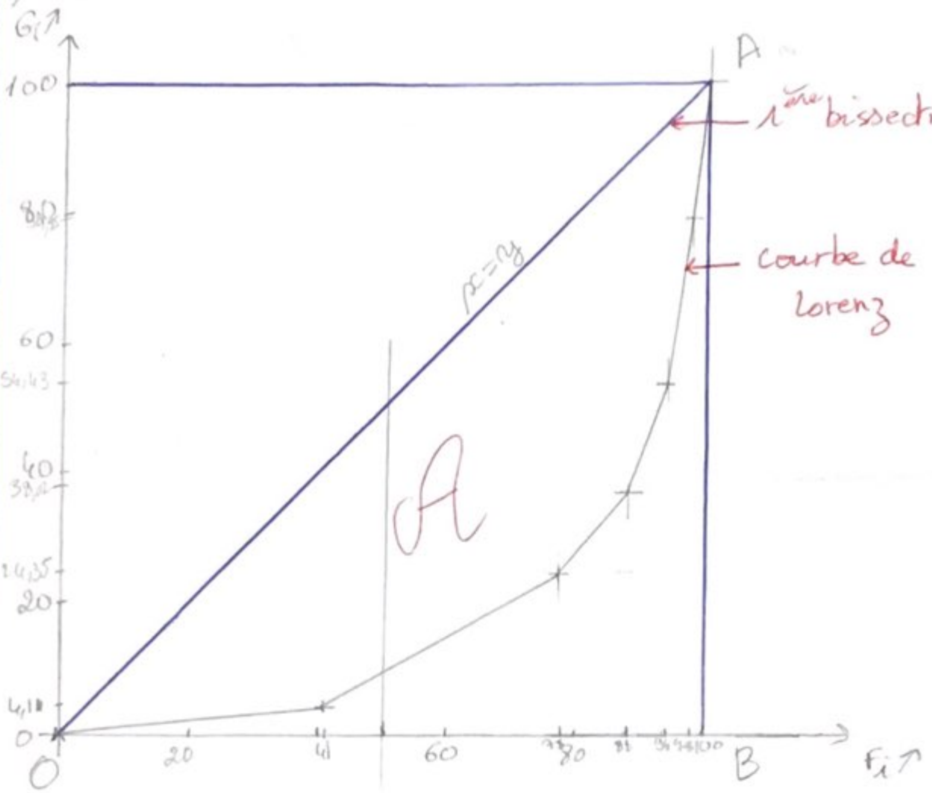
$$ML - 100000 = \frac{11{,}98}{16{,}41} \times 100000$$
$$ML = 173004{,}27$$

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 abcisse 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 confondu avec la 1ère 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 concentrat° c'est le cas ici

L'indice de Gini = indice de concentration

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

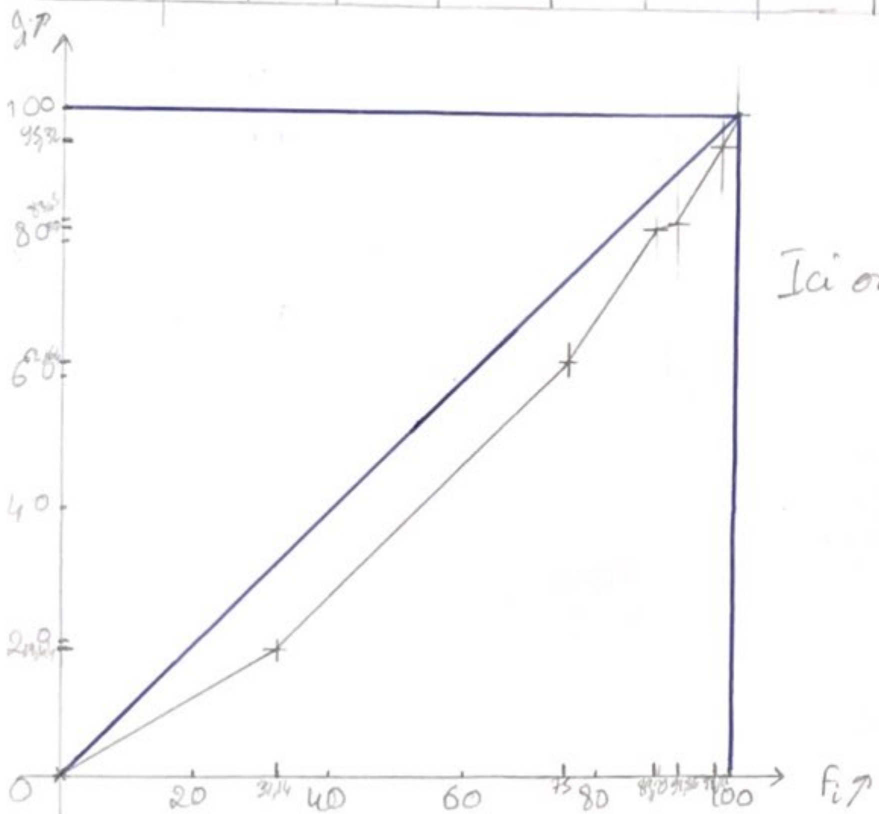
$\frac{\ell \times H}{2}$

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

$y = ax + b$ ou $ax + by + c = 0$

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

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



Ici on a une concentration faible