MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE BATNA 2
FACULTE DE TECHNOLOGIE
DEPARTEMENT DE SCIENCE TECHNOLOGIQUE
2^{éme} Année Socle Commun ST

Résumé de cours de Statistiques

Année universitaire 2020/2021
Chargés de cours :
Pr. M. ZIDANI
Pr. S. DERFOUF
Pr. K. MEFTAH

I- les indicateurs de position et de tendance centrale

1- Le mode

2- La médiane

3- les quantiles

4- la moyenne

II- les paramètres de dispersion

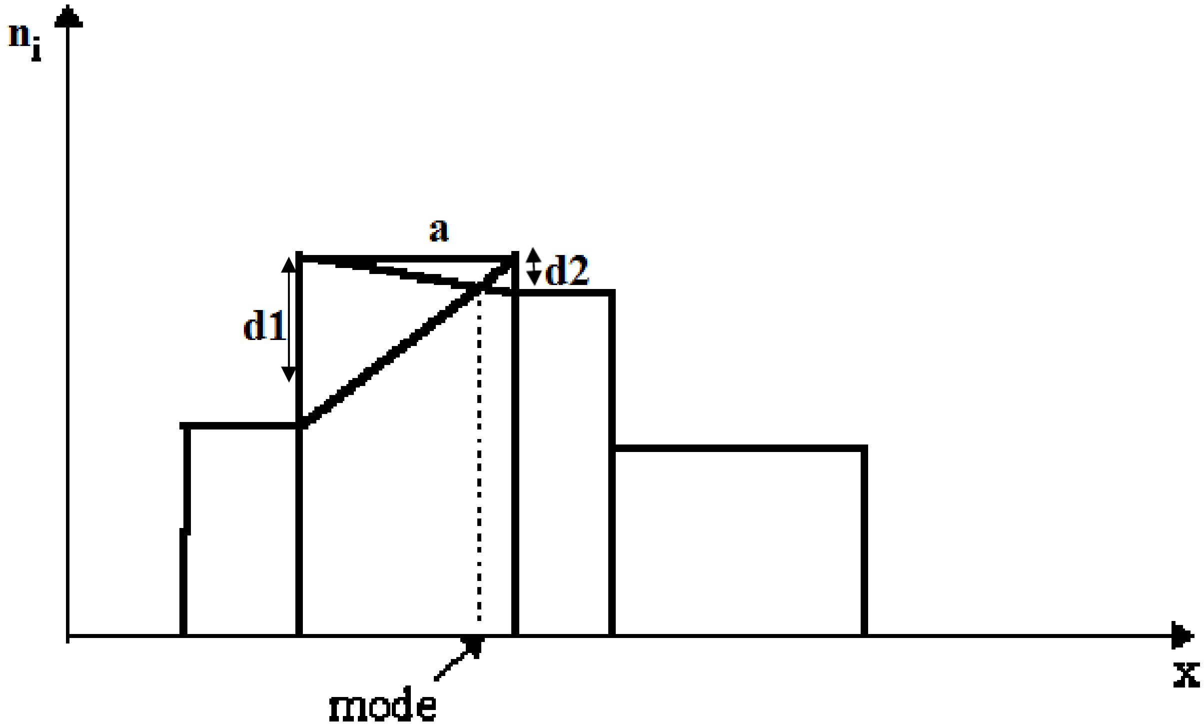
I- les indicateurs de position et de tendance centrale:

Ces indicateurs sont des paramètres calculés à partir de la série statistique dans le but de donner un résumé interprétable et exhaustif de l'information contenue dans cette série.

1- Le mode :

Le mode correspond à la modalité la plus fréquente. Pour un caractère continu pour le quel les données sont groupées en classes, la classe modale correspond à celle associée à l'effectif (corrigé) le plus élevé ou graphiquement au plus haut rectangle de l'histogramme.

Dans ce cas le mode est calculé à partir du centre de la classe modale selon la méthode suivante :

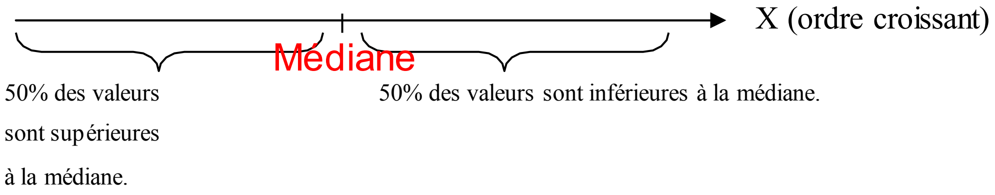


Si le mode appartient à la classe $[e_i ; e_{i+1}[$ alors :

$$M_O = e_i \times (\frac{d_1}{d_1 + d_2} \times a_i)$$

2- La médiane :

La médiane est la modalité qui divise la série des données statistiques en deux parties égales après avoir ranger ces données en ordre croissant (ou décroissant).



a- Cas d'un caractère discret :

Lorsqu'on possède la série des données brutes et distribution (non groupée), on doit ranger les n observations en ordre croissant.

Si n est **impair,** la médiane est la $\left(\frac{N+1}{2}\right)^{ième}$ bservation.

Si n est **pair,** la médiane est habituellement définie comme étant le point milieu entre la $\left(\frac{N}{2}\right)^{ième}$ et la $\left(\frac{N+1}{2}\right)^{ième}$ bservation.

b- cas d'un caractère continue :

la médiane est la modalité *x* tel que

$F(M_e) = P(X \leq M_e) = 0{,}5$

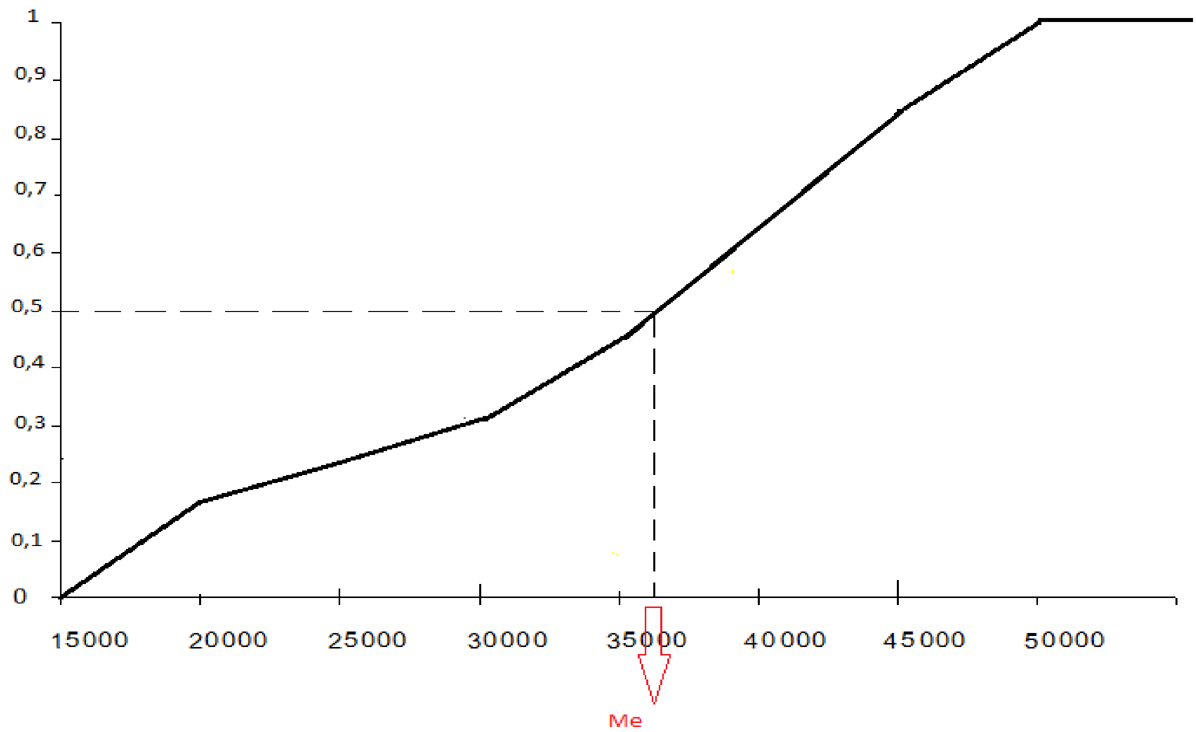
Pour calculer la médiane on doit déterminer la classe médiane à partir des fréquences cumulées croissant, puis on calcule la valeur ponctuelle de la médiane selon l'hypothèse de l'uniformité de la répartition des individus à l'intérieur de la classe médiane.

$M_e \in [e_i, e_{i+1}[$

$$M_e = e_i + (\frac{0,5 - F_{i-1}}{F_i - F_{i-1}} \times a_i)$$

La médiane se caractérise par le fait que sa valeur n’est pas influencée par les observations aberrantes ou les observations extrêmes.

Exemple



Me ϵ [35000,40000[

$$M_e = 35000 + (\frac{0,5 - 0,451}{0,873 - 0,451} \times 5000)$$

50% des salariés possèdent un salaire inferieur à la médiane.

3- les quantiles :

Les quantiles sont des indicateurs qui divisent la distribution en quatre parties égales. le premier quantile est indicateur noté Q_1 tel que

$F(Q_1) = P(X \leq Q_1) = 0{,}25$

Si $Q_1 \in [e_i, e_{i+1}[$

$$Q_1 = e_i + (\frac{0{,}25 - F_{i-1}}{F_i - F_{i-1}} \times a_i)$$

Le troisième quantile est noté Q_3

$F(Q_3) = P(X \leq Q_3) = 0{,}75$

Si $Q_3 \in [e_i, e_{i+1}[$

Alors

$$Q_3 = e_i + (\frac{0{,}75 - F_{i-1}}{F_i - F_{i-1}} \times a_i)$$

4- la moyenne :

La moyenne est un indicateur de tendance centrale qui permet de déterminer le centre de la distribution, la moyenne arithmétique est la moyenne est la plus utilisée, mais il existe d'autres types de moyennes utilisées dans le calcul de la tendance centrale de distributions statistiques telles que la moyenne géométrique et la moyenne quadratique.

La moyenne arithmétique :

La moyenne arithmétique est la somme de toutes les données observées divisées par le nombre des individus de l'échantillon.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{Ou bien} \quad \frac{x_1 + x_2 + \ldots.. + x_n}{n}$$

Si les données sont présentées dans un tableau statistique dans le quel chaque modalité est associée à fréquence absolue ou relative alors on calcule la moyenne arithmétique pondérée ainsi :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \quad \text{ou} \quad \bar{x} = \sum_{i=1}^{k} f_i x_i$$

Exemple : calcule du nombre de pièce moyen à partir de la distribution des logements selon le nombre des pièces :

Xi	ni	ni xi
0	4	0
1	5	5
2	9	18
3	3	9
4	7	28
5	2	10
total	30	70

$\bar{x} = \frac{\sum_{i=1}^{6} n_i s_i}{n} = \frac{70}{30} = 2{,}3$ le nombre de pièces moyen par logement est égale à 2

Dans le cas d'un tableau d'un caractère continu on remplace X_i par le centre de la classe $[e_i ; e_{i+1}[$ noté C_i

$$C_i = \frac{e_i + e_{i+1}}{2} \text{ et dans ce cas } \bar{x} = \frac{\sum_{i=1}^{k} n_i C_i}{n}$$

Calcule de salaire moyen

classe	n_i	Ci	ni Ci
[100-150[	120	125	15000
[150-250[	340	200,00	68000
[250-300[	200	275,00	55000
[300-400[	160	300,00	48000
[400-500[	120	450,00	54000
[500-700[	60	600,00	36000
TOTAL	1000		276000

$\bar{x} = \frac{\sum_{i=1}^{k} n_i C_i}{n} = 276$

La moyenne arithmétique correspond au centre d'inertie ou centre de gravité de la distribution puisqu'elle vérifie toujours cette égalité :

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

La moyenne arithmétique est un paramètre qui peut être influencé par les observations extrêmes ou aberrantes.

II- les paramètres de dispersion

Pour analyser une distribution on peut utiliser en plus des indicateurs de tendance centrale, telles que la médiane ou la moyenne, d'autres indicateurs qui permettent de mesurer la dispersion ou l'éparpillement de la série dans le but de bien décrire la distribution d'une variable.

Par exemple, les deux séries d'observations suivantes :

-20, -10, 0, 10, 20

-2000, -1000, 0, 1000, 2000

Possèdent la même moyenne et la même médiane (0) mais se diffèrent selon un autre indicateur qui mesure l'écart des ses observations par rapport à la valeur centrale. On va présenter dans cette partie les mesures de dispersion les plus utilisées : l'étendue, l'écart interquartile, la variance, l'écart-type et le coefficient de variation.

a- L'étendue:

L'étendue est un paramètre qui mesure l'écart entre la valeur la plus élevée et la valeur la plus faible de la distribution :

$$E = X_{max} - X_{min}$$

b- l’écart interquartile :

l’intervalle interquartile est l’intervalle [Q_1 ; Q_3[, cet intervalle contient 50% des observations.

L’écart interquartile est l’amplitude de l’intervalle interquartile :

$$EIQ = Q_3 - Q_1$$

L’écart interquartile est un indicateur qui a l’avantage d’écarter les observations extrêmes.

c- l’écart type :

L’écart type est l’indicateur de dispersion le plus utilisé et le plus simple à interpréter. Il permet de comparer les distributions dont la tendance centrale est identique. Il donne la variation moyenne de la distribution autour de la moyenne arithmétique. Pour calculer l’écart type on doit d’abord calculer la variance de X qui est égale à la somme des carrés des écarts à la moyenne divisée par l’effectif n, par la suite l’écart-type est égal à la racine de la variance.

La variance de X est calculée ainsi :
Pour des données brutes la variance est égale à :

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Le développement de cette formule permet de donner une formule plus simple à manipuler dans le calcul pratique de la variance.

$$V(X) = \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Lorsque les données sont groupées alors :

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{x})^2$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2$$

Exemple de calcule de la variance pour un caractère discret :

Xi	n_i	$n_i x_i$	X_i^2	$n_i X_i^2$
0	4	0	0	0
1	5	5	1	5
2	9	18	4	36
3	3	9	9	27
4	7	28	16	112
5	2	10	25	50
total	30	70		230

La répartition des logements selon le nombre des pièces

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \frac{230}{30} - (2{,}333)^2 = 2{,}223$$

$\sigma_x = 1{,}44$

calcule de la variance d'un caractère continu :

classe	n_i	Ci	ni Ci	C_i^2	$n_i C_i^2$
[100-150[	120	125	15000	15625	1875000
[150-250[	340	200	68000	40000	13600000
[250-300[	200	275	55000	75625	15125000
[300-400[	160	300	48000	90000	14400000
[400-500[	120	450	54000	202500	24300000
[500-700[	60	600	36000	360000	21600000
TOTAL	1000		276000		90900000

La répartition des salariés selon le salaire mensuel

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2 = \frac{90900000}{1000} - (276)^2 = 14724$$