

Université A- mira de Bejaia
Faculté des sciences économiques, commerciales et des sciences de gestion
Département SEGC(LMD)
Module stat I
Enseignant Dr. Mousli
Chapitre 4 : Les paramètres de dispersion

4.1. Position du problème et définition

Deux distributions de fréquence peuvent avoir la même valeur concernant les paramètres de tendance centrale et être assez différentes l'une de l'autre.

Exemple : supposons que 10 étudiants aient obtenu les résultats suivant lors d'un examen de mathématique (les données sont ordonnées par ordre croissant) :

0 7 7 7 7 7 7 7 10 11

Nous vérifions aisément que : $\bar{X} = M_e = M_o = 7$

Dans un deuxième examen, les résultats sont les suivants :

3 4 5 6 7 7 8 9 10 11

Encore ici, nous avons : $\bar{X} = M_e = M_o = 7$

L'examen attentif des deux séries fait ressortir une différence entre les deux distributions. Comment cette dissemblance peut-elle être traduite numériquement ?

La connaissance de la valeur centrale ne nous donne pas l'information complète. Il convient donc de se donner des mesures pour rendre compte de cet aspect important que constitue la variabilité ou la dispersion dans une série de données. Ceci nous permettra de faire des comparaisons concluantes entre diverses séries.

Définition :

Une caractéristique de dispersion est un nombre qui permet d'estimer dans quelle mesure des observations s'écartent les unes des autres (ou s'écartent de leur valeur centrale).

Nous étudierons ici les caractéristiques de dispersion selon le classement suivant :

-Différences entre deux valeurs d'une série

-Valeurs centrales de variables centrées.

4.2. Différences entre deux valeurs d'une série

4.2.1. L'étendue

Définition :

L'étendue notée E est égale à la différence entre la plus grande et la plus petite des valeurs prises par la variable statistique, classées préalablement dans l'ordre croissant ou décroissant.

4.2.1.1. Cas d'une variable discrète

Dans le cas d'une série discontinue, l'étendue ne pose pas de problèmes de calcul.
Nous avons :

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

Exemple 1 :

Reprenons les 2 séries de notes obtenues par 10 étudiants lors de deux examens (séries proposées précédemment).

0	7	7	7	7	7	7	7	10	11
3	4	5	6	7	7	8	9	10	11

- Dans la première distribution, l'étendue est égale à : $E = (11-0) = 11$
- Dans la deuxième distribution, l'étendue est égale à : $E = (11-3) = 8$.

Exemple 2 :

La série suivante dans le tableau ci-après donne la répartition des 500 familles selon le nombre de filles :

Tableau 19 : Distribution des 500 familles selon le nombre de filles

X_i	n_i	$n_i \uparrow$	$n_i \downarrow$
0	50	50	500
1	120	170	450
2	130	300	330
3	110	410	200
4	90	500	90
Total	500	-	-

Dans ce cas, l'étendue est égale à :

$$E = (4 - 0) = 4 \text{ Filles par famille.}$$

4.2.1.2. Cas d'une variable continue

Dans le cas d'une distribution continue, deux cas se présentent :

• 1^{er} cas : Les classes extrêmes sont bornées

Dans ce cas, l'étendue est égale à $E = e_k - e_1$

Exemple : le tableau suivant donne la répartition de 50 ouvriers selon leurs salaires :

Tableau 20 : Répartition des 50 ouvriers selon leur salaire en MDA

Classes	n_i
[10 - 50 [	15
[50 -100 [	20
[100-200[	10
[200-300[	5
Total	50

Dans ce cas, l'étendue est égale à :

$$E = 300 - 10 = 290 \text{ MDA}$$

- **2^{ème} cas : Les classes extrêmes ne sont pas bornées**

La détermination de l'étendue pose problème. En pratique, nous bornons les classes extrêmes selon la règle étudiée au chapitre 2 et nous calculons l'étendue comme ci-dessus.

Exemple2 :

Le tableau suivant donne la répartition des primes mensuelles des fonctionnaires d'une entreprise publique en milliers de dinars (MDA) :

Primes mensuelles (MDA)	Nombre d'ouvriers	Intervalle utilisé
-5	13	[04-05]
[05-06[	15	[05-06[
[06-07[	20	[06-07[
[07-09[	40	[07-09[
[09-10[	25	[09-10[
[10-15[	12	[10-15[
+15	10	[15-20]

Pour calculer l'étendue, nous bornons d'abord le 2 classes en tenant compte des classes adjacentes.

La classe « moins de 5 » a donc une amplitude de 1 car la classe adjacente est la classe [05-06[d'amplitude égale à « 1 ».

De même, la classe « plus de 15 » a une amplitude égale à « 5 » car la classe adjacente est la classe [10-15[d'amplitude égale à 5.

Dans ce cas, l'étendue est égale : **$E = (20-4) = 16 \text{ MDA}$.**

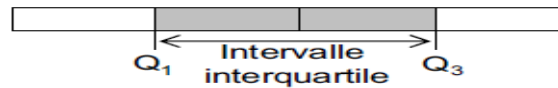
4.2.2. Écart interquartile Q

Définition :

L'intervalle interquartile $[Q_1; Q_3]$ est une mesure de la variation (dispersion) qui n'est pas influencée par les valeurs extrêmes, contrairement à l'étendue.

L'intervalle interquartile est l'intervalle qui contient 50% des observations, en laissant 25% des observations de part et d'autre de l'intervalle.

L'écart interquartile noté Q est égale à : $Q = Q_3 - Q_1$



4.2.3. Écart interdécile D

Définition :

L'intervalle interdécile $[D_1; D_9]$ est une mesure de dispersion qui contient 80% des observations, en laissant 10% des observations de part et d'autre de l'intervalle.

L'écart interdécile noté D est égale à : $D = D_9 - D_1$



4.3. Valeurs centrales de variables centrées

L'opération consiste, ici, à s'appuyer sur des variables centrées (différences des valeurs à leur moyenne ou, très rarement, à leur médiane). Nous retenons ici :

- L'écart absolu moyen
- La variance et l'écart-type (ou standard-déviations)
- Le coefficient de dispersion relative (coefficient de variation).

4.3.1. Ecart absolu moyen

Définition

L'écart moyen absolu que l'on note $e_a(X)$ est la moyenne arithmétique des valeurs absolues des écarts entre la valeur prise par la variable statistique X et une valeur type choisie qui est généralement la moyenne $\bar{X}$.

4.3.1.1. Cas d'une variable discrète

- Dans le cas d'une distribution discontinue de N observation, l'écart moyen des valeurs à la moyenne est : $e_a(X) = \frac{\sum |x_i - \bar{X}|}{N}$

Il s'agit d'une moyenne des valeurs absolues des variables centrées.

Exemple1 :

La série suivante représente le nombre d'absences d'un échantillon de 6 étudiants, à un cours de mathématique.

4 5 6 2 1 0

Calculons d'abord la moyenne : $\bar{X} = \frac{\sum x_i}{N} = \frac{18}{6} = 3$ absences

$$\text{D'où : } e_a(X) = \frac{\sum |x_i - \bar{X}|}{N} = \frac{|4-3| + |5-3| + |6-3| + |2-3| + |1-3| + |0-3|}{6} = \frac{12}{6} = 2 \text{ absences}$$

- Dans le cas d'une variable discontinue pondérée par les effectifs, la formule est :

$$e_a(X) = \frac{\sum_{i=1}^n n_i |x_i - \bar{X}|}{N}$$

Exemple2 :

La série suivante représente le nombre d'absences, d'un échantillon de 10 étudiants, à un cours de statistique.

9 2 1 1 7 7 7 4 6 7

Tableau 21 : Répartition des 10 étudiants selon le nombre d'absences

X_i	n_i	$n_i \times x_i$	$n_i \times x_i - \bar{X} $
1	2	2	8
2	1	2	3
4	1	4	1
6	1	6	1
7	4	28	8
9	1	9	4
Total	10	51	25

$$\bar{X} = \frac{\sum_{i=1}^n n_i \times x_i}{N} = \frac{51}{10} = 5,1 = 5$$

$$e_a(X) = \frac{\sum_{i=1}^n n_i |x_i - \bar{X}|}{N} = \frac{25}{10} = 2,5$$

$$e_a(X) \approx 3 \text{ absences}$$

4.3.1.2. Cas d'une variable continue

- Dans le cas d'observation groupées par classe, l'écart absolu moyen s'exprime par :

$$e_a(X) = \frac{\sum_{i=1}^n n_i |c_i - \bar{X}|}{N}$$

Avec c_i : centre de la classe d'une variable continue

Et n_i : effectif de la classe i

Exemple :

Le tableau suivant donne la répartition des jeunes chômeurs selon leurs allocations mensuelles en MDA.

Primes (MDA)	n_i	c_i	$n_i \times c_i$	$ c_i - \bar{X} $	$n_i \times c_i - \bar{X} $
[4-5 [	12	4,5	54	4,95	59,40
[5-6 [	23	5,5	126,5	3,95	90,85
[6-7 [	42	6,5	273	2,95	123,90
[7-9 [	56	8	448	1,45	81,20
[9-11[	34	10	340	0,55	18,70
[11-15[	32	13	416	3,55	113,60
[15-23[	16	19	304	9,55	152,80
[23-31[	4	27	1080	17,55	70,20
Total	219	-	2069,5	-	710,65

$$\bar{X} = \frac{\sum_{i=1}^n n_i c_i}{N} = \frac{2069,5}{219} = 9,45 MDA$$

$$D'où : e_a(X) = \frac{\sum_{i=1}^n n_i |c_i - \bar{X}|}{N} = \frac{710,65}{219} = 3,24 MDA$$

L'écart absolu moyen et donc de 3240 DA.

4.3.2. La variance et l'écart-type (ou standard-déviation)**a. La variance**

On appelle la variance de la variable X , la moyenne des carrés des écarts des valeurs observées par rapport à leur moyenne arithmétique $\bar{X}$.

$$\text{On note : } V(X) = \frac{\sum_{i=1}^n n_i (x_i - \bar{X})^2}{N} = \sum f_i (x_i - \bar{X})^2,$$

Elle signifie la fluctuation (ou la variation) des valeurs observées autour de leur moyenne. Ce carré permet de mesurer l'aptitude de la variable X à varier ou à fluctuer autour de sa moyenne.

La variance est toujours positive, et elle peut aussi se calculer à partir d'une formule développée déduite de la première formule.

$$V(X) = \frac{\sum_{i=1}^n n_i \times x_i^2}{N} - \bar{X}^2$$

b. L'écart type

L'écart type noté $\sigma(X)$ est la racine au carrée de la variance. On note : $\sigma(X)$

4.3.2.1. Cas d'une variable discrète

Exemple : Si nous reprenons l'exemple des données représentant le nombre d'absences d'un échantillon de 10 étudiants.

9 2 1 1 7 7 7 4 6 7

Tableau 22 : Répartition des 10 étudiants selon le nombre d'absences

X_i	n_i	$n_i \times x_i$	$n_i(x_i - \bar{X})^2$	$n_i \times x_i^2$
1	2	2	32	2
2	1	2	9	4
4	1	4	1	16
6	1	6	1	36
7	4	28	16	196
9	1	9	16	81
Total	10	51	75	335

$$\bar{X} = \frac{\sum_{i=1}^n n_i x_i}{N} = \frac{51}{10} \approx 5$$

$$V(X) = \frac{\sum_{i=1}^n n_i X_i^2}{N} - \bar{X}^2$$

$$V(X) = \frac{335}{10} - (5,1)^2 = 7,49$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{7,49} = 2,73$$

$$\sigma(X) \approx 3 \text{ absences}$$

4.3.2.2. Cas d'une variable continue

Dans ce cas, La variance est donnée par la formule suivante :

$$V(X) = \frac{\sum_{i=1}^n n_i c_i^2}{N} - \bar{X}^2$$

Avec : c_i = centre de la classe et n_i = effectif de la classe i.

Exemple : le tableau ci-après donne la répartition des ouvriers d'une firme multinationale selon leurs salaires mensuels en milliers de dinars (MDA)

Tableau 23 : Répartition des ouvriers selon leurs salaires mensuels.

Salaires (MDA)	n_i	c_i	$n_i \times c_i$	$n_i \times c_i^2$
[40-50 [	12	45	540	24300
[50-60 [	23	55	1265	69575
[60-70 [	42	65	2730	177450
[70-90 [	56	80	4480	358400
[90-110[	34	100	3400	340000
[110-150[	32	130	4160	540800
[150-230[	16	190	3040	577600
[230-310[	4	270	1080	291600
Total	219	-	20695	2379725

$$\bar{X} = \frac{\sum_{i=1}^n n_i c_i}{N} = \frac{20695}{219} \approx 94,5$$

$$\text{D'où : } V(X) = \frac{\sum_{i=1}^n n_i \times c_i^2}{N} - \bar{X}^2$$

$$V(X) = \frac{2379725}{219} - (94,5)^2$$

$$V(X) = 1936,5$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{1936,5}$$

$$\sigma(X) = 44 \text{ MDA}$$

4.4. Coefficient de dispersion relative : coefficient de variation

L'écart-type s'exprimant dans la même unité que les modalités du caractère, peut rendre difficiles les comparaisons de dispersion si les variables sont exprimées en unités différentes. Le coefficient de variation (CV) est un nombre sans unité de mesure, c'est le rapport de l'écart type $\sigma(X)$ sur la moyenne $\bar{X}$ de l'échantillon étudié.

$$CV = \frac{\sigma(X)}{\bar{X}} \times 100$$

Le CV est également indépendant des unités choisies, il permet de comparer les séries statistiques exprimées en unités de mesure différentes.

L'usage du CV répond donc à la faiblesse de l'écart type qui est un paramètre simple de dispersion. Un autre paramètre permet aussi d'écarter dans son calcul l'unité de mesure. Il s'agit du coefficient interquartile noté C_Q , dont la formule est la suivante :

$$C_Q = \frac{Q_3 - Q_1}{M_e} \times 100$$