Université A- mira de Bejaia
Faculté des sciences économiques, commerciales et des sciences de gestion
Département SEGC(LMD)
Module stat I
Enseignante : Dr. BERRAH

Chapitre 4 : Les paramètres de dispersion

Les mesures de tendance centrale, abordées dans le chapitre précédent, fournissent une vue d'ensemble du comportement des données. Cependant, elles ne révèlent pas comment ces données sont regroupées, ni leur variabilité au sein de l'échantillon. C'est là qu'interviennent les mesures de dispersion, qui renseignent sur la dispersion ou l'éparpillement des données autour notamment des paramètres de tendance centrale.

Ce chapitre se concentrera sur les principales mesures de dispersion, en mettant l'accent sur la variance et l'écart-type. Nous examinerons :

- L'étendue.
- L'intervalle interquartile.
- L'écart absolu moyen.
- La variance et l'écart-type.
- Le coefficient de variation.

1. L'étendue

L'étendue notée **E** est la différence entre la plus grande et la plus petite des valeurs observées de la variable statistique. Elle est utilisée pour obtenir très rapidement une première idée de la dispersion de la série. On note : $E=X_{max} - X_{min}$

Le principal avantage de ce paramètre est sa simplicité de calcul, mais il est, par contre, extrêmement peu robuste.

Exemple 1 (Cas d'une variable discrète) : Soit la distribution suivante qui indique les notes obtenues par 10 étudiants lors de deux examens.

xi	ni
3	1
4	1
5	2
7	2
8	1
9	2
11	1

Solution : L'étendue est égale à : $E=X_{max} - X_{min}$ = (11-3) = 8.

Exemple 2 (Cas d'une variable continue) : soit la distribution suivante dans le tableau ci-après donne la répartition de 24 étudiants selon leur taille.

X	n_i
[1,5 ; 1,6 [	6
[1,6 ; 1,7 [	7
[1,7 ; 1,8 [	8
[1,8 ; 1,9 [	2
[1,9 ; 2 [	1
Total	24

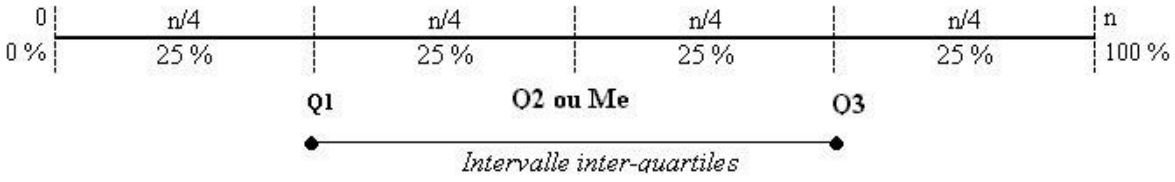
Solution : Dans ce cas, L'étendue est la différence entre la borne supérieure de la dernière classe et la borne inférieure de la première classe.

L'étendue est égale à : $E=X_{max} - X_{min}$ = (2 - 1,5) = 0.5.

2. Les intervalles et écarts interquartiles

Il existe trois intervalles et écarts interquartiles :

- **L'intervalle interquartile** $[Q_1;Q_3]$ est l'intervalle qui contient 50% des observations ; l'amplitude de l'intervalle est appelé **écart interquartile**, noté *EIQ* est égale à : $EIQ = Q_3 - Q_1$



- **L'intervalle interdécile** $[D_1;D_9]$ est l'intervalle qui contient 80% des observations ; l'amplitude de l'intervalle est appelé **écart interdécile**, noté *EID* est égale à : $EID = D_9 - D_1$

- **L'intervalle intercentile** $[C_1;C_{99}]$ est l'intervalle qui contient 98% des observations ; l'amplitude de l'intervalle est appelé **écart intercentile**, noté *EIC* est égale à : $EIC = C_{99} - C_1$

Plus ces intervalles sont importants, plus la dispersion est forte.

3. Ecarts absolus moyens

L'écart absolu moyen est la moyenne arithmétique des valeurs absolues des écarts à un indicateur de tendance centrale, en général la moyenne ou la médiane.

- **Ecart absolu moyen par rapport à la moyenne arithmétique** : $e_{\overline{X}} = \dfrac{\sum_{i=1}^{k} n_i \left| X_i - \overline{X} \right|}{N}$
- **Ecart absolu moyen par rapport à la médiane** : $e_{Mé} = \dfrac{\sum_{i=1}^{k} n_i \left| X_i - Mé \right|}{N}$

Remarque : Lorsque la variable est continue, on remplace les valeurs de xi par les centres de classe et on applique le même principe pour calculer l'écart absolu moyen.

Exemple 1 (cas variable discrète) : la distribution suivante représente les notes obtenues par les étudiants à l'examen statistique.

Notes xi	7	8	9	10	11	12	14
Etudiants ni	10	15	20	25	15	10	5

- Calculer l'écart absolu moyen par rapport à la moyenne arithmétique.

Solution : Pour rendre les calculs simples et facile, on dresse le tableau suivant :

X_i	n_i	$n_i \cdot X_i$	$\left\|X_i - \overline{X}\right\|$	$ni \cdot \left\|X_i - \overline{X}\right\|$
7	10	70	2,75	27,5
8	15	120	1,75	26,25
9	20	180	0,75	15
10	25	250	0,25	6,25
11	15	165	1,25	18,75
12	10	120	2,25	22,5
14	5	70	4,25	21,25
Total	100	975	-	137,5

On calcule d'abord la moyenne arithmétique :

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i X_i}{N} = \frac{975}{100} = 9{,}75$$

L'écart absolu moyen par rapport à la moyenne arithmétique =

$$e_{\bar{X}} = \frac{\sum_{i=1}^{k} n_i \left| X_i - \bar{X} \right|}{N} = \frac{137,5}{100} = 1,375$$

Exemple 2 (cas variable continue): on considère la distribution suivante :

Classes	[15-25[	[25-35[	[35-45[	[45-55[	[55-65[	[65-75[	[75-85[
Effectifs	9	15	22	29	17	6	2

- Calculer l'écart absolu moyen par rapport à la moyenne.
- Calculer l'écart absolu moyen par rapport à la médiane.

Solution : on dresse le tableau suivant :

Classes	n_i	c_i	$n_i \cdot c_i$	$\left\|c_i - \bar{X}\right\|$	$n_i \cdot \left\|c_i - \bar{X}\right\|$	$n_i \uparrow$	$\left\|c_i - Mé\right\|$	$n_i \cdot \left\|c_i - Mé\right\|$
[15-25[	9	20	180	25,6	230,4	9	26,38	237,42
[25-35[	15	30	450	16,6	234,0	24	16,38	245,70
[35-45[	22	40	880	5,6	123,2	46	6,38	104,36
[45-55[	29	50	1450	4,4	127,6	75	3,62	104,98
[55-65[	17	60	1020	14,4	244,8	92	13,62	231,54
[65-75[	6	70	420	24,4	146,4	98	23,62	141,72
[75-85[	2	80	160	34,4	68,8	100	33,62	67,24
Total	**100**	**-**	**4560**	**-**	**1175,2**	**-**	**-**	**1168,96**

- $\bar{X} = \frac{\sum_{i=1}^{n} n_i c_i}{N} = \frac{4560}{100} = 45,6 \Rightarrow e_{\bar{X}} = \frac{\sum_{i=1}^{k} n_i \left| c_i - \bar{X} \right|}{N} = \frac{1175,2}{100} = 11,75$

- $Mé \in [45;55[\Rightarrow Mé = X_{\min} + \frac{n/2 - N_{i\ précédent}^{\uparrow}}{ni_{Mé}} \times ai = 45 + \frac{50-46}{29} \times 10 = 46,38$

$$\Rightarrow e_{Mé} = \frac{\sum_{i=1}^{k} n_i \left| X_i - Mé \right|}{N} = \frac{1168,96}{100} = 11,69$$

4. La variance

La variance de la variable X est la moyenne des carrés des écarts des valeurs observées par rapport à la moyenne arithmétique. Elle signifie la variation des valeurs observées autours de leur moyenne. On la note généralement V(X). Sa formule correspondant à sa définition :

$$V(X)=\frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{N} \quad ou \quad V(X) = \sum f_i (X_i - \overline{X})^2 \text{..... } \textit{formule de définition}$$

Elle peut aussi se calculer à partir d'une formule développée déduite de la formule précédente : $V(X)=\frac{\sum_{i=1}^{n} n_i x_i^2}{N} - \overline{X}^2$ *ou* $V(X) = \sum_{i=1}^{n} fi\ x_i^2 - \overline{X}^2$ ..... *formule développée*

Propriétés de la variance

1. $V(X) \geq 0$
2. Soit Y une variable telle que : $y_i = a\chi_i + b$, sachant a et b sont des constantes réelles, alors : $V(Y) = a^2 V(X)$

Remarque : Pour calculer la variance dans le cas continu, on remplace les valeurs (xi) par les centres de classe ci, comme suite :

$$V(X)=\frac{\sum_{i=1}^{n} n_i (c_i - \overline{X})^2}{N} \quad \Rightarrow \quad V(X) = \sum f_i (c_i - \overline{X})^2$$

O u $V(X)=\frac{\sum_{i=1}^{n} n_i c_i^2}{N} - \overline{X}^2 \quad \Rightarrow \quad V(X) = \sum_{i=1}^{n} fi\ ci^2 - \overline{X}^2$

5. L'écart type

L'écart type est un indicateur de dispersion. Il est défini comme la racine carrée de la variance, il s'exprime dans la même unité que la variable xi. On le note généralement : $\sigma(X)$. Donc : $\sigma(X) = \sqrt{v(x)}$.

Cet indicateur mesure la distance moyenne entre $\overline{X}$ et les valeurs de X. Il sert à mesurer la dispersion d'une série statistique autour de sa moyenne.

- Plus il est petit, plus les caractères sont concentrés autour de la moyenne (on dit que la série est homogène).
- Plus il est grand, plus les caractères sont dispersés autour de la moyenne (on dit que la série est hétérogène).

6. Le coefficient de variation (coefficient de dispersion relative)

Malgré la pertinence de l'écart-type dans la mesure de la dispersion d'une distribution, il possède un inconvénient majeur : il est exprimé dans la même unité de la variable à laquelle il se rapporte, il est alors impossible de comparer les dispersions de deux distributions ayant un lien entre elles et dont les valeurs s'expriment dans des unités différentes.

Pour comparer la dispersion de deux séries qui ne sont pas exprimées dans les mêmes unités, on utilise le coefficient de variation (CV). Cette statistique est une mesure sans unité, qui s'exprime la plupart du temps en pourcentage. Il se calcule en divisant l'écart-type par la moyenne arithmétique et s'écrit donc : $CV = \frac{\sigma}{\overline{X}}$

Plus la valeur de ce coefficient est grande, plus la dispersion est grande.

Exemple1 (cas d'une variable discrète) : Le tableau suivant indique la distribution de notes obtenues par les étudiants à l'examen de statistique.

Notes	5	6	7	8	9	10	11	12	13	14	15
Etudiants	2	9	14	20	18	15	9	6	4	2	1

- Calculer la variance, l'écart type et le coefficient de variation.

Solution

On calcule la variance par la formule de définition, donc on dresse le tableau de cette manière :

X_i	n_i	$n_i X_i$	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$n_i (X_i - \overline{X})^2$
5	2	10	-4	16	32
6	9	54	-3	9	81
7	14	98	-2	4	56
8	20	160	-1	1	20
9	18	162	0	0	0
10	15	150	1	1	15
11	9	99	2	4	36
12	6	72	3	9	54
13	4	52	4	16	64
14	2	28	5	25	50
15	1	15	6	36	36
Total	100	900	-	-	444

- **La moyenne arithmétique :**

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \frac{900}{100} = 9 .$$

- **La variance :**

$$V(X) = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{N} = \frac{444}{100} = 4{,}44 \ .$$

- **L'écart-type :**

$$\sigma(X) = \sqrt{V(X)} = \sqrt{4{,}44} = 2{,}107 .$$

- **Le coefficient de variation :**

$$CV = \frac{\sigma}{\overline{X}} = \frac{2{,}107}{9} = 0{,}23$$

Exemple 2 (Cas d'une variable continue)

Le tableau ci-dessous indique la distribution des salaires mensuels des travailleurs d'une entreprise.

Classes	[50-60[	[60-70[	[70-80[	[80-90[	[90-100[
Effectifs	25	40	20	10	5

- Calculer la variance, l'écart-type et le coefficient de variation.

Solution

Dans ce cas, on calcule la variance par la formule développée, on doit construire alors le tableau de la manière suivante :

Classes	n_i	c_i	$n_i \times c_i$	$n_i \cdot c_i^2$
[50-60[	25	55	1375	75625
[60-70[	40	65	2600	169000
[70-80[	20	75	1500	112500
[80-90[	10	85	850	72250
[90-100[	5	95	475	45125
Total	100	-	6800	474500

On calcule d'abord **la moyenne arithmétique** :

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i c_i}{N} = \frac{6800}{100} = 68$$

La variance est égale à :

$$V(X) = \frac{\sum_{i=1}^{n} n_i c_i^2}{N} - \overline{X}^2 = \frac{474500}{100} - (68)^2 = 121$$

L'écart – type :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{121} = 11.$$

Le coefficient de variation :

$$CV = \frac{\sigma}{\overline{X}} = \frac{11}{68} = 0{,}16$$

Série des exercices

Exercice 1 : Le tableau suivant indique la distribution d'une population de 16 femmes d'un village selon leur nombre d'enfants.

Nombre d'enfants par femme (xi)	Nombre de femmes (ni)
0	1
1	3
2	8
3	4

1- Déterminer la population statistique.
2- Quel est le caractère de cette distribution ? préciser sa nature.
3- Calculer le nombre moyen d'enfants par femme.
4- Calculer la variance ainsi que l'écart type.
5- Calculer le coefficient de variation.

Exercice 2

Une enquête auprès de 500 visiteurs d'un musée âgés d'au moins 15 ans permet d'obtenir la distribution statistique ci-après :

Ages (en année)	[15 - 25[	[25 - 35[	[35 - 50[	[50- 65[	[65 - 75[
Nombre de visiteurs	96	118	138	101	47

1- Déterminer la population statistique, le caractère et sa nature.

2- Présenter graphiquement cette distribution.

3- Calculer la moyenne arithmétique.

4- Calculer les caractéristiques de dispersion suivantes : l'étendue, l'écart absolu moyen, l'écart type et coefficient de variation.