4. Caractéristiques des distributions à une variable

4.1. Les indicateurs de tendance centrale

Ce sont des valeurs qui permettent de résumer une série statistique pour une variable quantitative et qualitative. Il s'agit principalement du mode, de la médiane et de la moyenne.

a. Le mode

Le mode d'une distribution statistique est la ou les valeurs que l'on rencontre le plus fréquemment dans un ensemble de données. On l'appelle aussi valeur dominante.

2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 **Mode** = 9

3, 5, 8, 9, 12, 15, 16, 18, 23, 32, 40 Mode = Pas de mode

1, 2, 2, 3, 3, 4, 4, 4, 5, 5, 6, 7, 7, 7 **Mode =** 4 et 7

Comment repérer le mode ?

- **Cas des variables discrètes** : sur le tableau statistique, c'est la variable xi pour laquelle la fréquence ou l'effectif est la plus élevée ; sur le diagramme en bâtons, c'est la valeur de la variable xi correspondant au bâton le plus haut.

Nombre d'enfants	Effectif
1	10
2	15
3	15
4 = Mode	30
5	10
6	12
7	5
8	3

- **Cas des variables continues** : on définit la classe modale, c’est-à-dire celle qui a la plus grande fréquence ou l’effectif le plus élevé.

Loyer annuel en euros (xi)	Effectif (ni)
[5000 - 6000[	50
[6000 - 7000[	120
[7000 - 8000[= classe modale	150
[8000 - 9000[	80
[9000 - 10000[	60
[10000 - 11000[	40

b. La médiane

La Médiane d'un ensemble de nombres rangé par ordre de grandeur croissante ou décroissante est la valeur du milieu ou la moyenne arithmétique des valeurs centrales. Elle partage la série en deux sous-ensembles égaux.

3, 4, 4, 5, 6, 8, 8, 8, 10 Médiane = 6

5, 5, 7, 9, 11, 12, 15, 18 Médiane = (9 + 11) / 2 = **10**

Calcul de la médiane : Lorsqu'on a un tableau de la forme (xi ; ni), le calcul se fait à partir des effectifs cumulés $N(x)$. On calcule d'abord les effectifs cumulés, puis on repère dans le tableau de calcul, la valeur n/2.

$$\boldsymbol{M\acute{e}diane} = \boldsymbol{xi} + (\boldsymbol{a}\frac{\boldsymbol{i}\ (\boldsymbol{n}/\boldsymbol{2} - \boldsymbol{N}(\boldsymbol{x})}{\boldsymbol{ni}})$$

xi = origine de la classe médiane (borne gauche de la classe médiane)

ai = amplitude de la classe médiane

N(x) = Valeur de l'effectif cumulé correspondant

ni = Effectif partiel de la classe médiane

1. On calcule d'abord n/2 = 500/2 = 250
2. On repère la position de la valeur (250) dans le tableau des effectifs cumulés
3. On applique la formule

xi	ni	Effectif cumulé N(x)
[5000 - 6000[	50	50
[6000 - 7000[	120	170
[7000 - 8000[	150	320
[8000 - 9000[	80	400
[9000 - 10000[	60	460
[10000 - 11000[	40	500

n = 500

Médiane = 7000 + [1000 $\frac{(250 - 170)}{150}$

Médiane = 7 533,33

Application : calculer la médiane

xi	ni
[47 - 52[	10
[52 - 57[	30
[57 - 60[	60
[60 - 63[	72
[63 - 67[	40
[67 - 73[	24
[73 - 83[	14

$$\boldsymbol{M\acute{e}diane = xi + [\frac{ai\,(n/2 - N(x)}{ni}]}$$

c. La moyenne

La moyenne arithmétique d'une série statistique est égale au rapport de la somme des valeurs observées par le nombre d'observations.

- Moyenne arithmétique **simple ou non pondérée** : lorsqu'à chaque valeur de xi ne correspond qu'une seule observation. Dans ce cas :

- $\bar{x} = \frac{\sum x\mathrm{i}}{\Sigma n\mathrm{i}}$ **(formule de la moyenne arithmétique simple)**

3 5 7 9 10 11 12 18 $\bar{x}$ = (3 + 5 +7 +9 +10 +11 + 12 + 18)/8 = 9,375

- Moyenne arithmétique **pondérée** : lorsqu'à chaque valeur de xi peut correspondre plusieurs observations. Dans ce cas :

- $\bar{x} = \frac{\sum (x_i n_i)}{\sum n_i}$ **(formule de la moyenne arithmétique pondérée)**

Exemple : 3 3 3 5 9 9 11 11 $\bar{x} =$ (3 x 3) + (5 x1) + (9 x 2) + (11 x 2)/8 = 6,75

- **Cas d’une variable discrète**

L’entreprise E compte 30 ouvriers salariés à 6000 frs, 15 à 7000 frs et 5 cadres à 10000 frs. Calculez le salaire moyen dans cette entreprise.

x_i	n_i	$n_i\, x_i$
6000	30	180 000
7000	15	105 000
10 000	5	50 000
Totaux	Σni = 50	Σnixi = 335 000

$$\bar{x} = \frac{\sum (nix_i)}{\sum n_i}$$

$$\bar{x} = \frac{335\ 000}{50} = 6700\text{frs}$$

- **Cas d'une variable continue**

Par convention, on choisit pour le calcul, le centre de classe de chaque classe de la variable. Le centre de classe est égal à la somme des bornes de la classe sur 2.

- $\boldsymbol{\bar{x} = \frac{\sum x_i n_i}{\sum n_i}}$ $\boldsymbol{x_i}$ **=** centre de classe

xi	$\boldsymbol{n_i}$	**Centre de classe (**$\boldsymbol{x_i}$ **ou ci)**	$\boldsymbol{n_i\, x_i}$
[3000-4000[	26	3500	91 000
[4000-5000[	33	4500	148 500
[5000-6000[	74	5500	407 000
[6000-7000[	7	6500	45 500
[7000-8000[	10	7500	75 000
	Σni = 150		Σnixi = 766 500

$\bar{x}$ **= 766 500 / 150 = 5110**

Application

Soit la distribution statistique suivante. Calculer la moyenne.

xi	ni
[10 - 20[	165
[20 - 40[	178
[40 - 50[	83
[50 - 60[	54
[60 - 80[	38
[80 - 100[	29

4.2. Les indicateurs de dispersion

Ils permettent de comprendre comment les données sont réparties. Ils donnent un aperçu de la variabilité des données. Une dispersion faible signifie que les données sont regroupées autour d'une valeur centrale, une dispersion élevée indique que les données sont plus étalées.

a. L'intervalle de variation Etendue ou Range

C'est la différence entre la plus grande et la plus petite des valeurs observées de la variable en d'autres termes c'est la différence entre les deux valeurs extrêmes.

Exemples : Salaires horaires de 50 surveillants d'examen d'un établissement A

1) 23 ; 24 ; 25 ; 26 ; 27 Etendue = 27 – 23 = 4 dispersion faible

2) 20 ; 26 ; 32 ; 34 ; 45 Etendue = 45 – 20 = 25 forte dispersion

b. L’écart absolu moyen par rapport à la moyenne

Il permet de quantifier la dispersion des valeurs dans un ensemble de données par rapport à la moyenne.

$$e_{\overline{x}} = \frac{\sum (n_i|x_i - moy|)}{\Sigma n_i}$$

xi	n_i
[47 - 52[	10
[52 - 57[	30
[57 - 60[	60
[60 - 63[	72
[63 - 67[	40
[67 - 73[	24
[73 - 83[	14

Σni=250

$$e_{\overline{x}} = \frac{\sum(n_i|x_i - moy|)}{\Sigma n_i} = 1109{,}44 / 250 = 4{,}43$$

moy = 15 440 / 250 = 61,76

Classe	n_i	Centre de classe (x_i)	($x_i n_i$)	(x_i – moy)	n_i (x_i – moy)
[47 - 52[	10	49,5	495	12,26	122,6
[52 - 57[	30	54,5	1635	7,26	217,8
[57 - 60[	60	58,5	3510	3,26	195,6
[60 - 63[	72	61,5	4428	0,26	18,72
[63 - 67[	40	65	2600	3,24	129,6
[67 - 73[	24	70	1680	8,24	197,76
[73 - 83[	14	78	1092	16,24	227,36

Σni=250

Σxin_i =15 440

Σn_i (xi – moy) =1109,44

c. L'écart absolu moyen par rapport à la médiane

Il permet de quantifier la dispersion des valeurs dans un ensemble de données par rapport à la médiane.

$e_{mé} = \frac{\sum n_i |x_i - méd|}{\Sigma n_i}$ **= 1098 / 250 = 4,392 dispersion faible**

Classe	n_i	N(x)	Centre de classe (x_i)	$(x_i$ – méd)	n_i $(x_i$ – méd)
[47 - 52[	10	10	49,5	11,54	115,4
[52 - 57[	30	40	54,5	6,54	186,2
[57 - 60[	60	100	58,5	2,54	152,4
[60 - 63[	72	172	61,5	0,46	33,12
[63 - 67[	40	212	65	3,96	158,4
[67 - 73[	24	236	70	8,96	215,04
[73 - 83[	14	250	78	16,96	237,44

Σni = 250

Σni(xi – méd) = 1098

Médiane = 60 + (3 (125 – 100)/72 = 61,04

d. La variance

La variance mesure à quel point les valeurs d'un ensemble de données sont dispersées autour de leur moyenne.

$$\mathrm{V(x)} = \frac{\sum(\boldsymbol{n_i}|\boldsymbol{x_i} - \boldsymbol{moy}|^2)}{\boldsymbol{\Sigma n_i}}$$

- Une variance élevée reflète un écart important entre les données.
- Une variance faible reflète des valeurs regroupées autour de la moyenne.

$\bar{x}$ **(moyenne) = 15 440 / 250 = 61,76**

Classe	n_i	Centre de classe (x_i)	($nixi$)	(x_i – moy)	(x_i – moy)²	n_i (x_i - moy)²
[47 - 52[	10	49,5	495	**12,26**	**150,3**	**1503**
[52 - 57[	30	54,5	1635	**7,26**	**52,7**	**1581**
[57 - 60[	60	58,5	3510	**3,26**	**10,6**	**636**
[60 - 63[	72	61,5	4428	**0,26**	**0,06**	**4,32**
[63 - 67[	40	65	2600	**3,24**	**10,4**	**416**
[67 - 73[	24	70	1680	**8,24**	**67,8**	**1627,2**
[73 - 83[	14	78	1092	**16,24**	**263,7**	**3691,8**

Σ n_i = 250

Σ n_i xi = 15440

Σ n_i (x_i - moy)² =

V(x) = 9459,32 / 250 = 37,83

$\sigma = \sqrt{v(x)} = \sqrt{37,83 = 6,15}$

e. L’écart-type (σ)

Il permet aussi d’indiquer dans quelle mesure les valeurs d'un ensemble sont éloignées de la moyenne de cet ensemble.

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{v}(\boldsymbol{x})}$$