Les paramètres de dispersion

Dr : RAIAH. M
Laboratoire de Biostatistique, Faculté de Médecine d'Oran

I- Introduction

- **Objectifs de la statistique descriptive :**
  Résumer, synthétiser l'information contenue dans la série statistique et mettre en évidence ses propriétés.
- **Paramètres utilisés:**
  - **Les paramètres de tendance centrale (de position):**
    Permettent d'obtenir une idée de l'ordre de grandeur des valeurs de la série et indiquent la position où semble se rassembler les valeurs de la série.
  - **Les paramètres de dispersion:**
    Quantifient les fluctuations des valeurs observées et leur étalement.

II- Les paramètres de dispersion

1 - La variance
2 - L'écart type
3 - Le coefficient de variation
4 - L'étendue
5 - L'écart interquartile

Citons un exemple concret

1ère **série** (notes sur 20) : **3 4 6 14 18 19 20**
2ème **série** (notes sur 20) : **5 7 10 14 15 16 17**

⇓

On constate que les deux séries ont
la même moyenne ($\bar{X}$ **= 12**)
la même médiane (**Me = 14**)
mais la 2ème série est moins **dispersée** que la 1ère

⇓

Donc il est nécessaire de disposer d'autres paramètres qui étudient la dispersion (structure interne)

1-La variance

La variance est une distance moyenne des observations à la moyenne arithmétique qui, plus précisément c'est la moyenne arithmétique des carrés des écarts à la moyenne arithmétique.

Remarque:

La variance est un intermédiaire statistique au calcul de l'écart-type.

Cas des données non groupées

On définit la variance S^2 d'un ensemble de n valeurs x1, x2,................,xn par :

$$S^2 = \frac{\Sigma (Xi - \overline{X})^2}{n} = \frac{\Sigma Xi^2}{n} - (\overline{X})^2$$

Cas des données groupées

Quand x1, x2, x3,.........xN ont respectivement des effectifs d'apparition n1, n2, n3 ....nN, la variance prend la forme :

$$S^2 = \frac{\Sigma ni(Xi - \overline{X})^2}{n} = \frac{\Sigma ni\, xi^2}{n} - (\overline{X})^2$$

2-L'écart-type

La variance est une quantité élevée au carré. Si la variable x représente un poids exprimé en kg, alors la variance sera exprimée en kg², unité qui n'a pas de sens.

On utilise l'écart- type qui est la racine carrée de la variance afin d'obtenir une unité homogène avec celle de la variable statistique x.

$$S=\sqrt{S^2}$$

Exemples

1- Données non groupées :

Soit la série (en kg) : 5, 7, 10, 14

- **La moyenne**

$\overline{X} = (1/n)\ \Sigma\, xi = (1/4)\ \Sigma\, xi = \frac{5+7+10+14}{4} = \mathbf{9\ kg}$

- **La variance**

$S^2 = (1/n)\ \Sigma\,(Xi - \overline{X})^2 = \frac{(5-9)^2+(7-9)^2+(10-9)^2+(14-9)^2}{4} = \mathbf{11{,}5\ kg^2}$

- **L'écart-type**

$S = \sqrt{11{,}5} = \mathbf{3{,}39\ kg}$

2- Données groupées :

Tableau 1 : Nombre d'enfants par famille

xi	ni	ni . xi	ni . xi²
0	4	0	0
1	5	5	5
2	10	20	40
3	16	48	144
4	18	72	288
5	14	70	350
6	7	42	252
7	6	42	294
	$\Sigma n_i = n = 80$	$\Sigma n_i . x_i = 299$	$\Sigma n_i . x_i^2 = 1373$

- $\overline{X} = (1/n)\ \Sigma\, ni\ xi = (1/80)\ \Sigma\, ni\ xi = \frac{299}{80} = 3{,}74$
- $S^2 = (1/n)\ \Sigma\, ni\ xi^2 - (\overline{X})^2 = 1373/80 - (3{,}74)^2 = 3{,}17$
- $S = \sqrt{3{,}17} = 1{,}78$

2- Données groupées en classes :

Tableau 2 : Taille de 20 étudiants

Taille (xi)	ni	Centre de classe (Xcc)	ni . Xcc	ni . Xcc²
[150-160[	3	155	465	72075
[160-170[	8	165	1320	217800
[170-180[	5	175	875	153125
[180-190[	4	185	740	136900
	$\Sigma n_i = n = 20$		$\Sigma n_i . x_{cc} = 3400$	$\Sigma n_i . x_{cc}^2 = 579900$

- $\overline{X} = (1/n)\ \Sigma\, ni\ xi = (1/20)\ \Sigma\, ni\ xi = \frac{3400}{20} = 170\ cm$
- $S^2 = (1/n)\ \Sigma\, ni\ xi^2 - (\overline{X})^2 = 579900/20 - (170)^2 = 95\ cm^2$
- $S = \sqrt{95} = 9{,}75\ cm$

3-Le coefficient de variation (CV)

C'est un paramètre de dispersion relative qui étudie la dispersion dans le cas d'une seule distribution

Il est défini par le rapport de l'écart type à la moyenne

$$CV = \frac{S}{\overline{X}}$$

- Le coefficient de variation est souvent exprimé sous forme de pourcentage, et est indépendant du choix des unités.
- Le coefficient de variation permet de comparer les dispersions de deux séries différentes soit par les unités, soit par leur nature (**exemple :** une série de tailles à une série de poids).

En général dans la pratique quand :

CV > 0,33 ⟹ dispersion importante
CV < 0,33 ⟹ dispersion faible

4 – L'étendue

L'étendue est la différence entre la plus grande et la plus petite valeur de la série statistique.

Exemple

taille (cm)
150
158
158
163
163
164
164
164
165
166
169
172
173
174
175
177
180
181
188
189

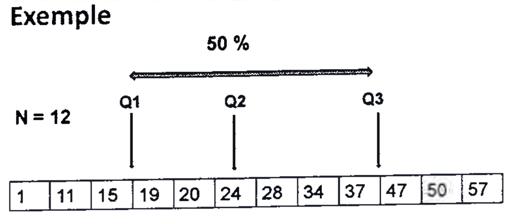
Étendue = 189-150=39 cm

5 – L'écart interquartile (IQ)

IQ : C'est la taille de l'intervalle situé au centre des données et incluant 50% des valeurs.

IQ = Q3 - Q1

Exemple



IQ = 42 - 17

IQ = 25

Références

- Schwartz D. Méthodes statistiques.1992
- Ancelle T. Statistique Épidémiologie. Édition 2002
- Mesli MF, Mokhtari A. Biostatistique. Édition mai 2007