Les paramètres de dispersion

TALHI. R
Service de Biostatistique, Faculté de Médecine d'Oran

Plan du cours

I- Introduction

Les objectifs de la statistique descriptive sont de résumer, synthétiser l'information contenue dans une série statistique et mettre en évidence ses propriétés.

Une distribution statistique est présentée sous forme de:

-tableaux statistiques,

-graphiques ou

-numérique (en utilisant des paramètres statistiques).

Paramètres utilisés:

-Les paramètres de tendance centrale (de position):

Permettent d'obtenir une idée de l'ordre de grandeur des valeurs de la série statistique 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.

Citons un exemple concret

1ère série (notes sur 2O) : **3 4 6 14 18 19 20**

2ème série (notes sur 2O) : **5 7 1O 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

En conclusion pour une même position ,nous avons une dispersion différente.

⬇

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

II- Les paramètres de dispersion

1 - La variance

2 - L’écart type

3 - Le coefficient de variation

4 - L’étendue

5 - L’écart interquartile

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.

1-a Cas des données non groupées:

On définit la variance S^2 d'un ensemble de n valeurs x_1, x_2,................,xn par :

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n} = \frac{\sum X_i^2}{n} - (\overline{X})^2$$

1-b Cas des données groupées:

Quand x_1, x_2, x_3,.........x_N ont respectivement des effectifs d'apparition n_1, n_2, n_3 ....n_N, la variance prend la forme :

$$S^2 = \frac{\sum n_i (X_i - \overline{X})^2}{n} = \frac{\sum n_i x_i^2}{n} - (\overline{X})^2$$

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:

Exemple 1: Données non groupées

Soit la série (en kg) : 5, 7, 1O, 14

-La moyenne

$\bar{X} = (1/n)\ \Sigma\ xi\ = (1/4)\ \Sigma\ xi\ = \dfrac{5+7+10+14}{4} =$ **9 kg**

-La variance

$S^2=(1/n)\ \Sigma\ (Xi-\ \bar{X})^2 = \dfrac{(5-9)^2+(7-9)^2+(1O-9)^2+(14-9)^2}{4} =$ **11,5 kg²**

-L'écart-type

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

En conclusion le poids moyen dans notre série statistique est de **9 ± 3,5 Kg**.

Exemple 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 = 4 enfants**

$S^2 = (1/n) \sum$ ni xi² $- (\overline{X})^2 = 1373/80 - (3,74)^2 =$ **3,17**

$S = \sqrt{3,17} =$ **1,78 = 2 enfants**

Le nombre d'enfants moyen par famille est de **4 ± 2 enfants.**

Exemple 3: Données groupées en classes

Tableau 2 : Taille de 20 étudiants

Taille (xi)	ni	Centre de classe (Xicc)	ni . xicc	ni . xicc²
[150-160[	3	155	465	72075
[160-170[	8	165	1320	217800
[170-180[	5	175	875	153125
[180-190[	4	185	740	136900
	$\Sigma n_i = n = 20$		**$\Sigma n_i . x_{icc} = 3400$**	**$\Sigma n_i . x_{icc}^2 = 579900$**

$\overline{X} = (1/n)\, \Sigma\, ni\, xi = (1/20)\, \Sigma\, ni\, xi = \frac{3400}{20}$ = **170 cm**

$S^2 = (1/n) \sum ni\, xi^2 - (\overline{X})^2 = 579900/20 - (170)^2 =$ **95 cm²**

$S = \sqrt{95} =$ **9,75 cm = 10 cm**

La taille moyenne des étudiants est de **170 ± 10 cm.**

Remarque

Plus l'écart type est **faible**, plus les valeurs de la série sont **concentrées** (moins dispersées) autour de leur moyenne.

-Le couple (moyenne et l'écart type) sont largement utilisés en statistique inductive (tests statistiques).

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.

$\bar{X} = 3S$ alors $S/\bar{X} = 1/3 = 0,33$ → situation « équilibrée »

Cette dernière est prise comme référence.

Pratiquement quand :

CV > O,33 →l’écart type est important (**dispersion importante**).

CV < O,33 → l’écart type est faible (**dispersion faible**).

4 – L’Etendue

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)

L’écart IQ représente la taille de l’intervalle situé au centre des données et incluant 50% des valeurs.

L'écart IQ représente la différence entre Q3 et Q1.

L’écart IQ = Q3 - Q1

Q1, Q2 et Q3 représentent **les quartiles.**

Les quartiles sont des **paramètres de position** qui divisent une série statistique ordonnée en **quatre groupes égaux** comprenant chacun 25% des données de la série.

On dit que :

-25% des données sont inférieures à Q1

-50% des données sont inférieures à Q2 →correspond aussi à **la médiane**

-75% des données sont inférieures à Q3

Le premier quartile (noté Q_1) est la valeur d'une série qui est supérieure ou égale à au moins **25 %** des données de la série ordonnée de valeurs statistiques.

Appelons N le nombre des valeurs d'une série, et calculons **0,25 ×N = N/4**.

Lorsque N/4 est **entier**, la valeur représentant le premier quartile est la **0,25$^{\text{ème}}$ valeur**.

Lorsque N/4 est un **décimal** non entier, on l'arrondit à l'entier supérieur p et alors la valeur représentant le premier quartile est la **p$^{\text{ème}}$ valeur**.

Exemple 1 :

Dans la série 10; 25; 30; 40; 41; 42; 50; 55; 70; 101; 110; 111.

Le premier quartile est **30**.

En effet, il y a **12** nombres dans cette série, et **12/4=3** donc

Le premier quartile est correspond à la 3ème valeur, soit **30**.

Exemple 2 :

Si **N/4=4,25** alors Q_1 est égale à la **cinquième** valeur.

Le troisième quartile (noté Q3) est la valeur d'une série qui est supérieure ou égale à au moins **75 %** des données de la série ordonnée de valeurs statistiques.

Lorsque **3N/4** est **entier**, la valeur représentant le troisième quartile est la **0,75ème** valeur.

Lorsque **3N/4** est un **décimal** non entier, on l'arrondit à l'entier supérieur p et alors la valeur représentant le troisième quartile est la **pième valeur**.

Exemple 1:

Dans la série: 10; 25; 30; 40; 41; 42; 50; 55; 70; 101; 110; 111.

Le troisième quartile est de 70.

En effet, il y a 12 nombres dans cette série, et **0,75×12 = 9.**

Le troisième quartile est donc la 9ème valeur, soit **70**.

Exemple 2 :

si **3N/4=15,25** alors Q3 est égale à la seizième valeur.

Attention : Pour le calcul des quartiles, la série doit d'abord être ordonnée selon l'ordre croissant (du plus petit au plus grand).

Exemple 3:

50 %

N = 12

Q1 **Q2** **Q3**

1	11	15	19	20	24	28	34	37	47	50	57

L'écart IQ = Q3 - Q1

12/4 = 3 donc **Q1** correspond à la 3ème valeur, soit 15
0,75×12 = 9 donc **Q3** correspond à la 9ème valeur, soit 37

L'écart IQ = Q3 - Q1= 37 – 15 = **25**

III- Bibliographie

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