

Objectifs

1-Résumer une distribution à l'aide de paramètres décrivant sa position,

2-Résumer une distribution à l'aide de paramètres décrivant son ampleur et sa dispersion,

3-Calculer déterminer les paramètres de tendance centrale : moyenne et médiane,

4-Calculer les paramètres de dispersion : variance et écart type.

Plan :

I- Paramètres de position :

1.1. Paramètres de tendance centrale : moyenne Arithmétique, médiane, mode.

1.2. Quantiles : quartiles, déciles, percentiles.

II- Paramètres de dispersion :

2.1. Extrêmes, Etendu

2.2. Variance et écart type

Conclusion

Références bibliographiques

I. PARAMETRES DE POSITION OU DE REDUCTION

Les paramètres ou statistiques de réduction sont les valeurs numériques permettant de résumer les caractéristiques principales de l'ensemble des mesures d'un **caractère quantitatif**.

1.1. PARAMETRES DE TENDANCE CENTRALE

Les paramètres de tendance centrale, sont des mesures qui localisent le centre de la distribution. Les plus utilisées sont : la moyenne arithmétique, la médiane et le mode.

1.1.1. Moyenne arithmétique :

La moyenne est une valeur calculée résultant de la somme algébrique des valeurs observées dans la série, divisée par le nombre de sujets.

Calcul de la moyenne :

x_i : correspond la valeur observée de l'individu (ou centre de la classe),

n_i : correspond à l'effectif de chaque valeur observée

N : correspond à l'effectif total de l'échantillon

m : moyenne.

$$m = \frac{\sum(n_i x_i)}{N}$$

Dans notre exemple (poids de 19 étudiants) elle est égale à 67,5 kg

La moyenne est centre de gravité des observations.

Elle est construite avec **toutes les valeurs** des observations.

Elle est sensible aux valeurs extrêmes

1.1.2. Médiane

La médiane est la **valeur** qui partage la série des individus en deux groupes d'effectifs égaux. Ainsi la moitié des sujets présente une valeur inférieure à la médiane, l'autre moitié présente une valeur supérieure à la médiane.

Lorsque les valeurs de la série statistiques sont connues, il suffit de les classer par ordre croissant (ou décroissant) et de prendre celle qui se trouve au milieu :

92,990 – 57,740 – 59,790 – 60,400 – 61,820 – 61,820- 64,100 – 64,990
– 65,450-66,330 – 68,280 – 69,120 – 70,130 – 70,560 – 72,880 – 76,340
– 76,360 – 79,650-83,450

Dans notre exemple = ?

Lorsque les données sont groupées en classes, la médiane correspond si on utilise les effectifs cumulés à $n/2$, ou si on utilise les fréquences relatives cumulées à 0.5 ou 50%.

Dans notre exemple = ?

1.1.3. Mode

Dans une distribution comportant de nombreuses données, le mode est la valeur qui revient le plus souvent, c'est la valeur la plus fréquente de la distribution. On l'appelle souvent pic de distribution.

1.2. QUANTILES

1.2.1. Quartiles

Ils divisent la série statistique en 4 parties égales comprenant le même nombre de sujets :

Le 1^{er} quartile ou quartile inférieur est la valeur du 25^{ème} sujet /100

Le 2^{ème} quartile ou médiane : est la valeur du 50^{ème} sujet /100

Le 3^{ème} quartile ou quartile supérieur : est la valeur du 75^{ème} sujet / 100

1.2.2. Déciles :

Ils divisent la série statistique ou la distribution en 10 groupes égaux. Chaque groupe comprend 10% des effectifs.

Il existe 09 valeurs des déciles

1.2.3. Percentiles :

Les percentiles sont les valeurs qui divisent la distribution en 100 parties égales. Il existe 99 valeurs des percentiles.

II.PARAMETRES DE DISPERSION :

Les statistiques de dispersion renseignent sur l'étalement de la série autour de la mesure de la tendance centrale (moyenne).

Les caractéristiques de tendance centrale sont insuffisantes pour caractériser complètement une série statistique.

Les paramètres de dispersion permettent donc d'estimer dans quelle mesure les observations s'écartent de la tendance centrale.

2.1. EXTREMES ET ETENDU :

2.1.1. Extrêmes : Ce sont les valeurs extrêmes de la distribution, valeurs minimum et maximum. Elles donnent une idée brute de la dispersion de la distribution de la part et d'autre de la médiane.

2.1.2. Etendu : C'est la différence entre les deux valeurs extrêmes. L'étendu donne en un seul chiffre une idée de la distribution autour de la médiane.

2.2. VARIANCE ET ECART TYPE

Considérons les 2 séries :

Série 1 : 15- 20- 25- 30- 35

Série 2 : 5- 15- 25- 35- 45

La moyenne de la : 1^{ère} série $(15+ 20+ 25+ 30+ 35) / 5 = 25$

: 2^{ème} série $(5+ 15+ 25+ 35+ 45) / 5 = 25$

Les 2 séries ont la même moyenne mais visiblement elles ne se ressemblent pas. La 2^{ème} série est plus **étalée** que la 1^{ère} **autour de la moyenne**. On dit que la 2^{ème} série est plus **dispersée**.

Les paramètres de dispersion permettent donc d'estimer dans quelle mesure les observations s'écartent de la tendance centrale dont les plus connus sont : l'écart type et la variance

Variance = moyenne des carrés des écarts à la moyenne.

Elle est notée s^2

$$s^2 = \frac{\sum ni (x_i - m)^2}{N}$$

$$s^2 = \frac{\sum(ni x_i^2) - (\sum ni x_i)^2 / N}{N}$$

L'écart type est tout simplement la racine carrée de la variance : $\sqrt{s^2} = s$
C'est la caractéristique de dispersion la plus utilisée car la plus satisfaisante.

La variance et l'écart type renseignent tous les deux sur l'étalement de la série statistique autour de la moyenne.

L'écart type d'une distribution s'exprime dans la même unité de mesure que les observations. Si par exemple une distribution s'exprime en kg, la variance s'exprime en kg^2 , l'écart type s'exprime en kg.

Dans l'exemple de la distribution du poids des 19 étudiants,

Centre de classe (xi)	Effectif (ni)	ni xi	xi ²	ni xi ²
52,5	1	52,5	2756,25	2756,25
57,5	2	115	3306,25	6612,5
62,5	5	312,5	3906,25	19531,25
67,5	4	270	4556,25	18225
72,5	3	217,5	5256,25	15768,75
77,5	3	232,5	6006,25	18018,75
82,5	1	82,5	6806,25	6806,25
Total	19	$\sum (ni xi)$ 1282,5	32593,75	$\sum (ni xi^2)$ 87718,75

$$S^2 = \frac{[(87718,75) - (1282,5)^2/19]}{19} = \frac{(87718,75 - 86568,75)}{19}$$

$$= 1150/18 = 60,52$$

La variance : $s^2 = 60,5 \text{ kg}^2$

L'écart type : $s = 7,77 \text{ kg}$

Références :

-Abdeldjellil Bezzaoucha. Epidémiologie et bio statistique à l'usage des étudiants en sciences médicales .Office des Publications Universitaires OPU 1996.

-Thierry Ancelle. Statistique Epidémiologie. Maloine 2002.