Université Paul Valéry

Arts - Lettres - Langues
Sciences humaines & sociales

Paramètres statistiques

Administration **É**conomique et **S**ociale

Mathématiques
XA100M

L'objet des paramètres statistiques est de résumer, à l'aide de quelques valeurs clés, l'information donnée par l'observation d'une **variable quantitative**.

On considère un caractère **quantitatif discret** X. On note n l'effectif total.

Modalités	m_1	m_2	$\cdots$	m_p
Effectifs	n_1	n_2	$\cdots$	n_p

La **moyenne** de X est

$$☞ \quad \mathrm{Moy}(X) = \frac{1}{n}\sum_{i=1}^{p} n_i m_i = \frac{n_1 m_1 + \cdots + n_p m_p}{n}.$$

P
Q

La moyenne peut être exprimée avec les fréquences :

$$\begin{aligned}\mathrm{Moy}(X) = \frac{n_1 m_1 + \cdots + n_p m_p}{n} &= \frac{n_1}{n} m_1 + \cdots + \frac{n_p}{n} m_p \\ &= f_1 m_1 + \cdots + f_p m_p.\end{aligned}$$

☞ $$\mathrm{Moy}(X) = \sum_{i=1}^{p} f_i m_i = f_1 m_1 + \cdots + f_p m_p.$$

Exemple

On étudie le nombre d'étages des demeures d'un quartier de Montpellier.

Modalités	0	1	2	3	4	5	6
Effectifs	50	58	112	138	106	55	16

$$\mathrm{Moy}(X) = \frac{50 \times 0 + 58 \times 1 + 112 \times 2 + 138 \times 3 + 106 \times 4 + 55 \times 5 + 16 \times 6}{50 + 58 + 112 + 138 + 106 + 55 + 16} = \frac{1491}{535} = 2,79.$$

La moyenne n'est pas nécessairement une modalité.

On considère un caractère **quantitatif continu** X. On note n l'effectif total.

Les modalités sont des intervalles. Pour calculer la moyenne, on remplace ces intervalles par leurs centres.

Modalités	$[a_1, a_2[$	$[a_2, a_3[$	$\cdots$	$[a_p, a_{p+1}[$
Centres	$c_1 = \dfrac{a_1 + a_2}{2}$	$c_2 = \dfrac{a_2 + a_3}{2}$	$\cdots$	$c_p = \dfrac{a_p + a_{p+1}}{2}$
Effectifs	n_1	n_2	$\cdots$	n_p

La **moyenne** de X est

$$☞ \quad \mathrm{Moy}(X) = \frac{1}{n} \sum_{i=1}^{p} n_i c_i = \frac{n_1 c_1 + \cdots + n_p c_p}{n}.$$

Comme précédemment, on peut aussi utiliser les fréquences

$$☞ \quad \mathrm{Moy}(X) = \sum_{i=1}^{p} f_i c_i = f_1 c_1 + \cdots + f_p c_p.$$

Exemple

On étudie le poids (en kg) de 50 enfants.

Modalités	$[10,15[$	$[15,20[$	$[20,25[$	$[25,30[$
Centres	$\frac{10+15}{2}=12,5$	$\frac{15+20}{2}=17,5$	$\frac{20+25}{2}=22,5$	$\frac{25+30}{2}=27,5$
Effectifs	10	18	13	9

$$\text{Moy}(X)=\frac{10\times 12,5+18\times 17,5+13\times 22,5+9\times 27,5}{10+18+13+9}=\frac{980}{50}=19,6.$$

Si X est un caractère : on note $m_1,...,m_i,...,m_p$
– ses modalités si le caractère est discret
– le centre de ses modalités si le caractère est continu.

Pour chaque modalité m_i, l'effectif associé est n_i, l'effectif total est n.

On appelle **série statistique de** X l'ensemble des couples $(m_1, n_1),...,(m_p, n_p)$.

☞ La moyenne de X est le nombre qui approche le mieux la série statistique de X.

Si a est un nombre, la distance de a à la série de X est

$$d(a) = \frac{1}{n} \sum_{i=1}^{p} n_i (m_i - a)^2.$$

– Valorisation des grands écarts entre a et les m_i
– Valorisation des modalités de grand effectif
– Dévalorisation des petits écarts entre a et les m_i
– Dévalorisation des modalités de petit effectif.

La moyenne de X est le seul nombre $\mathrm{Moy}(X)$ tel que

$$d\left(\mathrm{Moy}(X)\right) \leq d(a)$$

pour tout nombre a.

Ce nombre $d\big(\mathrm{Moy}(X)\big)$ est la plus petite erreur possible en approchant la série statistique de X par un nombre. On l'appelle **variance** de X :

$$\begin{aligned} ☞ \quad \mathrm{Var}(X) &= d\big(\mathrm{Moy}(X)\big) \\ &= \frac{1}{n}\sum_{i=1}^{p} n_i\big(m_i - \mathrm{Moy}(X)\big)^2 \\ &= \sum_{i=1}^{p} f_i\big(m_i - \mathrm{Moy}(X)\big)^2 \end{aligned}$$

où f_i est la fréquence de m_i.

En utilisant

$$\big(m_i - \mathrm{Moy}(X)\big)^2 = m_i^2 - 2m_i\,\mathrm{Moy}(X) + \mathrm{Moy}(X)^2$$

on montre

$$☞ \quad \begin{aligned} \mathrm{Var}(X) &= \mathrm{Moy}(X^2) - \mathrm{Moy}(X)^2 \\ &= \frac{1}{n}\sum_{i=1}^{p} n_i m_i^2 - \mathrm{Moy}(X)^2 \\ &= \sum_{i=1}^{p} f_i m_i^2 - \mathrm{Moy}(X)^2. \end{aligned}$$

En résumé :

☞ La moyenne est la valeur qui approche le mieux une série statistique.

☞ La variance mesure l'erreur faite en remplaçant la série par la moyenne.

☞ Si on remplaçait la série par nombre différent de la moyenne, on ferait une erreur plus grande.

Si X est un caractère **quantitatif** de modalités

$$m_1 < m_2 < \cdots < m_i < \cdots < m_p.$$

La **médianne** est un nombre tel que
– l'effectif de l'ensemble des modalités supérieures à ce nombre
– l'effectif de l'ensemble des modalités inférieures à ce nombre
sont égaux.

Un tel nombre n'existe pas toujours et on donne des solutions approchées.

2.1. Médianne d'un caractère quantitatif discret.

On note F la fonction de distribution cumulative **en fréquence** du caractère quantitatif discret X.

☞ La médianne est la modalité m_i telle que

$$F(m_{i-1}) < 0,5 \text{ et } F(m_i) \geq 0,5.$$

La médianne d'un caractère quantitatif discret est donc une modalité. Si on imagine qu'on classe les individus par modalités croissantes puis arbitrairement à l'intérieur de chaque modalité, la médianne est la modalité associée à le individu qui a plus (ou autant) d'individus avant lui que d'individu après.

2.2. Médianne d’un caractère quantitatif continu.

On note F la fonction de distribution cumulative **en fréquence** du caractère quantitatif continu X.

Dans ce cas, il existe toujours un unique nombre x tel que $F(x) = 0,5$. C’est ce nombre qu’on appelle médianne de X.

2.3. Fractile d'un caractère quantitatif discret.

Le but d'un fractile d'ordre p est de partager les individus en p parties de même effectifs.

🛑 pour $p = 2$ on retrouve la notion de médianne.

On note F la fonction de distribution cumulative **en fréquence** du caractère quantitatif discret X.

Soit j un nombre entier parmi $1, 2, \ldots, p-1$:

☞ Le j^{e} fractile d'ordre p est la modalité m_i telle que

$$F(m_{i-1}) < \frac{j}{p} \text{ et } F(m_i) \geq \frac{j}{p}.$$

P

Q

2.4. Fractile d'un caractère quantitatif continu.

Soit j un nombre entier parmi $1, 2, \ldots, p-1$:

On note F la fonction de distribution cumulative **en fréquence** du caractère quantitatif continu X.

Dans ce cas, il existe toujours un unique nombre x tel que $F(x) = \frac{j}{p}$. C'est ce nombre qu'on appelle j^{e} fractile d'ordre p de X.

2.5. Cas particuliers.

☞ Les fractiles d'ordre 10 s'appellent déciles

☞ Les fractiles d'ordre 100 s'appellent centiles

☞ Les fractiles d'ordre 4 s'appellent quartiles

Si Q_1 est la premier quartile et Q_3 le troisième, on appelle **espace interquartiles** le nombre $Q_3 - Q_1$.

Il y a 50% de la population associée à des modalités comprises entre Q_1 et Q_3.