TEMA IIC MEDIDAS DE DISPERSIÓN E FORMA

Estas medidas utilízan-se para medir a variabilidade ou esparcemento dos valores da distribución en torno a un valor central. Serven para determinar a bondade dunha medida de posición tomada. Por exemplo, á parte de ter a média mostral interésa-nos saber se en xeral os valores están concentrados en torno á média ou non.

Hai dous tipos:

Absolutas Dependen da dimensión da variábel. Teñen unidades de medida e inflúe a unidade na que estamos traballando. Non permiten facer comparación entre várias variábeis.

son percorrido ou rango, percorrido intercuartílico, desviación absoluta respeito á média ou á mediana, **varianza**, **desviación típica e MEDA**.

Relativas Dependen da dimensión da variábel. Non teñen unidades de medida, son adimensionais. Permiten facer comparación entre várias variábeis.

son coeficiente de apertura, percorrido relativo, percorrido semi-intercuartílico e coeficiente de variación.

Outros contidos deste tema: Desigualdade de Tchebychev. Tipificación de variábeis.

Medidas de forma

Coeficiente de asimetria de Fisher. Coeficiente de aplastamento ou curtose.

Medidas de dispersión ABSOLUTAS

- **PERCORRIDO OU RANGO**

  Diferéncia entre o maior valor e o menor

$$\boxed{R = x_{(n)} - x_{(1)}}$$

- **PERCORRIDO INTERCUARTÍLICO**

  Diferéncia entre o terceiro e o primeiro cuartil

$$\boxed{RI = Q_3 - Q_1}$$

- **DESVIACIÓN ABSOLUTA** respeito a MÉDIA ou MEDIANA

  Suma das diferéncias en valor absoluto entre os valores da variábel e a medida de posición que queremos, multiplicado cada sumando pola frecuéncia relativa. Canto menor sexa, mellor. Quere dicer que a medida de posición é representativa.

$$D_{\bar{x}} = \sum_{i=1}^{m} |xi - \bar{x}| \cdot f_i$$

$$D_{Me} = \sum_{i=1}^{m} |xi - Me| \cdot f_i$$

- **VARIANZA**

  Suma das diferéncias ao cuadrado entre os valores e a média, cada sumando multiplicado pola frecuéncia relativa respectiva. É o momento respeito á média de orde 2.

$$\boxed{Var = \sum_{i=1}^{m} (x_i - \bar{x})^2 \cdot f_i}$$

$=m_2$

$Var = m_2 = a_2 - a_1^2$ desviación típica $= \sigma_n = S = +\sqrt{var}$

Propriedades

a) Var(a+x)=Var(x)

b) Var(ax)=a^2Var(x)

é dicer $Var(ax+b)=a^2 \cdot Var(x)$

demostración a)

x+a quere dicer que á variábel x súma-se-lle unha constante a; ousexa que cada observación é x_i+a

$$Var(a+x) = \sum_{i=1}^{m} (a+x_i - \overline{x+a})^2 \cdot f_i = \sum_{i=1}^{m} (\cancel{a}+x_i - \bar{x}+\cancel{a})^2 \cdot f_i = Var(x)$$

$\uparrow$ $E(ax+b)=a \cdot E(x)+b$

demostración b)

ax quere dicer que a variábel x multiplíca-se por unha constante a; ousexa que cada observación é $a \cdot x_i$

$$Var(ax) = \sum_{i=1}^{m} (a \cdot x_i - \overline{a \cdot x})^2 \cdot f_i = \sum_{i=1}^{m} (\underline{a} \cdot x_i - \bar{x} \cdot \underline{a})^2 \cdot f_i = a^2 \cdot Var(x)$$

$\uparrow$

$E(\underline{ax}+b)=\underline{a\cdot E(x)}+b$

- **DESVIACIÓN TÍPICA**

É a raiz cuadrada positiva da varianza. É a medida de disperasión mais utilizada.

$\boxed{\sigma_n = +\sqrt{var}}$

Propriedades

a) $\sigma(a+x)=\sigma(x)$ porque $Var(a+x)=Var(x)$ e $\sigma=\sqrt{var}$

b) $\sigma(ax)=|a|\sigma(x)$ porque $Var(ax)=a^2Var(x)$ e $\sigma=\sqrt{var}$

Exercício: Sexan N_1 e N_2 os tamaños das povoacións 1 e 2. Calcular a varianza da povoación total en función das subpovoacións.

Varianza = Média das varianzas + Varianza das médias

$$\sigma_x^2= \sum_{i=1}^{2} \sigma_i^2\cdot(N_i/n) + \sum_{i=1}^{2} (\bar{x}_i - \bar{x})^2\cdot(N_i/n)$$

$$\sigma_x^2= \sum_{i=1}^{2} (x_i - \bar{x})^2\cdot(n_i/n) \quad (^*)$$

$$n= \sum_{i=1}^{2} n_i$$

$N_1+N_2=n$ $x_1, x_2, ..., x_m$ $N_1=x_1, x_2, ..., x_h$; $N_2= x_{h+1}, ..., x_m$

$$\bar{x} = \frac{N_1 \bar{x}_1 + N_2 \bar{x}_2}{n}$$ substituir isto en (*)

e separar o sumatório (*) en 1...h, h+1...n e

- **MEDA**

Defíne-se como a mediana de $|x_i\text{-mediana}|$

$\boxed{MEDA = Me\ |x_i - Me|}$

A interpretación é que en |Me-MEDA,Me+MEDA| atópan-se o 50% das observacións.

Aplicacións de média e desviación típica:

Desigualdade de Tchebychev

"No intervalo $(\bar{x} \pm k\sigma) = (\bar{x} - k\sigma, \bar{x} + k\sigma)$ con k constante

$100\cdot(1-\frac{1}{k^2})\%$ das observacións"

Tipificación de variábeis

Dada unha variábel calquera x, de média $\overline{x}$ e desviación típica σ podemos definir unha nova variábel z que tén média 0 e varianza 1.

$$\mathbf{z = \frac{x - \overline{x}}{\sigma}} \qquad \mathbf{\overline{z} = 1} \qquad \mathbf{var(z) = 0}$$

demostracións:

$$\overline{z} = E\left(\frac{x - \overline{x}}{\sigma}\right) = \frac{1}{\sigma} \cdot E(x-\overline{x}) = 1/\sigma \cdot [E(x) - E(\overline{x})] = 1/\sigma \cdot [\overline{x} - \overline{x}] = 0$$

$$var(z) = var\left(\frac{x - \overline{x}}{\sigma}\right) = 1/\sigma^2 \cdot var(x-\overline{x}) = 1/\sigma^2 \cdot var(\overline{x}) = \sigma^2/\sigma^2 = 1$$

Exercício: "Nun exame realízan-se duas probas A e B, obténdo-se no curso os seguintes resultados:

$\overline{x}_A = 14'2$ $\qquad \overline{x}_B = 80'5$

$\sigma_A = 4'4$ $\qquad \sigma_B = 32'4$

O aluno 1 obtivo: $x_{1A} = 14'2$ $\qquad x_{1B} = 74'2$
O aluno 2 obtivo: $x_{2A} = 12'4$ $\qquad x_{2B} = 91'2$

Converter as notas de cada aluno nunha nota global e dicer cal dos alunos supera en puntuación ao outro."

Tipificamos cada nota respeito á proba á que corresponde:

O aluno 1 obtivo: $z_{1A} = (14'2-14'2)/4'4$ $\qquad z_{1B} = (74'2-80'5)/32'4$
O aluno 2 obtivo: $z_{2A} = (12'4-14'2)/4'4$ $\qquad z_{2B} = (91'2-80'5)/32'4$

ousexa

O aluno 1 obtivo: $z_{1A} = 0$ $\qquad z_{1B} = -0'19444$
O aluno 2 obtivo: $z_{2A} = -0'40909$ $\qquad z_{2B} = 0'33025$

Unha vez tipificadas as variábeis póden-se sumar

O aluno 1 obtivo: $z_{1A}+z_{1B} = 0 -0.19444 = -0'19444$
O aluno 2 obtivo: $z_{2A}+z_{2B} = -0.40909 +0.33025 = -0'07884$

O aluno 2 levou mellor nota que o aluno 1.

Isto podémo-lo facer porque supoñemos que as duas probas valen o mesmo! É dicer: en realidade non estamos sumando $z_{1A}+z_{1B}$ senón $0'50 \cdot z_{1A}+0'50 \cdot z_{1B}$ e o mesmo para o aluno 2. Pero se por exemplo a proba A valera o 70% da nota final, teriamos que calcular $0'70 \cdot z_{1A}+0'30 \cdot z_{1B}$ (e o mesmo para o aluno 2) para facer a comparación final.

Medidas de dispersión RELATIVAS

- **COEFICIENTE DE APERTURA**

  É o cociente do maior valor entre o menor valor. Adimensional.

$x_{(n)}/x_{(1)}$

- **PERCORRIDO RELATIVO**

  É o percorrido dividido entre a média aritmética. Adimensional.

$R/\bar{x} = [x_{(n)}-x_{(1)}]/\bar{x}$

- **PERCORRIDO INTERCUARTÍLICO**

  Adimensional.

$$\frac{Q_3-Q_1}{Q_3+Q_1}$$

- **COEFICIENTE DE VARIACIÓN**

  É a mais utilizada. Adimensional. Vé-se afectado polo tamaño dos valores.

$$CV = \frac{\sigma}{\bar{x}}$$

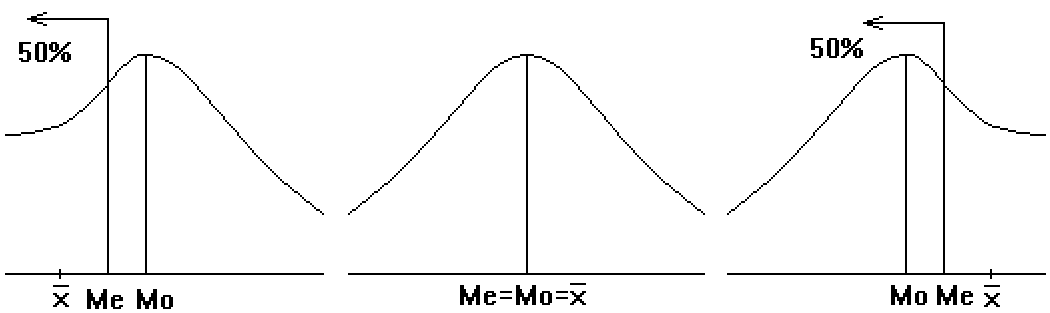
Medidas de FORMA

- **COEFICIENTE DE ASIMETRIA DE FISHER**

Mide se as observacións están dispostas simetrica ou asimetricamente respeito á média e o seu grau de asimetria. É adimensional.

$$CA = \frac{m_3}{S^3} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^3 \cdot f_i}{\sigma^3}$$

Exemplo: unha distribución unimodal



Asimétrica á esquerda	Simétrica	Asimétrica á dereita
CA < 0	CA = 0	CA > 0
Média<Mediana<Moda	Média=Mediana=Moda	Moda<Mediana<Média

- **COEFICIENTE DE APUNTAMENTO OU CURTOSE**

Tráta-se de medir o grau de apuntamento dunha distribución respeito á distribución denominada *normal* que se toma como padrón e cuio coeficiente de curtose é cero. É adimensional.

$$CA = \frac{m_4}{S^4} = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^4 \cdot f_i}{\sigma^4}$$

*tamén se usa esta expresión restándolle 3.

CP > 0 $\Rightarrow$ distribución mais apuntada que a normal LEPTOCÚRTICA
CP = 0 $\Rightarrow$ distribución igual de apuntada que a normal MESOCÚRTICA
CP < 0 $\Rightarrow$ distribución menos apuntada que a normal PLATICÚRTICA