

ESTADÍSTICA I · FORMULARIO

Formulario del Tema 3

Medidas de síntesis: posición, dispersión y forma

NOVA

Cómo usar este formulario

- **Qué entra (plan docente):** 3.1 Medidas de posición · 3.2 Medidas de dispersión · 3.3 Medidas de forma.
- **Notación:** x_i valores o marcas de clase · n_i frecuencias · N_i acumuladas · L_{i-1} límite inferior · a_i amplitud.

3.1 · Media aritmética

DATOS SUELTOS

$$\bar{x} = \frac{\sum x_i}{n}$$

TABLA DE FRECUENCIAS

$$\bar{x} = \frac{\sum x_i n_i}{n}$$

Agrupados: con marcas de clase

MEDIA DE VARIOS GRUPOS

$$\bar{x} = \frac{\bar{x}_A n_A + \bar{x}_B n_B}{n_A + n_B}$$

💡 Sensible a los valores extremos · $\sum (x_i - \bar{x}) n_i = 0$

3.1 · Mediana

SIN AGRUPAR · n IMPAR

$$Me = x_{(n+1)/2}$$

Datos ordenados

SIN AGRUPAR · n PAR

$$Me = \frac{x_{n/2} + x_{n/2+1}}{2}$$

TABLA DE FRECUENCIAS

Primer valor con $N_i \geq \frac{n}{2}$

DATOS AGRUPADOS

$$Me = L_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} a_i$$

Intervalo: el primero con $N_i \geq n/2$

3.1 · Cuartiles, deciles y percentiles

DATOS AGRUPADOS

$$C_k = L_{i-1} + \frac{k n - N_{i-1}}{n_i} a_i$$

Intervalo: el primero con $N_i \geq k n$

QUÉ k USAR

$$Q_1, Q_2, Q_3: k = 0,25 \cdot 0,50 \cdot 0,75$$

$$D_j: k = j/10 \quad P_j: k = j/100$$

💡 «Mínimo para estar en el 25 % con más» = Q_3 · «Máximo del 15 % con menos» = P_{15}

3.1 · Moda

DENSIDAD DEL INTERVALO

$$h_i = d_i = \frac{n_i}{a_i}$$

DATOS AGRUPADOS

$$Mo = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} a_i$$

Intervalo modal: el de mayor h_i

💡 Amplitudes iguales: basta con mirar n_i · Única medida para variables nominales

3.2 · Varianza, desviación y coeficiente de variación

VARIANZA

$$S^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n - 1}$$

Entre $n - 1$, como en R-Commander

PARA CALCULAR MÁS RÁPIDO

$$S^2 = \frac{\sum x_i^2 n_i - n \bar{x}^2}{n - 1}$$

DESVIACIÓN TÍPICA

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

COEFICIENTE DE VARIACIÓN

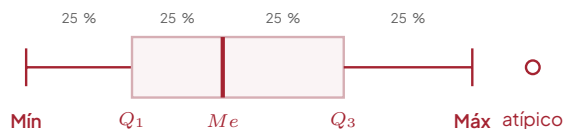
$$CV = \frac{S}{|\bar{x}|}$$

Sin unidades · menor CV = media más representativa

CHEBYSHEV

En $(\bar{x} - k S, \bar{x} + k S)$ hay al menos el $(1 - \frac{1}{k^2}) \cdot 100$ % de los datos $k = 2: 75 \% \cdot k = 3: 89 \%$

3.2 · Rango, rango intercuartílico y box-plot



RANGO

$$R = x_{\text{máx}} - x_{\text{mín}}$$

RANGO INTERCUARTÍLICO

$$RIC = Q_3 - Q_1$$

ATÍPICOS

Fuera de
 $Q_1 - 1,5 RIC$ y
 $Q_3 + 1,5 RIC$

Extremos: con 3 RIC

DATOS EN CADA TRAMO

$$0,25 \cdot n$$

3.2 · Variable tipificada

TIPIFICAR

$$z_i = \frac{x_i - \bar{x}}{S}$$

PROPIEDADES

$$\bar{z} = 0 \quad S_z = 1$$

Sin unidades

PARA QUÉ SIRVE

Comparar la **posición**
 en grupos distintos

💡 «Más es mejor»: gana el z más alto · «Menos es mejor» (tiempos, costes): el más bajo
 · No compara dispersiones: para eso, el CV

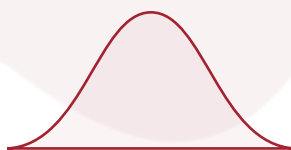
3.3 · Asimetría

$$\bar{x} < Me$$



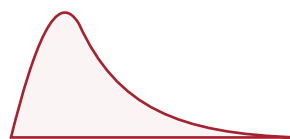
Negativa · cola a la izquierda

$$\bar{x} = Me = Mo$$



Simétrica

$$\bar{x} > Me$$



Positiva · cola a la derecha

PEARSON

$$A_P = \frac{\bar{x} - Mo}{\frac{S}{3(\bar{x} - Me)}} \approx \frac{\bar{x} - Me}{S}$$

FISHER

$$g_1 = \frac{\sum (x_i - \bar{x})^3 n_i / n}{S^3}$$

CÓMO LEER EL SIGNO

> 0 positiva · = 0
 simétrica · < 0 negativa

Rápido: compara $\bar{x}$ con Me

3.3 · Curtosis (apuntamiento)

FISHER

$$g_2 = \frac{\sum (x_i - \bar{x})^4 n_i / n}{S^4} - 3$$

CÓMO LEER EL SIGNO

> 0 leptocúrtica · $= 0$ mesocúrtica · < 0 platicúrtica

Solo si se parece a una normal

Transformaciones lineales: $Y = a + bX$

MEDIDAS DE POSICIÓN

$$\bar{y} = a + b\bar{x}$$

VARIANZA

$$S_Y^2 = b^2 S_X^2$$

No le afecta a

DESVIACIÓN, RANGO, RIC

$$S_Y = |b| S_X$$

No le afecta a

COEFICIENTE DE VARIACIÓN

Cambia si $a \neq 0$

Igual si solo hay escala

TIPIFICADA

z no cambia

Con $b > 0$

💡 $+20\%$ es $b = 1,2$ · -18% es $b = 0,82$