

Tema 1

Alejandro Zubiri

Wed Oct 09 2024

Contents

1	Definiciones Básicas	2
2	Tabla de distribución	2
3	Clases	2
4	Representación gráfica	3
4.1	Función de distribución empírica	3
5	Medidas de centralización	3
5.1	Media aritmética	3
5.2	Media geométrica	3
5.3	Media armónica	3
5.4	Mediana	3
5.5	Moda	3
6	Medidas de dispersión	4
6.1	Desviación media	4
6.2	Varianza	4
6.3	Cuasivarianza	4
6.4	Desviación típica	4
6.5	Cuasidesviación típica	4
6.6	Coefficiente de variación	4
6.7	MEDA	4
6.8	Cuantiles	5
6.9	Rango Intercuartílico	5
7	Medidas de forma	5
7.1	Coefficiente de asimetría Fisher	5

8	Medidas de concentración	5
8.1	Curva de Lorenz	5
8.2	Índice de Gini	6
8.3	Coeficiente de Asimetría Bowley	6
8.4	Curtosis	6
8.5	Diagrama de Caja	6
8.6	Números Índice	7
8.7	Índice de Laspeyres	7
8.8	Índice de Paasche	7

1 Definiciones Básicas

- **Muestra:** datos recopilados
- **Escalas:** intervalo (existe 0 pero no indica ausencia) y razón (0 es ausencia).

2 Tabla de distribución

Se trabaja con la **frecuencia** con la que aparecen los datos:

- Absoluta (f): cuánto aparece.
- Total (n): número de datos.
- Relativa (f_r): $\frac{f}{n}$.
- Acumulada: con datos ordenados, sumar todos los anteriores a x_i :
- Absoluta (F)
- Relativa (F_r)

3 Clases

- Clase modal: clase que tiene más frecuencia por unidad de amplitud (Densidad de frecuencia).
se calcula mediante

$$h_j = \frac{f_j}{b_j - b_{j-1}} \quad (1)$$

Es decir, la frecuencia de la clase entre su amplitud.

- Intervalo / Clase mediana: clase donde se encuentra la mediana.

4 Representación gráfica

4.1 Función de distribución empírica

La función de distribución empírica se define por intervalos. En una gráfica, haremos tantos puntos en el eje X como extremos de intervalos haya. Al final del intervalo, su altura será su frecuencia relativa acumulada. Después, uniremos todos los puntos.

5 Medidas de centralización

5.1 Media aritmética

$$\bar{x} = \frac{\sum x_i}{n} \quad (2)$$

Para datos agrupados en clases, siendo m_j el valor central y $f_r(m_j)$ la frecuencia relativa de esta:

$$\bar{x} = \sum m_j f_r(m_j) \quad (3)$$

Para **transformaciones a la media**, esta es lineal, puesto que para una suma de k , se le suma k a la media, y una transformación por k multiplica por k la media.

5.2 Media geométrica

$$\bar{x}_G = (\prod x_i)^{\frac{1}{n}} \quad (4)$$

5.3 Media armónica

$$\bar{x}_H = \frac{n}{\sum \frac{1}{x_i}} \quad (5)$$

Es fácilmente demostrable que:

$$\bar{x}_H \leq \bar{x}_G \leq \bar{x} \quad (6)$$

5.4 Mediana

La mediana es el valor que está “en medio” en un número de datos impar, o la media aritmética de los dos datos del centro. Cumple que el 50% de los valores es menor a la mediana y el otro 50% es mayor.

5.5 Moda

Es el valor más frecuente.

6 Medidas de dispersión

Miden la separación de los datos entre sí

6.1 Desviación media

$$D_{\bar{x}} = \frac{1}{n} \sum |x_i - \bar{x}| \quad (7)$$

6.2 Varianza

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2 f(x_i)}{n} = \frac{1}{n} \sum x_i^2 f(x_i) - \bar{x}^2 \quad (8)$$

Algunas propiedades de la varianza:

- Es acotado y positivo
- No se ve afectado por cambios de origen
- Se ve afectado por cambios de escala k en un factor k^2 .

6.3 Cuasivarianza

$$\hat{s}_x^2 = \frac{\sum (x_i - \bar{x})^2 f(x_i)}{n - 1} \quad (9)$$

6.4 Desviación típica

$$\bar{s}_x = \frac{\sqrt{\sum (x_i - \bar{x})^2 f(x_i)}}{n} \quad (10)$$

6.5 Cuasidesviación típica

$$\hat{s}_x = \frac{\sqrt{\sum (x_i - \bar{x})^2 f(x_i)}}{n - 1} \quad (11)$$

6.6 Coeficiente de variación

Mide la dispersión relativamente a los datos.

$$CV = \frac{s_x}{|\bar{x}|} / \bar{x} \neq 0 \quad (12)$$

6.7 MEDA

”Mediana de las desviaciones absolutas respecto a la media”

$$MEDA = \text{median}(|x_i - Med_x|) \quad (13)$$

6.8 Cuantiles

Dividen la distribución en c partes. Los más comunes son:

- Cuartiles (Q): cuatro partes
- Quintiles (K): cinco partes
- Percentiles (p): cien partes

La posición de un cuantil viene dada por:

$$C_i = \frac{i \cdot n}{c} \quad (14)$$

El cuantil i , con n datos y c divisiones.

Para encontrar el valor del cuantil, que definiremos como C_i^v , tenemos que encontrar el valor cuya frecuencia absoluta acumulada sea mayor o igual al valor de la posición cuantil:

$$C_i^v \rightarrow F_i \geq C_i \quad (15)$$

6.9 Rango Intercuartílico

$$RI = Q_3 - Q_1 \quad (16)$$

7 Medidas de forma

7.1 Coeficiente de asimetría Fisher

Caracteriza la deformación en el eje X:

$$CA_f = \gamma_1 = \frac{\sum (x_i - \bar{x})^3 f(x_i)}{ns^3} \quad (17)$$

- Si $\gamma_1 = 0 \Rightarrow$ distribución simétrica. La media es igual a la mediana.
- Si $\gamma_1 > 0 \Rightarrow$ distribución asimétrica a derechas. La media es mayor a la mediana.
- Si $\gamma_1 < 0 \Rightarrow$ distribución asimétrica a izquierdas. La media es menor a la mediana.

8 Medidas de concentración

8.1 Curva de Lorenz

Definimos los montos acumulados S_i como

$$S_i = x_i f_i \quad (18)$$

Y S_n como

$$S_n = \sum x_i f_i \quad (19)$$

Definimos q_i

$$q_i = \frac{S_i}{S_n} \cdot 100 \quad (20)$$

Y p_i

$$p_i = \frac{F_i}{n} \cdot 100 \quad (21)$$

Ahora, las coordenadas de la curva serán

$$\frac{p_i}{q_i} \quad (22)$$

8.2 Índice de Gini

Oscila entre 0 y 1, el 0 indicando uniformidad, y el 1 máxima desigualdad. Es la medida más apropiada para la concentración.

$$I_G = \frac{\sum^{n-1} (p_i - q_i)}{\sum^{n-1} p_i} = 1 - \frac{\sum^{n-1} q_i}{\sum^{n-1} p_i} \quad (23)$$

8.3 Coeficiente de Asimetría Bowley

$$CA_B = \frac{Q_1 + Q_3 - 2Q_2}{RI} \quad (24)$$

8.4 Curtosis

Agrupamiento respecto a la media

$$CA_P = \frac{\sum (x_i - \bar{x})^4}{ns^4} - 3 \quad (25)$$

8.5 Diagrama de Caja

Necesitamos los mínimos, máximos y cuartiles.

Definimos el límite inferior como

$$L_I = Q_1^v - 1.5RI \quad (26)$$

Y el límite superior como

$$L_S = Q_3^v + 1.5RI \quad (27)$$

Todos los datos por debajo del inferior o por encima del superior son atípicos.

8.6 Números Índice

Cambio relativo de una variable o variables respecto al tiempo. Se relaciona el valor actual con el valor en un período base.

$$I = \frac{x_t}{x_0} \quad (28)$$

Donde x_t es el valor ahora y x_0 el valor en el período base.

8.7 Índice de Laspeyres

Sea

- q_{i0} la cantidad comprada en el período origen.
- p_{i0} el precio de producto i en tiempo origen.
- q_{it} la cantidad de i comprada ahora.
- p_{it} el precio de i ahora.

$$IPL_t = \frac{\sum p_{it}q_{i0}}{\sum p_{i0}q_{i0}} \quad (29)$$

8.8 Índice de Paasche

Con la mismas variables

$$IPP_t = \frac{\sum p_{it}q_{it}}{\sum p_{i0}q_{it}} \quad (30)$$