ANÁLISIS DE LOS DATOS

MEDIDAS DE SINTESIS PARA VARIABLES UNIDIMENSIONALES

Son números que, al igual que las tablas o las gráficas vistas anteriormente, pretenden resumir la información recogida a la vez que ponen de manifiesto las principales características de una distribución.

Aunque la clasificación no es definitiva ni excluyente veremos:

- Medidas de posición
- Medidas de dispersión
- Momentos
- Medidas de forma
  - Simetría / Asimetría
  - Curtosis
- Medidas de concentración
  - Indice Gini
  - Lorentz

* MEDIDAS DE POSICIÓN
- CENTRADAS
  - Media aritmética
  - Mediana
  - * "Moda"
- NO CENTRADAS
  - Cuartiles
  - Deciles
  - Percentiles

* MEDIDAS DE DISPERSIÓN
- ABSOLUTAS
  - Recorrido
  - Varianza
  - Desviación
  - IQR
- RELATIVAS
  - Coef. de apertura
  - Coef. de variación

MEDIDAS DE POSICIÓN Y CENTRALIZACIÓN

Tienen por objeto dar una idea del valor o valores de la variable, alrededor de los cuales se agrupa una cantidad de datos.

- **MEDIDAS DE POSICIÓN CENTRADAS**

Determinan el valor de X que ocupa la posición central, es decir, buscan el "centro" alrededor del cual se agrupan todos los datos. Intentan representar los valores de una muestra o población indicando dónde se localizan pero no cómo se localizan. Las más importantes son :

- MEDIA ARITMÉTICA → V. cuantitativas { Discretas / Continuas

- Se calcula sumando todos los valores observados, y dividiéndolos por el tamaño muestral.

$$\boxed{\bar{X} = \frac{\sum x_i \cdot n_i}{n} = \sum x_i \cdot f_i}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} Xi}{n} \quad ó \quad \bar{X} = \frac{\sum_{i=1}^{n} Xi \cdot n_i}{n}$$

- Es la suma ponderada de todas las modalidades de la variable por sus respectuvas frecuencias relativas.

$$\bar{x} = \sum_{i=1}^{k} x_i \cdot f_i = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$$

x_i	n_i
0	3
1	2
2	5
3	0

$\sum n_i = 10$

$$\bar{X} = \frac{0 \cdot 3 + 1 \cdot 2 + 2 \cdot 5 + 3 \cdot 0}{} = \frac{11}{10} = 1'1$$

- Si la tabla de frecuencias está agrupada en intervalos, se calcula con las marcas de clase.

$$\bar{x} = \frac{\sum_{i=1}^{m} n_i c_i}{n}$$

Propiedades:

- Si a todos los valores observados les sumamos una constante K (cambio de origen), la media de los nuevos valores se obtiene sumando a la media de los valores originales esta constante K.

$$X_A = \{x_1, x_2, \ldots, x_n\} \Rightarrow \overline{X}_A$$

$$X_B = \{x_1 + k, x_2 + k, \ldots, x_n + k\} \Rightarrow \overline{X}_B = \overline{X}_A + k$$

- Si todas las observaciones se multiplican por una constante K (cambio de escala), la nueva media es el producto de la anterior por la constante K.

- La media de una constante es la misma constante.

- Es centro de gravedad de la distribución $\sum_{i=1}^{n}\left(Xi - \overline{X}\right) = 0.$

- La media es la cantidad que hace mínima la suma de los cuadrados de las desviaciones respecto a un valor.

- $\sum_{i=1}^{n}\left(Xi - \overline{X}\right)^2$: esta diferencia es mínima cuando la constante es la media aritmética.

Ventajas e inconvenientes de la media aritmética sobre otras medidas de síntesis

- **Ventajas**: siempre es calculable, es única (para una distribución dada, siempre toma el mismo valor), y tiene en cuenta a todos los valores de la distribución.

- **Inconvenientes**: al tener en cuenta todos los valores, es sensible a valores extremos. Estos valores pueden distorsionar el sentido de la media aritmética, es poco robusta a errores.

- MEDIANA

- cuantitativas { Discretas / continuas
- cualitativos ordinales

Es el valor que quedaría en medio de la distribución si ordenásemos de menor a mayor todas las observaciones. La mediana deja un 50% de las observaciones a cada lado.

Xmin — 50% — Me — 50% — Xmax

• **Si n es par** la mediana se calcula como la media aritmética de los dos valores centrales. La media aritmética de los valores n/2 y (n/2)+1.

$$\boxed{Me = \frac{x\left(\frac{n}{2}\right) + x\left(\frac{n}{2}+1\right)}{2}}$$

$Me = \frac{7+9}{2} = 8$

* Ejemplo: $X = \{1,2,5,7,9,10,13,14\}$

n=8 (par) ⇒ n/2=4 (número que ocupa la 4ª posición, en nuestro ej. es el 7).

(n/2) + 1=5 (número que ocupa la 5ª posición, en nuestro ej. es el 9).

Me = (7+9)/2 = 8.

• **Si n es impar** la mediana es la posición (n+1)/2.

$$\boxed{Me = x\left(\frac{n+1}{2}\right)}$$

Me

* Ejemplo: $X=\{1,2,5,7,9,10,13\}$ ⟶ Me = 7

n = 7 (impar) ⇒ (n+1)/2 = 4 (valor que ocupa la 4ª posición, en nuestro ej. es el 7).

Me = 7.

- MODA

$\begin{cases} \text{- cualitativas} \begin{cases} \text{ordinales} \\ \text{nominales} \end{cases} \\ \text{- cuantitativos} \begin{cases} \text{Discretos} \\ \text{continuos} \end{cases} \end{cases}$

La moda es el valor que más veces se repite en la distribución. La moda de una variable estadística es el valor/es que tiene/n asociada la frecuencia máxima. Podría no considerarse una medida de posición ya que no nos da ni información sobre el centro de la distribución.

Puede suceder que en los datos no se repita ningún valor, en este caso podemos decir que no hay moda. Por otra parte, si hay dos valores que se repiten el mismo número de veces, hablamos de datos bimodales (o datos multimodales si se repiten un mayor número de veces).

¡¡ ATENCIÓN !!

Media y mediana coincidirán únicamente en el caso de que la distribución sea simétrica (la moda no tiene por qué coincidir con ellas).
Si la distribución es simétrica y unimodal, media, mediana y moda coincidirán

- RANGO MEDIO

Es la media entre el mayor valor y el menor valor de la distribución.

$$\boxed{RM = \frac{Xn + Xo}{2} = \frac{X_{MAX} + X_{MIN}}{2}}$$

• **MEDIDAS DE POSICIÓN NO CENTRADAS**

CUARTILES, DECILES Y PERCENTILES

Cuartiles
- Se dividen los datos en cuatro partes iguales
- Q1= 25%, Q2=50%, Q3=75%

$Me = Q_2$

Deciles
- Se dividen los datos en 10 partes iguales
- Se calcula desde el D1 al D9

$Me = D_5$

Percentiles
- Se dividen los datos en 100 partes iguales
- Se calcula del P1 al P99

$Me = P_{50}$

- CUARTILES

El cuartil es una medida en estadística que divide un conjunto de datos ordenados en cuatro segmentos iguales. Cada cuartil muestra un valor específico bajo el cual cae un cierto porcentaje de los datos.

El primer cuartil, Q_1, es un número tal que a lo mucho 25% de los datos son menores en el valor que Q_1 y a lo sumo 75% son mayores. El segundo cuartil, Q_2 , es la mediana (50%). El tercer cuartil, Q3, es un número tal que a lo mucho 75% de los datos son menores en valor que Q_3 y a lo sumo 25% son mayores.

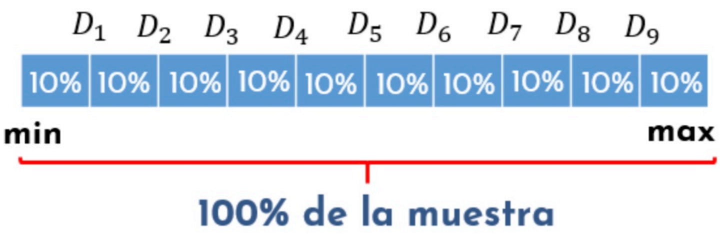
Fórmulas para calcular las posiciones de los cuartiles:

$$Pos\,Q_1 = \frac{(n+1)}{4},\ Pos\,Q_2 = \frac{2(n+1)}{4},\ Pos\,Q_3 = \frac{3(n+1)}{4}$$

Este cuartil equivale al 50% por lo tanto también debe de ser igual a la mediana.

- DECILES

Son valores de la variable que dividen los datos ordenados en diez partes iguales (9 divisiones).



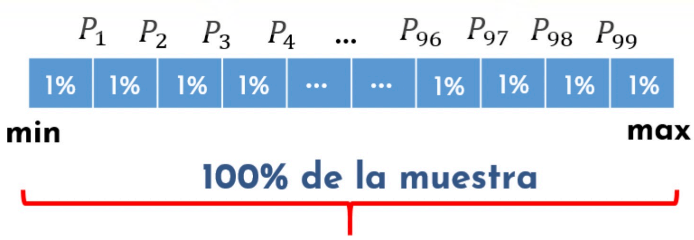
Fórmulas para calcular las posiciones de los deciles:

$$D_1 = \frac{(n+1)}{10},\; D_5 = \frac{5(n+1)}{10},\; D_9 = \frac{9(n+1)}{10}$$

Este cuartil equivale al 50% por lo tanto también debe de ser igual a la mediana.

- PERCENTILES

Son los valores de la variable que dividen un conjunto de datos en 100 subconjuntos iguales; cada conjunto de datos tiene 99 percentiles. El k-ésimo percentil, P_k, es un valor que a lo mucho k% de los datos son menores en valor que P_k y a lo sumo (100-k)% de los datos son mayores.



Fórmulas para calcular las posiciones de los percentiles:

$$P_1 = \frac{(n+1)}{100},\; P_{50} = \frac{50(n+1)}{100},\; P_{99} = \frac{99(n+1)}{100}$$

Este cuartil equivale al 50% por lo tanto también debe de ser igual a la mediana.

Ejemplo:

- De 20 estudiantes tenemos sus evaluaciones de una examen calcular el Q_1, D_5, P_{75}.

- 5,5,8,7,9,10,7,6,8,7,8,9,10,10,8,7,6,5,9,6.

$Q_1 \longrightarrow Pos\,Q_1 = \frac{n+1}{4}$

$D_5 \longrightarrow D_5 = Me$

$P_{75} = Q_3 \longrightarrow Pos\,Q_3 = \frac{3(n+1)}{4}$

$Q_1 = 6$ $\quad Me = D_5 = 7'5$ $\quad Q_3 = 9$

5, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 9, 9, 9, 10, 10, 10

$n = 20 \qquad Me = \frac{X_{(n/2)} + X_{(n/2+1)}}{2} = \frac{X_{10} + X_{11}}{2} = \frac{7+8}{2} = 7'5$

$Q_1 \qquad Pos\,Q_1 = \frac{20+1}{4} = 5'25 \longrightarrow Q_1 = \frac{X_5 + X_6}{2} = \frac{6+6}{2} = 6$

$Q_3 \qquad Pos\,Q_3 = 3\left(\frac{20+1}{4}\right) = 15'75 \longrightarrow Q_3 = \frac{X_{15} + X_{16}}{2} = 9$

x_i	n_i	N_i
5	3	3
6	3	6
7	4	10
8	4	14
9	3	17
10	3	20

$\Sigma n_i = 20$