

Estadística Descriptiva e Inferencial

Módulo 2. Medidas estadísticas



Medidas estadísticas

1. Tendencia central
2. Variabilidad
3. Posición



1. Medidas estadísticas de tendencia central



Medidas estadísticas de tendencia central

Son valores numéricos que tienden a ubicarse en la parte central de un conjunto de datos (o en su distribución de frecuencias).

❖ Media

❖ Mediana

❖ Moda



Media aritmética (\bar{x})

Se define como la suma de n observaciones ($x_1, x_1 \dots x_n$) dividido entre n .

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$



Media aritmética - Características

- Todas las variables de nivel de intervalo y de razón tienen una media.
- La media es única para un conjunto de datos.
- La media es sensible a valores extremos.



Mediana (m)

Se define como el valor central de los datos después de que se ordenan (de menor a mayor o viceversa).

- Si el número de datos n es un número no o impar, la mediana será el valor central de los datos.
- Si el número de datos n es un número par, hay dos valores centrales y la mediana se calcula obteniendo la media aritmética de estos valores.



Mediana - Características

- Se puede obtener para variables de nivel intervalo, de razón y en algunos casos para ordinales.
- La mediana es única para un conjunto de datos.
- Para un conjunto de datos con distribución de frecuencias asimétrica o sesgada, la mediana es mejor medida de centralización que la media aritmética, ya que no se ve afectada por valores extremos.



Moda (M_o)

Se define como el dato que ocurre con mayor frecuencia.

La moda es una medida poco usada por las siguientes razones:

- a. Puede ocurrir que no exista.
- b. A menudo no es un valor único.



Moda - Características

- Se puede obtener para toda variable (nominal, ordinal, intervalo y razón o proporción).
- No se ve afectada por valores extremos.
- En un conjunto de datos que han sido agrupados en un histograma, la clase con la mayor frecuencia se llama **clase modal**, y el valor medio (v_i) de esa clase se toma como la moda.



2. Medidas estadísticas de variabilidad o dispersión



Medidas estadísticas de variabilidad o dispersión

Son valores numéricos que nos indican la variabilidad que existe en un conjunto de datos. Nos indican en qué medida los datos se asemejan entre sí.

❖ Rango

❖ Varianza

❖ Desviación estándar



Rango (R)

Se define como la diferencia entre el dato mayor o máximo (max) y el dato menor o mínimo (min).

$$R = Max - Min$$



Mediana - Características

- Se puede obtener para variables de nivel de intervalo y de razón



Varianza (S^2)

Desviación estándar (S)

La varianza se define como la suma del cuadrado de las diferencias entre las observaciones y su media aritmética, dividida entre $(n - 1)$.

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

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



Varianza

Características

Desviación estándar

- Se puede calcular para variables con escala de medición de intervalo y de razón.
- Es no negativa y es cero sólo si todas las observaciones son iguales.
- Las unidades de la varianza son las unidades originales de la variable al cuadrado.
- La desviación estándar tiene las mismas unidades que la variable en estudio.



3. Medidas estadísticas de posición



Medidas estadísticas de posición

Son valores que dividen a un conjunto de datos en partes iguales. Nos ayudan a describir la localización de un dato respecto al resto de los mismos.

❖ Cuartiles

❖ Deciles

❖ Percentiles



Cuartiles (Q_k)

Son números que dividen a los datos ordenados en 4 partes. Cada conjunto de datos tiene 3 cuartiles:

- **El primer cuartil (Q_1)**, es un número tal que el 25% de los datos (una cuarta parte) es menor o igual al Q_1 y el 75% de ellos es mayor o igual al Q_1 .
- **El segundo cuartil (Q_2)**, es la mediana, es un número tal que el 50% de los datos (dos cuartas partes) es menor o iguales al Q_2 y el 50% de ellos es mayor o igual al Q_2 .
- **El tercer cuartil (Q_3)**, es un número tal que el 75% de los datos (tres cuartas partes) es menor o igual al Q_3 , y el 25% de ellos es mayor o igual al Q_3 .



Deciles (D_k)

Son números que dividen a los datos ordenados en 10 partes. Cada conjunto de datos tiene 9 deciles:

- **El primer decil (D_1)**, es un número tal que el 10% de los datos (una décima parte) es menor o igual al D_1 y el 90% de ellos es mayor o igual al D_1 .
- **El segundo decil (D_2)**, es un número tal que el 20% de los datos (dos décimas partes) es menor o igual al D_2 y el 80% de ellos es mayor o igual al D_2 .
- . . .
- **El noveno decil (D_9)**, es un número tal que el 90% de los datos (nueve décimas partes) es menor o igual al D_9 y el 10% de ellos es mayor o igual al D_9 .



Percentiles (P_k)

Son números que dividen a los datos ordenados en 100 partes. Cada conjunto de datos tiene 99 percentiles:

- **El P_1** , es un número tal que el 1% de los datos es menor o igual al P_1 y el 99% de ellos es mayor o igual al P_1 .
- **El P_2** , es un número tal que el 2% de los datos es menor o igual al P_2 y el 98% de ellos es mayor o igual al P_2 .
- . . .
- **El P_{99}** , es un número tal que el 99% de los datos es menor o igual al P_{99} y el 1% de ellos es mayor o igual al P_9 .



Pasos para calcular las medidas de posición

1. Ordenar los datos en orden creciente.
2. Encontrar la posición de la medida buscada, sustituyendo en la fórmula según corresponda:

$$Posición = \left(\frac{k}{i} \right) * n + \frac{1}{2}$$

n = número total de datos

k = cuartil (1, 2, 3), decil (1, 2, ..., 9) o percentil (1, 2, ..., 99) en cuestión

i = 4, 10 o 100 (cuartil, decil, percentil; de acuerdo a la medida de posición buscada)



Pasos para calcular las medidas de posición

- 3.1 Si el valor obtenido en la posición es un entero, la medida de posición buscada será el dato que se encuentre en esa posición.
- 3.2 Si el valor obtenido en la posición es un número no entero (ejemplo, 7.6), la medida de posición buscada estará entre el dato de la posición entera inmediata inferior y el dato de la posición entera inmediata superior (7º y 8º datos, en el ejemplo). Finalmente, la medida de posición se obtiene de la siguiente forma:

$$\text{Medida de posición} = DPm + (DPM - DPm) * f$$

DPm: Dato que se encuentra en la posición entera inmediata inferior

DPM: Dato que se encuentra en la posición entera inmediata superior

f: Parte decimal (fraccionaria) obtenida al calcular la posición



Relación entre las medidas de posición



