

# Varianza



# ¿Qué es?

La varianza es una medida de cómo se dispersan o se distribuyen los valores de un conjunto de datos en relación con su media (promedio). En otras palabras, la varianza nos dice qué tan lejos están los datos del valor promedio.



# Conceptos clave

**Media (Promedio):** La media es el valor promedio de un conjunto de datos. Se calcula sumando todos los valores y dividiéndolos por la cantidad de valores.

$$\text{Media}(\mu) = \frac{\sum_{i=1}^n x_i}{n}$$

donde  $x_i$  son los valores del conjunto de datos y  $n$  es el número total de valores.

**Desviación de cada valor respecto a la media:** Cada valor del conjunto de datos tiene una desviación con respecto a la media. Esta desviación se calcula restando la media de cada valor.

$$\text{Desviación} = x_i - \mu$$

**Desviación al cuadrado:** Para evitar que las desviaciones positivas y negativas se cancelen entre sí, se eleva al cuadrado cada desviación.

$$\text{Desviación al cuadrado} = (x_i - \mu)^2$$

# Varianza

La varianza es el promedio de estas desviaciones al cuadrado. Se calcula sumando todas las desviaciones al cuadrado y dividiéndolas por el número total de valores.

$$\text{Varianza}(\sigma^2) = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

# Ejemplo

Supongamos que tienes los siguientes datos: [2, 4, 4, 4, 5, 5, 7, 9]

Calcula la media:

$$\mu = \frac{2 + 4 + 4 + 4 + 5 + 5 + 7 + 9}{8} = \frac{40}{8} = 5$$

Calcula la desviación de cada valor respecto a la media:

$$\begin{aligned}(2 - 5)^2 &= 9 \\(4 - 5)^2 &= 1 \\(4 - 5)^2 &= 1 \\(4 - 5)^2 &= 1 \\(5 - 5)^2 &= 0 \\(5 - 5)^2 &= 0 \\(7 - 5)^2 &= 4 \\(9 - 5)^2 &= 16\end{aligned}$$

Calcula la varianza:

$$\sigma^2 = \frac{9 + 1 + 1 + 1 + 0 + 0 + 4 + 16}{8} = \frac{32}{8} = 4$$

**Interpretación:**

- Una varianza baja indica que los valores del conjunto de datos están muy cerca de la media.
- Una varianza alta indica que los valores están muy dispersos y lejos de la media.

La varianza es una herramienta fundamental en estadística y análisis de datos para entender la variabilidad y la dispersión de los datos.