MEDIDAS DE DISPERSIÓN

Las **medidas de dispersión** son fundamentales para entender la variabilidad de un conjunto de datos, ya que nos indican qué tan dispersos o concentrados están los valores respecto a la **media** o cualquier otra medida de tendencia central. Dos de las medidas de dispersión más comunes son la **varianza** y la **desviación estándar**.

MEDIDAS DE DISPERSIÓN PARA DATOS NO AGRUPADOS

Cuando tenemos **datos no agrupados**, es decir, cuando cada valor individual está disponible y no está organizado en intervalos, las fórmulas para calcular la **varianza** y la **desviación estándar** se realiza de manera directa.

Varianza para Datos No Agrupados (σ^2)

La **varianza** mide el promedio de las distancias al cuadrado de cada dato con respecto a la media del conjunto de datos. Es decir, la varianza mide qué tan dispersos están los datos alrededor de la media. La varianza es igual a la desviación estándar elevada al cuadrado.

La fórmula para calcular la varianza de n datos es:

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

Donde:
Xi es cada valor individual de los datos.
μ es la **media** de los datos.
n es el número total de observaciones.

Si los datos representan una **muestra** en lugar de toda la población, la fórmula se ajusta dividiendo entre n−1 en lugar de n, para corregir el sesgo en la estimación de la varianza de la población (s^2):

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Donde:

$\bar{X}$ es la **media muestral**.

Desviación Estándar para Datos No Agrupados

La **desviación estándar** es simplemente la raíz cuadrada de la varianza. Al tomar la raíz cuadrada, convertimos las unidades de la varianza (que están al cuadrado) a las mismas unidades de los datos originales, lo que la hace más interpretable.

La fórmula para la **desviación estándar** de una población es:

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

Mientras, que la fórmula para la desviación estándar de una muestra es:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Nota: Toma en cuenta que, si tienes una serie de valores de una muestra y necesitas calcular su varianza y su desviación estándar, deberás calcular primero la media poblacional $\bar{x}$ con la siguiente fórmula:

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

Ejemplo:

Del siguiente conjunto de datos: 2, 4, 6 y 8 determina la varianza y la desviación estándar:

Paso 1: **Calcular la media**:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + x_4}{N} = \frac{2+4+6+8}{4} = \frac{20}{4} = 5$$

Paso 2: **Calcular la varianza**:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + (x_4 - \mu)^2}{N}$$

$$\sigma^2 = \frac{(2-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2}{4} = \frac{(-3)^2 + (-1)^2 + (1)^2 + (3)^2}{4}$$

$$\sigma^2 = \frac{9+1+1+9}{4} = \frac{20}{4} = 5$$

Paso 3: **Calcular la desviación estándar**:

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

$$\sigma = \sqrt{5} = 2{,}236$$

MEDIDAS DE DISPERSIÓN PARA DATOS AGRUPADOS

Cuando los datos están agrupados en clases o intervalos, se utilizan las frecuencias de las clases y los puntos medios de las mismas para calcular las medidas de dispersión.

Varianza para datos agrupados.

La fórmula de la varianza en datos agrupados se calcula de manera similar a la de los datos no agrupados, pero utilizando los puntos medios de las clases (x_i) y las frecuencias de las clases (f_i):

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

Donde:

x_i es el punto medio de la i-ésima clase.

f_i es la frecuencia de la i-ésima clase.

n es el número total de observaciones ($n=\sum f_i$).

*μ es la **media** del conjunto de datos.*

Si se trata de una **muestra**, la fórmula se ajusta dividiendo por n−1 en lugar de n.

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

Desviación estándar para datos agrupados.

La **desviación estándar** para datos agrupados es simplemente la raíz cuadrada de la **varianza**.

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

Ejemplo: calcula **la Varianza y Desviación Estándar** para los siguientes datos agrupados:

Intervalo de clase	Frecuencia (fi)
0-10	5
10-20	8
20-30	12
30-40	10
40-50	5

Paso 1: Calcular los puntos medios.

Intervalo de clase	Frecuencia (fi)	Punto medio (x_i)
0-10	5	5
10-20	8	15
20-30	12	25
30-40	10	35
40-50	5	45

Paso 2: Calcular la media.

La media para los datos agrupados se calcula como:

$$\mu = \frac{\sum_{i=1}^{k} f_i * x_i}{n}$$

Donde n = 5+8+12+10+5=40

$$\mu = \frac{5(5) + 8(15) + 12(25) + 10(35) + 5(45)}{40} = 25.5$$

Paso 3: Calcular la varianza.

$$\sigma^2 = \frac{1}{40}[5(420.5) + 8(110.25) + 12(0.25) + 10(90.25) + 5(380.25)] = 144.725$$

Paso 4: Calcular la desviación estándar.

$$\sigma = \sqrt{\sigma^2} = \sqrt{144.725} = 12.02$$

Por lo tanto, la varianza es aproximadamente **144.73** y la desviación estándar es aproximadamente **12.02.**

Las medidas de dispersión como la varianza y la desviación estándar son esenciales para comprender la distribución de los datos, ya que nos indican cuán dispersos o concentrados están los datos alrededor de la media. Aunque los cálculos para datos agrupados requieren el uso de los puntos medios y las frecuencias, los principios básicos para calcular la varianza y la desviación estándar son los mismos. Ambas medidas nos permiten tomar decisiones más informadas sobre la variabilidad y la estabilidad de los datos en estudios estadísticos, investigaciones y análisis de datos.

Referencias:

Walpone, Ronald. (2012) Probabilidad y estadística para ingenierías y ciencias. México. Pearson.

Redacción. (2022) Varianza y desviación estándar, ejemplos y ejercicios. Mate móvil. Recuperado de: https://matemovil.com/varianza-y-desviacion-estandar-ejemplos-y-ejercicios/

Estrada Rivera, Martha Rocío; Hernández, Xóchitl Itzel. (2015) Probabilidad y estadística I. Telebachillerato. SEP. Recuperado de: https://libros.conaliteg.gob.mx/2021/TB5EA.htm#page/1

Redacción. (2023) Entendiendo la Varianza: Claves para Interpretar la Dispersión de Datos. El mundo de los datos. Recuperado de: https://elmundodelosdatos.com/entendiendo-la-varianza-claves-para-interpretar-la-dispersion-de-datos/