CURSO: ESTADÍSTICAS I

Unidad I: Estadística descriptiva.

Clase 2: Estadísticos de Dispersión.

Profesor: Diego Miranda

Data Scientist

INTRODUCCIÓN

Imagina que estás a punto de tomar una decisión importante basada en datos, ya sea en el ámbito financiero, científico o incluso en tu vida cotidiana. ¿Te has preguntado alguna vez qué tan confiables son los datos que estás utilizando? ¿Cómo se extienden y distribuyen esos valores en torno a un punto central? Aquí es donde entran en juego las medidas de dispersión. Estas medidas no solo cuantifican la amplitud y la variabilidad de los datos, sino que también arrojan luz sobre patrones, tendencias y posibles incertidumbres en la información que manejamos.

Imagina que eres un gerente financiero y tienes que evaluar dos inversiones potenciales para tu empresa. Ambas inversiones pueden tener rendimientos promedio similares, pero ¿cómo se distribuyen los retornos individuales alrededor de esos promedios? Una medida de dispersión te ayudaría a entender qué tan riesgosa es cada inversión. O considera un estudio médico en el que se miden los niveles de glucosa en pacientes. Saber cuánto varían estos niveles podría ser crucial para determinar la eficacia de un tratamiento.

MEDIDAS DE DISPERSIÓN

Las medidas de dispersión, también conocidas como medidas de variabilidad o dispersión estadística, son estadísticas utilizadas para cuantificar la extensión en la que los datos de un conjunto se distribuyen alrededor de un valor central, como la media o la mediana. En otras palabras, las medidas de dispersión indican cuánto se separan los datos individuales del valor central, proporcionando información sobre la amplitud o la variabilidad de los datos en un conjunto.

VARIANZA DE LA POBLACIÓN

Cada población tiene una varianza, su símbolo es σ^2 (sigma cuadrada). Para calcular la varianza de una población, la suma de los cuadrados de las distancias entre la media y cada elemento de la población se divide entre el número total de observaciones en población. Al elevar al cuadrado cada distancia, logramos que todos los números sean positivos y, al mismo tiempo, asignamos más peso a las desviaciones más grandes (desviación es la distancia entre la media y un valor).

VARIANZA DE LA POBLACIÓN PARA DATOS AGRUPADOS

Varianza de población	
$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \frac{\Sigma x^2}{N} - \mu^2$$	[3-12]

DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN

La desviación estándar de la población, , es simplemente la raíz cuadrada de la varianza de la población. Como la varianza es el promedio de los cuadrados de las distancias de las observaciones a la media, la desviación estándar es la raíz cuadrada del promedio de los cuadrados de las distancias entre las observaciones y la media. Mientras que la varianza se expresa con el cuadrado de las unidades utilizadas para medir los datos, la desviación estándar está en las mismas unidades que las que se usaron para medir los datos.

La fórmula para la desviación estándar es:

Desviación estándar de la población
$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} = \sqrt{\frac{\Sigma x^2}{N} - \mu^2}$ [3-13]

COEFICIENTE DE VARIACIÓN

El coeficiente de variación es una de estas medidas relativas de dispersión. Relaciona la desviación estándar y la media, expresando la desviación estándar como porcentaje de la media. La unidad de medida, entonces, es “porcentaje”, en lugar de las unidades de los datos originales. Para una población, la fórmula para el coeficiente de variación es:

Coeficiente de variación

Desviación estándar de la población

$$\text{Coeficiente de variación de la población} = \frac{\sigma}{\mu}(100) \qquad [3\text{-}20]$$

Media de la población