Utel UNIVERSIDAD

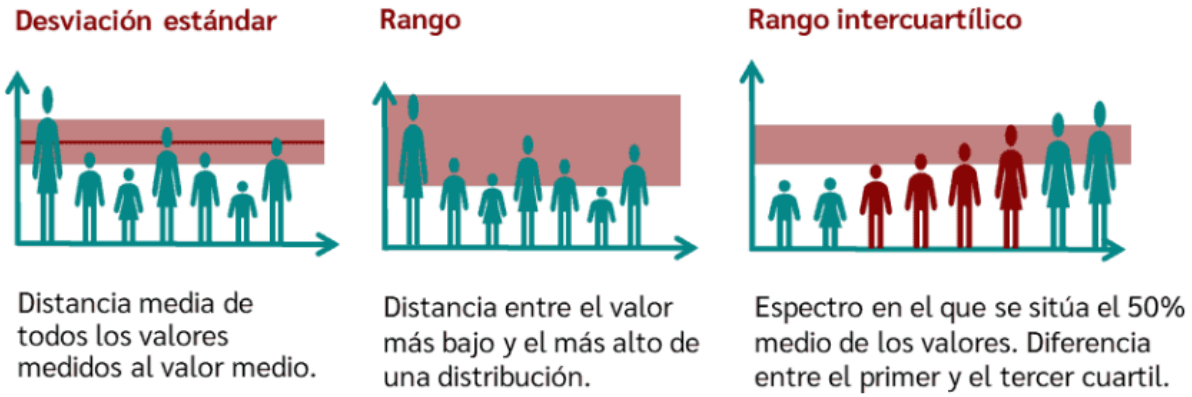
Material de Consulta

Estadística para las ciencias sociales

Semana 4

Medidas de dispersión

En estadística descriptiva, los parámetros de dispersión son medidas que nos permiten comprender la variabilidad de los valores de una muestra con relación a un parámetro de ubicación. Estas medidas nos brindan información sobre cuánto se dispersan los datos alrededor de un valor medio, complementando así la información proporcionada por las medidas de tendencia central.



¿Qué es el rango?

El rango es una medida de dispersión que nos indica la distancia entre el valor máximo y el valor mínimo de una distribución de datos. Es decir, el rango nos muestra la amplitud total de los valores en una muestra. Para calcular el rango, simplemente restamos el valor mínimo al valor máximo.

R= Máx - Mín

Por ejemplo, si tenemos una muestra de alturas donde el valor máximo es 1.90 metros y el valor mínimo es 1.50 metros, el rango sería de 0.40 metros.

El rango es una medida sencilla y fácil de entender, ya que nos da una idea de qué tan dispersos están los valores extremos con relación al valor medio. Sin embargo, el rango no nos proporciona información sobre cómo se distribuyen los valores dentro del rango.

Varianza y desviación estándar

La varianza y la desviación estándar son dos medidas de dispersión comunes, utilizadas para variables métricas. Estas medidas nos dan información sobre cuánto se dispersan los valores individuales con relación al valor medio.

¿Qué es la varianza?

La varianza es una medida que describe la distancia promedio al cuadrado de cada valor con respecto al valor medio. Para calcular la varianza, sumamos las desviaciones al cuadrado de cada valor con respecto al valor medio y dividimos esta suma por el número de valores. La varianza se expresa en unidades al cuadrado, lo que dificulta su interpretación directa con relación a los datos originales.

¿Qué es la desviación estándar?

La desviación estándar es la raíz cuadrada de la varianza y se utiliza para obtener una medida de dispersión más fácil de interpretar. La desviación estándar nos indica la dispersión promedio de los valores con respecto al valor medio. Al calcular la desviación estándar, obtenemos una medida en las mismas unidades que los datos originales, lo que facilita su interpretación.

La desviación estándar se calcula tomando la raíz cuadrada de la varianza. En la fórmula, utilizamos la letra "s" para representar la desviación estándar de una muestra y la letra griega "σ" para representar la desviación estándar de una población.

Cálculo de la varianza y la desviación estándar

Existen diferentes fórmulas para calcular la varianza y la desviación estándar, dependiendo de si tenemos datos de toda la población o sólo una muestra de la misma. En el caso de tener datos de toda la población, podemos utilizar las siguientes fórmulas para calcular la varianza y la desviación estándar:

Varianza poblacional:

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$$

Donde:

- σ^2 es la varianza poblacional.
- x_i son los valores individuales de la población.
- μ es el valor medio de la población.
- N es el tamaño de la población.

Desviación estándar poblacional:

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

Donde:

- σ es la desviación estándar poblacional.

Por otro lado, si sólo tenemos una muestra, utilizamos otra fórmula ligeramente diferente para calcular la varianza y la desviación estándar:

Varianza de la muestra:

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

Donde:

- s^2 es la varianza de la muestra.
- x_i son los valores individuales de la muestra.
- $\bar{x}$ es el valor medio de la muestra.
- *N* es el tamaño de la muestra.

Desviación estándar de la muestra:

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

Donde:

- s es la desviación estándar de la muestra.

Es importante tener en cuenta la diferencia entre los denominadores de las fórmulas para la varianza, la desviación estándar de la población y la muestra. En el caso de la muestra, dividimos la suma de las desviaciones al cuadrado por (*n*-1) en lugar de (*n*) para tener en cuenta la estimación de la población basada en la muestra.

Interpretación y uso de la varianza y la desviación estándar

Tanto la varianza como la desviación estándar nos proporcionan información sobre la dispersión de los valores con relación al valor medio. Cuanto mayor sea la varianza o la desviación estándar, mayor será la dispersión de los valores en la muestra o población.

Estas medidas de dispersión son útiles para comparar la variabilidad entre diferentes muestras o poblaciones. Por ejemplo, si estamos estudiando el rendimiento académico de diferentes grupos de estudiantes, podemos utilizar la desviación estándar para determinar qué grupo tiene una mayor variabilidad en los resultados.

Coeficiente de variación

Otro de los usos de la desviación estándar es su utilidad para calcular el coeficiente de variación, que es una medida relativa de la dispersión. El coeficiente de variación nos permite comparar la variabilidad de diferentes variables o muestras, teniendo en cuenta sus respectivas escalas.

El coeficiente de variación se representa mediante la expresión CV y se obtiene dividiendo la desviación estándar entre la media, multiplicando este resultado por 100, no importando si se trata de datos no agrupados o de datos agrupados, o que provengan de una muestra o de una población.

El coeficiente de variación se puede calcular mediante la fórmula siguiente:

$CV = \left(\frac{\sigma}{\mu}\right) 100\%$	En caso de una población
$CV = \left(\frac{S}{\bar{X}}\right) 100\%$	En caso de una muestra

Donde:

CV = Coeficiente de variación.
S = Desviación estándar de la muestra.
X = Media de los datos.
σ = Desviación estándar de la población.
μ = Media poblacional.

A diferencia de otras medidas de variabilidad, el coeficiente de variación mide la dispersión en términos de porcentaje y no en unidades de medida. Por ejemplo, si los analistas de un despacho de bienes raíces están interesados en determinar si el valor de un avalúo tiene mayor variabilidad que el tamaño del lote, resultaría imposible comparar en forma directa la dispersión mediante el rango, la varianza o la desviación estándar, pues el valor del avalúo se mide en unidades monetarias; por ejemplo en miles de pesos, mientras que el tamaño del lote se mide en metros cuadrados. En este caso, los analistas pueden utilizar el coeficiente de variación, expresado en porcentajes, y así comparar la dispersión de dos variables expresadas en distintas unidades de medida.

En general, los parámetros de dispersión como rango, varianza y desviación estándar son herramientas fundamentales en la estadística descriptiva. Nos permiten comprender cómo se distribuyen los valores alrededor del valor medio y nos brindan información valiosa sobre la variabilidad de los datos. Al utilizar estas medidas correctamente, podemos obtener una visión más completa y precisa de nuestras muestras o poblaciones de interés.

Referencias

DATAtab (2023). ***Parámetros de dispersión***. https://datatab.es/tutorial/dispersion-parameter

INITE (2011). ***Medidas de tendencia central y dispersión.*** https://gc.scalahed.com/recursos/files/r161r/w24762w/3/Medidasdetendenciacentralydispersion.pdf