

Conceptos Esenciales

- La mediana de un conjunto de datos puede ser computado ordenando los valores de los datos e identificando el valor en “el medio”.
- La media representa el punto de equilibrio del conjunto de datos y la mediana representa el percentil 50, o el valor que divide los datos en mitad.
- Los siguientes pasos se pueden aplicar para calcular una desviación estándar a mano:
 1. Calcula la media de la población o muestra
 2. Tome la diferencia entre cada valor de los datos y la media, luego tome el cuadrado de cada diferencia
 3. Suma todas las diferencias cuadradas
 4. Dividir entre el número total de observaciones en el caso de una población, o entre 1 menos que el total en el caso de una muestra
 5. Toma la raíz cuadrada del resultado del paso 4
- La variabilidad se puede juzgar a partir de un histograma examinando la distancia de las barras desde el centro estadístico (media o mediana) del gráfico. Si la variabilidad es alta, aparecerán barras de igual tamaño o más altas barras aparecieran lejos del centro del gráfico. Si la variabilidad es baja, los datos aparecerán agrupados alrededor del centro.
- Los valores más grandes del rango indican más variabilidad en los datos, pero el valor del rango solo utiliza dos observaciones en todo el conjunto de datos para medir la variabilidad. Esta no es una medida ideal de propagación, pero cuando se usa en combinación con otras medidas de propagación, puede ayudarte a obtener una comprensión más clara de la propagación de una distribución.
- La mediana se mantiene relativamente fija en un conjunto de datos si un valor cambia en una gran cantidad, la media no. Esta es una indicación de que la media es sensible a la presencia de valores extremos en el conjunto de datos.
- Cuando una distribución es simétrica, la media y la mediana ocupan el mismo valor. Bajo un sesgo, la media se “tira” en la dirección de los valores atípicos:
 - Sesgado a la derecha: la media es mayor que la mediana
 - Sesgado a la izquierda: la media es menor que la mediana
- La media, bajo ciertas condiciones, puede ser un indicador engañoso de un valor de observación “típico”.
- Un diagrama de caja captura solo la mediana del conjunto de datos, no la media, como medida del centro. Proporciona un vistazo rápido (o un resumen de cinco números) de

los datos para hacer comparaciones basadas en la mediana, el sesgo, los valores atípicos y los cuartiles.

- Rango intercuartílico (*IQR*) es el mejor método para determinar si una observación es un valor atípico en la distribución. Este valor es igual que cualquier distancia $1.5(IQR)$ menos de $Q1$ o mayor que $Q3$.
- La estandarización del valor incluye encontrar la diferencia entre el valor dado y la media, y dividir esa distancia por la desviación estándar. El valor resultante es una serie de desviaciones estándar y no tiene unidades asociadas a él.
- Las puntuaciones estandarizadas pueden dar lugar a valores positivos y negativos. Se puede pensar que un negativo indica un valor que se encuentra a la izquierda de la media, y un positivo indica un valor que se encuentra a la derecha de la media.
- Se puede hacer una estimación de cuántas observaciones están dentro de un cierto número de desviaciones estándar si una distribución tiene forma de campana, unimodal y simétrica.
- La regla empírica establece que:
 - Alrededor del 68% de las observaciones en un conjunto de datos estarán dentro de una desviación estándar de la media
 - Alrededor del 95% de las observaciones en un conjunto de datos estarán dentro de dos desviaciones estándar de la media
 - Alrededor del 99,7% de las observaciones en un conjunto de datos estarán dentro de las tres desviaciones estándar de la media

Ecuaciones clave

Conversión de valores en puntuaciones estandarizadas

$z = \frac{x - \mu}{\sigma}$, donde x representa el valor de la observación, μ representa la media de la población, σ representa la desviación estándar de la población, y z representa el valor estandarizado, o puntuación z .

Desviación de la media

$(x - \bar{x})$, donde (x) es la observación en el conjunto de datos, y (\bar{x}) es la media de la muestra.

Rango intercuartílico (*IQR*)

$$Q3 - Q1$$

Menor valor atípico

$Q1 - 1.5(IQR)$, recuerda multiplicar 1.5 por IQR primero, luego resta de $Q1$

Media

$$\frac{\text{suma de valores de datos}}{\text{número total de valores de datos}} \quad \text{o} \quad \bar{x} = \frac{\sum x}{n},$$

dónde x es el medio, $\sum x$ es el símbolo de "suma de", x representa los valores de los datos, y n es el número total de valores de datos.

Desviación estándar de una población

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n}}, \text{ donde } \sum \text{ es la suma de } (x - \mu)^2 \text{ para cada observación, } (x) \text{ es la observación en el conjunto de datos, } (\mu) \text{ es el medio, y } (n) \text{ es el número de observaciones.}$$

Desviación estándar de una muestra

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}, \text{ donde } \sum \text{ es la suma de } (x - \bar{x})^2 \text{ para cada observación, } (x) \text{ es la observación en el conjunto de datos, } (\bar{x}) \text{ es el medio, y } (n) \text{ es el número de observaciones.}$$

Atípico superior

$Q3 + 1.5(IQR)$, recuerda multiplicar 1.5 por IQR primero, luego añadir a $Q3$

Variación de una población

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n}, \text{ donde } \sum \text{ es la suma de } (x - \mu)^2 \text{ para cada observación, } (x) \text{ es la observación en el conjunto de datos, } (\mu) \text{ es el medio, y } (n) \text{ es el número de observaciones.}$$

Variación de una muestra

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}, \text{ donde } \sum \text{ es la suma de } (x - \bar{x})^2 \text{ para cada observación, } (x) \text{ es la observación en el conjunto de datos, } (\bar{x}) \text{ es el medio, y } (n) \text{ es el número de observaciones.}$$

Glossario

s

la desviación estándar de una muestra de observaciones.

sigma (σ)

la desviación estándar de una población de observaciones.

S^2

la variación de una muestra de observaciones.

σ^2

la varianza de una población de observaciones.

desviación de la media

la distancia entre una observación (x) en un conjunto de datos y la media (\bar{x}) del conjunto de datos.

reglas empíricas

Una guía que predice el porcentaje de observaciones dentro de un cierto número de desviaciones estándar. También conocido como el 68 – 95 – 99.7 regla que establece que en una distribución unimodal en forma de campana, casi todos los valores de datos observados, x , se encuentran dentro de tres desviaciones estándar, σ , a cualquier lado de la media, μ , más específicamente, sobre 68% de las observaciones en un conjunto de datos estarán dentro de una desviación estándar de la media ($\mu \pm \sigma$), sobre 95% de las observaciones en un conjunto de datos estarán dentro de dos desviaciones estándar de la media ($\mu \pm 2\sigma$), y sobre 99.7% de las observaciones en un conjunto de datos estarán dentro de tres desviaciones estándar de la media ($\mu \pm 3\sigma$).

primer cuartil

El valor por debajo del cual se encuentra una cuarta parte de los datos, también igual al percentil 25. A veces se denota $Q1$.

resumen de cinco números

La colección del mínimo, primer cuartil, mediana, tercer cuartil y máximo de la variable.

rango intercuartílico

La cantidad $Q3 - Q1$. A veces denotado como IQR .

sesgo hacia la izquierda (sesgo negativo)

La mayoría de los datos están agrupados a la derecha del gráfico con una "cola" de valores poco frecuentes en el extremo izquierdo (inferior) de la distribución.

valor atípico más bajo

una observación que es menos de $Q1 - 1.5(IQR)$

medio

despliega los conjuntos de datos cuales tienen un gran número de observaciones por agrupando "contenedores" de tamaño iguales

mediana

el valor "medio" de un conjunto de valores enumerados en orden numérico.

caso aparte

un valor inusual o extremo, dados los otros valores del conjunto de datos.

alcance

el valor máximo (o más grande) - el valor mínimo (o más pequeño).

resistente

no es afectado por la oblicuidad de un gráfico.

sesgado a la derecha (sesgo positivo)

la mayoría de los datos se agrupan a la izquierda del gráfico con una "cola" de valores infrecuentes en el extremo derecho (arriba) de la distribución.

desviación estándar

una medida de cuán dispersas son las observaciones a partir de la media.

valor estandarizado

el número de desviaciones estándar de una observación está lejos de la media. También se conoce como puntuación z .

simétrico

los lados izquierdo y derecho de la distribución (de cerca) se reflejan entre sí. Si dibujas una línea vertical por el centro de la distribución y doblas la distribución por la mitad, los lados izquierdo y derecho se igualarán estrechamente.

tercer cuartil

el valor por debajo del cual se encuentran tres cuartas partes de los datos, también igual al percentil 75. A veces es denotado como $Q3$

valor atípico superior

una observación que es mayor que $Q3 + 1.5(IQR)$

variabilidad

una medida de cuán dispersos (propagación) están los datos. A menudo se le conoce como la propagación o dispersión de un conjunto de datos

varianza

la desviación estándar al cuadrado