



# Medidas de Forma

## Asimetría

Aaric Llerena Medina





# Índice

## 1 Introducción

- ▶ **Introducción**
- ▶ Tipos de Asimetría
- ▶ Coeficientes de Asimetría
- ▶ Corrección de Datos Asimétricos
- ▶ Preguntas y Respuestas



# ¿Qué es la asimetría?

## 1 Introducción

La asimetría es una medida estadística que describe la falta de simetría en una distribución de datos. En una distribución simétrica, los datos están igualmente distribuidos alrededor de la media y la curva de distribución es simétrica respecto a un eje vertical que pasa por la media. Sin embargo, si la distribución es asimétrica, entonces la curva de distribución no es simétrica, y la media, mediana y moda pueden tener diferentes valores.

Figura: Curva Simétrica

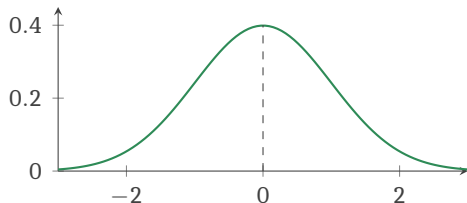
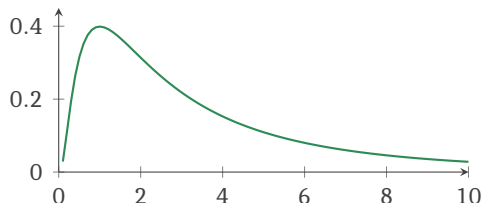


Figura: Curva Asimétrica





# ¿Por qué es importante entender la asimetría?

## 1 Introducción

### Interpretación adecuada

Si los datos no están distribuidos de manera simétrica, la media puede ser una medida inadecuada para resumir la distribución y puede ser necesario usar la mediana u otra medida de tendencia central. Además, la asimetría puede afectar la interpretación de otras medidas estadísticas, como la desviación estándar.

### Inferencia estadística

La asimetría puede afectar la inferencia estadística, ya que muchas pruebas estadísticas se basan en supuestos de distribuciones simétricas. Si los datos no son simétricos, puede ser necesario utilizar pruebas estadísticas diferentes o ajustar los modelos estadísticos para tener en cuenta la asimetría.



# ¿Por qué es importante entender la asimetría?

## 1 Introducción

### Toma de decisiones

Si los datos tienen una distribución asimétrica, puede ser importante tener en cuenta los valores atípicos o extremos en la cola de la distribución. Esto es especialmente relevante en la toma de decisiones críticas, como en la evaluación de riesgos financieros o médicos.

### Modelo adecuado

En muchos análisis estadísticos y modelos predictivos, se asume que los datos siguen una distribución normal o simétrica. Sin embargo, si los datos son asimétricos, estos supuestos pueden no ser válidos y pueden conducir a resultados incorrectos o inexactos. Comprender la asimetría permite seleccionar y aplicar los modelos adecuados que se ajusten mejor a la naturaleza de los datos.



# ¿Por qué es importante entender la asimetría?

## 1 Introducción

### Detección de sesgos

La asimetría puede indicar la presencia de sesgos en los datos. Un sesgo puede surgir cuando los datos están sesgados hacia un extremo de la distribución.

### Selección de métodos de transformación

En algunos casos, la asimetría en los datos puede afectar la interpretación de los resultados o violar los supuestos de algunos métodos estadísticos. En tales situaciones, se pueden aplicar transformaciones a los datos para hacerlos más simétricos. Comprender la asimetría ayuda a identificar cuándo y cómo aplicar métodos de transformación apropiados.



# Índice

2 Tipos de Asimetría

- ▶ Introducción
- ▶ **Tipos de Asimetría**
- ▶ Coeficientes de Asimetría
- ▶ Corrección de Datos Asimétricos
- ▶ Preguntas y Respuestas



## Asimetría Positiva

### 2 Tipos de Asimetría

También conocida como sesgo hacia la derecha, se produce cuando la cola derecha de la distribución es más larga que la cola izquierda. En otras palabras, hay más valores en la parte inferior de la distribución y menos valores en la parte superior. En una distribución asimétrica positiva, la media es mayor que la mediana y la moda, y la mayoría de los valores se concentran en la parte inferior de la distribución.

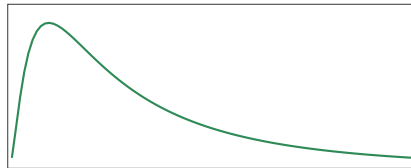
### Comparación de Estadísticos Centrales

$\text{Moda} < \text{Mediana} < \text{Promedio}$

### Comparación de Coeficiente de Asimetría

$\text{Asimetría} > 0$

Figura: Asimetría Positiva





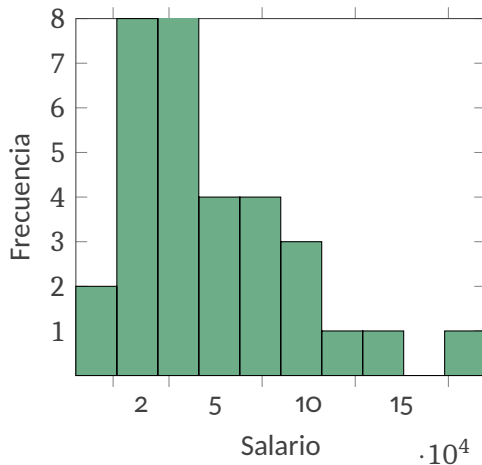


## Asimetría Positiva

2 Tipos de Asimetría

### Ejemplo

Imagina que estás analizando la distribución de los salarios de una empresa y descubres que la mayoría de los empleados ganan salarios bajos, mientras que un pequeño grupo de empleados de alto nivel ganan salarios muy altos. En este caso, la distribución de los salarios tiene una asimetría positiva, ya que la cola derecha es más larga que la cola izquierda.





## Asimetría Negativa

### 2 Tipos de Asimetría

También conocida como sesgo hacia la izquierda, se produce cuando la cola izquierda de la distribución es más larga que la cola derecha. En otras palabras, hay más valores en la parte superior de la distribución y menos valores en la parte inferior. En una distribución asimétrica negativa, la media es menor que la mediana y la moda, y la mayoría de los valores se concentran en la parte superior de la distribución.

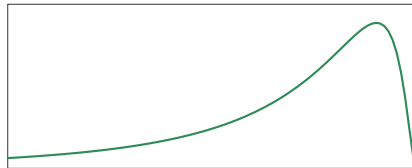
#### Comparación de Estadísticos Centrales

$\text{Moda} > \text{Mediana} > \text{Promedio}$

#### Comparación de Coeficiente de Asimetría

$\text{Asimetría} < 0$

Figura: Asimetría Negativa



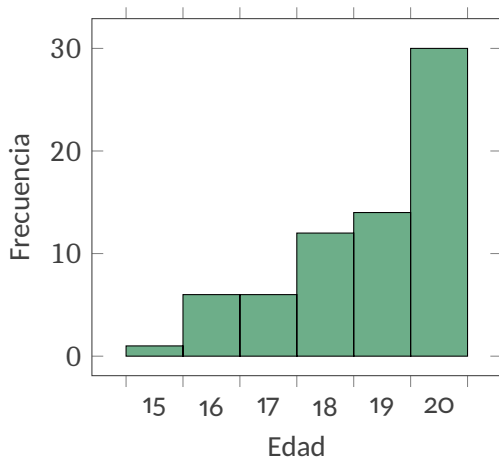


## Asimetría Negativa

2 Tipos de Asimetría

### Ejemplo

Suponga que estás analizando la distribución de las edades de un grupo de estudiantes universitarios y descubres que la mayoría de los estudiantes son mayores de edad, mientras que un pequeño grupo de estudiantes menores de edad está presente en la muestra. En este caso, la distribución de las edades tiene una asimetría negativa, ya que la cola izquierda es más larga que la cola derecha.





## Asimetría Cero

### 2 Tipos de Asimetría

También conocida como simetría, se produce cuando la distribución es perfectamente simétrica, lo que significa que la cola izquierda y la cola derecha son iguales en longitud y forma. En una distribución simétrica, la media, la mediana y la moda son iguales y se encuentran en el centro de la distribución.

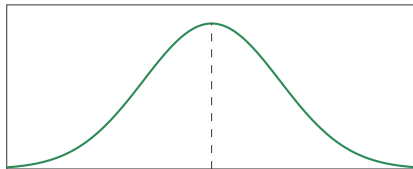
#### Comparación de Estadísticos Centrales

Moda = Mediana = Promedio

#### Comparación de Coeficiente de Asimetría

Asimetría = 0

Figura: Asimetría Cero



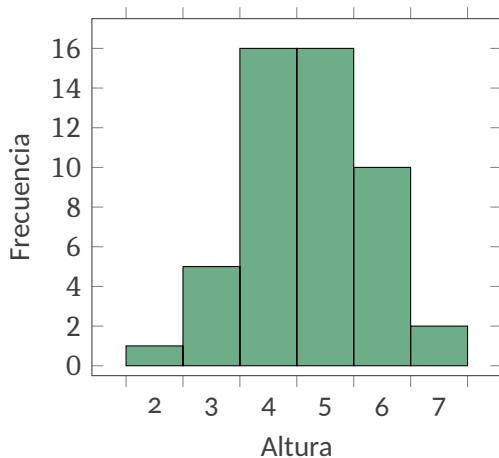


## Asimetría Cero

2 Tipos de Asimetría

### Ejemplo

Suponga que estás analizando la distribución de las alturas de un grupo de árboles y descubres que la distribución es perfectamente simétrica. Es decir, hay un número igual de árboles de diferentes alturas a ambos lados de la distribución. En este caso, la distribución de las alturas de los árboles tiene una asimetría cero.





# Índice

3 Coeficientes de Asimetría

► Introducción

► Tipos de Asimetría

► Coeficientes de Asimetría

► Corrección de Datos Asimétricos

► Preguntas y Respuestas



## Coeficiente de Asimetría de Pearson

### 3 Coeficientes de Asimetría

El coeficiente de asimetría de Pearson, también conocido como coeficiente de asimetría de tercer orden, es una medida estadística utilizada para medir la asimetría de una distribución de datos. Fue desarrollado por el matemático Karl Pearson y se define como la tercera desviación estándar dividida por la desviación estándar.

El coeficiente de asimetría de Pearson es una medida estadística que se utiliza para evaluar la asimetría de una distribución de datos. Este coeficiente mide la diferencia entre la media y la moda de la distribución, dividida por la desviación estándar. El coeficiente de asimetría de Pearson se puede calcular utilizando la siguiente fórmula:

$$As = \frac{3 (\text{Media} - \text{Mediana})}{\text{Desviación Estándar}}$$



## Coeficiente de Asimetría de Pearson - Interpretación

### 3 Coeficientes de Asimetría

El coeficiente de asimetría de Pearson puede tomar valores positivos o negativos. Si el coeficiente es positivo, la distribución de datos es asimétrica hacia la derecha, lo que significa que la cola derecha de la distribución es más larga y la mayoría de los datos se concentran en la parte izquierda de la distribución. Si el coeficiente es negativo, la distribución es asimétrica hacia la izquierda, lo que significa que la cola izquierda de la distribución es más larga y la mayoría de los datos se concentran en la parte derecha de la distribución. Si el coeficiente es igual a cero, la distribución es simétrica.

Criterio	Interpretación
$As = 0$	Entonces la distribución es simétrica
$As > 0$	Entonces la distribución es asimétrica hacia la derecha (sesgo positivo)
$As < 0$	Entonces la distribución es asimétrica hacia la izquierda (sesgo negativo)





## Coeficiente de Asimetría de Pearson - Ventajas

### 3 Coeficientes de Asimetría

- **Es fácil de calcular y de interpretar:** El coeficiente de asimetría de Pearson es una medida estadística relativamente fácil de calcular, lo que lo hace accesible incluso para aquellos que no tienen un conocimiento avanzado en estadística.
- **Ampliamente utilizado:** El coeficiente de asimetría de Pearson es una medida estadística ampliamente utilizada en diferentes campos, incluyendo las ciencias sociales, la biología, la economía y la psicología, lo que permite la comparación de diferentes estudios y la interpretación de los resultados en un contexto más amplio.
- **Informa sobre la dirección de la asimetría:** El coeficiente de asimetría de Pearson proporciona información sobre la dirección de la asimetría en una distribución de datos, lo que puede ser útil para comprender la naturaleza de los datos y para la interpretación de los resultados.



# Coeficiente de Asimetría de Pearson - Ventajas

## 3 Coeficientes de Asimetría

- **Ayuda a identificar distribuciones no normales:** El coeficiente de asimetría de Pearson puede ayudar a identificar distribuciones no normales, lo que puede ser importante para seleccionar los métodos estadísticos adecuados para analizar los datos.
- **Proporciona información complementaria:** El coeficiente de asimetría de Pearson puede proporcionar información complementaria a otras medidas estadísticas, como la media y la desviación estándar, para comprender mejor la distribución de los datos y para identificar patrones y tendencias en los datos.



## Coeficiente de Asimetría de Pearson - Desventajas

### 3 Coeficientes de Asimetría

- **Sensibilidad a valores atípicos:** El coeficiente de asimetría de Pearson puede verse afectado por valores atípicos en los datos, lo que puede influir en la precisión de la medida. Si la distribución de datos contiene valores atípicos, el coeficiente de asimetría de Pearson puede no reflejar con precisión la verdadera asimetría de la distribución.
- **Supuestos subyacentes:** El coeficiente de asimetría de Pearson se basa en varios supuestos subyacentes, como la normalidad de la distribución y la homogeneidad de la varianza. Si estos supuestos no se cumplen, el coeficiente de asimetría de Pearson puede no ser una medida precisa de la asimetría de la distribución.
- **No es adecuado para distribuciones bimodales:** El coeficiente de asimetría de Pearson solo mide la asimetría en una dirección (hacia la izquierda o hacia la derecha). Por lo tanto, no es adecuado para distribuciones bimodales que presentan asimetría en ambas direcciones.



## Coeficiente de Asimetría de Pearson - Desventajas

### 3 Coeficientes de Asimetría

- **Limitaciones a distribuciones simétricas o moderadamente asimétricas:** El coeficiente de asimetría de Pearson es más adecuado para distribuciones simétricas o moderadamente asimétricas. Para distribuciones altamente asimétricas, puede ser necesario utilizar otras medidas estadísticas, como el coeficiente de asimetría de Fisher.
- **No proporciona información sobre la magnitud de la asimetría:** El coeficiente de asimetría de Pearson solo indica si una distribución es asimétrica hacia la izquierda o hacia la derecha, pero no proporciona información sobre la magnitud de la asimetría. Para obtener una medida más detallada de la asimetría de la distribución, puede ser necesario utilizar otras medidas estadísticas, como el coeficiente de curtosis.



# Coeficiente de Asimetría de Pearson - Función en RStudio

3 Coeficientes de Asimetría

## Código para la Asimetría de Pearson

```
asimetria_pearson <- function(data){  
  media = mean(data)  
  mediana = median(data)  
  desv = sd(data)  
  aspearson = 3*(media - mediana)/desv  
  print(round(aspearson,4))  
}
```

## Solución Manual

$A = \{3, 4, 4, 4, 7, 8, 12, 13, 13, 14, 15, 18, 22, 24, 25, 26, 26, 27, 29, 32\}$

$\bar{X} = 16.30$

Mediana = 14.50

$\sigma = 9.498$

As. Pearson = 0.5685

## Solución RStudio

```
asimetria_pearson(a)
```

```
[1] 0.5685
```



## Coeficiente de Asimetría de Fisher

### 3 Coeficientes de Asimetría

El coeficiente de asimetría de Fisher es una medida de asimetría que se utiliza para medir la diferencia entre la media y la moda de una distribución de datos, normalizada por la desviación estándar de los datos.

La fórmula para el coeficiente de asimetría de Fisher es la siguiente:

$$As = \frac{1}{n} \times \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\sigma^3}$$

donde  $\bar{x}$  es la media aritmética de los datos,  $\sigma$  es la desviación estándar de los datos y  $n$  es el tamaño de la muestra.



## Coeficiente de Asimetría de Fisher - Interpretación

### 3 Coeficientes de Asimetría

El valor del coeficiente de asimetría de Fisher puede ser positivo, negativo o cero. Un valor de cero indica que la distribución es simétrica. Un valor positivo indica que la distribución es asimétrica hacia la derecha, lo que significa que hay una cola más larga a la derecha de la media que a la izquierda de la media. Un valor negativo indica que la distribución es asimétrica hacia la izquierda, lo que significa que hay una cola más larga a la izquierda de la media que a la derecha de la media.

Criterio	Interpretación
$As = 0$	Entonces la distribución es simétrica
$As > 0$	Entonces la distribución es asimétrica hacia la derecha (sesgo positivo)
$As < 0$	Entonces la distribución es asimétrica hacia la izquierda (sesgo negativo)



## Coeficiente de Asimetría de Fisher - Ventajas

### 3 Coeficientes de Asimetría

- **Normalización:** El coeficiente de asimetría de Fisher es una medida normalizada de asimetría, lo que significa que se puede comparar fácilmente la asimetría entre diferentes distribuciones de datos, incluso si tienen diferentes escalas y desviaciones estándar.
- **Sensibilidad a la cola:** El coeficiente de asimetría de Fisher es sensible a la presencia de colas en una distribución de datos. Cuando una distribución tiene una cola larga en una dirección, el coeficiente de asimetría de Fisher aumenta en esa dirección. Esto lo convierte en una medida útil para identificar distribuciones que tienen colas largas y/o que están sesgadas.
- **Utilización en pruebas estadísticas:** El coeficiente de asimetría de Fisher se utiliza a menudo en pruebas estadísticas para evaluar la normalidad de los datos. Si los datos no son normales, el coeficiente de asimetría de Fisher se puede utilizar como una medida de la asimetría de los datos en lugar de pruebas que requieren la normalidad, como la prueba t.





## Coeficiente de Asimetría de Fisher - Desventajas

### 3 Coeficientes de Asimetría

- **Sensibilidad a valores extremos:** El coeficiente de asimetría de Fisher puede verse afectado por valores extremos o atípicos en la distribución de datos. Los valores extremos pueden hacer que la distribución parezca más asimétrica de lo que realmente es, especialmente en muestras pequeñas.
- **Limitado a distribuciones simétricas unimodales:** El coeficiente de asimetría de Fisher solo es aplicable a distribuciones simétricas unimodales, lo que significa que no se puede utilizar para evaluar la asimetría de distribuciones bimodales o multimodales.
- **No proporcionar información completa sobre la distribución:** El coeficiente de asimetría de Fisher solo proporciona información sobre la asimetría de la distribución de datos, pero no proporciona información sobre otros aspectos importantes de la distribución, como la forma y la concentración de los datos alrededor de la media.



# Coeficiente de Asimetría de Fisher - Función en RStudio

3 Coeficientes de Asimetría

## Código para la Asimetría de Pearson

```
library(e1071)
```

```
e1071::skewness(a)
```

## Solución Manual

$A = \{3, 4, 4, 4, 7, 8, 12, 13, 13, 14, 15, 18, 22, 24, 25, 26, 26, 27, 29, 32\}$

$$\sum (x_i - \bar{x})^3 = 16.30$$

$$n = 20$$

$$\sigma = 9.498$$

$$\text{As. Fisher} = 0.0465$$

## Solución RStudio

```
e1071::skewness(a)
```

```
[1] 0.04648278
```



## Coeficiente de Asimetría de Bowley

### 3 Coeficientes de Asimetría

El coeficiente de asimetría de Bowley es una medida estadística que mide la asimetría de una distribución. Fue desarrollado por el estadístico inglés Arthur Bowley en 1920 como una alternativa al coeficiente de Pearson, que a menudo se ve afectado por valores extremos.

La fórmula del coeficiente de asimetría de Bowley es:

$$As = \frac{Q_1 + Q_3 + 2Q_2}{Q_3 - Q_1}$$

Donde  $Q_1$  es el primer cuartil, el  $Q_2$  es el segundo cuartil o también conocido como mediana, mientras que  $Q_3$  es el tercer cuartil.



## Coeficiente de Asimetría de Bowley - Interpretación

### 3 Coeficientes de Asimetría

El resultado del coeficiente de asimetría de Bowley puede ser positivo o negativo, lo que indica la dirección de la asimetría. Si el valor es positivo, significa que la distribución tiene una cola positiva o sesgo hacia la derecha, lo que significa que la mayoría de los valores están en la parte inferior de la distribución. Si el valor es negativo, indica que la distribución tiene una cola negativa o sesgo hacia la izquierda, lo que significa que la mayoría de los valores están en la parte superior de la distribución.

En general, se considera que una distribución es simétrica si el coeficiente de asimetría de Bowley es cercano a cero (entre -0,5 y 0,5). Un valor superior a 0,5 o inferior a -0,5 indica una asimetría moderada a fuerte en la distribución. Sin embargo, es importante tener en cuenta que el coeficiente de asimetría de Bowley no proporciona información sobre la magnitud de la asimetría, por lo que es posible que se requiera una evaluación visual adicional de la distribución.



## Coeficiente de Asimetría de Bowley - Ventajas

### 3 Coeficientes de Asimetría

- El coeficiente de asimetría de Bowley es robusto frente a valores extremos, lo que significa que no se ve muy afectado por valores atípicos en la distribución.
- Es fácil de calcular y entender, ya que solo requiere los tres cuartiles y la mediana de la distribución.
- Proporciona información sobre la dirección de la asimetría (es decir, si la distribución tiene una cola positiva o negativa) y, por lo tanto, puede ayudar a identificar patrones en los datos.
- Permite una comparación fácil de la asimetría entre diferentes distribuciones.



## Coeficiente de Asimetría de Bowley - Desventajas

### 3 Coeficientes de Asimetría

- No proporciona información sobre la magnitud de la asimetría, sino solo sobre la dirección de la cola de la distribución.
- Puede no ser muy sensible a pequeñas desviaciones de la simetría, lo que significa que puede no detectar asimetrías leves en la distribución.
- Requiere datos ordenados y puede ser difícil de calcular para distribuciones con muchos valores repetidos.
- No es adecuado para distribuciones bimodales o multimodales, que tienen más de dos modas o picos.
- Puede ser influenciado por la elección de los cuartiles y la mediana, especialmente si los datos son sesgados o tienen valores extremos.



# Coeficiente de Asimetría de Bowley - Función en RStudio

3 Coeficientes de Asimetría

## Código para la Asimetría de Bowley

```
asimetria_bowley <- function(data){  
  quartiles <- quantile(data)  
  q1 <- quartiles[2]  
  q2 <- quartiles[3]  
  q3 <- quartiles[4]  
  asbowley <- (q1+q3+2*q2)/(q3-q1)  
  print(round(as.numeric(asbowley),4))  
}
```

## Solución Manual

$A = \{3, 4, 4, 4, 7, 8, 12, 13, 13, 14, 15, 18, 22, 24, 25, 26, 26, 27, 29, 32\}$

$Q_1 = 7.75$

$Q_2 = 14.50$

$Q_3 = 25.25$

As. Bowley = 3.5429

## Solución RStudio

```
asimetria_bowley(a)
```

```
[1] 3.5429
```



## ¿Cuándo utilizar cada coeficiente?

### 3 Coeficientes de Asimetría

El **coeficiente de asimetría de Pearson** se recomienda utilizar cuando la distribución es simétrica o casi simétrica y solo hay una cola en la distribución de frecuencias. El coeficiente de Pearson puede ser utilizado para medir la asimetría de una variable continua.

El **coeficiente de asimetría de Fisher** se utiliza cuando la distribución es sesgada y tiene una cola más larga que la otra. Este coeficiente es útil para medir la asimetría de variables continuas y se considera más robusto que el coeficiente de Pearson, especialmente cuando la distribución tiene valores atípicos.

El **coeficiente de asimetría de Bowley** es una medida de asimetría que se utiliza para variables discretas o discretizadas. Se recomienda su uso cuando la distribución tiene valores extremos que generan problemas al usar el coeficiente de Pearson o el de Fisher. También se utiliza para medir la asimetría en variables ordinales.





# Índice

## 4 Corrección de Datos Asimétricos

- ▶ Introducción
- ▶ Tipos de Asimetría
- ▶ Coeficientes de Asimetría
- ▶ Corrección de Datos Asimétricos
- ▶ Preguntas y Respuestas



# Transformación de datos

## 4 Corrección de Datos Asimétricos

Una forma común de corregir la asimetría es aplicar una transformación matemática a los datos. Al aplicar una transformación matemática, se cambia la escala de los datos, lo que puede alterar la forma de la distribución y reducir su asimetría.

Existen diferentes tipos de transformaciones que se pueden aplicar, dependiendo del tipo de asimetría presente y de las características específicas de los datos. Algunas de las transformaciones más comunes son la raíz cuadrada, el logaritmo o la inversa.

La elección de la transformación dependerá de las características de los datos y de los objetivos del análisis. En algunos casos, una transformación simple como la raíz cuadrada puede ser suficiente para corregir la asimetría, mientras que en otros casos más complejos, puede ser necesario probar varias transformaciones diferentes para encontrar la que mejor se ajuste a los datos.



## Eliminación de valores atípicos

### 4 Corrección de Datos Asimétricos

Si la asimetría es causada por valores atípicos, la eliminación de estos valores puede reducir la asimetría. Los valores atípicos son observaciones que se alejan significativamente de los demás datos en un conjunto y pueden tener un efecto negativo en el análisis estadístico. Al eliminar valores atípicos, la distribución puede parecer más simétrica y puede ayudar a mejorar la precisión de los análisis estadísticos.

Sin embargo, es importante tener cuidado al eliminar valores atípicos ya que puede afectar la representatividad de la muestra. En algunos casos, los valores atípicos pueden ser datos válidos y precisos que deben mantenerse en el análisis. Además, la eliminación de valores atípicos puede afectar la precisión de las estimaciones y conclusiones que se derivan de los datos, especialmente si la muestra es pequeña. Por lo tanto, es importante analizar cuidadosamente los datos y determinar si los valores atípicos son realmente errores o simplemente valores extremos. Si se decide eliminar valores atípicos, es importante documentar los criterios utilizados para identificarlos y justificar la decisión.



## Cambio de escala

### 4 Corrección de Datos Asimétricos

Si la asimetría se debe a diferencias en las escalas de medida, se puede considerar la posibilidad de cambiar la escala de medida. El cambio de escala es una técnica que se utiliza para transformar una variable de medida de una escala a otra escala. Esta técnica puede ser útil en situaciones en las que la escala de medida original de una variable no es adecuada para el análisis estadístico o cuando se desea comparar variables medidas en diferentes escalas.

El cambio de escala puede ser una técnica útil para reducir la asimetría, ya que puede ayudar a homogeneizar los datos y mejorar la comparabilidad de las variables. Sin embargo, es importante tener en cuenta que el cambio de escala puede alterar la interpretación de los datos y debe ser justificado cuidadosamente en función de los objetivos del análisis.



## Cambio de escala - Ejemplo

### 4 Corrección de Datos Asimétricos

Un ejemplo de cambio de escala es el siguiente: suponga que se tiene una variable que mide el ingreso anual de los trabajadores en dólares estadounidenses (USD). Sin embargo, la distribución de ingresos está sesgada hacia la derecha, lo que dificulta el análisis estadístico. En este caso, se puede considerar la posibilidad de cambiar la escala de medida de dólares a porcentajes. Para hacer esto, se divide el ingreso anual de cada trabajador por el ingreso promedio anual de la población y se multiplica por 100 para obtener el porcentaje correspondiente. De esta manera, se obtiene una nueva variable que mide el ingreso de los trabajadores en porcentaje del ingreso promedio de la población, lo que puede facilitar el análisis estadístico y la comparación con otras variables medidas en diferentes escalas.

Es importante tener en cuenta que el cambio de escala debe realizarse cuidadosamente, ya que puede alterar la interpretación de los resultados y no siempre es la solución óptima para corregir la asimetría de los datos.



# Re-codificación de variables

## 4 Corrección de Datos Asimétricos

La re-codificación de variables es una técnica que se utiliza para modificar o agrupar las categorías de una variable. Esta técnica se puede utilizar para corregir la asimetría de los datos, simplificar la variable o hacerla más fácil de interpretar. La re-codificación puede ser especialmente útil en variables nominales o categóricas que presentan una asimetría hacia ciertas categorías o valores.

La re-codificación de variables puede ser una técnica útil para corregir la asimetría, ya que puede ayudar a homogeneizar los datos y mejorar la precisión de los análisis estadísticos. Además, la re-codificación puede simplificar las variables y hacerlas más fáciles de interpretar. Sin embargo, es importante tener en cuenta que la re-codificación puede alterar la información original de los datos y debe ser justificada cuidadosamente en función de los objetivos del análisis.



## Re-codificación de variables - Ejemplo

### 4 Corrección de Datos Asimétricos

Un ejemplo de re-codificación de variables es el siguiente: suponga que se tiene una variable nominal que mide el nivel de educación de los trabajadores. La variable tiene las categorías “primaria”, “secundaria”, “universidad” y “posgrado”, pero se observa una asimetría hacia la categoría “secundaria”, que es la más frecuente. Para corregir la asimetría, se puede considerar la posibilidad de agrupar las categorías “primaria” y “secundaria” en una sola categoría llamada “educación básica”, mientras que las categorías “universidad” y “posgrado” se agrupan en una sola categoría llamada “educación avanzada”. De esta manera, se obtiene una nueva variable que mide el nivel de educación de los trabajadores con dos categorías más equilibradas.

Es importante tener en cuenta que la re-codificación de variables debe realizarse cuidadosamente y con base en un análisis previo de los datos y de la interpretación de las nuevas categorías resultantes. Además, la re-codificación debe tener en cuenta las características específicas de la variable y del análisis que se desea realizar.



## Uso de modelos específicos

### 4 Corrección de Datos Asimétricos

En algunos casos, se pueden utilizar modelos estadísticos específicos que sean robustos a la asimetría de los datos. Por ejemplo, se pueden utilizar modelos lineales generalizados (GLM) en lugar de modelos lineales simples para variables con asimetría.

Otro ejemplo de modelo específico que puede utilizarse para datos asimétricos es el modelo de regresión robusta. Este modelo utiliza una función de pérdida que es menos sensible a los valores atípicos que los modelos de regresión tradicionales, lo que los hace más adecuados para analizar datos con valores atípicos o distribuciones asimétricas.





# Índice

5 Preguntas y Respuestas

- ▶ Introducción
- ▶ Tipos de Asimetría
- ▶ Coeficientes de Asimetría
- ▶ Corrección de Datos Asimétricos
- ▶ Preguntas y Respuestas



## ¿Y la función “skewness” del paquete “moments”?

### 5 Preguntas y Respuestas

La función en mención utiliza el coeficiente de asimetría del momento de Pearson, se basa en los momentos centrales de orden 2 y 3 de la distribución. La fórmula del coeficiente de asimetría del momento de Pearson es la siguiente:

$$As = \frac{m_3}{m_2^{3/2}}$$

$$m_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

donde  $m_2$  es el segundo momento central de la distribución (la varianza) y  $m_3$  es el tercer momento central de la distribución.

En general, el coeficiente de asimetría del momento de Pearson se considera más robusto que el coeficiente de asimetría de Pearson, especialmente en presencia de valores atípicos en la distribución. Sin embargo, el coeficiente de asimetría de Pearson es más fácil de calcular y se utiliza comúnmente en muchos análisis estadísticos.



## ¿Datos asimétricos = Distribución no normal?

5 Preguntas y Respuestas

No necesariamente. Si bien es cierto que la asimetría es una característica común en muchas distribuciones no normales, no implica automática que los datos no sigan una distribución normal.

La asimetría puede surgir en diferentes tipos de distribuciones, incluida la distribución normal. En una distribución normal, se espera que la asimetría sea cercana a cero, lo que significa que la distribución es aproximadamente simétrica. Sin embargo, incluso en una distribución normal, puede haber pequeñas desviaciones de la simetría debido a variaciones aleatorias en los datos.



## ¿Datos asimétricos = Distribución no normal?

5 Preguntas y Respuestas

Por otro lado, existen distribuciones no normales que también pueden ser simétricas. Por ejemplo, la distribución uniforme es simétrica pero no es normal. Además, hay distribuciones no normales que pueden ser asimétricas positivas o negativas, como la distribución exponencial, la distribución logarítmica y muchas otras.

Por lo tanto, la asimetría por sí sola no determina si los datos siguen una distribución normal o no. Es importante realizar pruebas y análisis más específicos, como pruebas de normalidad, para evaluar si los datos se ajustan a una distribución normal o a otro tipo de distribución.



## ¿Datos asimétricos = Distribución no normal?

5 Preguntas y Respuestas

No necesariamente. Si bien es cierto que la asimetría es una característica común en muchas distribuciones no normales, no implica automática que los datos no sigan una distribución normal.

La asimetría puede surgir en diferentes tipos de distribuciones, incluida la distribución normal. En una distribución normal, se espera que la asimetría sea cercana a cero, lo que significa que la distribución es aproximadamente simétrica. Sin embargo, incluso en una distribución normal, puede haber pequeñas desviaciones de la simetría debido a variaciones aleatorias en los datos.



## ¿Qué herramientas se pueden utilizar para evaluar la normalidad de una distribución?

### 5 Preguntas y Respuestas

- **Histograma:** Es una representación gráfica de la distribución de frecuencias de los datos. Si la distribución es simétrica y normal, el histograma tendrá una forma de campana. Si la distribución es asimétrica, el histograma tendrá una forma sesgada.
- **Gráfico de probabilidad normal (QQ-plot):** Es una herramienta gráfica para evaluar si una muestra de datos proviene de una distribución normal. Si los datos se ajustan a una distribución normal, los puntos en el gráfico deberían seguir una línea diagonal. Caso contrario, los puntos se desvían de la línea diagonal.
- **Prueba de normalidad:** Existen varias pruebas estadísticas para evaluar la normalidad de una distribución, como la prueba de Shapiro-Wilk y de Kolmogorov-Smirnov. Estas pruebas calculan una estadística de prueba y la comparan con una distribución conocida, como la distribución normal. Si la estadística de prueba es menor que el valor crítico, se rechaza la hipótesis nula de que los datos siguen una distribución normal.



## ¿Es posible que una distribución tenga una asimetría positiva y una curtosis negativa, o viceversa?

5 Preguntas y Respuestas

La asimetría y la curtosis son dos medidas que describen diferentes aspectos de la forma de una distribución. La asimetría mide la falta de simetría de una distribución, mientras que la curtosis mide la concentración de los datos en torno a la media. En general, se puede decir que una distribución con alta asimetría tiende a tener una forma sesgada hacia un lado, y una distribución con alta curtosis tiende a tener una forma más puntiaguda en el centro. Sin embargo, no existe una relación directa entre la asimetría y la curtosis.

Es posible que una distribución tenga una asimetría positiva y una curtosis negativa, o viceversa. Por ejemplo, una distribución con una asimetría positiva y una curtosis negativa tendría una forma sesgada hacia la derecha y sería más plana en el centro. Esto se puede observar en distribuciones como la distribución t de Student con un pequeño número de grados de libertad.



# ¿Cómo se puede detectar y corregir los sesgos de selección y respuesta en estudios con datos asimétricos?

## 5 Preguntas y Respuestas

En estudios con datos asimétricos, es importante detectar y corregir la presencia de sesgos de selección y de respuesta, ya que pueden afectar la validez de los resultados.

- **Sesgo de selección:** Este sesgo puede ocurrir cuando los participantes del estudio no son representativos de la población de interés. Para detectar este sesgo, es recomendable comparar las características de los participantes del estudio con las de la población de interés. Si se identifica un sesgo de selección, se pueden utilizar técnicas de muestreo estratificado o de ponderación de datos para corregir el sesgo.
- **Sesgo de respuesta:** Este sesgo puede ocurrir cuando los participantes del estudio no dan información precisa o completa debido a factores como: olvido, falta de comprensión o tendencia a responder de manera socialmente deseable. Para detectar este sesgo, se pueden utilizar técnicas como la comparación de las respuestas de los participantes con las de otra fuente o el uso de preguntas de control o de verificación.





# Medidas de Forma

*Gracias por su atención!*

*¿Alguna pregunta?*