

¿QUÉ ES?

En estadística, **la correlación** es una medida que cuantifica el grado y la dirección de la relación entre dos variables. Indica cómo una variable cambia en relación con otra. Sin embargo, la correlación **no implica causalidad**, es decir, que una variable influya directamente en la otra.

¿CÓMO LO ENTIENDO?

La correlación es una medida que cuantifica el grado de relación de dos datos.

Características clave de la correlación

1. Rango de valores:

- Va de **-1** a **1**:
 - **+1**: Correlación positiva perfecta. Cuando una variable aumenta, la otra también lo hace de forma proporcional.
 - **0**: No hay correlación (no hay relación lineal).
 - **-1**: Correlación negativa perfecta. Cuando una variable aumenta, la otra disminuye de forma proporcional.

2. Dirección:

- **Positiva**: Ambas variables tienden a moverse en la misma dirección.
- **Negativa**: Las variables tienden a moverse en direcciones opuestas.

3. Magnitud:

- Indica la fuerza de la relación. Valores más cercanos a **-1** o **1** indican una relación más fuerte; valores cercanos a **0** indican una relación más débil.
-

Tipos de correlación

1. Lineal:

- Relación que sigue una línea recta.
- Ejemplo: Altura y peso suelen tener una correlación lineal positiva.

2. No lineal (o curvilínea):

- Relación que no sigue una línea recta pero aún existe una asociación.

3. Nula:

- No hay ningún patrón aparente entre las variables.

Métodos para medir la correlación

1. Coeficiente de correlación de Pearson:

- Evalúa relaciones lineales entre variables continuas.
- Es el más común.
- Ejemplo: Altura y peso.

2. Coeficiente de correlación de Spearman:

- Evalúa relaciones monotónicas (no necesariamente lineales) entre variables.
- Se usa para datos ordinales o cuando hay valores atípicos.

3. Coeficiente de Kendall Tau:

- Evalúa relaciones de concordancia entre pares de datos.
- Es útil para conjuntos pequeños.

Ejemplo práctico

Imagina que estudias la relación entre el tiempo de estudio y las calificaciones de los estudiantes:

- Si al aumentar las horas de estudio las calificaciones también aumentan, hay una correlación positiva.
- Si al aumentar las horas de estudio las calificaciones disminuyen (por ejemplo, por exceso de fatiga), hay una correlación negativa.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Cargar los datos desde el archivo CSV
```

```
df = pd.read_csv("Libro6.csv") # Asegúrate de usar el nombre correcto de tu archivo CSV
```

```
# Seleccionar las columnas relevantes para la correlación
```

```
correlation_data = df[['Median Base Salary', 'Job Satisfaction', 'Job Openings']]
```

```
# Calcular la matriz de correlación
```

```
correlation_matrix = correlation_data.corr()
```

```
# Mostrar la matriz de correlación
```

```
print(correlation_matrix)
```

```
# Graficar la matriz de correlación
```

```
plt.figure(figsize=(8, 6))
```

```
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
```

```
plt.title("Matriz de Correlación entre Salario, Satisfacción Laboral y Ofertas de Trabajo")
```

```
plt.show()
```

