



- Dashboard
- Career Path
- Forms
- Profile



Análisis de Relaciones y Correlaciones - Día 4

Change Status

in-progress 40 min

Learning Objectives

- 1 Calcular e interpretar coeficientes de correlación entre variables
- 2 Construir y analizar matrices de correlación para datasets multivariados
- 3 Visualizar relaciones mediante scatter plots y mapas de calor

Theory

Practice

Quiz

Evidence

Actividades y Aprendizajes

Aprende todo sobre funciones y módulos en Python con ejemplos prácticos.

Task 1: Fundamentos de Correlación (10 minutos)

La correlación mide la **relación lineal** entre dos variables, respondiendo "¿cambian juntas estas variables?" de manera cuantificada.

Coeficiente de Correlación de Pearson

El coeficiente de Pearson mide la **fuerza y dirección** de la relación lineal entre dos variables continuas.

Fórmula: $r = \frac{\sum[(x_i - \mu_x)(y_i - \mu_y)]}{\sqrt{\sum(x_i - \mu_x)^2} \sqrt{\sum(y_i - \mu_y)^2}}$

Rango: $-1 \leq r \leq +1$

r = +1: Correlación positiva perfecta (línea recta ascendente)

r = 0: No hay correlación lineal

r = -1: Correlación negativa perfecta (línea recta descendente)

Interpretación práctica:

$0.8 \leq |r|$: Correlación fuerte

$0.6 \leq |r| < 0.8$: Correlación moderada

$0.3 \leq |r| < 0.6$: Correlación débil

$|r| < 0.3$: Correlación muy débil o inexistente

Limitaciones de la Correlación

No implica causalidad: Dos variables pueden estar correlacionadas sin que una cause la otra.

Sensibilidad a outliers: Un solo punto extremo puede distorsionar completamente el coeficiente.

Solo mide relaciones lineales: Variables pueden tener relaciones no lineales fuertes que no se detectan.

Requiere variables continuas: Para variables categóricas se necesitan otros enfoques.

Task 2: Matrices de Correlación (10 minutos)

Las matrices de correlación permiten **análisis multivariado** sistemático, revelando patrones complejos en datasets con muchas variables.

Construcción de Matrices

```
# Matriz de correlación completa
correlation_matrix = df.corr()

# Correlaciones con una variable específica
correlaciones_target = df.corr()['variable_objetivo'].sort_values(ascending=False)

# Correlaciones por pares significativas
strong_correlations = correlation_matrix[abs(correlation_matrix) > 0.7]
```

Patrones a Identificar

Correlaciones altas positivas: Variables que "se mueven juntas"

Ejemplo: precio y calidad percibida, ingresos y gastos de lujo

Correlaciones altas negativas: Variables que "se mueven opuestas"

Ejemplo: precio y volumen de ventas, edad y capacidad de aprendizaje

Grupos de variables correlacionadas: Indican dimensiones latentes

Ejemplo: Variables de "satisfacción del cliente" que correlacionan fuertemente



Visualización de Matrices

Mapas de calor: Representación visual intuitiva de correlaciones.

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

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix,
            annot=True,
            cmap='coolwarm',
            center=0,
            fmt='.2f')

plt.title('Matriz de Correlación')
plt.tight_layout()
```

Task 3: Análisis Visual de Relaciones (10 minutos)

La visualización complementa los coeficientes numéricos con **representaciones intuitivas** de las relaciones.

Scatter Plots para Dos Variables

```
# Scatter plot básico
plt.scatter(df['variable_x'], df['variable_y'], alpha=0.6)
plt.xlabel('Variable X')
plt.ylabel('Variable Y')
plt.title('Relación entre X e Y')

# Con línea de tendencia
sns.regplot(x='variable_x', y='variable_y', data=df, scatter_kws={'alpha':0.6})
```

Interpretación visual:

Nube de puntos alineada: Correlación fuerte

Nube dispersa: Correlación débil

Curvatura: Relación no lineal

Outliers: Puntos alejados del patrón general

Pair Plots para Múltiples Variables

```
# Matriz de scatter plots
sns.pairplot(df[['var1', 'var2', 'var3', 'var4']], diag_kind='kde')
```



- Dashboard
- Career Path
- Forms
- Profile

```
plt.suptitle('Relaciones entre múltiples variables', y=1.02)
```

Ventajas: Permite detectar patrones complejos de interdependencia entre múltiples variables simultáneamente.

Correlaciones Espurias vs Causales

Correlación espuria: Asociación que parece real pero es coincidental o debida a una tercera variable.

Ejemplos clásicos:

- Correlación entre consumo de helado y ahogamientos (ambos aumentan en verano)
- Correlación entre número de iglesias y crimen (ambos correlacionan con tamaño poblacional)

Cómo identificar:

- Contexto temporal:** ¿La correlación tiene sentido causal?
- Variables de control:** ¿Persiste controlando por terceras variables?
- Mecánismo causal:** ¿Existe explicación lógica para la relación?