

Visualización de Datos y Pruebas de Normalidad

Prof. David Martínez Salazar
Colegio Universitario de Cartago

3 de diciembre de 2025

Objetivos de la Presentación

- Explicar la construcción e interpretación de un *heatmap*.
- Exponer el uso y lectura de un gráfico de violín.
- Analizar las pruebas de normalidad más comunes.
- Interpretar correctamente el valor p en KolmogorovSmirnov y otras pruebas.

Heatmap: Concepto

- Un **heatmap** representa valores mediante colores.
- Permite observar patrones, relaciones y concentraciones.
- Se utiliza comúnmente en matrices de correlación.

Heatmap: Interpretación

- Los colores cálidos suelen representar valores altos.
- Los colores fríos suelen representar valores bajos.
- En correlación:
 - Colores intensos indican relaciones fuertes.
 - Colores neutros indican relaciones débiles.
- Ayuda a detectar multicolinealidad en modelos estadísticos.

Gráfico de Violín: Concepto

- Combina un *boxplot* con una estimación de densidad.
- Representa la distribución de datos de manera detallada.
- Permite visualizar asimetrías, modos y dispersión.

Gráfico de Violín: Interpretación

- Zonas más anchas indican mayor densidad de datos.
- La línea central muestra la mediana.
- La forma del violín revela si la distribución es simétrica o sesgada.
- Permite comparar varias distribuciones simultáneamente.

Pruebas de Normalidad

- **KolmogorovSmirnov (KS)**
- **ShapiroWilk (SW)**
- **AndersonDarling (AD)**
- **JarqueBera (JB)**

Todas comparan los datos contra una distribución normal teórica.

- Compara la distribución acumulada de los datos con la teórica.
- Se basa en la diferencia máxima entre ambas curvas.
- Es sensible en toda la distribución, pero menos en colas.

- **Si** $p < \alpha$: se rechaza la normalidad. La persona concluye que la distribución es distinta de la normal.
- **Si** $p \geq \alpha$: no existe evidencia suficiente para rechazar la normalidad.
- Recordatorio: no se confirma normalidad; solo no se rechaza.

- Es una de las pruebas más potentes para muestras pequeñas y medianas.
- Evalúa la correlación entre datos ordenados y valores esperados bajo normalidad.

- **Si** $p < \alpha$: se rechaza la normalidad.
- **Si** $p \geq \alpha$: no se rechaza la normalidad.
- Alta sensibilidad a asimetrías y colas.

- Da más peso a los datos ubicados en las colas.
- Detecta anomalías o comportamientos extremos.
- El estadístico se compara con puntos críticos.

- Se basa en la asimetría y la curtosis.
- Mide desviaciones respecto a la forma normal ideal.
- Se utiliza frecuentemente en econometría.

- **H0:** los datos provienen de una distribución normal.
- **H1:** los datos no provienen de una distribución normal.
- La decisión depende del valor p, complementado con inspección gráfica.

- Los *heatmaps* y violines ofrecen una visualización clara de patrones y densidades.
- Las pruebas de normalidad permiten validar el uso de métodos paramétricos.
- La interpretación siempre debe acompañarse de sentido estadístico y análisis gráfico.