



Análisis e Interpretación de Datos

Dra. Mariana-Edith Miranda-Varela

18-agosto-2025

Variable MEDV

- Distribución de la variable MEDV no es normal
- Existe un sesgo positivo (cola larga hacia los valores altos)
- Solución
 - Aplicar log para comprimir esa cola

Ventajas de aplicar log

1. Comprime la escala de los valores grandes
2. Mejora la homocedasticidad
3. Ajusta mejor relaciones no lineales: RM, LSTAT
4. Datos económicos
 - Logaritmo permite una mejor interpretación (cambio porcentual)

Colinealidad

- Dos o más variables independientes de un modelo de regresión están correlacionadas entre sí.
- Aportan casi la misma información al modelo
- ¿Cómo afecta?
 - β inestables
 - Dificulta detectar qué variables son significativas
- Colinealidad – 2 variables
- Multicolinealidad – más de 2 variables

Factor de Inflación de la Varianza

- Métrica que detecta la multicolinealidad
- Incremento de la varianza de un coeficiente de regresión
- Valores
 - $VIF \approx 1$ no hay correlación con otras variables
 - $1 < VIF < 5$ correlación moderada, lo cual es aceptable
 - $VIF \geq 5$ o 10 alta multicolinealidad, eliminar o transformar variables

「 muchas gracias. 」