

Introducción al modelado con R

Curso Básico R

CEPAL - Unidad de Estadísticas Sociales

2025-10-22

Tabla de contenidos I

Introducción

Sección 2: Familia de modelos y modelo ajustado

Introducción

El Modelo: Señal y Ruido

El modelado es una herramienta para la Exploración de Datos (EDA).

Definición Fundamental

- ▶ Un **modelo** es una representación simplificada de la realidad que captura patrones relevantes (señal) y deja fuera la variabilidad irrelevante (ruido).
- ▶ Propósitos: **describir**, **resumir** y **predecir**.

💡 George Box:

“Todos los modelos están equivocados, algunos son útiles.”.

Definición y propósito

Su propósito es distinguir:

- ▶ Señal ($f(X)$): Los patrones verdaderos generados por el fenómeno de interés.
- ▶ Ruido (ϵ): Las variaciones aleatorias o el error que no nos interesa.



Ecuación del Modelo:

$$Y = f(X) + \epsilon$$

En este apartado, nos enfocamos en modelos predictivos (que generan \hat{Y}).

El Rol Exploratorio del Modelo

El foco en esta sección es usar modelos para *construir intuición* y encontrar patrones sutiles.

Función en EDA

- ▶ Se utiliza para extraer primero la Señal conocida, y luego examinar lo que queda, es decir, el Ruido (ϵ) o residuales.
- ▶ Si encontramos patrones en los residuales, hemos descubierto una nueva señal que el modelo inicial omitió.

La Regla Crítica: Exploración vs. Confirmación

Es fundamental distinguir la fase de búsqueda de patrones de la fase de prueba formal.

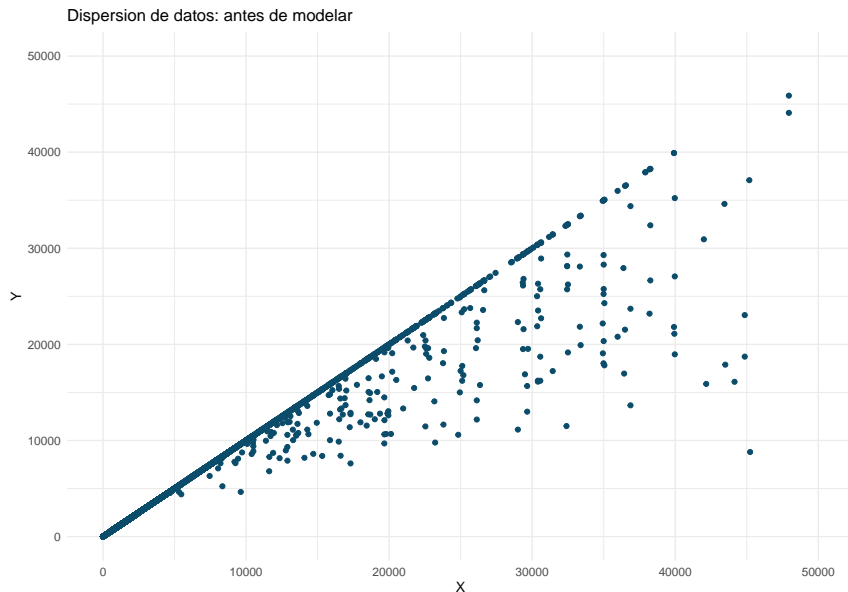
Principio de la Independencia

- ▶ Exploración (Generación de Hipótesis): Puedes usar un dato muchas veces.
- ▶ Confirmación (Inferencia Formal): Solo puedes usar una observación **UNA SOLA VEZ**.

! Advertencia:

Si usas los mismos datos para generar y confirmar una hipótesis, serás **demasiado optimista**.

Ejemplo práctico: visualización previa al ajuste



Sección 2: Familia de modelos y modelo ajustado

Concepto: familia vs miembro ajustado

- ▶ **Familia de modelos:** conjunto de funciones posibles (ej. lineales, polinomiales, logísticas).
- ▶ **Modelo ajustado:** parámetros estimados que describen el miembro concreto de la familia.

! Importante:

seleccionar familia adecuada según pregunta de interés y naturaleza de los datos.

Código: ajuste lineal y resumen

```
library(broom)
mod1 <- lm(gasto ~ ingreso, data = datos)
tidy(mod1)
```

term	estimate	std.error	statistic	p.value
(Intercept)	762.2737293	10.1274951	75.26775	0
ingreso	0.7513918	0.0016143	465.45436	0

```
glance(mod1)[,c(1:4,8:9)]
```

r.squared	adj.r.squared	sigma	statistic	AIC	BIC
0.8820561	0.882052	1447.88	216647.8	503913.7	503938.5

Código: ajuste lineal y resumen

