Modelos de Regresión y Series de Tiempo (MRST)
2025 - 02

Clase 2 – Estimación de parámetros del MRLS

Docente: Natalia Jaramillo Quiceno

Escuela de Ingenierías

natalia.jaramilloq@upb.edu.co

Regresión lineal simple
Definición

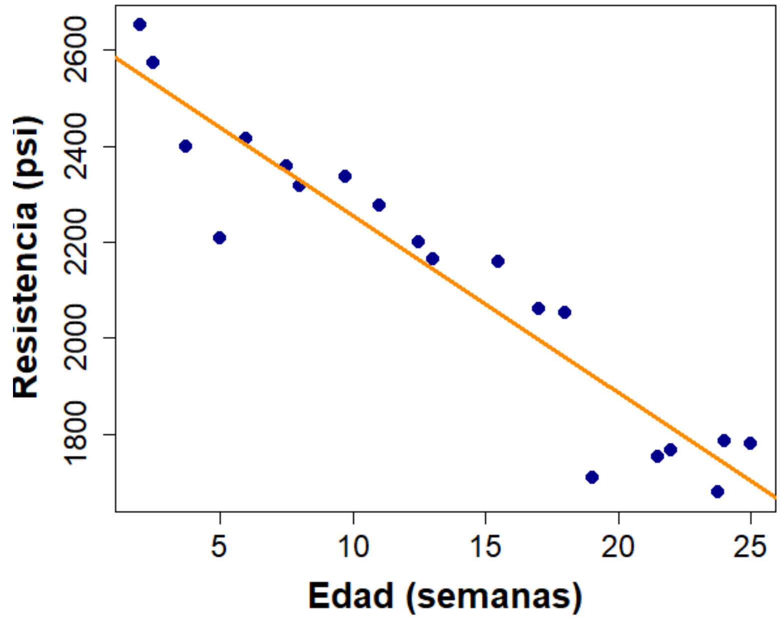
Análisis de regresión

Técnica estadística que permite modelar la relación entre dos variables relacionadas en una forma no determinística.

Modelo de regresión lineal simple (RLS)

El modelo de RLS se define de la siguiente forma:

$$Y = \beta_0 + \beta_1 x + \epsilon$$



Donde,

$\boldsymbol{Y}$: Variable respuesta (variable aleatoria)

$\boldsymbol{x}$: Variable independiente (variable fija)

$\boldsymbol{\beta_0}$ **y** $\boldsymbol{\beta_1}$: Parámetros del modelo. Deben ser estimados a partir de datos muestrales.

$\boldsymbol{\epsilon}$: Componente de error aleatorio con, $E(\epsilon) = 0$ y $V(\epsilon) = \sigma^2$ **Miremos este concepto primero…**

Residuales
Definición

Residual, error residual o error aleatorio

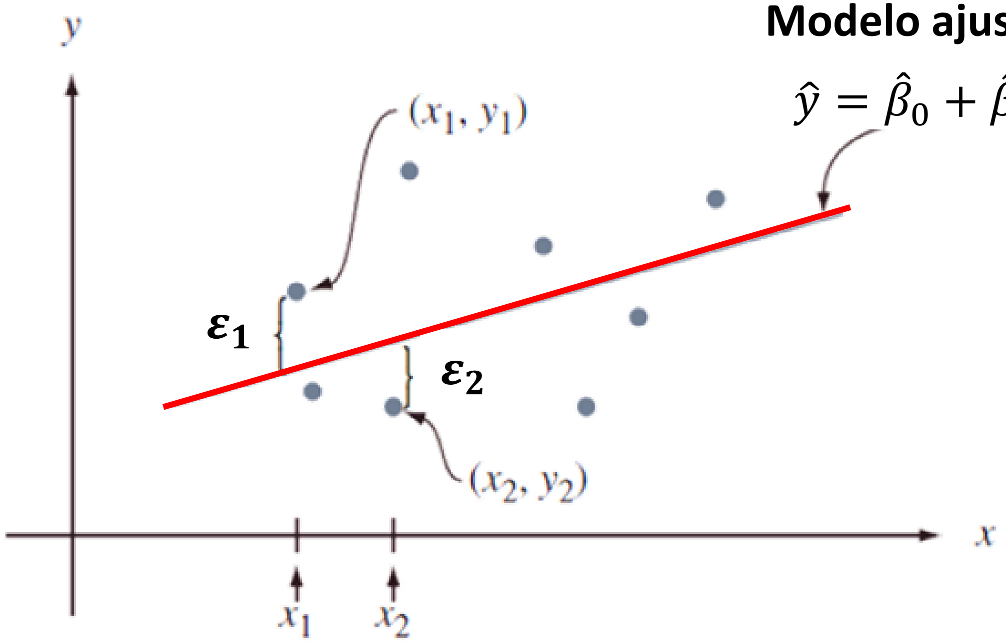
- Restos del modelo ajustado
- Denotados por la letra griega $\boldsymbol{\varepsilon}$ o $\boldsymbol{\epsilon}$
- Cada dato (x, y) tiene un residual asociado, denotado por $\boldsymbol{\varepsilon_i}$ o $\boldsymbol{\epsilon_i}$
- **Cada dato está dado por:** valor ajustado + error
- Diferencia entre el **valor observado** y el **valor ajustado** (dado por el modelo) de $\boldsymbol{y}$:

Error o Residual

$$\varepsilon_i = \underbrace{y_i}_{\text{observado}} - \overbrace{\hat{y}_i}^{\text{ajustado}}$$

Modelo ajustado

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$



El mejor modelo → Minimiza los residuales!!!

Residuales
Preguntas de interpretación

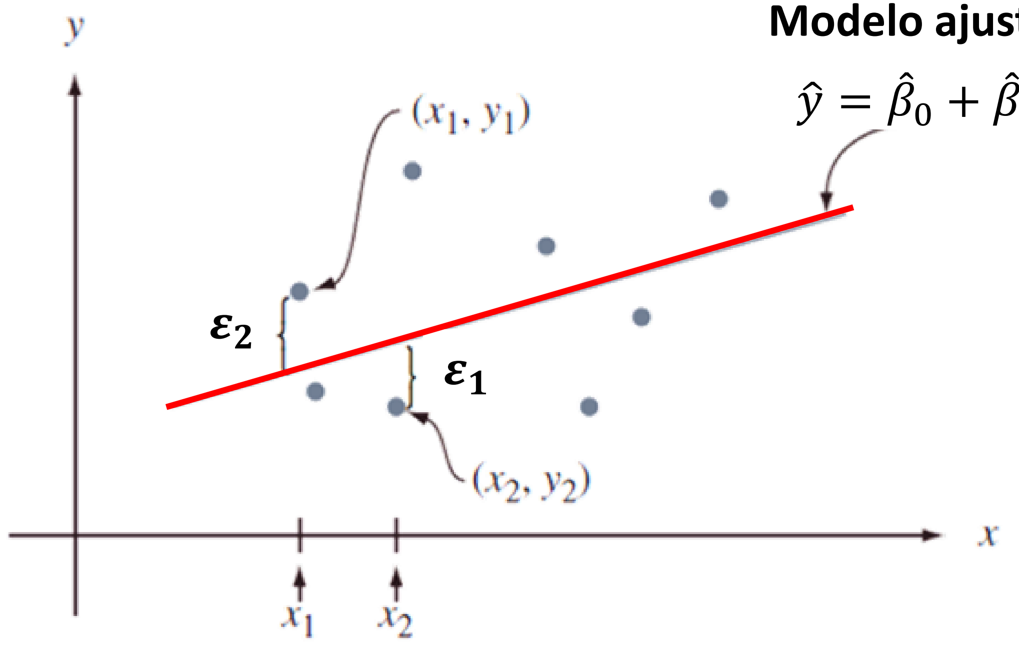
Residuales o errores

1. Si se tiene que, para un dato específico, el valor predicho o ajustado de la variable respuesta es superior al valor observado ¿el residual resultante sería positivo o negativo?

2. En caso de que, para un dato específico, el valor predicho o ajustado de la variable dependiente es menor que el valor observado ¿se diría que el modelo subestima o sobrestima dicho dato?

Modelo ajustado

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$



Regresión lineal simple
Principio de los mínimos cuadrados

Mínimos cuadrados

Idea general – Minimizar la suma de los residuales al cuadrado

$$\varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2$$

Modelo general

$$Y = \beta_0 + \beta_1 x + \boxed{\epsilon}$$

despejando

Esta sumatoria también la podemos expresar como la función L:

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \beta_1 x_i\right)^2$$

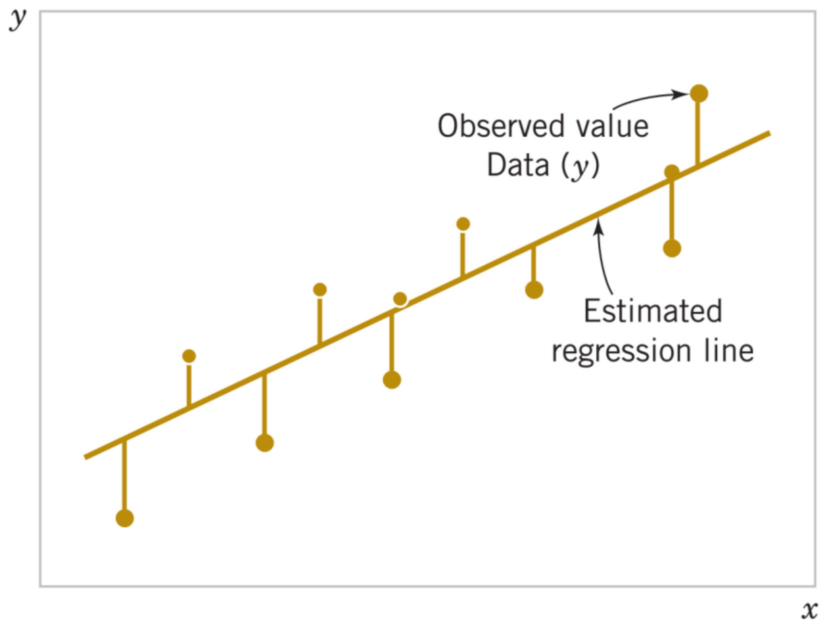
- Las estimaciones puntuales de β_0 y β_1, denotadas $\hat{\beta}_0$ y $\hat{\beta}_1$, son aquellos valores que **minimizan L**.
- La línea de regresión estimada es entonces la línea cuya ecuación es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Regresión lineal simple

Principio de los mínimos cuadrados

Se tiene entonces que los estimadores por mínimos cuadrados de β_0 y β_1, estos son $\hat{\beta}_0$ y $\hat{\beta}_1$, deben satisfacer



$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0$$

Simplificando se obtiene el siguiente sistema de ecuaciones, llamado **ecuaciones normales de mínimos cuadrados**:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad \boxed{1}$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i\, x_i \qquad \boxed{2}$$

Regresión lineal simple
Principio de los mínimos cuadrados

La solución de las ecuaciones normales de mínimos cuadrados es la siguiente:

Pendiente

$$\hat{\beta}_1 = R * \frac{S_y}{S_x}$$

S_y → Desviación estándar de las $\boldsymbol{y}$

S_x → Desviación estándar de las $\boldsymbol{x}$

R ↑ Coeficiente de correlación

Intercepto

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

← Este resultado además demuestra que la línea de regresión siempre pasa por los valores medios de $\boldsymbol{x}$ y $\boldsymbol{y}$

Regresión lineal simple

El vicepresidente de I+D de una empresa cree que las ganancias anuales de la empresa dependen de la cantidad de dinero invertida en I+D. El nuevo presidente no está de acuerdo y ha solicitado pruebas.
Los datos de los últimos años son:

año	Inversión en I+D [Millones USD anual]	Ganancias empresa [Millones USD anual]
1	2	20
2	3	25
3	5	34
4	4	30
5	11	40
6	5	31

¿Cuál es la variable dependiente y cuál la independiente?

Regresión lineal simple

¿Cómo interpretar el modelo?

- El modelo de RLS es: $y = 20 + 2x$

- **β_0 es el intercepto →**

Cuántas ganancias se esperan si no hay inversión en I+D.

En promedio, las ganancias anuales de la compañía serán de 20 millones de USD si la inversión en I+D es de 0 millones de USD anuales.

Ojo, el intercepto no siempre tiene una interpretación práctica.

- **β_1 es la pendiente →**

Cuánto se espera que aumenten las ganancias por cada millón USD adicional invertido en I+D.

Por cada millón USD adicional invertido en I+D, **en promedio** la empresa incrementará sus ganancias anuales en 2 millones USD.

MUCHAS GRACIAS

Natalia Jaramillo Quiceno
e-mail: natalia.jaramilloq@upb.edu.co