

Modelos de Regresión y Series de Tiempo (MRST)

2025 - 02

Descomposición de la variabilidad de y en sumas cuadráticas

Varianza de los errores del modelo

Docente: Natalia Jaramillo Quiceno

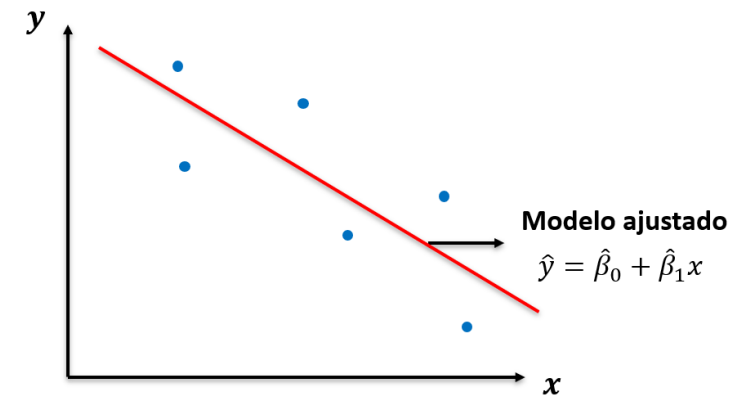
Escuela de Ingenierías

natalia.jaramilloq@upb.edu.co

Regresión lineal simple

Paso a paso muy simplificado

- Análisis **descriptivo** de lo datos
 - Gráfico de dispersión, coeficiente de correlación de Pearson (**R**)
- A la **ecuación** ajustada se le llama **modelo de regresión**
 - El ajuste se realiza mediante el método de los **mínimos cuadrados**
 - Se debe evaluar si el modelo es útil para describir la variable Y en función de X
- Es necesario evaluar la calidad del ajuste que presenta el modelo:
 - **Coeficiente de determinación r^2 o R^2** : cuánta variabilidad de los datos explica el modelo lineal
- Se debe comprobar que el modelo cumple unos supuestos:
 - Validación de supuestos sobre los residuos (errores): **normalidad, varianza constante**



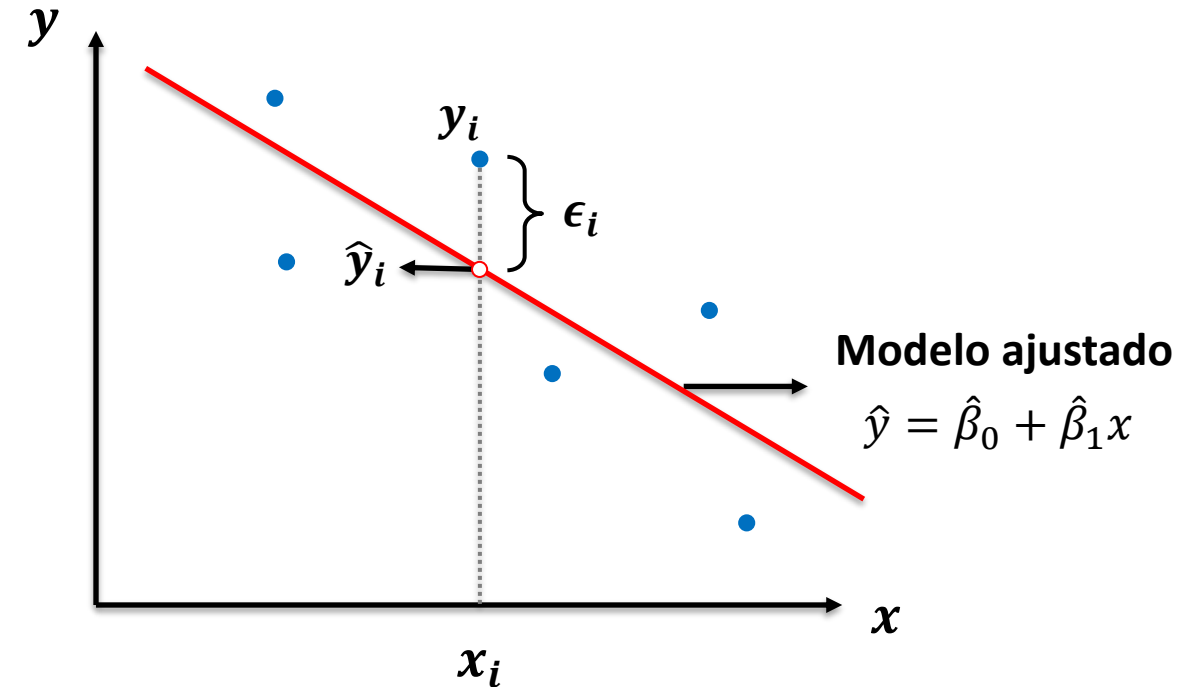
Regresión lineal simple

Descomposición de la variabilidad en sumas cuadráticas

Recordemos la siguiente idea...

Cada dato (y_i) es el resultado de:

valor ajustado (\hat{y}_i) + error aleatorio (ϵ_i)



Así, de manera análoga, podemos plantear lo siguiente...

La variabilidad de y es el resultado de:

Variabilidad explicada por el modelo + variabilidad explicada por el error aleatorio

Regresión lineal simple

Descomposición de la variabilidad en sumas cuadráticas

La variabilidad de y es el resultado de:

Variabilidad explicada por el modelo + variabilidad explicada por el error aleatorio

¿Cómo estimamos estas variabilidades?

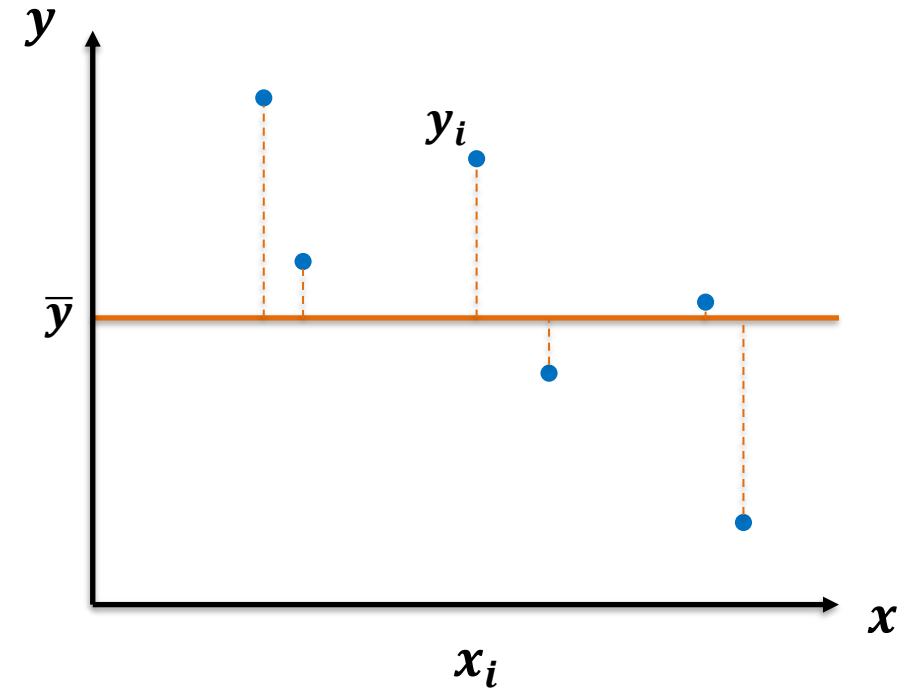
Nos vamos a apoyar de las sumas cuadráticas...

Estimación de la variabilidad de y

Primer paso - Hallar suma de cuadrados totales

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \textcolor{red}{SST}$$

Clave: recordar cómo se calcula la varianza muestral S^2



Regresión lineal simple

Descomposición de la variabilidad en sumas cuadráticas

La variabilidad de y es el resultado de:

Variabilidad explicada por el modelo + variabilidad explicada por el error aleatorio

Estimación de la variabilidad explicada por el modelo

Primer paso - Hallar suma de cuadrados debida a la regresión

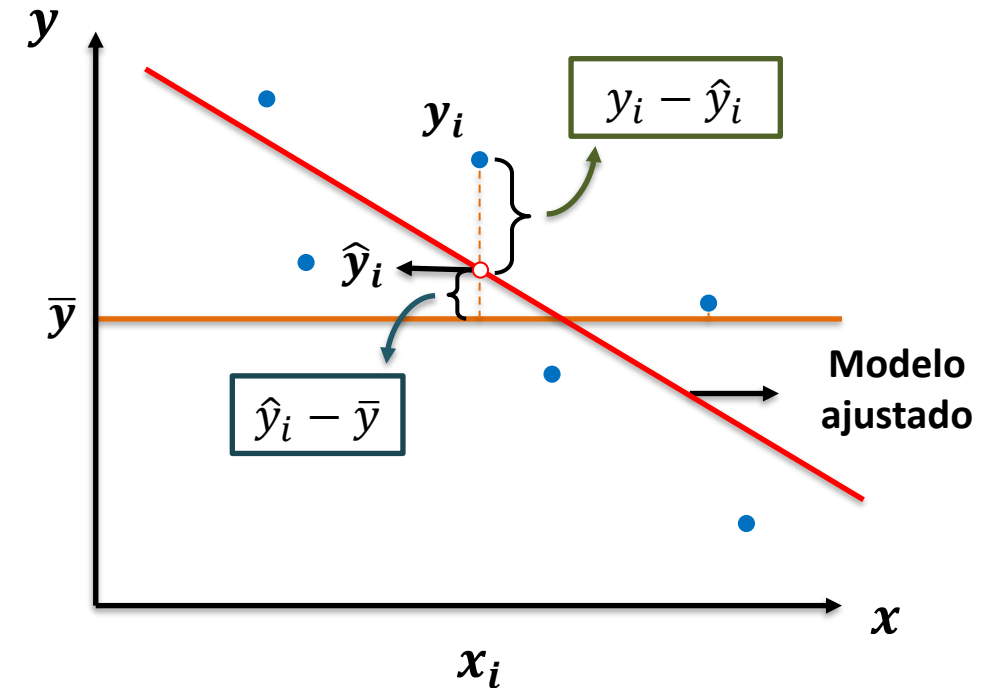
$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \text{SSR}$$

Estimación de la variabilidad explicada por el error aleatorio

(lo que el modelo no pudo describir)

Primer paso - Hallar suma de cuadrados debido al error

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \varepsilon_i^2 = \text{SSE}$$

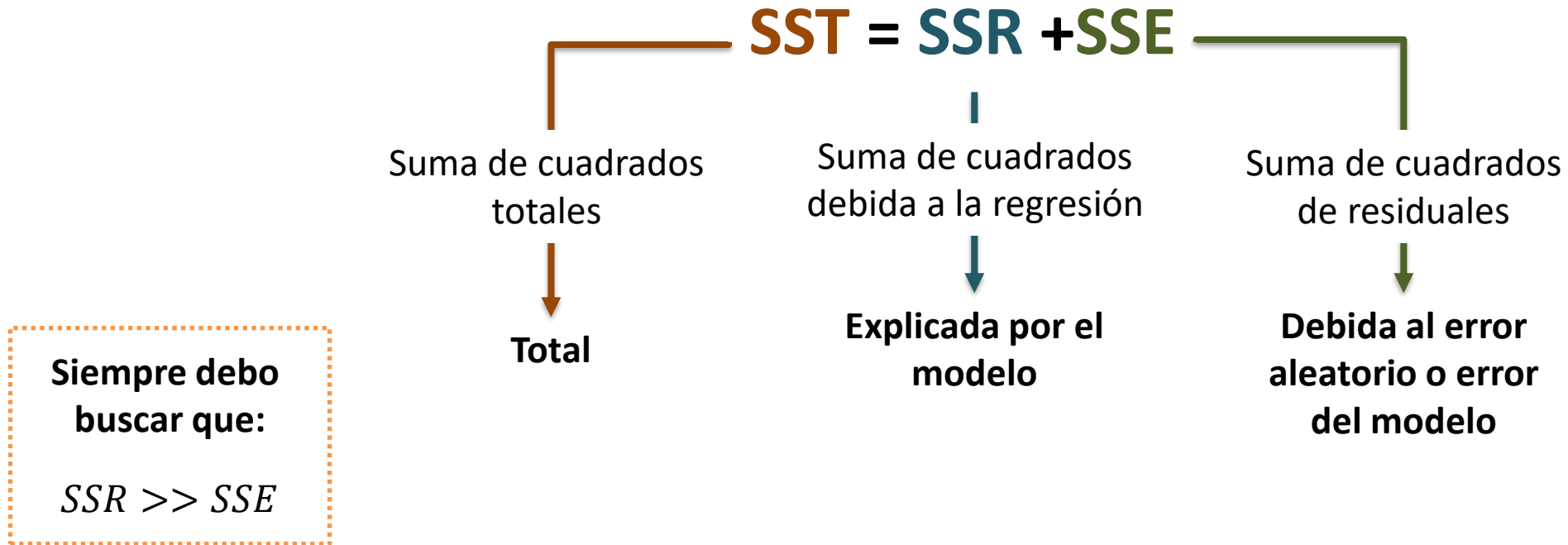


Regresión lineal simple

Descomposición de la variabilidad en sumas cuadráticas

La variabilidad de y es el resultado de:

Variabilidad explicada por el modelo + variabilidad explicada por el error aleatorio (error del modelo)

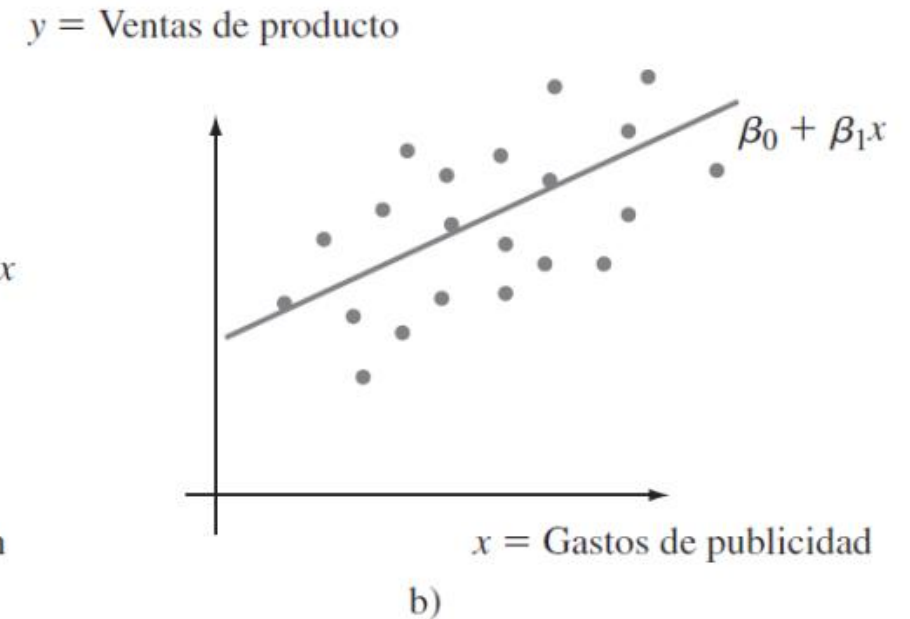
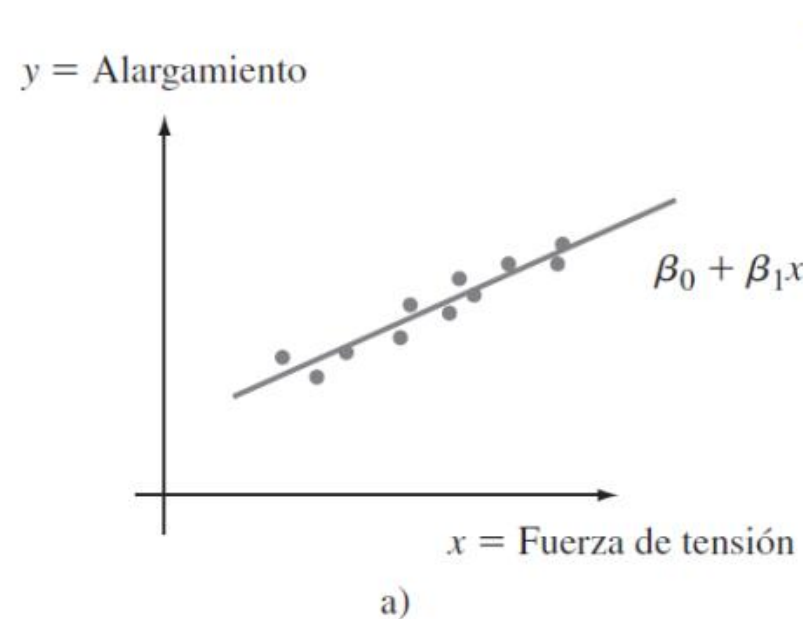


Regresión lineal simple

Varianza asociada a los errores del modelo



Errores
o Residuales



Varianza de los errores del modelo == σ^2

Desviación típica o estándar de los errores del modelo == σ

Regresión lineal simple

Estimación de σ^2 y σ



El estimador de σ^2 se puede hallar a partir del modelo ajustado, partiendo de la **suma de cuadrados de error (SSE)**:

$$SS_E = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Se puede demostrar que el valor esperado de SSE es $\rightarrow E(SS_E) = (n - 2)\sigma^2$

Por lo tanto, un **estimador insesgado de σ^2** es: $\sigma^2 = \frac{SS_E}{n - 2}$

Cuadrado medio
de los errores
MSE

Regresión lineal simple - Seguimiento

Estimación de σ^2 y σ



x	y
2	20
3	25
5	34
4	30
11	40
5	31

La **varianza estimada** de los errores del modelo es:

Por lo tanto, la **desviación estándar estimada** de los errores del modelo es:

↑
SSE

Regresión lineal simple

Ejemplo ganancias en R

> `summary(modelo)`  Comando en R para generar resumen del modelo

Coefficients:

	Estimate		Std. Error	t value	Pr(> t)	
(Intercept)	20.0000	← $\hat{\beta}_0$	2.6458	7.559	0.00164	**
inv	2.0000	← $\hat{\beta}_1$	0.4583	4.364	0.01202	*

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.24 on 4 degrees of freedom

Multiple R-squared: 0.8264, Adjusted R-squared: 0.7831

F-statistic: 19.05 on 1 and 4 DF, p-value: 0.01202

MUCHAS GRACIAS

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