

Regresión Múltiple

Alejandro Zubiri

Wed Dec 18 2024

Contents

1	Diagnosis del modelo	2
2	Transformaciones	2
3	Variables binarias	2
4	Contraste de parámetros	2

Vamos a crear una recta de la forma:

$$y = \beta_0 + \beta_1 x_{1i} + \dots + \beta_i x_{ik} + e_i \quad (1)$$

Que podemos expresar de forma matricial de la siguiente forma:

$$Y = X\beta + e \quad (2)$$

Obtendremos los coeficientes mediante:

$$\beta = (X^T X)^{-1} X^T Y \quad (3)$$

Para estimar la varianza σ^2 :

$$S_e^2 = \frac{\sum e^2}{n} \quad (4)$$

Que es la varianza residual. Para que sea **sesgado**, usamos:

$$S_R^2 = \frac{\sum e^2}{n - p} \quad (5)$$

Donde p es el número de parámetros beta. Se cumple que:

$$\begin{aligned} E(S_R^2) &= \sigma^2 \\ e &\rightarrow N(0, \sigma^2) \\ y &\rightarrow N(\beta_0 + \beta_1 x_{1i} + \dots, \sigma^2) \end{aligned} \quad (6)$$

El coeficiente de determinación corregido es aquel que funcione para muchos valores:

$$R^2 = 1 - \frac{S_R^2}{S_y^2} \quad (7)$$

1 Diagnóstico del modelo

Debemos comprobar las hipótesis del modelo:

- Linealidad: residuos frente a valores. La distribución debe ser aleatoria.
- Homocedasticidad: residuos frente a x , que debe ser aleatoria.
- Independencia.
- Normalidad.

2 Transformaciones

Son las vistas previamente. Solo se deben transformar las variables sin comportamiento lineal.

3 Variables binarias

Son aquellas que toman valores de 0 o 1, y representan ausencia o presencia de propiedad. Deben ser excluyentes.

4 Contraste de parámetros

Queremos ver si las variables son significativas, y existen tres métodos:

- Si $p < 0.05$, β es significativa.
- t -valor: si $|t| > 2$, β es significativa.
- Error estándar: calculamos un intervalo de confianza para β con $\alpha = 0.05$:

$$\beta \pm 2SE(\beta) \quad (8)$$

Si el 0 está incluido, β **NO** es significativa