

1. Ingredientes y datos

- $k \geq 1$ variables explicativas X_1, \dots, X_k .
- Variable respuesta Y .
- Serie de n datos (cada uno de longitud $k + 1$).
- $n \geq k + 2$.
- No colinealidad (columnas X_1 a X_k).

| X_1 | X_2 | \dots | X_k | Y |
|-----------|-----------|----------|-----------|----------|
| $x_{1,1}$ | $x_{1,2}$ | \dots | $x_{1,k}$ | y_1 |
| $x_{2,1}$ | $x_{2,2}$ | \dots | $x_{2,k}$ | y_2 |
| \vdots | \vdots | \ddots | \vdots | \vdots |
| $x_{n,1}$ | $x_{n,2}$ | \dots | $x_{n,k}$ | y_n |

2. Modelo

El vector $\mathbf{y} = (y_1, \dots, y_n)^\top$ es una realización del vector aleatorio $\mathbb{Y} = (Y_1, \dots, Y_n)^\top$ dado por

$$\mathbb{Y} = X \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

donde

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,k} \end{pmatrix}$$

es la matriz de diseño (de rango $k + 1$), y donde

- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$ es el vector de parámetros,
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$ tiene vector de medias $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ y matriz de covarianzas $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$, donde σ^2 es otro parámetro e I_n es la matriz identidad $n \times n$.
- (Hipótesis de normalidad). Para buena parte de lo que sigue (intervalos de confianza, contrastes de hipótesis), supondremos que $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$. Es decir, las variables ε_i son normales independientes de media 0 y varianza σ^2 .

Bajo esta hipótesis de normalidad, el vector \mathbb{Y} se distribuye como una $\mathcal{N}(X \cdot \boldsymbol{\beta}, \sigma^2 I_n)$.

3. Estimación de parámetros

a) Dada la muestra $\mathbf{y} = (y_1, \dots, y_n)^\top$, la estimación (por mínimos cuadrados, y también por máxima verosimilitud, si hay normalidad) de los parámetros es

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y}.$$

Para el caso de la regresión lineal simple ($k = 1$), llamando $\mathbf{x} = (x_1, \dots, x_n)^\top$ a la (única) columna de observaciones,

$$\hat{\beta}_1 = \frac{\text{cov}_{\mathbf{x}, \mathbf{y}}}{V_{\mathbf{x}}}, \quad \hat{\beta}_0 = \bar{y} - \frac{\text{cov}_{\mathbf{x}, \mathbf{y}}}{V_{\mathbf{x}}} \bar{x},$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad V_{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{cov}_{\mathbf{x}, \mathbf{y}} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

b) Valores pronosticados y residuos. Dada la muestra $\mathbf{y} = (y_1, \dots, y_n)^\top$, los pronósticos $\hat{\mathbf{y}} = (\hat{y}_1, \dots, \hat{y}_n)^\top$ y los residuos $\mathbf{e} = (e_1, \dots, e_n)^\top$ son

$$\begin{aligned}\hat{\mathbf{y}} &= X\hat{\boldsymbol{\beta}} = X(X^\top X)^{-1}X^\top \mathbf{y} := H\mathbf{y}, \\ \mathbf{e} &= \mathbf{y} - \hat{\mathbf{y}} = (I_n - H)\mathbf{y}.\end{aligned}$$

La matriz H es $n \times n$, simétrica, definida positiva e idempotente de rango $k + 1$. La matriz $I_n - H$ es asimismo simétrica, definida positiva e idempotente, pero de rango $n - k - 1$.

c) Sumas de cuadrados: TSS = MSS + RSS, con

$$\begin{aligned}(\text{total}) \quad \text{TSS} &= \sum_{i=1}^n (y_i - \bar{y})^2 = nV_{\mathbf{y}} = \mathbf{y}^\top (I_n - \tfrac{1}{n}J_n) \mathbf{y}, \\ (\text{explicada por modelo}) \quad \text{MSS} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \mathbf{y}^\top (H - \tfrac{1}{n}J_n) \mathbf{y}, \\ (\text{residual}) \quad \text{RSS} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2 = \mathbf{y}^\top (I_n - H) \mathbf{y},\end{aligned}$$

donde J_n denota la matriz $n \times n$ con unos.

d) Estimación para σ^2 :

$$\widehat{\sigma^2} = s_R^2 = \frac{1}{n - k - 1} \sum_{i=1}^n e_i^2 = \frac{\text{RSS}}{n - k - 1}.$$

e) Coeficiente R^2 :

$$R^2 = \frac{\text{MSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}.$$

Obsérvese que $\text{MSS}/\text{RSS} = R^2/(1 - R^2)$.

4. Propiedades de los estimadores

Consideramos los estimadores (estadísticos asociados a $\mathbb{Y} = (Y_1, \dots, Y_n)^\top$)

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top \mathbb{Y} \quad \text{y} \quad s_R^2 = \frac{1}{n - k - 1} \mathbb{Y}^\top (I_n - H) \mathbb{Y}.$$

En el caso $k = 1$, llamando $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$,

$$\hat{\beta}_1 = \frac{1}{V_{\mathbf{x}}} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}), \quad \hat{\beta}_0 = \bar{Y} - \frac{\bar{x}}{V_{\mathbf{x}}} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}).$$

Se consideran también los estadísticos:

- vector de pronósticos $\hat{\mathbb{Y}} = H\mathbb{Y}$,
- vector de residuos $\mathbf{e} = (I_n - H)\mathbb{Y} = (I_n - H)\boldsymbol{\varepsilon}$.

Medias y varianzas/covarianzas.

El vector $\hat{\boldsymbol{\beta}}$ tiene

- vector de medias $\mathbb{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ (estimadores insesgados),
- y matriz de covarianzas $\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(X^\top X)^{-1}$.

En el caso $k = 1$,

$$\mathbf{V}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{nV_{\mathbf{x}}} \right], \quad \mathbf{V}(\hat{\beta}_1) = \sigma^2 \frac{1}{nV_{\mathbf{x}}}, \quad \text{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\bar{x}}{nV_{\mathbf{x}}}.$$

Por su parte, $\mathbb{E}(s_R^2) = \sigma^2$ (también estimador insesgado). La variable aleatoria $(n-k-1)s_R^2/\sigma^2$ tiene media $(n-k-1)$ y varianza $2(n-k-1)$.

Distribución bajo normalidad.

Si suponemos que $\mathbb{Y} \sim \mathcal{N}(X\beta, \sigma^2 I_n)$, entonces

- $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 (X^\top X)^{-1})$,
- $(n-k-1)s_R^2/\sigma^2 \sim \chi_{n-k-1}^2$,
- y s_R^2 es independiente de $\hat{\beta}$.

En particular, para $j = 0, \dots, k$, y llamando q_{jj} al elemento que está en la posición $j+1$ de la diagonal de la matriz $(X^\top X)^{-1}$ (que tiene dimensiones $(k+1) \times (k+1)$),

$$\frac{\hat{\beta}_j - \beta_j}{s_R \sqrt{q_{jj}}} \sim t_{n-k-1}.$$

(Nota: todo lo que sigue es válido bajo la hipótesis de normalidad).

5. Intervalos de confianza para los parámetros

Dado α , y para $j = 0, \dots, k$,

$$\text{IC}_{1-\alpha}(\beta_j) = \hat{\beta}_j \pm t_{\{n-k-1; \alpha/2\}} s_R \sqrt{q_{jj}}.$$

Para el caso $k = 1$,

$$\text{IC}_{1-\alpha}(\beta_0) = \hat{\beta}_0 \pm t_{\{n-2; \alpha/2\}} s_R \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{nV_{\mathbf{x}}}}, \quad \text{IC}_{1-\alpha}(\beta_1) = \hat{\beta}_1 \pm t_{\{n-2; \alpha/2\}} s_R \sqrt{\frac{1}{nV_{\mathbf{x}}}}.$$

Para σ^2 ,

$$\text{IC}_{1-\alpha}(\sigma^2) = \left(\frac{(n-k-1)s_R^2}{\chi_{\{n-k-1; \alpha/2\}}^2}, \frac{(n-k-1)s_R^2}{\chi_{\{n-k-1; 1-\alpha/2\}}^2} \right).$$

6. Contrastes de hipótesis

a) Hipótesis individuales $H_0 : \beta_j = 0$, con $j \in \{1, \dots, k\}$.

Región de rechazo con nivel de significación α :

$$\mathcal{R}_j = \left\{ \left| \frac{\hat{\beta}_j}{s_R \sqrt{q_{jj}}} \right| > t_{\{n-k-1; \alpha/2\}} \right\}.$$

b) Hipótesis global $H_0 : \beta_1 = \dots = \beta_k = 0$.

Bajo H_0 , se tiene que

$$\frac{\text{MSS}/k}{\text{RSS}/(n-k-1)} \sim F_{k, n-k-1}.$$

Región de rechazo con nivel de significación α :

$$\mathcal{R} = \left\{ \frac{\text{MSS}/k}{\text{RSS}/(n-k-1)} > F_{\{k, n-k-1; \alpha\}} \right\}.$$

Tabla ANOVA:

| Fuente | suma cuadrados | g.l. | varianza | estadístico F |
|-------------------------|----------------|---------|------------------------------|------------------------|
| explicada por regresión | MSS | k | MSS/k | $(\text{MSS}/k)/s_R^2$ |
| residual | RSS | $n-k-1$ | $\text{RSS}/(n-k-1) = s_R^2$ | |
| total | TSS | $n-1$ | | |

7. Predicciones

Condicionando sobre una observación $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,k})^\top$, y si llamamos $\tilde{\mathbf{x}}_0 = (1, x_{0,1}, \dots, x_{0,k})^\top$, la predicción, tanto sobre la media de Y como sobre el valor de Y , es

$$\hat{y}_0 = \tilde{\mathbf{x}}_0^\top \cdot \hat{\beta}.$$

Intervalos de confianza:

$$\begin{aligned} \text{IC}_{1-\alpha}(\text{media de } Y \mid \mathbf{x}_0) &= \hat{y}_0 \pm t_{\{n-k-1; \alpha/2\}} \cdot s_R \cdot \sqrt{\tilde{\mathbf{x}}_0^\top (X^\top X)^{-1} \tilde{\mathbf{x}}_0} \\ \text{IC}_{1-\alpha}(\text{valor de } Y \mid \mathbf{x}_0) &= \hat{y}_0 \pm t_{\{n-k-1; \alpha/2\}} \cdot s_R \cdot \sqrt{1 + \tilde{\mathbf{x}}_0^\top (X^\top X)^{-1} \tilde{\mathbf{x}}_0} \end{aligned}$$

En el caso $k = 1$, dada la observación x_0 ,

$$\begin{aligned} \text{IC}_{1-\alpha}(\text{media de } Y \mid x_0) &= \hat{y}_0 \pm t_{\{n-2; \alpha/2\}} \cdot s_R \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n V_{\mathbf{x}}}}, \\ \text{IC}_{1-\alpha}(\text{valor de } Y \mid x_0) &= \hat{y}_0 \pm t_{\{n-2; \alpha/2\}} \cdot s_R \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n V_{\mathbf{x}}}}. \end{aligned}$$