Modelo de regresión múltiple

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K + \varepsilon$$

K = número de variables

De forma extendida:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_K X_{iK} + \varepsilon_i$$

$i = 1, 2, \ldots, n \qquad n$ = número de observaciones

Matricialmente

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1K} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2K} \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nK} \end{bmatrix}$$

Así

$$Y = X\beta + \varepsilon$$

Definamos el modelo ajustado

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}$$

y el error

$$\hat{u}_i = y_i - \hat{y}_i ; \boxed{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{u}_i = S}$$

Se puede ver como

$$S = \hat{u}^t \hat{u} = [\hat{u}_1 \ \hat{u}_2 \ \cdots \ \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

Recordando que el modelo ajustado tambien se ve como:

$$\hat{y} = X\hat{\beta}$$

Entonces

$$\begin{aligned} S &= (y - X\hat{\beta})^t (y - X\hat{\beta}) \\ &= y^t y - \hat{\beta}^t x^t y - y^t X\hat{\beta} + \hat{\beta}^t x^t x \hat{\beta} \\ &= y^t y - 2\hat{\beta}^t x^t y + \hat{\beta}^t x^t x \hat{\beta} \end{aligned}$$

De cálculo

$$\min S \Rightarrow \frac{\partial S}{\partial \hat{\beta}} = -2x^t y + 2x^t x\hat{\beta} = 0$$

Por lo tanto debe solucionarse $x^t x \hat{\beta} = x^t y$

Recordemos

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

Cuyo modelo ajustar es

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}$$

Sea

$$\hat{u}_i = y_i - \hat{y}_i \Rightarrow \min \sum_{i=1}^{\hat{n}} \hat{u}_i^2 \text{ ó } \min \sum_{i=1}^{\hat{n}} SEC$$

$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{\hat{n}} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}))^2$$

De esta forma minimizando

$$\frac{\partial SEC}{\partial \hat{\beta}_j} = 0 \qquad j\ 1, 2, \dots, k$$

$$\frac{\partial SEC}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}))(-1)$$

$$\frac{\partial SEC}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{\hat{n}} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}))(-X_{i1})$$

$$\frac{\partial SEC}{\partial \hat{\beta}_2} = 2 \sum_{i=1}^{\hat{n}} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}))(-X_{i2})$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial SEC}{\partial \hat{\beta}_k} = 2 \sum_{i=1}^{\hat{n}} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}))(-X_{ik})$$

ahora

$$\frac{\partial SEC}{\partial \beta_j} = 0 \qquad j = 0, 1, 2, \dots, k$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}\hat{\beta}_1 + \sum_{i=1}^{n} x_{i2}\hat{\beta}_2 + \sum_{i=1}^{n} x_{i3}\hat{\beta}_3 + \dots + \sum_{i=1}^{n} x_{ik}\hat{\beta}_k$$

$$\sum_{i=1}^{n} x_{i1} y_i = \sum_{i=1}^{n} x_{i1}\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}^2\hat{\beta}_1 + \sum_{i=1}^{n} x_{i2}x_{i1}\hat{\beta}_2 + \sum_{i=1}^{n} x_{i3}x_{i1}\hat{\beta}_3 + \dots + \sum_{i=1}^{n} x_{ik}x_{i1}\hat{\beta}_k$$

$$\sum_{i=1}^{n} x_{i2} y_i = \sum_{i=1}^{n} x_{i2}\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}x_{i2}\hat{\beta}_1 + \sum_{i=1}^{n} x_{i2}^2\hat{\beta}_2 + \sum_{i=1}^{n} x_{i3}x_{i2}\hat{\beta}_3 + \dots + \sum_{i=1}^{n} x_{ik}x_{i2}\hat{\beta}_k$$

$$\sum_{i=1}^{n} x_{i3} y_i = \sum_{i=1}^{n} x_{i3}\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}x_{i3}\hat{\beta}_1 + \sum_{i=1}^{n} x_{i2}x_{i3}\hat{\beta}_2 + \sum_{i=1}^{n} x_{i3}^2\hat{\beta}_3 + \dots + \sum_{i=1}^{n} x_{ik}x_{i3}\hat{\beta}_k$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{ik} y_i = \sum_{i=1}^{n} x_{ik}\hat{\beta}_0 + \sum_{i=1}^{n} x_{i1}x_{ik}\hat{\beta}_1 + \sum_{i=1}^{n} x_{i2}x_{ik}\hat{\beta}_2 + \sum_{i=1}^{n} x_{i3}x_{ik}\hat{\beta}_3 + \dots + \sum_{i=1}^{n} x_{ik}^2\hat{\beta}_k$$

Notemos

$$\underbrace{\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ \sum_{i=1}^{n} x_{i2}y_i \\ \sum_{i=1}^{n} x_{i3}y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik}y_i \end{bmatrix}}_{X^t Y} = \underbrace{\begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i3} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i3}x_{i1} & \cdots & \sum_{i=1}^{n} x_{ik}x_{i1} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i1}x_{i2} & \sum_{i=1}^{n} x_{i2}^2 & \sum_{i=1}^{n} x_{i3}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik}x_{i2} \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1}x_{ik} & \sum_{i=1}^{n} x_{i2}x_{ik} & \sum_{i=1}^{n} x_{i3}x_{ik} & & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix}}_{X^t X} \underbrace{\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}}_{\hat{\beta}}$$

Por propiedades de matrices

$$[x^t x \hat{\beta} = x^t y](x^t x)^{-1}$$

$$\hat{\beta} = (x^t x)^{-1} x^t y$$

Demostrando que $E[\hat{\beta}] = \beta$

$$E[\hat{\beta}] = E[(x^t x)^{-1} x^t y]$$

$$= E[(x^t x)^{-1} x^t [x\beta + \varepsilon]]$$

$$= E[\underbrace{(x^t x^{-1})\, x^t x}_{I\text{: matriz identidad}} \beta + (x^t x)^{-1} x^t \varepsilon]$$

$$= E[I\beta + (x^t x)^{-1} x^t \varepsilon]$$

$$= E[I\beta] + E[(x^t x)^{-1} x^t \varepsilon]$$

$$= E[\beta] + (x^t x)^{-1} x^t \cancel{E[\varepsilon]}^{0} \; ; \; E[\varepsilon] = 0$$

β = constantes

$$E[\hat{\beta}] = \beta$$

Para probar $VAR(\hat{\beta})$. Recordemos que, para una variable aleatoria

$$VAR(X) = E[(X - E[X])^2],$$

$$E[\hat{\beta}] = \beta$$

$$VAR(\hat{\beta}) = E[(\hat{\beta} - \beta)^2]$$

$$= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)]$$

Recordemos

$$\hat{\beta} = (X^t X)^{-1} X^t Y$$

$$\hat{\beta} = (X^t X)^{-1} X^t [X\beta + \varepsilon]$$

$$\hat{\beta} = \underbrace{(X^t X)^{-1} X^t X}_{I} \beta + (X^t X)^{-1} X^t \varepsilon$$

$$\hat{\beta} = \beta + (X^t X)^{-1} X^t \varepsilon$$

$\hat{\beta} - \beta = (X^t X)^{-1} X^t \varepsilon$, de este modo

$$VAR(\hat{\beta}) = E[[(X^t X)^{-1} X^t \varepsilon][(X^t X)^{-1} X^t \varepsilon]^t]$$

$$= E[\underbrace{(X^t X)^{-1} X^t X}_{I} \varepsilon\varepsilon^t (X^t X)^{-1}]$$

$$= E[\; I \; \varepsilon\varepsilon^t (X^t X)^{-1}]$$

$$VAR(\hat{\beta}) = (X^t X)^{-1} E(\varepsilon\varepsilon^t)$$

Recordemos

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$E[\varepsilon\varepsilon^t] = E\left[\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}[\varepsilon_1 \ \varepsilon_2 \cdots \ \varepsilon_n]\right]$$

$$= E\begin{bmatrix} \varepsilon_1\varepsilon_1 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2\varepsilon_2 & \cdot & \cdot & \varepsilon_2\varepsilon_n \\ \varepsilon_3\varepsilon_1 & \vdots & \varepsilon_{33} & \ddots & \vdots \\ \vdots & & & & \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \cdots & \varepsilon_n\varepsilon_n \end{bmatrix}$$

Si las variables son independientes e identicamente distribuidas

$E[\varepsilon_i\varepsilon_j] = 0$ para toda $i \neq j$, de esta forma

$E[\varepsilon] = 0$ y $VAR[\varepsilon] = E[(\varepsilon - 0)^2] = E[\varepsilon^2] = \sigma^2$

$$E[\varepsilon\varepsilon^t] = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdot & \cdots & 0 \\ 0 & \vdots & \sigma^2 & & \vdots \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \sigma^2 \end{bmatrix} = \sigma^2$$

Por lo tanto

$$\underline{VAR(\hat{\beta}) = \sigma^2(X^tX)^{-1}}$$

Supuestos de regresión lineal múltiple

1) Lineal en los parámetros: el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

donde $\beta_0, \beta_1, \ldots \beta_k$ son los parametros (constantes) desconocidos de interes y u es un error aleatorio no observable o término de perturbación

2) Muestreo aleatorio

Se tiene una muestra aleatoria de n observaciones $\{(x_{i1}, x_{i2}, \ldots, x_{in}, y_i) : i=1,2,\ldots n\}$ que satisface el primer supuesto. Nos dice ademas que se tienen datos que pueden emplearse para estimar las β_i y que estos datos han sido elegidos de manera que sean representativos de la población.

3) Colinealidad no perfecta.

En la muestra ninguna de las variables independientes es constante y no hay relación lineal constante entre las variables independientes

Si en cada variable independiente hay variaciones muestrales y no existe una relación lineal exacta entre las variables independientes, β_j se calcula (se puede)

4) Media condicional cero

El error u tiene un valor esperado de cero. Es decir,

$$E[u \mid x_1, x_2, \ldots, x_k] = 0.$$

Si se cumple este supuesto los estimadores serán insesgados es decir

$$E[\hat{\beta}] = \beta$$

5) Homocedasticidad

Para cualesquiera valores de las variables explicativas, el error u tiene la misma varianza

$$VAR(u \mid x_1, x_2, \dots, x_k) = \sigma^2$$

Este supuesto no tiene relación con el insesgamiento de $\hat{\beta}_j$, sin embargo

1) permite obtener formulas para la varianza de muestreo cuyos componentes son faciles de caracterizar
2) Bajo los supuestos de Gauss-Markov los estimadores son los que tienen menor varianza entre todos los estimadores lineales insesgados.

6) Normalidad

El error poblacional u es independiente de las variables explicativas $x_1, x_2, \dots, x_k$ y esta distribuido normalmente con media cero y varianza σ^2: es decir

$$u \sim Normal(0, \sigma^2)$$

Para la matriz de varianza covarianza, se deben cumplir.

$E[\varepsilon_j] = 0 \rightarrow$ media cero para toda $j = 1, n, \dots, i$

$E[\varepsilon_i \varepsilon_j] = 0 \rightarrow$ para toda $i \neq j$

$VAR[\varepsilon_j] = \sigma^2 \Rightarrow E[(\varepsilon - E[\varepsilon])^2] = E[(\varepsilon - 0)^2] = E[\varepsilon^2] = \sigma^2$

Si se cumplen estos supuestos los estimadores son insesgados y eficientes