

MAT-266: Modelos lineales

Felipe Osorio

`fosorios.mat.utfsm.cl`

Departamento de Matemática, UTFSM



Definición 1 (modelo lineal):

Considere la variable aleatoria Y , decimos que sigue un **modelo lineal** si

$$E(Y) = \sum_{j=1}^p x_j \beta_j = \mathbf{x}^\top \boldsymbol{\beta},$$

donde $\mathbf{x} = (x_1, \dots, x_p)^\top$ representa un vector de p **variables regresoras**, mientras que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ denota un vector de **parámetros desconocidos**.

Suponga que tenemos n observaciones recolectadas desde Y , entonces podemos considerar el modelo

$$E(Y_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \dots, n.$$

Es conveniente escribir lo anterior como:

$$E(\mathbf{Y}) = \begin{pmatrix} \mathbf{x}_1^\top \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_n^\top \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \boldsymbol{\beta} = \mathbf{X} \boldsymbol{\beta},$$

donde $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n \times p}$ se denomina **matriz de diseño**.



Observación:

Cuando

$$\text{Cov}(\mathbf{Y}) = \sum_{r=1}^K \phi_r \mathbf{\Sigma}_r = \mathbf{\Sigma}(\boldsymbol{\phi}),$$

donde $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)^\top$ son parámetros desconocidos y $\mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_K$ son matrices (simétricas) conocidas, decimos que \mathbf{Y} sigue un **modelo lineal general**.

Típicamente, asumiremos que Y_1, \dots, Y_n son independientes con varianza constante, en cuyo caso

$$\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I},$$

y decimos que \mathbf{Y} sigue un **modelo lineal simple**.



Definición 2:

El vector \mathbf{Y} se dice seguir un **modelo lineal** si

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \quad \text{Cov}(\mathbf{Y}) = \sum_{r=1}^K \phi_r \boldsymbol{\Sigma}_r,$$

o equivalentemente

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

con

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\epsilon}) = \sum_{r=1}^K \phi_r \boldsymbol{\Sigma}_r.$$



Observación:

Estos supuestos de momentos, suelen ser llamados **condiciones de Gauss-Markov**. Aunque es usual que sean expresados en términos del modelo lineal simple, como

$$Y = X\beta + \epsilon,$$

con

$$E(\epsilon) = 0, \quad \text{Cov}(\epsilon) = \sigma^2 I_n.$$



Definición 3:

Se dice que el vector \mathbf{Y} sigue un **modelo lineal normal** si $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\phi))$, o bien

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma}(\phi)).$$

Mientras que sigue un **modelo normal simple**, si

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}). \quad (1)$$

Observación:

Se debe destacar que el modelo en (1) puede ser escrito como

$$Y_i \stackrel{\text{ind}}{\sim} N_1(\mathbf{x}_i^\top \boldsymbol{\beta}, \sigma^2), \quad i = 1, \dots, n. \quad (2)$$

En efecto, la inferencia estadística para los modelos definidos en Ecuaciones (1) y (2) es equivalente.¹

¹Aunque esto **NO** es verdad en general.



Supuestos (Modelo lineal normal simple):

El modelo lineal descrito por la ecuación:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

tiene los siguientes supuestos:

- A1:** \mathbf{X} es una matriz no aleatoria $n \times p$ con $n > p$.
- A2:** La matriz \mathbf{X} tiene rango p , es decir, \mathbf{X} es rango columna completo.
- A3:** El vector aleatorio n -dimensional \mathbf{Y} tiene elementos que son observables.
- A4:** El vector aleatorio no observable $\boldsymbol{\epsilon}$ satisface

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}, \quad \sigma^2 > 0.$$

- A4*:** El vector aleatorio $\boldsymbol{\epsilon}$ satisface $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, o equivalentemente

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$$



Inferencia en el modelo lineal

Considere la función de verosimilitud proveniente del modelo $\mathbf{Y} \sim \mathbf{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$, dada por

$$L(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2\right),$$

con $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \sigma^2)^\top$. De este modo, podemos llevar a cabo la **estimación ML**, considerando

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 \\ &= -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} Q(\boldsymbol{\beta}),\end{aligned}$$

donde

$$Q(\boldsymbol{\beta}) = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \sum_{i=1}^n (Y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2,$$

se denomina **suma de cuadrados del error**.



Diferenciando con relación a β y σ^2 , obtenemos

$$d_{\beta} \ell(\theta) = -\frac{1}{2\sigma^2} d_{\beta} \|Y - X\beta\|^2 = \frac{1}{\sigma^2} (Y - X\beta)^{\top} X d\beta,$$

y

$$d_{\sigma^2} \ell(\theta) = -\frac{n}{2\sigma^2} d\sigma^2 + \frac{1}{2\sigma^4} \|Y - X\beta\|^2 d\sigma^2.$$

Es decir,²

$$\frac{\partial \ell(\theta)}{\partial \beta} = \frac{1}{\sigma^2} X^{\top} (Y - X\beta),$$

$$\frac{\partial \ell(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \|Y - X\beta\|^2.$$

Desde la condición de primer orden, tenemos las [ecuaciones de verosimilitud](#):

$$X^{\top} (Y - X\hat{\beta}) = 0$$

$$n\hat{\sigma}^2 - \|Y - X\hat{\beta}\|^2 = 0.$$

²usando el 1er Teorema de identificación dado en Magnus y Neudecker, 2007, pag. 98

Resolviendo con relación a β obtenemos las **ecuaciones normales**

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{Y},$$

dado que $\text{rg}(\mathbf{X}) = \text{rg}(\mathbf{X}^\top \mathbf{X}) = p$, entonces el sistema anterior admite solución única dada por:³

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}, \\ \hat{\sigma}^2 &= \frac{1}{n} Q(\hat{\beta}) = \frac{1}{n} \|\mathbf{Y} - \mathbf{X} \hat{\beta}\|^2.\end{aligned}$$

Se define el vector de **valores predichos** como

$$\hat{\mathbf{Y}} = \mathbf{X} \hat{\beta} = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{H} \mathbf{Y},$$

donde $\mathbf{H} = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$.

³Note que $\hat{\beta} = \mathbf{X}^+ \mathbf{Y}$ con $\mathbf{X}^+ = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ la inversa Moore-Penrose de \mathbf{X} .

Es fácil notar que \mathbf{H} es simétrica e idempotente, en cuyo caso

$$\text{rg}(\mathbf{H}) = \text{tr}(\mathbf{H}) = \text{tr}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) = \text{tr}((\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X}) = p.$$

Además, el vector de diferencias entre \mathbf{Y} y $\hat{\mathbf{Y}}$ se denomina el **vector de residuos**, es decir

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

Además, tenemos que $\mathbf{I} - \mathbf{H}$ también es simétrica e idempotente. Con esta notación podemos escribir

$$\begin{aligned} Q(\hat{\boldsymbol{\beta}}) &= \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = \mathbf{e}^\top \mathbf{e} = \mathbf{Y}^\top (\mathbf{I} - \mathbf{H})\mathbf{Y} \\ &= \sum_{i=1}^n (Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}})^2, \end{aligned}$$

que es conocido como **suma de cuadrados residual**.



Resultado 1:

Considere el modelo:

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n),$$

con los supuestos A1 a A4*. Entonces, tenemos que:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \sim N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}), \quad (3)$$

de donde sigue que $\hat{\boldsymbol{\beta}}$ es insesgado y $\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$. Además, la variable aleatoria

$$\frac{Q(\hat{\boldsymbol{\beta}})}{\sigma^2} = \frac{\mathbf{Y}^\top (\mathbf{I} - \mathbf{H}) \mathbf{Y}}{\sigma^2} \sim \chi^2(n - p).$$

Finalmente $\hat{\boldsymbol{\beta}}$ y $s^2 = Q(\hat{\boldsymbol{\beta}})/(n - p)$ son independientes.



Demostración:

Notando que $\hat{\beta}$ es una transformación lineal de un vector aleatorio normal y

$$E(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\mathbf{Y}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} \beta = \beta,$$

$$\text{Cov}(\hat{\beta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \text{Cov}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1},$$

y el resultado en (3) sigue. Por otro lado,

$$\frac{Q(\hat{\beta})}{\sigma^2} = \frac{(n-p)s^2}{\sigma^2} = \frac{\mathbf{Y}^\top (\mathbf{I} - \mathbf{H}) \mathbf{Y}}{\sigma^2},$$

sigue una distribución chi-cuadrado pues $\mathbf{I} - \mathbf{H}$ es matriz idempotente con

$$\text{rg}(\mathbf{I} - \mathbf{H}) = \text{tr}(\mathbf{I} - \mathbf{H}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{H}) = n - p.$$

Además, el parámetro de no centralidad es dado por

$$\lambda = \frac{1}{2\sigma^2} \beta^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{H}) \mathbf{X} \beta = 0.$$



En efecto,

$$\mathbf{H}\mathbf{X} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{X} = \mathbf{X} \Rightarrow (\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}.$$

La independencia entre $\hat{\beta}$ y s^2 sigue desde $(\mathbf{I} - \mathbf{H})\mathbf{X} = \mathbf{0}$

Observación:

Es fácil notar que

$$\begin{aligned} \mathbb{E}\{Q(\hat{\beta})\} &= \mathbb{E}\{\mathbf{Y}^\top (\mathbf{I} - \mathbf{H})\mathbf{Y}\} = \sigma^2 \operatorname{tr}(\mathbf{I} - \mathbf{H}) + \beta^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{H})\mathbf{X}\beta \\ &= \sigma^2(n - p), \end{aligned}$$

es decir,

$$\mathbb{E}(\hat{\sigma}^2) = \left(\frac{n - p}{n}\right)\sigma^2,$$

lo que permite sugerir el **estimador insesgado**:

$$s^2 = \frac{1}{n - p} \|\mathbf{Y} - \mathbf{X}\hat{\beta}\|^2.$$



Resultado 2:

El vector de valores predichos y el vector de residuos son independientemente distribuidos como

$$\hat{Y} \sim N_n(X\beta, \sigma^2 H), \quad e \sim N_n(0, \sigma^2(I - H)).$$

Demostración:

Considere

$$\begin{pmatrix} \hat{Y} \\ e \end{pmatrix} = \begin{pmatrix} H \\ I - H \end{pmatrix} Y,$$

luego,

$$E \begin{pmatrix} \hat{Y} \\ e \end{pmatrix} = \begin{pmatrix} H \\ I - H \end{pmatrix} X\beta = \begin{pmatrix} HX\beta \\ 0 \end{pmatrix} = \begin{pmatrix} X\beta \\ 0 \end{pmatrix}.$$



Además,

$$\begin{aligned}\text{Cov} \begin{pmatrix} \hat{\mathbf{Y}} \\ \mathbf{e} \end{pmatrix} &= \sigma^2 \begin{pmatrix} \mathbf{H} \\ \mathbf{I} - \mathbf{H} \end{pmatrix} (\mathbf{H}^\top, (\mathbf{I} - \mathbf{H})^\top) \\ &= \sigma^2 \begin{pmatrix} \mathbf{H}\mathbf{H}^\top & \mathbf{H}(\mathbf{I} - \mathbf{H})^\top \\ (\mathbf{I} - \mathbf{H})\mathbf{H}^\top & (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})^\top \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} \mathbf{H}^2 & \mathbf{0} \\ \mathbf{0} & (\mathbf{I} - \mathbf{H})^2 \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{H} \end{pmatrix},\end{aligned}$$

lo que permite establecer el resultado.

Observación:

Debemos resaltar que \mathbf{H} y $\mathbf{I} - \mathbf{H}$ son matrices de rango incompleto y por tanto $\hat{\mathbf{Y}}$ y \mathbf{e} siguen distribuciones normales **singulares**.



Considere el siguiente problema

$$\min_{\beta} Q_n(\beta) = \min_{\beta} \|Y_n - X_n \beta\|^2,$$

cuya solución, $\hat{\beta}_n$ es llamado estimador **mínimos cuadrados** ordinarios (**OLS**), dado por:

$$\hat{\beta}_n = (X_n^\top X_n)^{-1} X_n^\top Y_n,$$

Es habitual asumir el modelo

$$Y_n = X_n \beta + \epsilon_n,$$

con $E(\epsilon_n) = \mathbf{0}$ y $\text{Cov}(\epsilon_n) = \sigma^2 \mathbf{I}_n$. Esto lleva a

$$E(\hat{\beta}_n) = (X_n^\top X_n)^{-1} X_n^\top E(X_n \beta + \epsilon_n) = \beta$$

$$\text{Cov}(\hat{\beta}_n) = \text{Cov}((X_n^\top X_n)^{-1} X_n^\top (X_n \beta + \epsilon_n)) = \sigma^2 (X_n^\top X_n)^{-1}.$$



Supuestos (Estimación mínimos cuadrados):

Considere el modelo,

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\epsilon}_n, \quad (4)$$

y suponga las condiciones:

B1: $E(\boldsymbol{\epsilon}_n) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\epsilon}_n) = \sigma^2 \mathbf{I}$.

B2: Sea $h_{kk} = \mathbf{x}_{k,n}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{k,n}$ el k -ésimo elemento de la diagonal de \mathbf{H}_n y considere

$$\max_{1 \leq k \leq n} h_{kk} \rightarrow 0, \quad \text{conforme } n \rightarrow \infty.$$

B3: $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}_n^\top \mathbf{X}_n = \mathbf{K}$ es una matriz no singular (no estocástica).



Usando el supuesto B3, tenemos

$$\lim_{n \rightarrow \infty} \text{Cov}(\hat{\beta}_n) = \sigma^2 \lim_{n \rightarrow \infty} \frac{1}{n} \left(\frac{\mathbf{X}_n^\top \mathbf{X}_n}{n} \right)^{-1} = \sigma^2 \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{K}^{-1} = \mathbf{0}.$$

Esto implica que $\hat{\beta}_n \xrightarrow{2\text{nd}} \beta$. Es decir, $\hat{\beta}_n$ es un estimador consistente de β .

Resultado 3 (Distribución asintótica del estimador LS):

Considere el modelo (4) bajo los supuestos B1 a B3. Entonces

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{D} N_p(\mathbf{0}, \sigma^2 \mathbf{K}^{-1}).$$

Demostración:

Ver Sen y Singer (1993),⁴ pág. 279.

⁴Large Sample Methods in Statistics. Chapman & Hall, London.



Resultado 4 (Teorema de Gauss-Markov):

Suponga $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ y sea $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ el estimador mínimos cuadrados. Asuma el supuesto B1. El estimador $\hat{\gamma} = \mathbf{h}^\top \hat{\boldsymbol{\beta}}$ de $\gamma = \mathbf{h}^\top \boldsymbol{\beta}$ es el **mejor estimador lineal e insesgado (BLUE)**.

Demostración:

Sea $\mathbf{c}^\top \mathbf{Y}$ cualquier otro estimador lineal e insesgado de $\gamma = \mathbf{h}^\top \boldsymbol{\beta}$. Dado que $\mathbf{c}^\top \mathbf{Y}$ es insesgado, sigue que

$$\mathbf{h}^\top \boldsymbol{\beta} = E(\mathbf{c}^\top \mathbf{Y}) = \mathbf{c}^\top E(\mathbf{Y}) = \mathbf{c}^\top \mathbf{X} \boldsymbol{\beta}, \quad \forall \boldsymbol{\beta},$$

luego, tenemos

$$\mathbf{c}^\top \mathbf{X} = \mathbf{h}^\top.$$



Ahora,

$$\text{var}(\mathbf{c}^\top \mathbf{Y}) = \mathbf{c}^\top \text{Cov}(\mathbf{Y}) \mathbf{c} = \sigma^2 \mathbf{c}^\top \mathbf{c}, \quad (5)$$

mientras que

$$\text{var}(\mathbf{h}^\top \hat{\boldsymbol{\beta}}) = \mathbf{h}^\top \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{h} = \sigma^2 \mathbf{h}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{h} = \sigma^2 \mathbf{c}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{c}, \quad (6)$$

desde (5) y (6), tenemos

$$\begin{aligned} \text{var}(\mathbf{c}^\top \mathbf{Y}) - \text{var}(\mathbf{h}^\top \hat{\boldsymbol{\beta}}) &= \sigma^2 (\mathbf{c}^\top \mathbf{c} - \mathbf{c}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{c}) \\ &= \sigma^2 \mathbf{c}^\top (\mathbf{I} - \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}) \mathbf{c} \geq 0, \end{aligned}$$

el resultado sigue pues $\mathbf{I} - \mathbf{H}$ es semidefinida positiva.

