TEMA 8: Introducción a la teoría general de modelos lineales: regresión y análisis de la varianza

- Descripción del modelo lineal general. Modelo de Gauss-Markov.
- Estimación de un modelo de Gauss-Markov.
- Inferencia bajo hipótesis de normalidad.
- Modelo de regresión lineal simple.
- Análisis de la varianza de una vía.

I. El modelo lineal general. Modelo de Gauss-Markov

MODELO LINEAL GENERAL

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\boldsymbol{Y} = \mathbf{x_1}\beta_1 + \cdots, \mathbf{x_k}\beta_k + \boldsymbol{\varepsilon})$$

- $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ es un vector aleatorio n-dimensional *observable*.
- $\mathbf{X} = (\mathbf{x_1}, \ldots, \mathbf{x_k})$ es una *matriz conocida* de dimensión $n \times k$ $(k < n)$, denominada *matriz de diseño:*

$$\mathbf{x_j} = (x_{1j}, \ldots, x_{nj})^T, \quad j = 1, \ldots, k$$

  El rango de $\mathbf{X}$ determina el *rango del modelo*; si el rango es k, el modelo es de *rango máximo o completo.*

- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T$ es un vector de *parámetros desconocidos*, denominado *vector de efectos*, cuyas componentes ponderan los efectos de los vectores columna de $\mathbf{X}$ en el vector $\boldsymbol{Y}$.
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ es un vector aleatorio *no observable*, llamado *vector de errores*, que representa el error que se comete si se describe $\boldsymbol{Y}$ por $\mathbf{X}\boldsymbol{\beta}$.

↓

MODELO DE GAUSS-MARKOV

Es un modelo lineal en el que las componentes del vector de errores son variables aleatorias de segundo orden, centradas, homocedásticas (igual varianza) e incorreladas:

$$E[\varepsilon_i] = 0, \quad Var[\varepsilon_i] = E[\varepsilon_i^2] = \sigma^2, \ i = 1 \ldots, n, \qquad Cov[\varepsilon_i, \varepsilon_j] = E[\varepsilon_i \varepsilon_j] = 0, \ i \neq j = 1 \ldots, n.$$

o, equivalentemente:

$$E[\boldsymbol{\varepsilon}] = 0, \quad Cov[\boldsymbol{\varepsilon}] = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 I_{n\times n},$$

↓

Objetivo

Realizar inferencia sobre los parámetros $\beta_1, \ldots, \beta_k, \sigma^2$ a partir de una observación del vector $\boldsymbol{Y}$.

Modelo de Gauss-Markov

$$\boldsymbol{Y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \longrightarrow \begin{cases} E[\boldsymbol{\varepsilon}]=0, \quad Cov[\boldsymbol{\varepsilon}]=E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]=\sigma^2 I_{n\times n} \\ E[\boldsymbol{Y}]=\mathbf{X}\boldsymbol{\beta}, \quad Cov[\boldsymbol{Y}]=\sigma^2 I_{n\times n}. \end{cases}$$

$$Y_i=\sum_{j=1}^{k} x_{ij}\beta_j+\varepsilon_i, \quad i=1,\ldots,n \longrightarrow \begin{cases} E[\varepsilon_i]=0, \quad Var[\varepsilon_i]=\sigma^2, \quad Cov[\varepsilon_i,\varepsilon_j]=0, \ i\neq j \\ E[Y_i]=\sum_{j=1}^{k} x_{ij}\beta_j, \quad Var[Y_i]=\sigma^2, \quad Cov[Y_i,Y_j]=0, \ i\neq j. \end{cases}$$

I a) ESTIMACIÓN DE MÍNIMOS CUADRADOS DEL VECTOR DE EFECTOS

$$\textit{Minimizar } \ S^2(\boldsymbol{\beta})=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\Big(Y_i-\sum_{j=1}^{k}x_{ij}\beta_j\Big)^2=||\boldsymbol{Y}-\mathbf{X}\boldsymbol{\beta}||^2$$

$$\bullet \ \frac{\partial S^2(\boldsymbol{\beta})}{\partial \beta_h}=-2\sum_{i=1}^{n}\Big(Y_i-\sum_{j=1}^{k}x_{ij}\beta_j\Big)x_{ih}=0, \quad h=1,\ldots,k.$$

Ecuaciones normales: $\sum_{i=1}^{n} Y_i x_{ih}=\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}x_{ih}\beta_j, \quad h=1,\ldots,k \longrightarrow \mathbf{X}^T\boldsymbol{Y}=(\mathbf{X}^T\mathbf{X})\boldsymbol{\beta}.$

$\downarrow$

Estimador de mínimos cuadrados de $\boldsymbol{\beta} \rightarrow \widehat{\boldsymbol{\beta}}(\boldsymbol{Y})=\Big(\widehat{\beta}_1(\boldsymbol{Y}),\ldots,\widehat{\beta}_k(\boldsymbol{Y})\Big)^T$.

- *Existencia*: existe, al menos, un estimador de mínimos cuadrados de $\boldsymbol{\beta}$.
- *Unicidad*: no está garantizada, salvo si *el modelo es de rango máximo*:

$$\widehat{\boldsymbol{\beta}}:=\widehat{\boldsymbol{\beta}}(\boldsymbol{Y})=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{Y} \rightarrow \text{Función lineal de } \boldsymbol{Y} \rightarrow \begin{cases} E[\widehat{\boldsymbol{\beta}}]=\boldsymbol{\beta}. \\ Cov[\widehat{\boldsymbol{\beta}}]=\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \end{cases}$$

Función estimable (escalar): *$\psi(\boldsymbol{\beta})$ es estimable si admite un estimador insesgado, función lineal de las componentes de $\boldsymbol{Y}$:*

$$\psi(\boldsymbol{\beta}) \textit{estimable} \ \Leftrightarrow \exists \ \boldsymbol{c}\in\mathbb{R}^n \ / \ E\left[\boldsymbol{c}^T\boldsymbol{Y}\right]=\psi(\boldsymbol{\beta}), \quad \forall\boldsymbol{\beta},$$

o, equivalentemente, $\psi(\boldsymbol{\beta})=\boldsymbol{c}^T\mathbf{X}\boldsymbol{\beta}, \quad \boldsymbol{c}\in\mathbb{R}^n$.

- *Si el modelo es de rango máximo, $\psi(\boldsymbol{\beta})$ es estimable $\Leftrightarrow \psi(\boldsymbol{\beta})=\boldsymbol{a}^T\boldsymbol{\beta}, \quad \boldsymbol{a}\in\mathbb{R}^k$.*

Teorema de Gauss-Markov: *Si $a^T\boldsymbol{\beta}$ es estimable, admite un único estimador lineal insesgado uniformemente de mínima varianza en la clase de estimadores lineales insesgados. Dicho estimador es $a^T\widehat{\boldsymbol{\beta}}(\boldsymbol{Y})$, y se denomina estimador de mínimos cuadrados de $a^T\boldsymbol{\beta}$.*

I b) MODELO ESTIMADO, RESIDUOS Y ESTIMACIÓN DE LA VARIANZA

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_1, \ldots, \widehat{\beta}_k\right)^T := \widehat{\boldsymbol{\beta}}(\boldsymbol{Y}) \text{ estimador de mínimos cuadrados de } \boldsymbol{\beta}$$

Modelo estimado: $\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \longrightarrow \widehat{Y}_i = \sum_{j=1}^{k} x_{ij}\widehat{\beta}_j, \quad i = 1, \ldots, n.$

Residuos mínimo-cuadráticos: $\boldsymbol{R} = \boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \longrightarrow R_i = Y_i - \widehat{Y}_i, \quad i = 1, \ldots, n.$

Propiedades:

- $\widehat{Y}_i$ *es el estimador lineal insesgado de mínima varianza (estimador de mínimos cuadrados) de* $E[Y_i] = \sum_{j=1}^{k} x_{ij}\beta_j, \ \forall i = 1, \ldots, n.$
- *Los residuos son variables aleatorias con media nula,* $E[R_i] = 0, \ \forall i = 1, \ldots, n.$
- *El vector de residuos es ortogonal a los vectores columna de* $\boldsymbol{X}$, $\mathbf{X}^T\boldsymbol{R} = 0$.[1]
- *El vector de residuos es ortogonal al vector estimado,* $\widehat{\boldsymbol{Y}}^T\boldsymbol{R} = 0.$

$$\textit{VARIANZA RESIDUAL}: \ S_R^2 = \frac{\sum_{i=1}^{n} R_i^2}{n-r} = \frac{||\boldsymbol{R}||^2}{n-r} = \frac{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{n-r} \longrightarrow \text{estimador insesgado de } \sigma^2$$

[1] Esto indica que existen k relaciones lineales entre los residuos $R_1, \ldots, R_n$, determinadas por las k columnas de $\mathbf{X}$, $\mathbf{x_j^T R} = \sum_{i=1}^{n} x_{ij}R_i = 0, \quad j = 1, \ldots, k$. Por tanto, si $\mathbf{X}$ es de rango r, el número de residuos R_i linealmente independientes es $n - r \longrightarrow$ *Los residuos tienen* $n - r$ *grados de libertad.*

I c) INFERENCIA BAJO HIPÓTESIS DE NORMALIDAD (Ver Apéndice)

$$\boxed{\begin{array}{l}\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \to \mathcal{N}_n(0,\ \sigma^2 I_{n\times n})\end{array}} \Leftrightarrow\ \boldsymbol{Y} \to \mathcal{N}_n(\boldsymbol{X\beta},\ \sigma^2 I_{n\times n})$$

▶ Estimadores de máxima verosimilitud:

$$\textit{Función de verosimilitud}: \ \boldsymbol{y} \in \mathbb{R}^n \ \to \ L_{\boldsymbol{y}}(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \ \exp\left\{-\frac{||\boldsymbol{y} - \boldsymbol{X\beta}||^2}{2\sigma^2}\right\}.$$

- *Estimador máximo verosímil de* $\boldsymbol{\beta} \longrightarrow \widehat{\boldsymbol{\beta}}$ (mínimos cuadrados). [2]
- *Estimador máximo verosímil de* $\sigma^2 \longrightarrow \widehat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} R_i^2}{n} = \dfrac{(n-r)S_R^2}{n}$. [3]

▶ Test de razón de verosimilitudes para la hipótesis lineal general:

Hipótesis lineal general: $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$, siendo $\mathbf{C}_{q\times k}$ una matriz conocida de rango q $(\leq k)$, tal que todas las componentes del vector $\mathbf{C}\boldsymbol{\beta}$ son estimables.

Test de razón de verosimilitudes de tamaño α

$$\boxed{\varphi(\boldsymbol{Y}) = \begin{cases} 1 & F(\boldsymbol{Y}) > F_{q,n-r;\ \alpha} \\ 0 & F(\boldsymbol{Y}) \leq F_{q,n-r;\ \alpha} \end{cases} \qquad F(\boldsymbol{Y}) = \frac{n-r}{q}\left(\frac{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^0||^2 - ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}\right)}$$

$\widehat{\boldsymbol{\beta}}^0$: estimador máximo verosímil de $\boldsymbol{\beta}$ bajo H_0

[2] Maximizar $L_{\boldsymbol{y}}(\boldsymbol{\beta}, \sigma^2)$ en $\boldsymbol{\beta}$ equivale a minimizar $||\boldsymbol{y} - \boldsymbol{X\beta}||^2$.

[3] $\dfrac{\partial ln L_{\boldsymbol{y}}(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}||\boldsymbol{y} - \boldsymbol{X\beta}||^2 = 0 \ \Rightarrow \ \widehat{\sigma}^2(\boldsymbol{y}) = \dfrac{||\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}||^2}{n} = \dfrac{\sum_{i=1}^{n} r_i^2}{n}$.

II. Modelo de regresión lineal simple

Hipótesis: X, Y variables aleatorias tales que $E[Y^2] < +\infty$ y, fijado un valor arbitrario, $X = x$:

$$E\left[Y/X = x\right] = \beta_0 + \beta_1 x, \qquad Var\left[Y/X = x\right] = \sigma^2.$$

$$\Downarrow$$

Formulación: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1 \ldots n.$

- $x_i, \quad i = 1 \ldots n$, valores arbitrarios (fijos) de X (al menos dos distintos).
- $Y_i, \quad i = 1 \ldots n$, variables aleatorias que describen las observaciones de Y, supuesto que $X = x_i$ $(Y_i \equiv Y/X = x_i)$.
- $\varepsilon_i, \quad i = 1 \ldots n$, variables aleatorias tales que $E[\varepsilon_i] = 0,\ Var[\varepsilon_i] = \sigma^2,\ i = 1, \ldots, n.$

$$\Downarrow\ Y_1, \ldots, Y_n \text{ independientes}$$

Modelo de Gauss-Markov de rango máximo (2).

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}_{n\times 2} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}_{2\times 1} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

II a) ESTIMACIÓN DEL MODELO [4]

▶ Estimador de mínimos cuadrados de $\boldsymbol{\beta}$:

Modelo de rango máximo $\longrightarrow \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ [5] $\begin{cases} \text{Único y función lineal de } \boldsymbol{Y}. \\ E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}, \;\; Cov[\widehat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}. \end{cases}$

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_0,\ \widehat{\beta}_1\right)^T \longrightarrow \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{x}, \quad \widehat{\beta}_1 = \frac{\sigma_{xY}}{\sigma_x^2}$$

$\widehat{\beta}_0$ y $\widehat{\beta}_1$ son los estimadores de mínimos cuadrados de β_0 y β_1, respectivamente.

$\Downarrow$

Son los de mínima varianza uniformemente en la clase de estimadores lineales insesgados:

$\star$ $E[\widehat{\beta}_0] = \beta_0, \;\; E[\widehat{\beta}_1] = \beta_1.$

$\star$ $Var[\widehat{\beta}_0] = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{n\sigma_x^2}\right), \qquad Var[\widehat{\beta}_1] = \sigma^2\frac{1}{n\sigma_x^2}, \qquad Cov[\widehat{\beta}_0, \widehat{\beta}_1] = -\sigma^2\frac{\overline{x}}{n\sigma_x^2}.$

▶ Modelo estimado, residuos mínimo cuadráticos y propiedades:

- $\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \longrightarrow \widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \longrightarrow \widehat{Y}_i = \overline{Y} + \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \overline{x}), \quad i = 1, \ldots, n.$
- $\boldsymbol{R} = \boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \longrightarrow R_i = Y_i - \widehat{Y}_i \longrightarrow R_i = Y_i - \overline{Y} - \frac{\sigma_{xY}}{\sigma_x^2}(x_i - \overline{x}), \quad i = 1, \ldots, n.$
- $\mathbf{X}^T\boldsymbol{R} = 0 \Rightarrow \sum_{i=1}^n R_i = 0 \Rightarrow \sum_{i=1}^n R_i/n = 0 \Rightarrow \sum_{i=1}^n \widehat{Y}_i/n = \overline{Y}.$
- $\widehat{\boldsymbol{Y}}^T\boldsymbol{R} = \sum_{i=1}^n \widehat{Y}_i R_i = 0.$

▶ Varianza residual: $S_R^2 = \dfrac{\sum_{i=1}^n R_i^2}{n-2}$.

[4] $\overline{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \sigma_x^2 = \frac{\sum_{i=1}^n (x_i-\overline{x})^2}{n}, \quad \overline{Y} = \frac{\sum_{i=1}^n Y_i}{n}, \quad \sigma_Y^2 = \frac{\sum_{i=1}^n (Y_i-\overline{Y})^2}{n}, \quad \sigma_{xY} = \frac{\sum_{i=1}^n (x_i-\overline{x})(Y_i-\overline{Y})}{n}.$

[5] $(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{pmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1} = \frac{1}{n\sigma_x^2}\begin{pmatrix} \sum_{i=1}^n x_i^2/n & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix} \quad (\sigma_x^2 \neq 0 \;\text{ ya que }\; \exists x_i \neq x_j).$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \frac{1}{n\sigma_x^2}\begin{pmatrix} \sum_{i=1}^n x_i^2/n & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}\begin{pmatrix} n\overline{Y} \\ \sum_{i=1}^n x_i Y_i \end{pmatrix} = \begin{pmatrix} \overline{Y} - \overline{x}\frac{\sigma_{xY}}{\sigma_x^2} \\ \frac{\sigma_{xY}}{\sigma_x^2} \end{pmatrix}.$$

II b) ANÁLISIS DE LA BONDAD DEL MODELO ESTIMADO

Descomposición de la variabilidad de las observaciones [6]

$$VT = VE + VNE$$

- $VT = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = n\sigma_Y^2 \longrightarrow$ *Variabilidad total* (de $Y_1, \ldots, Y_n$).
- $VE = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 = \dfrac{n\sigma_{xY}^2}{\sigma_x^2} \longrightarrow$ *Variabilidad explicada por el modelo* (de $\widehat{Y}_1, \ldots, \widehat{Y}_n$).
- $VNE = \sum_{i=1}^{n} R_i^2 \longrightarrow$ *Variabilidad no explicada por el modelo* (de $R_1, \ldots, R_n$).

Coeficiente de determinación lineal

$$R^2 = \frac{VE}{VT} = \frac{\sigma_{xY}^2}{\sigma_x^2 \sigma_Y^2}.$$

$\downarrow$

Proporción de variabilidad total explicada por el modelo

---

[6] $Y_i = \widehat{Y}_i + R_i, \quad i = 1, \ldots, n.$

$\Downarrow$

$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (\widehat{Y}_i + R_i - \overline{Y})^2 = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} R_i^2 + 2\underbrace{\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y}) R_i}_{0} = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 + \sum_{i=1}^{n} R_i^2.$$

II c) PREDICCIÓN A PARTIR DEL MODELO ESTIMADO

$$X = x_p \longrightarrow Y_p = \beta_0 + \beta_1 x_p + \varepsilon_p; \quad E[\varepsilon_p] = 0, \quad Var[\varepsilon_p] = \sigma^2$$

$$\Downarrow$$

$$\textit{Predicción de } Y \textit{ cuando } X = x_p \longrightarrow \widehat{Y}_p = \widehat{\beta}_0 + \widehat{\beta}_1 x_p = \overline{Y} + \frac{\sigma_{xY}}{\sigma_x^2}(x_p - \overline{x}).$$

- $E[\widehat{Y}_p] = \beta_0 + \beta_1 x_p = E[Y_p]$.
- $Var[\widehat{Y}_p] = E[(\widehat{Y}_p - E[\widehat{Y}_p])^2] = \sigma^2 \left[\frac{1}{n} + \frac{(x_p - \overline{x})^2}{n\sigma_x^2}\right]$.
- $ECM[\widehat{Y}_p] = E[(\widehat{Y}_p - Y_p)^2] = Var[\widehat{Y}_p] + Var[Y_p] = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{n\sigma_x^2}\right]$.

II d) APLICACIÓN PRÁCTICA: RECTA DE REGRESIÓN ESTIMADA

$\boxed{(x_1, y_1), \ldots, (x_n, y_n) \text{ observaciones independientes del vector aleatorio } (X, Y).}$

Modelo lineal correspondiente a $x_1, \ldots, x_n \longrightarrow Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$

Modelo estimado a partir de $Y_1, \ldots, Y_n \longrightarrow \widehat{Y}_i = \overline{Y} + \dfrac{\sigma_{xY}}{\sigma_x^2}(x_i - \overline{x}), \quad i = 1, \ldots \ n.$

$$\downarrow Y_1 = y_1, \ldots, Y_n = y_n$$

$$\widehat{y}_i = \overline{y} + \frac{\sigma_{xy}}{\sigma_x^2}(x_i - \overline{x}), \quad i = 1, \ldots \ n.$$

Recta de regresión estimada [7] *a partir de* $(x_1, y_1), \ldots, (x_n, y_n) \longrightarrow \boxed{y = \overline{y} + \dfrac{\sigma_{xy}}{\sigma_x^2}(x - \overline{x})}$

Coeficiente de determinación lineal estimado [8] *a partir de* $(x_1, y_1), \ldots, (x_n, y_n) \longrightarrow \boxed{r^2 = \dfrac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2}}$

Predicción a partir de la recta de regresión estimada $\longrightarrow \widehat{y}_p = \overline{y} + \dfrac{\sigma_{xy}}{\sigma_x^2}(x_p - \overline{x}).$

Estimación del error cuadrático medio de la predicción $\longrightarrow \widehat{ECM}(\widehat{Y}_p) = s_R^2 \left[1 + \dfrac{1}{n} + \dfrac{(x_p - \overline{x})^2}{n\sigma_x^2}\right].$

[7] Recta de regresión teórica: $y = E[Y] + \dfrac{Cov[X,Y]}{Var[X]}(x - E[X])$

[8] Coeficiente de determinación lineal de X e Y : $\rho_{XY}^2 = \dfrac{Cov^2[X,Y]}{Var[X]Var[Y]}$

II e) CONTRASTE DE REGRESIÓN BAJO HIPÓTESIS DE NORMALIDAD

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1 \dots n, \quad \varepsilon_i \to \mathcal{N}(0,\ \sigma^2) \ \left(\Leftrightarrow Y_i \to \mathcal{N}(\beta_0 + \beta_1 x_i,\ \sigma^2) \right)$$

$Y_1, \dots, Y_n$ independientes.

$H_0 : \beta_1 = 0 \ \longleftrightarrow \ H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ con $\mathbf{C} = (0,1)_{1\times 2}$ ($\mathbf{C}\boldsymbol{\beta}$ estimable y rango $q = 1$).

Estadístico de contraste del test de razón de verosimilitudes

$$F(\boldsymbol{Y}) = \frac{n-r}{q}\left(\frac{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^0||^2 - ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2}\right) \longrightarrow \begin{cases} ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}||^2 = \sum_{i=1}^n R_i^2 = VNE \\ ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^0||^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2 = VT^{(*)} \end{cases}$$

$^{(*)}$ Bajo H_0, $Y_i = \beta_0 + \varepsilon_i \to \mathcal{N}(\beta_0, \sigma^2)$, $i = 1, \dots, n \Rightarrow (Y_1, \dots, Y_n)$ m.a.s. de $\mathcal{N}(\beta_0, \sigma^2) \ \Rightarrow \ \widehat{\boldsymbol{\beta}}_0^0 = \overline{Y}$.

$$\Downarrow$$

$$\widehat{\boldsymbol{\beta}}^0 = \left(\widehat{\beta}_0^0, \widehat{\beta}_1^0\right)^T = (\overline{Y}, 0)^T \ \Rightarrow \ \mathbf{X}\widehat{\boldsymbol{\beta}}^0 = (\overline{Y}, \dots, \overline{Y})^T.$$

$$\Downarrow$$

$$F(\boldsymbol{Y}) = (n-2)\frac{VT - VNE}{VNE} = \frac{VE}{VNE/(n-2)} = \frac{VE}{S_R^2} \underset{H_0}{\longrightarrow} F(1, n-2).$$

Test de razón de verosimilitudes de tamaño α $\longrightarrow \varphi(\boldsymbol{Y}) = \begin{cases} 1 & F(\boldsymbol{Y}) > F_{1,n-2;\ \alpha} \\ 0 & F(\boldsymbol{Y}) \leq F_{1,n-2;\ \alpha}. \end{cases}$

Nota: $F(\boldsymbol{Y}) = (n-2)\dfrac{R^2}{1-R^2} \longrightarrow$ función creciente de R^2. Así, grandes valores de R^2 conducen al rechazo de H_0, lo que concuerda con la definición de R^2 como medida de bondad del modelo estimado.

III. Análisis de la varianza de una vía

Técnica estadística (Fisher) para *determinar si un supuesto factor de variación afecta al comportamiento de una cierta variable aleatoria*. Se aplica bajo los siguientes supuestos:

- *La variable de interés no está afectada por factores distintos al que es objeto de estudio.*
- *El factor de variación tiene un número finito de niveles (k) y, en cada uno de ellos, la variable tiene distribución normal, con la misma varianza:*

$$Y_i: \text{ variable de interés en el nivel } i\text{-ésimo } \longrightarrow \mathcal{N}(\mu_i, \sigma^2), \quad i = 1, \ldots, k.$$

Hipótesis a contrastar

El supuesto factor de variación no afecta al comportamiento de la variable:

$$H_0 : \mu_1 = \cdots = \mu_k$$

Igualdad de medias de k poblaciones normales con varianza común.

Resolución:

- $(Y_{i1}, \ldots, Y_{in_i})$ muestra aleatoria simple de Y_i, $i = 1, \ldots, k$, todas independientes.
- $Y_{ij} \to \mathcal{N}(\mu_i, \sigma^2) \Rightarrow \underbrace{Y_{ij} = \mu_i + \varepsilon_{ij},\ i = 1 \ldots, k,\quad j = 1, \ldots, n_i,}$ con $\varepsilon_{ij} \longrightarrow \mathcal{N}(0, \sigma^2)$.

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \text{ modelo lineal } n\text{-dimensional } (n = \sum_{i=1}^{k} n_i)$$

$$\boldsymbol{Y} = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \hline Y_{21} \\ \vdots \\ Y_{2n_2} \\ \hline \vdots \\ \hline Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix}_{n\times 1} \quad \mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \\ \hline \mathbf{0} & \mathbf{1} & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & 0 \\ \hline & \vdots & & \\ \hline \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n\times k} \to \text{rango} = k, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}_{k\times 1} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \hline \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \hline \vdots \\ \hline \varepsilon_{k1} \\ \vdots \\ \varepsilon_{kn_k} \end{pmatrix}_{n\times 1}$$

$$E[\varepsilon_{ij}] = 0,\ i = 1 \ldots, k,\quad j = 1, \ldots, n_i; \quad E[\varepsilon_{ij}\varepsilon_{i'j'}] = 0,\ i \neq i' \text{ ó } j \neq j'$$

$$\Downarrow$$

Modelo de Gauss-Markov de rango máximo (k) .

III a) ESTIMACIÓN DEL MODELO

▶ Estimador máximo verosímil (mínimos cuadrados) de $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)^T$ [9]

$$\boxed{\widehat{\mu}_i = \overline{Y}_{i\cdot}, \quad i = 1, \ldots, k \longrightarrow \widehat{\boldsymbol{\mu}} = (\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{k\cdot})^T}$$

▶ Modelo estimado, residuos mínimo cuadráticos y propiedades:

- $\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{\mu}} = \left(\overline{Y}_{1\cdot}, \ldots, \overline{Y}_{1\cdot} | \cdots | \overline{Y}_{k\cdot}, \ldots, \overline{Y}_{k\cdot}\right)^T \longrightarrow \widehat{Y}_{ij} = \overline{Y}_{i\cdot}, \quad i = 1, \ldots, k, \ j = 1, \ldots, n_i.$
- $\boldsymbol{R} = \boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}} \longrightarrow R_{ij} = Y_{ij} - \widehat{Y}_{ij} = Y_{ij} - \overline{Y}_{i\cdot}, \quad i = 1, \ldots, k, \ j = 1, \ldots, n_i.$
- $\mathbf{X}^T\boldsymbol{R} = 0 \Rightarrow \sum_{j=1}^{n_i} R_{ij} = 0, \ i = 1, \ldots, k \Rightarrow \begin{cases} \sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}/n = 0 \\ \sum_{i=1}^{k}\sum_{j=1}^{n_i} \widehat{Y}_{ij}/n = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \overline{Y}_{i\cdot}/n = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}/n = \overline{Y}_{\cdot\cdot}. \end{cases}$
- $\widehat{\boldsymbol{Y}}^T\boldsymbol{R} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \widehat{Y}_{ij}R_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \overline{Y}_{i\cdot}R_{ij} = 0.$

[9] $S^2(\boldsymbol{\mu}) = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 \Rightarrow \dfrac{\partial S^2(\boldsymbol{\mu})}{\partial \mu_h} = \dfrac{d\sum_{j=1}^{n_h} (Y_{hj} - \mu_h)^2}{d\mu_h} = -2\sum_{j=1}^{n_h} (Y_{hj} - \mu_h) = 0 \Rightarrow \widehat{\mu}_h = \dfrac{\sum_{j=1}^{n_h} Y_{hj}}{n_h} = \overline{Y}_{h\cdot}$

III b) DESCOMPOSICIÓN DE LA VARIABLIDAD DE LAS OBSERVACIONES[10]

$$VT = VE + VNE$$

- $VT = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}\right)^2 \rightarrow$ *Variabilidad total* (de $Y_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, n_i$).
- $VE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\widehat{Y}_{ij} - \overline{Y}\right)^2 \rightarrow$ *Variabilidad explicada por el modelo* (de $\widehat{Y}_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, n_i$).
- $VNE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}^2 \rightarrow$ *Variabilidad no explicada por el modelo* (de $R_{ij},\ i = 1, \ldots, k,\ j = 1, \ldots, n_i$).

$\downarrow\ \widehat{Y}_{ij} = \overline{Y}_i,\ R_{ij} = Y_{ij} - \overline{Y}_i$

- $VE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\overline{Y}_i - \overline{Y}\right)^2 \longrightarrow$ Variabilidad de $\overline{Y}_1, \ldots, \overline{Y}_k \longrightarrow$ *Variabilidad entre grupos.*
- $VNE = \sum_{i=1}^{k} \underbrace{\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2}_{\text{Variabilidad de las observaciones de la muestra } i\text{-ésima}} \longrightarrow$ *Variabilidad dentro de grupos.*

[10] $Y_{ij} = \widehat{Y}_{ij} + R_{ij}$

$\Downarrow$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\widehat{Y}_{ij} + R_{ij} - \overline{Y}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\widehat{Y}_{ij} - \overline{Y}\right)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}^2 + 2\underbrace{\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(\widehat{Y}_{ij} - \overline{Y}\right) R_{ij}}_{0}$$

III c) PROBLEMA DE CONTRASTE $H_0 : \mu_1 = \cdots = \mu_k$

$$H_0 : \mu_1 - \mu_2 = 0,\ \mu_1 - \mu_3 = 0, \ldots, \mu_1 - \mu_k = 0$$

$$\updownarrow$$

$$H_0 : \mathbf{C}\boldsymbol{\mu} = 0,\ \text{con } \mathbf{C} = \underset{(k-1)\times k}{\begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}} \rightarrow \text{rango} = k-1 \ \textbf{(Hipótesis lineal general)}$$

Estadístico de contraste del test de razón de verosimilitudes

$$F(\boldsymbol{Y}) = \frac{n-r}{q}\left(\frac{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}^0||^2 - ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}||^2}{||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}||^2}\right) \longrightarrow \begin{cases} ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}||^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij}^2 = VNE \\ ||\boldsymbol{Y} - \mathbf{X}\widehat{\boldsymbol{\mu}}^0||^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y})^2 = VT^{(*)} \end{cases}$$

$^{(*)}$ Bajo H_0, $Y_{ij} \rightarrow \mathcal{N}(\mu, \sigma^2)$, $\forall i,j \Rightarrow (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k})$ m.a.s. de $\mathcal{N}(\mu, \sigma^2) \Rightarrow \widehat{\mu} = \overline{Y}$.

$$\Downarrow$$

$\widehat{\mu}_i^0 = \overline{Y},\ \ i = 1, \ldots, k \longrightarrow \widehat{\boldsymbol{\mu}}^0 = (\overline{Y}, \ldots, \overline{Y})^T \longrightarrow \mathbf{X}\widehat{\boldsymbol{\mu}}^0 = (\overline{Y}, \ldots, \overline{Y}|\cdots|\overline{Y}, \ldots, \overline{Y})^T$.

$$\Downarrow$$

$$F(\boldsymbol{Y}) = \frac{n-k}{k-1}\frac{VT - VNE}{VNE} = \frac{VE/(k-1)}{VNE/(n-k)} \underset{H_0}{\longrightarrow} F(k-1, n-k).$$

TABLA ANOVA DE UNA VÍA

Fuentes de Variación	Variabilidad	Grados de libertad	Varianzas
Entre grupos	$VE = \sum_{i=1}^{k} n_i \left(\overline{Y}_i - \overline{Y}\right)^2$	$k-1$	$S_E^2 = VE/(k-1)$
Dentro de grupos	$VNE = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}_i\right)^2$	$n-k$	$S_R^2 = VNE/(n-k)$
Total	$VT = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y}\right)^2$	$n-1$	

$\downarrow$

Test de razón de verosimilitudes de tamaño α

$$\varphi(\boldsymbol{Y}) = \begin{cases} 1 & F(\boldsymbol{Y}) > F_{k-1,n-k;\ \alpha} \\ 0 & F(\boldsymbol{Y}) \leq F_{k-1,n-k;\ \alpha} \end{cases} \qquad F(\boldsymbol{Y}) = \frac{S_E^2}{S_R^2}.$$

Nota: En la práctica, si el problema planteado no especifica un nivel de significación se trabaja con el denominado *p-nivel o p-valor asociado a los datos:*

$$p = P(F(k-1, n-k) > F_{exp}),$$

siendo F_{exp} el valor del estadístico $F(\boldsymbol{Y})$ obtenido de los datos concretos:

- Si el *p-nivel* es pequeño (usualmente, 0.05 o menor) se rechaza H_0.
- Si el *p-nivel* es grande (0.15 o mayor), se acepta H_0.
- Para valores intermedios hay que tratar cada situación en particular aunque, normalmente, es aconsejable tomar más datos y rehacer los cálculos.

Apéndice: Distribución normal n-dimensional

Un vector aleatorio $Y = (Y_1, \ldots, Y_n)^T$ tiene distribución normal (n-dimensional), lo que se denota $Y \to \mathcal{N}_n(\mu, \Sigma)$, si su función de densidad es de la forma:

$$f(y) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{(y-\mu)^T\Sigma^{-1}(y-\mu)}{2}}, \quad y \in \mathbb{R}^n,$$

donde $\mu = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^n$ y $\Sigma = ((\sigma_{ij}))_{n\times n}$ es una matriz definida positiva.

Propiedades:

- $E[Y] = \mu$, $Cov[Y] = E[(Y-\mu)(Y-\mu)^T] = \Sigma$.
- Las distribuciones marginales de cualquier dimensión son normales y, en particular, $Y_i \to \mathcal{N}(\mu_i, \sigma_{ii})$, $i = 1, \ldots, n$.
- $Y \to \mathcal{N}_n(\mu, \Sigma) \Rightarrow \forall \gamma \in \mathbb{R}^n$ (vector constante), $Y + \gamma \to \mathcal{N}_n(\mu + \gamma, \Sigma)$.
- $Y = (Y_1, \ldots, Y_n)^T \to \mathcal{N}_n(\mu, \Sigma)$:
  $Y_1, \ldots, Y_n$ son independientes $\Leftrightarrow$ Σ es diagonal $\Leftrightarrow \rho_{Y_i,Y_j} = 0$, $\forall i, j = 1, \ldots, n$, $i \neq j$.