Cuarto taller

sábado, 27 de septiembre de 2025 11:58 a. m.

De respuesta a las preguntas formuladas a continuación en base a la teoría tratada en clase. **Provea una interpretación de ser necesario.**

1. Considere el siguiente modelo de regresión lineal múltiple con k variables regresoras, $p = (k+1)$ parámetros asociados $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

   - (a) Escriba el modelo de forma matricial junto con sus supuestos. **Especifique las dimensiones de cada componente.**
   - (b) Demuestre que el estimador $\hat{\boldsymbol{\beta}}$ que se obtiene a través del método de mínimos cuadrados es un estimador insesgado para $\boldsymbol{\beta}$. Analice $\hat{\boldsymbol{\beta}}$.

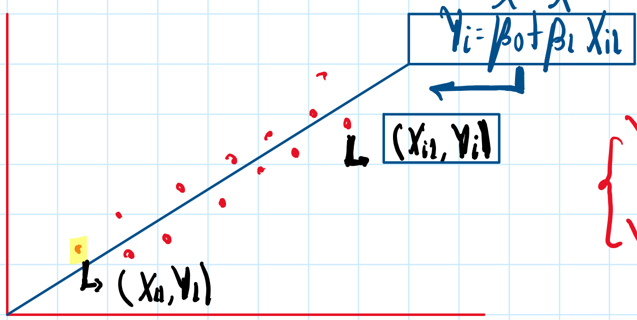
$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)}$

k: # covariables ; j: Alusión coeficiente $\boxed{j = 1, \dots, k}$

p: # parámetros

$\boxed{p = k+1}$ $\mathbb{R}^p$

$\boxed{X_{ij} : i: \text{Observación} \quad i = 1, \dots, n}$



$\boxed{Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i}$ $i = 1, \dots, n$

$$\begin{cases} Y_1 = \beta_0 + \beta_1 X_{11} + \dots + \beta_k X_{1k} + \varepsilon_1 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_{n1} + \dots + \beta_k X_{nk} + \varepsilon_n \end{cases} \qquad Y_n = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n \times 1}$$

$$\varepsilon_n = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n \times 1} ; \quad X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ \vdots & & \ddots & \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}_{\substack{n \times p \\ n \times (k+1)}} ; \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1) \times 1} : \quad \boxed{Y_n = X\beta + \varepsilon} \quad \boxed{\varepsilon_n \sim N_n(0_n, \sigma^2 I)}$$

$\boxed{\hat{\beta} \rightarrow \beta}$: $\boxed{\text{MCO}}$ $\boxed{\hat{\beta} = (\underbrace{X^T X}_{\text{Debe ser invertible}})^{-1} X^T Y_n}$: $\hat{\theta}$ es insesgado si $E[\hat{\theta}] = \theta$

: No multicolinealidad $\quad A^{-1}A = A \cdot A^{-1} = I$

$$E[\hat{\beta}] = E[(X^T X)^{-1} X^T Y_n]$$
$$= (X^T X)^{-1} X^T E[Y_n]$$

$\boxed{Y_n \sim N_n(X\beta, \dots)}$: $\boxed{Y \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)}$: $(X^T X)^{-1} \underbrace{X^T X}_{I} \beta \therefore \boxed{E[\hat{\beta}] = \beta}$

Construir el modelo ajustado:

(1). $Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i : \quad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$

(2). $Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i : \quad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \dots + \hat{\beta}_k X_{ik}$: $\boxed{\hat{Y}_n = X\hat{\beta}}$

Varianza utilizada en el modelo

(1). MSE: $\quad MSE = \dfrac{SSE}{n-2} = \dfrac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} = \dfrac{\sum_{i=1}^n (Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_{i1})^2}{n-2}$

(2). MSE: $\quad MSE = \dfrac{SSE}{n-p} = \dfrac{(Y_n - \hat{Y}_n)^T (Y_n - \hat{Y}_n)}{n-p} = \dfrac{(Y_n - X\hat{\beta})^T (Y_n - X\hat{\beta})}{n-p} \quad \hat{\sigma} = \dfrac{MSE}{n-p}$

2. Determine el valor de verdad de las siguientes afirmaciones.

   - (a) Los modelos de regresión múltiple (1).$Y_i = \beta_0 + \sum_{j=1}^{k} \beta_j (X_{ij})^{1/2} + \varepsilon_i$; (2).$\ln(Y_i) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(X_{ij}) + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ no son modelos lineales, puesto que se

(1). MCO: Método matemático
(2). MLE: Método estadístico

2. Determine el valor de verdad de las siguientes afirmaciones.

- (a) Los modelos de regresión múltiple (1).$\mathrm{Y}_i = \beta_0 + \sum_{j=1}^{k} \beta_j (X_{ij})^{1/2} + \varepsilon_i$; (2).$\ln(\mathrm{Y}_i) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(X_{ij}) + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ no son modelos lineales, puesto que se altera la forma de la superficie de respuesta.

- (b) Bajo los supuestos del modelo de regresión lineal múltiple, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, el estimador para los parámetros $\hat{\boldsymbol{\beta}}$ obtenido a través del método de mínimos cuadrados ordinarios es el mismo que el de máxima verosimilitud, así como para la varianza $\hat{\sigma}^2$. **Explique**.

(1). MCO : Método matemático

(2). MLE : Método estadístico

$\hat{\sigma}_{MCO} = \frac{MSE}{n-p}$; $\hat{\sigma}_{MLE} = \frac{MSE}{n}$

- (c) Se requiere que la matriz $(\mathbf{X}^t\mathbf{X})$ sea singular. Además, la matriz $\mathbf{H}$ es simétrica e idempotente, al igual que $(\mathbf{I_n} - \mathbf{H})$.

$\hat{Y}_n = X\hat{\beta} : X(X^tX)^{-1}X^tY$

$\hat{Y}_n = HY_n$

$H^n = H$

$n \in \mathbb{N}$

- (d) La matriz de varianzas- covarianzas siempre es simétrica respecto a su diagonal principal, además, siempre tiene unos en su diagonal principal.

σ^2 :

$\boxed{Cov[X_1, X_1]}$: $\boxed{Corr[X_1, X_1] = 1}$

$\boxed{Cov[X_1, X_2]}$: Medida de relación

$\boxed{-1 \le Corr[X_1, X_2] \le 1}$: Grado de relación lineal entre X_1 y X_2

- (e) El parámetro β_0, como la respuesta media de Y es únicamente interpretable si el conjunto de observaciones $(X_1, X_2, ..., X_k) = (0, 0, ..., 0)$ está incluido en el rango experimental.

- (f) Los parámetros $\beta_j, j = 1, ..., k$ indican el cambio en la respuesta media de Y por unidad de incremento en la respectiva variable X_j.

Todas las demás variables fijas : Efectos parciales

- $X_1, X_2, \dots X_n$: X_1, X_2, X_3 :

$Cov[X_1, X_1]$	$Cov[X_1, X_2]$
$Cov[X_2, X_2]$ ...	$Cov[X_1, X_3]$
$Cov[X_3, X_3]$	$Cov[X_2, X_3]$

$$\begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline X_1 & \sigma_{11} & \sigma_{12} & \sigma_{13} \\ X_2 & \sigma_{21} & \sigma_{22} & \sigma_{23} \\ X_3 & \sigma_{31} & \sigma_{32} & \sigma_{33} \end{array}$$

- La diagonal no es 1 necesariamente