Primer taller

miércoles, 4 de diciembre de 2024 2:03 p. m.

$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$; $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$; k: # covariables ; $p = k+1$ parámetros

$Cov[X_i, X_i] = Var[X_i]$

$i = 1, \dots, n$:

$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \dots + \beta_k X_{1k} + \varepsilon_1$

$Y_n = \beta_0 + \beta_1 X_{n1} + \beta_2 X_{n2} + \dots + \beta_k X_{nk} + \varepsilon_n$

$Y = X\beta + \varepsilon$; $\varepsilon \sim N_n(0_{n\times 1}, \sigma^2 I_n)$

$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1}$; $X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ & \vdots & & & \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}_{n\times(k+1)}$ $n \times p$; $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1}$; $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$

$\hat{\beta}_j \to \beta_j$: $\hat{\beta} \to \beta$; $\hat{\beta} = (X'X)^{-1}X'Y$; $E[\hat{\theta}] = \theta$

$E[\hat{\beta}] = E[(X'X)^{-1}X'Y] = A\,E[Y]$; $Y \sim N_n(X\beta, \sigma^2 I_n) \to Y \sim N(\beta_0 + \beta_1 X_{i1}, \sigma^2)$

Determinística

$E[\hat{\beta}] = (X'X)^{-1}X'X\beta$: $A \cdot A^{-1} = I = A^{-1}A$

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

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

- (a) Bajo los supuestos del modelo de regresión lineal múltiple, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, el estimador para los parámetros $\hat{\beta}$ es el mismo que el de máxima verosimilitud, así como para la varianza $\hat{\sigma}^2$.
- (b) Se requiere que la matriz $(\mathbf{X}^t\mathbf{X})$ sea singular.
- (c) La matriz de varianzas- covarianzas siempre es simétrica respecto a su diagonal principal, además, siempre tiene unos en su diagonal principal.
- (d) La matriz $\mathbf{H}$ es simétrica e idempotente, al igual que $(I_n - H)$.

Falso: MCO: $\hat{\sigma}^2 = \frac{SSE}{n-p}$

MLE: $\hat{\sigma}^2 = \frac{SSE}{n}$

No se puede $\hat{\beta}$

Invertible: Multicolinealidad

Inversa: $\frac{Ad(A^t)}{Det(A)}$

Falso

Matriz var-cov: $\sigma^2 I_n = \sigma^2 \begin{bmatrix} 1 & Cov[X_1,X_2] & \cdots \\ Cov[X_1,X_2] & 1 & \\ \vdots & & 1 \end{bmatrix}_{n\times n}$

Verdadero

$H = X(X'X)^{-1}X'$

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

H

$A^n = H^2 = H$; $n \in \mathbb{N}$

Parte práctica: $n = 500$

$Y_{500\times 1}$ X_1 X_2 X_3

Performance	Strength	Skills	Speed
71.98394	46.64215	62.54602	51.64761
69.90818	66.69022	67.26147	50.50888
64.82243	47.40409	55.15063	51.73342
72.17802	34.96857	69.47489	52.43953
74.75588	47.54257	57.83499	47.76798

$K = 3$ covariables

$X_{500\times 4}$

Intercept	Strength	Skills	Speed
1	46.64215	62.54602	51.64761
1	66.69022	67.26147	50.50888
1	47.40409	55.15063	51.73342

Para el vector de parámetros est.

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

$\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 = \hat{Y}$

64.82243	47.40409	55.15063	51.73342
72.17802	34.96857	69.47489	52.43953
74.75588	47.54257	57.83499	47.76798
52.53104	47.27276	53.28930	64.89863
64.61239	23.03113	73.87992	33.99096
67.45053	49.45705	73.19110	41.71503
72.56171	47.69065	81.74311	48.96745
81.93311	56.96206	63.53312	33.56811
62.28259	68.48956	76.12079	48.24146
65.02038	61.26565	78.26196	66.61470
64.87466	47.31111	76.67161	50.20886
61.98472	38.93474	57.61519	52.30701
66.52814	75.73360	72.24045	37.39835

⇒

1	66.69022	67.26147	50.50888
1	47.40409	55.15063	51.73342
1	34.96857	69.47489	52.43953
1	47.54257	57.83499	47.76798
1	47.27276	53.28930	64.89863
1	23.03113	73.87992	33.99096
1	49.45705	73.19110	41.71503
1	47.69065	81.74311	48.96745
1	56.96206	63.53312	33.56811
1	68.48956	76.12079	48.24146
1	61.26565	78.26196	66.61470
1	47.31111	76.67161	50.20886
1	38.93474	57.61519	52.30701

↓

solve(t(X) %*% X) %*% t(X) %*% Y

	V1
Intercept	70.69331750
Strength	-0.05090263
Skills	-0.01680885
Speed	0.04666696

β_0 β_1 β_2 4x1 β_3

Para la respuesta media: $\hat{Y} = X(X'X)^{-1}X'Y = HY = X\hat{\beta} = X(X'X)^{-1}X'Y$

H

Consultar en R

Para el cálculo errores est: $\hat{\varepsilon}_i = y_i - \hat{y}_i$; $\hat{\varepsilon} = Y - \hat{Y}$

Consultar en R

$\hat{\varepsilon}^t$ $\hat{\varepsilon}$

- $\hat{\sigma}^2 = \frac{SSE}{n-p} : \frac{\sum_{i=1}^{n}(Y_i - \hat{y}_i)^2}{n-p} = \frac{\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_k X_{ik})^2}{n-p} = \frac{(Y - X\hat{\beta})^t (Y - X\beta)}{n-p}$

$Y - X\hat{\beta}$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 70.69332    3.94693  17.911   <2e-16 ***
Strength    -0.05090    0.04580  -1.112    0.267
Skills      -0.01681    0.03095  -0.543    0.587
Speed        0.04667    0.04565   1.022    0.307
---
```

	V1
Intercept	70.69331750
Strength	-0.05090263
Skills	-0.01680885
Speed	0.04666696

β_0 β_1 β_2 4x1 β_3