EXERCISE

Consider the following multiple linear model:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \qquad i = 1, \dots, 20$$

with $\varepsilon_1, \dots, \varepsilon_{20}$ independent and identically distributed normal random variables with distribution $N(0, \sigma^2)$. Moreover, let

$x_{i1} = 0$ for $i = 1, \dots, 5$ and $x_{i1} = 1$ otherwise

$x_{i2} = 0$ for $i = 1, \dots, 10$ and $x_{i2} = 1$ otherwise

$x_{i3} = -1$ for $i = 1, \dots, 15$ and $x_{i3} = +1$ otherwise.

(a) indicate the sample and parameter space

(b) represent the model in matrix form $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, specifying $\underline{Y}, X, \underline{\beta}, \underline{\varepsilon}$, and the distribution of $\underline{\varepsilon}$.

(c) what is the dimension of the subspace $C(X)$ of $\mathbb{R}^n$ spanned by the columns of X?

(d) obtain the expressions of the matrix X^TX and of the vector X^Ty. Explain how they are used to derive the maximum likelihood estimate $\hat{\underline{\beta}}$ of $\underline{\beta}$.

(e) write the exact distribution of the estimators $\hat{\underline{\beta}}$ and $\hat{\beta}_1$.

(f) sketch how you would perform a test with significance level 0.05 to test the hypothesis $H_0: \beta_1 = 0$ vs $H_1: \beta_1 < 0$

(g) let $\underline{e} = \underline{y} - X\hat{\underline{\beta}}$ be the vector of residuals. Indicate which of the following equivalences are true (motivate).

(i) $\sum_{i=1}^{20} e_i = 0$ (iii) $\sum_{i=11}^{20} e_i = 0$

(ii) $\sum_{i=1}^{5} e_i = 0$ (iv) $\sum_{i=1}^{15} e_i = \sum_{i=16}^{20} e_i$

(a) indicate the sample and parameter space

Sample space: we have $n = 20$ realizations of Y_i

$\Rightarrow \mathcal{Y} = \mathbb{R}^{20}$

Parameter space: the parameters are $(\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2)$

$\Rightarrow \Theta = \mathbb{R}^4 \times (0, +\infty)$

(b) represent the model in matrix form $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$, specifying $\underline{Y}, X, \underline{\beta}, \underline{\varepsilon}$, and the distribution of $\underline{\varepsilon}$.

$\underline{Y}$ is a vector of random variables, dimension = 20 $\qquad \underline{Y} = [Y_1 \; Y_2 \; \dots \; Y_{20}]^T$

$\underline{\beta}$ is a vector of unknown constants, dimension = 4 $\qquad \underline{\beta} = [\beta_1 \; \beta_2 \; \beta_3 \; \beta_4]^T$

X is a $(n \times p) = (20 \times 4)$ matrix of known constants $\qquad X = [\underline{1}_{20} \; \underline{x}_1 \; \underline{x}_2 \; \underline{x}_3]$

$$= \begin{array}{c} \left[\begin{array}{cccc} 1 & 0 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & -1 \\ \hline 1 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & -1 \\ \hline 1 & 1 & 1 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & -1 \\ \hline 1 & 1 & 1 & +1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & +1 \end{array}\right] \end{array} \begin{array}{l} i=1 \\ \\ i=5 \\ i=6 \\ \\ i=10 \\ i=11 \\ \\ i=15 \\ i=16 \\ \\ i=20 \end{array} = \begin{bmatrix} \underline{1}_5 & \underline{0}_5 & \underline{0}_5 & -\underline{1}_5 \\ \underline{1}_5 & \underline{1}_5 & \underline{0}_5 & -\underline{1}_5 \\ \underline{1}_5 & \underline{1}_5 & \underline{1}_5 & -\underline{1}_5 \\ \underline{1}_5 & \underline{1}_5 & \underline{1}_5 & \underline{1}_5 \end{bmatrix}$$

$\underline{\varepsilon}$ is a vector of random variables, dimension = 20 $\qquad \underline{\varepsilon} \sim N_{20}(\underline{0}, \sigma^2 I_{20})$

(c) what is the dimension of the subspace $C(X)$ of $\mathbb{R}^n$ spanned by the columns of X?

The dimension of the column space of X, $C(X)$ is equal to the number of linearly independent vectors. Here, $\underline{1}, \underline{x}_1, \underline{x}_2$ and $\underline{x}_3$ are linearly independent $\Rightarrow \dim(C(X)) = 4$.

(Notice that if $\dim(C(X)) < 4$ it means that the covariates are collinear and you wouldn't be able to obtain $\hat{\beta}$)

(d) obtain the expressions of the matrix X^TX and of the vector X^Ty. Explain how they are used to derive the maximum likelihood estimate $\hat{\underline{\beta}}$ of $\underline{\beta}$.

$$\underset{\substack{4\times 20 \;\; 20\times 4 \\ 4\times 4}}{X^TX} = \begin{bmatrix} \underline{1}^T \\ \underline{x}_1^T \\ \underline{x}_2^T \\ \underline{x}_3^T \end{bmatrix} \cdot [\underline{1} \; \underline{x}_1 \; \underline{x}_2 \; \underline{x}_3] = \begin{bmatrix} 1^T1 & 1^Tx_1 & 1^Tx_2 & 1^Tx_3 \\ x_1^T1 & x_1^Tx_1 & x_1^Tx_2 & x_1^Tx_3 \\ x_2^T1 & x_2^Tx_1 & x_2^Tx_2 & x_2^Tx_3 \\ x_3^T1 & x_3^Tx_1 & x_3^Tx_2 & x_3^Tx_3 \end{bmatrix}$$

remember $a^Tb = \sum_{i=1}^{n} a_i b_i$

symmetric

$$= \begin{bmatrix} 20 & 15 & 10 & -10 \\ 15 & 15 & 10 & -5 \\ 10 & 10 & 10 & 0 \\ -10 & -5 & 0 & 20 \end{bmatrix}$$

$$\underset{\substack{4\times 20 \;\; 20\times 1 \\ 4\times 1}}{X^Ty} = \begin{bmatrix} \underline{1}^T \\ \underline{x}_1^T \\ \underline{x}_2^T \\ \underline{x}_3^T \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{20} \end{bmatrix} = \begin{bmatrix} 1^Ty \\ x_1^Ty \\ x_2^Ty \\ x_3^Ty \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{20} y_i \\ \sum_{i=6}^{20} y_i \\ \sum_{i=11}^{20} y_i \\ -\sum_{i=1}^{15} y_i + \sum_{i=16}^{20} y_i \end{bmatrix}$$

The MLE $\hat{\underline{\beta}}$ is obtained as $\hat{\underline{\beta}} = (X^TX)^{-1}X^T\underline{y}$

(e) write the exact distribution of the estimators $\hat{\underline{\beta}}$ and $\hat{\beta}_1$.

$\hat{\underline{\beta}}(\underline{Y}) \sim N_4(\underline{\beta}, \sigma^2(X^TX)^{-1})$ $\qquad$ the marginal $\hat{\beta}_1(\underline{Y}) \sim N(\beta_1, \sigma^2[(X^TX)^{-1}]_{(2,2)})$

(f) sketch how you would perform a test with significance level 0.05 to test the hypothesis $H_0: \beta_1 = 0$ vs $H_1: \beta_1 < 0$

We want to test $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 < 0 \end{cases}$

value assumed under $H_0 \rightarrow b = 0$

The test statistic is $\dfrac{\hat{\beta}_1 - b}{\sqrt{\hat{V}(\hat{\beta}_1)}} \overset{H_0}{\sim} t_{n-p}$

$20 - 4 = 16$

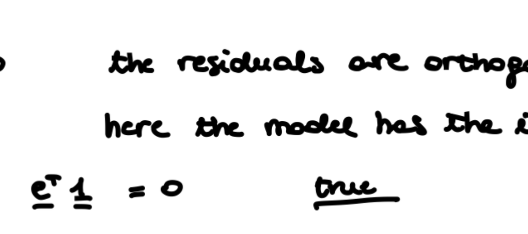
estimator of the variance of the estimator $\hat{\beta}_1$ $= s^2[(X^TX)^{-1}]_{22}$

Hence, $T = \dfrac{\hat{\beta}_1}{\sqrt{s^2[(X^TX)^{-1}]_{2,2}}} \overset{H_0}{\sim} t_{16}$

In this case we only reject for negative values of $\hat{\beta}_1 \Rightarrow$ negative (large) values of T

Critical region: $T < k$

$\Rightarrow \alpha = 0.05 = \mathbb{P}_{H_0}(T < k) \Rightarrow k = t_{16;\alpha}$ = quantile of level α of a Student's t distribution with 16 degrees of freedom.



$R = (-\infty; t_{16;0.05})$ critical region.

$\Rightarrow$ reject H_0 if $t^{obs} \in R$

(g) let $\underline{e} = \underline{y} - X\hat{\underline{\beta}}$ be the vector of residuals. Indicate which of the following equivalences are true (motivate).

(i) $\sum_{i=1}^{20} e_i = 0$ $\qquad$ the residuals are orthogonal to the vectors $\in C(X)$: if $\underline{a} \in C(X) \Rightarrow \underline{e}^T\underline{a} = 0$

here the model has the intercept $\Rightarrow \underline{1}_{20} \in C(X)$

$\sum_{i=1}^{20} e_i = \underline{e}^T\underline{1} = 0$ $\qquad$ <u>true</u>

(ii) $\sum_{i=1}^{5} e_i = 0$ $\qquad$ we have $\sum_{i=1}^{5} e_i = \sum_{i=1}^{20} e_i - \sum_{i=6}^{20} e_i = \sum_{i=1}^{20} e_i \cdot 1 - \sum_{i=1}^{20} e_i \cdot x_{i1} =$

$= \underbrace{\underline{e}^T\underline{1}}_{=0} - \underbrace{\underline{e}^T\underline{x}_1}_{=0} = 0$ $\qquad$ <u>true</u>

(iii) $\sum_{i=11}^{20} e_i = 0$

$\sum_{i=11}^{20} e_i = \sum_{i=1}^{20} e_i x_{i2} = \underline{e}^T\underline{x}_2 = 0$ $\qquad$ <u>true</u>

(iv) $\sum_{i=1}^{15} e_i = \sum_{i=16}^{20} e_i$

$\sum_{i=1}^{15} e_i = \sum_{i=1}^{20} e_i - \sum_{i=16}^{20} e_i = \cancel{\underline{e}^T\underline{1}} - \sum_{i=16}^{20} e_i = -\sum_{i=16}^{20} e_i$

(iv) is true $\Leftrightarrow \sum_{i=1}^{15} e_i = \sum_{i=16}^{20} e_i = 0$

$\sum_{i=16}^{20} e_i = \sum_{i=1}^{20} e_i (1 + x_{i3}) \cdot \frac{1}{2} = \frac{1}{2}(\underline{e}^T\underline{1} + \underline{e}^T\underline{x}_3) = 0$

indeed, $\underline{1} + \underline{x}_3 = \begin{bmatrix} \underline{1}_{15} \\ \underline{1}_5 \end{bmatrix} + \begin{bmatrix} -\underline{1}_{15} \\ \underline{1}_5 \end{bmatrix} = \begin{bmatrix} \underline{0}_{15} \\ \underline{2}_5 \end{bmatrix}$