Regression Analysis Assignment1

November 1, 2021

Assumption. $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^t$, $\mathbf{1} = (1, \ldots, 1)^t, \mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^t$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^t$ are n-dimensional vectors, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_n)^t$ is a $(p+1)$-dimensional vector, $\mathbf{X} = (\mathbf{1}, \mathbf{x_1}, \ldots, \mathbf{x_p})$ is a $n \times (p+1)$ known matrix of full column rank and $\boldsymbol{\epsilon} \sim \boldsymbol{N}_n(0, \sigma^2 I_n)$ with $\sigma^2 > 0$.

Problem1. Let $\mathbf{A}$ be a $(p+1) \times q$ matrix of full column rank and $\mathbf{c}$ be q-dimensional vector($q < p+1$). Let $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}$ be the least squared estimator with constraint $\boldsymbol{A^t\beta = c}$. It means

$$\hat{\beta}_{\mathbf{r}} = \underset{\boldsymbol{A^t\beta=c}}{\arg\min} ||\boldsymbol{Y} - \boldsymbol{X\beta}||$$

Find the $\hat{\boldsymbol{\beta}}_{\boldsymbol{r}}$ using the Lagrangian multiplier method.

Problem2. Let $\mathbf{X}_i$ be i-th row of $\mathbf{X}$ and $\hat{\boldsymbol{\beta}}$ be the least squared estimator based on $(\mathbf{X}_1, Y_1), \ldots (\mathbf{X}_n, Y_n)$. Suppose $Y_i = \mathbf{X}_i^t\hat{\boldsymbol{\beta}}$ for $i = 1, 2, \ldots, k$. Then, show that the least squared estimator $\hat{\boldsymbol{\beta}}_{(-)}$ based on $(\mathbf{X}_{k+1}, Y_{k+1}), \ldots (\mathbf{X}_n, Y_n)$ equals to $\hat{\boldsymbol{\beta}}$.

Problem3. Derive $100(1-\alpha)\%$ confidence interval for mean response and out-of-sample response, referring to slide 16.