625.661 Statistical Models and Regression

Module 4 Discussion Questions

H.M. James Hung

In a multiple linear regression model, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where x_1 and x_2 are non-random independent variables, and ε is the random error. Now a set of n items give data, $(y_1, x_{11}, x_{21}), \ldots, (y_n, x_{1n}, x_{2n})$, which follow this model. Decompose SS_T into SS_R and SS_{Res} using the "hat" matrix, discuss and **state the assumptions** in your discussion.

Assumption: X is of full rank (e.g., all columns and all rows are linearly independent).

Let $\boldsymbol{\beta} = (\boldsymbol{\beta_0}, \ \boldsymbol{\beta_1}, \ \boldsymbol{\beta_2})'$.

Then, the uncorrected $SS_R = SS_R(\boldsymbol{\beta_0}, \ \boldsymbol{\beta_1}, \ \boldsymbol{\beta_2}) = \hat{\boldsymbol{\beta}}' X' y = y' X (X'X)^{-1} X' y = y' H y$.

SS_T = y'y = y'Iy = y'Hy + y'(I – H)y ***SS_R = y'Hy*** ***SS_{Res} = y'(I – H)y***

H = X(X'X)$^{-1}$X' **Assumption: *X'X* is invertible.**

The corrected *SS_T* = $y'y - \frac{(\sum y_i)^2}{n}$. ***SS_R* = $y'Hy - \frac{(\sum y_i)^2}{n}$.**

***SS_{Res}* = *y'(I – H)y* .**