Due 6 September 2015

Graded Homework Problem (70 points total)

For this assignment use the following matrix definition for the components of a *multiple linear regression* using *ordinary least squares* (OLS) and calculating several measures of statistical interest.

The formulas are as follows

$$\begin{aligned} H &= (X^T \cdot X)^{-1} \\ B &= H \cdot (X^T \cdot Y) \\ P &= X \cdot B \\ E &= Y - P \\ SSE &= E^T \cdot E \\ V &= \frac{1}{n-m} \cdot SSE \\ SE_{i,1} &= \sqrt{V \cdot H_{i,i}} \end{aligned}$$

For this data, $n = 20$ (the number of data rows in X and Y) and $m = 4$ (the number of data columns in X). Y has 1 column.

In the case where one or more subscripts are used in a formula, you should identify all the possible values of each subscript to determine the size of the final result.

(HINT: SE is a matrix with 1 column.)

Be prepared to identify the following values in eLearning (5 points for each answer)

(a) the size of B (rows and columns)

(b) the size of P (rows and columns)

(c) the size of E (rows and columns)

(d) the size of SSE (rows and columns)

(e) the size of V (rows and columns)

(f) the size of SE (rows and columns)

(g) the size of H (rows and columns)

Each of the above listed values for each entry will get full credit of no credit, no partial credit will be given.