**Statistics 108**, Homework 7

*Due: December 4th, 2017, In Class* (turn in paper form)

*You need to show the steps to get the full credits.

This homework is to practice on model selection. Attach the complete R codes for Problem 2 at the end of the homework. Total: 90 points.

1. (30 points) *Understanding model selection criteria and procedures.* Say whether the following statements are true or false and explain why.
   (a) The more number of predictor variables in the model, the larger the $R^2$.
   (b) For model of the same size, their $C_p, AIC_p, BIC_p$ values are monotonically increasing with $SSE_p$.
   (c) For model of the same size, their $PRESS_p$ values are monotonically increasing with $SSE_p$.
   (d) Compared with $AIC_p$, $BIC_p$ criterion tends to select smaller models because it puts more penalty on model size.
   (e) The best subsets procedure is guaranteed to find the "best" model under a given criterion.
   (f) The forward stepwise procedure is guaranteed to find the "best" model under a given criterion.

2. (40 points) *Practice model selection on an example data set.* Data set "HW7Q2.txt" contains 4 variables with the response variable $Y$ on the first column followed by 3 predictor variables. We consider all first-order models.
   (a) (5 pt) How many first-order models are there?
   (b) (20 pt) Among all the first-order models, report the "best" models according to each of the following criteria: $R^2_{a,p}, AIC_p, BIC_p, C_p, PRESS_p$, as well as their corresponding values according to the criterion.
   (c) (15 pt) Start from the none-model (consisting of no predictor, only the intercept), if we use the forward stepwise procedure, what model you end up with? How about if we use the forward selection procedure?

3. (20 points) *Rigorous proof.* Consider the multiple linear regression model in the matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I_n$. Let $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ be the hat matrix. Show that the diagonal elements of $H$ are all between 0 and 1.