Math 583 HW 5 Fall 2017. Due Wednesday, Oct. 4.
Quiz 5 on Friday, Oct. 6 is similar to HW 5. Use 3 sheets of notes.

Problem numbers are from the Olive text.

**A) 3.4.** Suppose $\hat{\boldsymbol{Y}} = \boldsymbol{HY}$ where $\boldsymbol{H}$ is an $n \times n$ hat matrix. Then the degrees of freedom $df(\hat{\boldsymbol{Y}}) = tr(\boldsymbol{H})$ = sum of the diagonal elements of $\boldsymbol{H}$. An estimator with low degrees of freedom is inflexible while an estimator with high degrees of freedom is flexible. If the degrees of freedom is too low, the estimator tends to underfit while if the degrees of freedom is to high, the estimator tends to overfit.

a) Find $df(\hat{\boldsymbol{Y}})$ if $\hat{\boldsymbol{Y}} = \overline{Y}\boldsymbol{1}$ which uses $\boldsymbol{H} = (h_{ij})$ where $h_{ij} \equiv 1/n$ for all $i$ and $j$. This inflexible estimator uses the sample mean $\overline{Y}$ of the response variable as $\hat{Y}_i$ for $i = 1, ..., n$.

b) Find $df(\hat{\boldsymbol{Y}})$ if $\hat{\boldsymbol{Y}} = \boldsymbol{Y} = \boldsymbol{I}_n\boldsymbol{Y}$ which uses $\boldsymbol{H} = \boldsymbol{I}_n$ where $h_{ii} = 1$. This bad flexible estimator interpolates the response variable.

**B) 3.5.** Suppose $\boldsymbol{Y} = \boldsymbol{X\beta}+\boldsymbol{e}$, $\boldsymbol{Z} = \boldsymbol{W\eta}+\boldsymbol{e}$, $\hat{\boldsymbol{Z}} = \boldsymbol{W}\hat{\boldsymbol{\eta}}$, $\boldsymbol{Z} = \boldsymbol{Y}-\overline{\boldsymbol{Y}}$, and $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{Z}}+\overline{\boldsymbol{Y}}$. Let the $n \times p$ matrix $\boldsymbol{W}_1 = [\boldsymbol{1} \;\; \boldsymbol{W}]$ and the $p \times 1$ vector $\hat{\boldsymbol{\eta}}_1 = (\overline{Y} \;\; \hat{\boldsymbol{\eta}}^T)^T$ where the scalar $\overline{Y}$ is the sample mean of the response variable. Show $\hat{\boldsymbol{Y}} = \boldsymbol{W}_1\hat{\boldsymbol{\eta}}_1$.

**C) 2.8** Consider the output above Olive (2017) Example 2.7 for the minimum $C_p$ forward selection model based on the residual bootstrap.

a) What is $\hat{\boldsymbol{\beta}}_{I_{min}}$?

b) What is $\hat{\boldsymbol{\beta}}_{I_{min},0}$?

c) The large sample 95% shorth CI for $H$ is [0,0.016]. Is $H$ needed is the minimum $C_p$ model given that the other predictors are in the model?

d) The large sample 95% shorth CI for $\log(S)$ is [0.324,0.913] for all subsets. Is $\log(S)$ needed is the minimum $C_p$ model given that the other predictors are in the model?

e) Suppose $x_1 = 1$, $x_4 = H = 130$, and $x_5 = \log(S) = 5.075$. Find $\hat{Y} = (x_1 \, x_4 \, x_5)\hat{\boldsymbol{\beta}}_{I_{min}}$. Note that $Y = \log(M)$.

**D) 3.19.** This simulation is similar to that used to form Table 2.2, but 1000 runs are used so coverage in [0.93,0.97] suggests that the actual coverage is close to the nominal coverage of 0.95.

The model is $Y = \boldsymbol{x}^T\boldsymbol{\beta} + e = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + e$ where $\boldsymbol{\beta}_S = (\beta_1, \beta_2, ..., \beta_{k+1})^T = (\beta_1, \beta_2)^T$ and $k = 1$ is the number of active nontrivial predictors in the population model. The output for *test* tests $H_0 : (\beta_{k+2}, ..., \beta_p)^T = (\beta_3, ..., \beta_p)^T = \boldsymbol{0}$ and $H_0$ is true. The output gives the proportion of times the prediction region method bootstrap test fails to reject $H_0$. The nominal proportion is 0.95.

After getting your output, make a table similar to Table 2.2 with 4 lines. If your $p = 5$ then you need to add a column for $\beta_5$. Two lines are for reg (the OLS full model) and two lines are for vs (forward selection with $I_{min}$). The $\beta_i$ columns give the coverage and lengths of the 95% CIs for $\beta_i$. If the coverage $\geq 0.93$, then the shorter CI length is more precise. Were the CIs for forward selection more precise than the CIs for the OLS full model for $\beta_3$ and $\beta_4$?

To get the output, copy and paste the source commands from (http://parker.ad.siu.edu/Olive/slrhw.txt) into $R$. Copy and past the library command for this problem into $R$.

If you are person $j$ then copy and paste the $R$ code for person $j$ for this problem into $R$.

You are person $j = 1$ .