Math 583 HW 9 Fall 2017. Due Wednesday, Nov. 8.
Quiz 9 on Friday, Nov. 10 is similar to HW 9. Use 4 sheets of notes.

Problem numbers and example numbers are from the Olive text.
A) 6.1. Suppose
$$\boldsymbol{R} = \begin{bmatrix} 1 & 0.4 & 0.8 \\ 0.4 & 1 & 0.5 \\ 0.8 & 0.5 & 1 \end{bmatrix}.$$

a) Find $\boldsymbol{R}_{\delta=1}$.

b) Find $\boldsymbol{R}(\delta = 1, \tau = 0.3)$.

Do the two source commands to get the data into R.

B) 5.14 For the Gladstone (1905) data, the response variable $Y = gender$, gives the group (0-F, 1-M). The predictors are $x_1 = age$, $x_2 = log(age)$, $x_3 = breadth$ of head, x_4 and x_5 are indicators for *cause* of death coded as a factor, $x_6 = cephalic\ index$ (a head measurement), $x_7 = circumference$ of head, $x_8 = height$ of the head, $x_9 = height$ of the person, $x_{10} = length$ of head, $x_{11} = size$ of the head, and $x_{12} = log(size)$ of head. The sample size is $n = 267$.

a) The R code for this part does backward elimination for logistic regression. Backward elimination should only be used if $n \geq Jp$ with $J \geq 5$ and preferably $J \geq 10$.

Include the coefficients for the selected model (given by the `summary(back)` command) in *Word*. (You may need to do some editing to make the table readable.)

b) The R code for this part gives the response plot for the backward elimination submodel I_B. Does the response plot look ok?

c) Use the R code for this part to give the AER for I_B.

d) Use the R code for this part to give a validation ER for I_B.

(Another validation ER would apply backward elimination on the cases not in the validation set. We just used the variables from the backward elimination model selected using the full data set. The first method is likely superior, but the second method is easier to code.)

e) These R commands will use lasso with a classification criterion. We got rid of the factor (two indicator variables) since `cv.glmnet` uses a matrix of predictors. Lasso can handle indicators like gender as a response variable, but will not keep or delete groups two or more indicators that are needed for a quantitative variable with 3 or more levels. These commands give the k-fold CV error rate for the lasso logistic regression. What is it?

f) Use the commands for this part to get the relaxed lasso response plot where relaxed lasso uses the lasso from part e). Include the plot in *Word*.

g) Use the commands from this plot to make the EE plot of the ESP from lasso variable selection (ESPRL) versus the ESP from lasso (ESPlasso).