1. (30 pts.) The data frame `HSWRESTLER`, (from package `PASWR2`) contains information on nine variables for a group of 78 high school wrestlers that was collected by the human performance lab at Appalachian State University. It is of interest to predict wrestler's hydrostatic fat (`hwfat`) using predictors `age,ht,wt,abs,triceps` and `subscap`. Split the data set into a training set and a test set (50%). Use `set.seed(1)` each time you need to use `sample()` or `cv.glmnet()` functions.

   (a) Fit a linear model using least squares on the training set, and report the test error obtained.

   (b) Fit a ridge regression model on the training set, with λ chosen by cross-validation (use 15-fold cross validation). Report the test mspe.

   (c) Fit a lasso model on the training set, with λ chosen by cross-validation (use 15-fold CV). Report the test error obtained.

2. (30 pts.) Generate simulated data, and will then use this data to perform best subset selection. Generate values of a predictor X of length $n = 100$, using `X = rnorm(n)`. Generate values of a noise vector ϵ of length $n = 100$ using ϵ `= 0.1*rnorm(n)` Generate a response vector Y of length n = 100 using

$$Y = 1 - 0.1X + +0.05X^2 + 0.75X^3 + \epsilon$$

   (a) Use `regsubsets()` to choose the best model containing the predictors $X, X^2, ..., X^{10}$. What is the best model obtained according to adjusted R^2?

   (b) Fit a lasso model to the simulated data, again using $X, X^2, ..., X^{10}$ as predictors. Use 10-fold cross-validation to select the optimal value of λ. Report the test error.

3. (40 pts.) A real estate appraiser is interested in predicting residential home prices in a mid-western city as a function of various features. For that purpose a regression model is to be constructed from a sample of 522 houses. Use the `homes.xls` data set from blackboard. Consider the predictors
   x_1: lot size (square feet), x_2: area (square feet), x_3: number of bedrooms,
   x_4: number of bathrooms, x_5: year of construction, x_6: garage size (number of cars).
   Split the data set into a training set and a test set (50%).

   (a) Fit a ridge regression model on the training set, with λ chosen by cross-validation (use 15-fold cross validation). Report the test mspe.

   (b) Fit a lasso model on the training set, with λ chosen by cross-validation (use 15-fold CV). Report the test error obtained.

   (c) Predict the price when all predictors are equal to their median values using both the ridge regression and the lasso models.