Due on October 5th, 2018 (No Quiz)

1. (30 pts.) The dataframe `VIT2005` in the `PASWR2` package contains information about apartments in Vitoria, Spain. Let us assume that it is of interest to predict `totalprice`.

   a) Use AIC criterion to find the best model, `m1` (use all the data set).
   Plot the fitted values (y-axis) vs. `totalprice`.

   b) Use BIC criterion to find the best model, `m2` (use all the data set).
   Plot the fitted values (y-axis) vs. `totalprice`.

   c) Use `cv.glm()` function to find the 5-fold cross validation MSPE of the models `m1` and `m2`.

2. (40 pts.) The maintenance of swimming pools is quite costly because of all the chlorine that is needed to keep the water clear and relatively free of germs. A chain of hotels (all with outdoor pools) seeking to reduce costs decided to analyze the factors that determine how much chlorine is needed. They commissioned a chemist to conduct an analysis. It is believed that the speed at which chlorine in a pool is depleted is dependent on the temperature of the water (higher temperature uses chlorine faster); pH level, which is a measure of the acidity of the water (pH ranges from 0 to 14, where 0 is very acidic and 14 is very alkaline; levels around 7.5 use the least chlorine); and weather (sunshine uses up chlorine). The chemist conducted the following experiment. The percentage of chlorine depletion during 8-hour days was recorded under varying conditions of pH level, water temperature, and weather conditions. The file `chlorine` on Blackboard contains the following data: Percentage of chlorine depletion over 8 hours, Temperature (Fahrenheit), pH level, and, weather, coded as, 1 = Mainly cloudy, 2 = Sunny, 3 = Partly sunny.

   a) Fit a full model (linear in all variables, but quadratic in the pH level, see p. 289 text).

   b) Can we infer that higher temperatures deplete chlorine more quickly?

   c) Is there evidence to conclude that the belief about the relationship between chlorine depletion and pH level is correct?

   d) Can we conclude that weather is a factor in chlorine depletion?

3. (30 pts.) A student analyzed the relationship between the viewers memory of the commercials product and two predictors, the length and type of the commercial. Three types of commercial were considered, humorous (1), musical (2), and serious (3). Use the file `commercial` on Blackboard.

   a) Fit a regression model using the numeric codes provided for the type of commercial. With this model can we conclude that the memory test score is related to the type of commercial? why?

   b) Fit a regression model where the type of commercial is included as a factor. With this model can we conclude that the memory test score is related to the type of commercial? why?

   c) Report the R^2 and adj-R^2. of the models in parts (a) and (b).