HW 4

1. (14pt) A national insurance organization wanted to study the consumption patter on cigarettes in all 50 states and the District of Columbia. The variables chosen for the study are

| Variable | Definition |
|---|---|
| Age | Median age of the a person living in a state |
| HS | Percentage of people over 25 years of age in a state who had completed high school |
| Income | Per capita personal income for a state in dollars |
| Black | Percentage of blacks living in a state |
| Female | Percentage of females living in a state |
| Price | Weighted average price (in cents) of a pack of cigarettes in a state |
| Sales | Number of packs of cigarettes sold in a state on a per capita basis |

   (a) (2pt) What would you expect the relationship between Sales and each of the explanatory variables to be? explain.

   (b) (2pt) Compute the pairwise correlation coefficient matrix and construct the corresponding scatter plot matrix

   (c) (2pt) Obtain the six variance inflation factors. What do these results suggest about the effect of multicollinearity?

   (d) (2pt) Are the there any outlying Sales observations in the regression model relating Sales to the six predictors?

   (e) (2pt) Obtain the leverages (the diagonal elements of the hat matrix). Are the there any outlying states in the six predictors?

   (f) (2pt) Are there any influential points?

   (g) (2pt) Use log(Sales) instead of Sales and repeat questions d), e) and f).

2. (6pt) Suppose that North American Oil Company is attempting to develop a regular gasoline that will deliver improved gasoline mileage. As part of its development process, the company would like to study the effect of two qualitative factors on the gasoline mileage obtained by an automobile called the Fire-Hawk. These factors are regular gasoline type (which has levels A, B and C) and gasoline additive type (which has levels M, N, O and P). To carry out the study, the company test drove three Fire-Hawks using each treatment. However upon completion of the experiment the company found that several

Fire-Hawks have not been driven under the proper test conditions. Rather than running more tests, the company decided (because of limited time) to analyze the data that had remained after the data for improperly tested Fire-Hawks was dropped from the data set. The remaining data are in Oildata.cvs. Let

$y_{ij,k}$ = the kth gasoline mileage obtained when using regular gasoline type i and additive type j

A reasonable model to use for this data is

$$y_{ijk} = \mu + \alpha_B D_{i,B} + \alpha_C D_{i,C} + \beta_N D_{j,N} + \beta_O D_{j,O} + \beta_P D_{j,P} + \epsilon_{ij,k}$$

where

$D_{i,B}$ = 1 if i = B, that is, if we are using gasoline type B and 0 otherwise
$D_{i,C}$ = 1 if i = C, that is, if we are using gasoline type C and 0 otherwise
$D_{j,N}$ = 1 if j = N, that is, if we are using additive type N and 0 otherwise
$D_{j,O}$ = 1 if j = O, that is, if we are using additive type O and 0 otherwise
$D_{j,P}$ = 1 if j = P, that is, if we are using additive type P and 0 otherwise

(a) (3pt) To compare the effects of the regular gasoline type, we need to test $H_0 : \alpha_B = \alpha_C = 0$ versus $H_a$ : at least one of $\alpha_B$ or $\alpha_C$ does not equal zero.

(b) (3pt) To compare the effects of the gasoline additive type, we need to test $H_0 : \beta_N = \beta_O = \beta_P = 0$ versus $H_a$ : at least one of $\beta_N, \beta_O, \beta_P$ does not equal zero.