HW 6

1. You are presented with data about the sale price of 26 apartments that have sold recently. Here is the first three rows of the data

```
> foo<-read.table(
+     "http://users.stat.umn.edu/~gmeeden/classes/5021/datasets/mnsales.txt",
+     header=TRUE)
> foo[1:3,]

  Code      y x1 x2    x3 x4    x5 Condition
1  229  90300  4 82  4365  0  4266         F
2   94 384000 20 13 17798  0 14391         G
3   43 157500  5 66  5913  0  6615         G
```

You are informed that y is the sale price, $x1$ is the number of units in the complex and $x2$ is the age of the complex. *Condition* takes on 3 possible values F (fair), G (good) and E (excellent) which described the unit at the time of the sale. What the other three variables are measuring is unknown to you. Using these data find a model that you think could be useful in predicting the sale price on an apartment.

2. Phosphorus used in soil fertilizers can contaminate freshwater sources during rainfall runoff. Hence it is of interest to investigate the relationship between x, soil loss, and y, dissolved phophorus in water samples for data collected in 20 watersheds. The data is in phosphor.txt at the class web site. Plot the data and discuss wheter or not a linear or quadratic model seems to fit the data.

3. A study was carried out to see if one can predict ice cream consumption using price, income and temperature. For three years Ice cream consumption was measured over 30 four-week periods. For each period the the following data were collected.

- IC: Ice cream consumption in pints per capita
- Pr: Price of ice cream per pint in dollars
- Inc: Average weekly family income in dollars
- Tmp: Mean temperature in degrees F.
- Yr: Year within the study (0 = 1996, 1 = 1997, 2 = 1998)

Consider the model

$$IC = \beta_0 + \beta_{Pr}Pr + \beta_{Inc}Inc + \beta_{Tmp}Tmp + \beta_{Yr}Yr + Z$$

i) In the full model at level $\alpha = 0.05$ test $H : \beta_{Tmp} = \beta_{Yr} = 0$ against K: At least one is not zero.

ii) In the full model at level $\alpha = 0.05$ test $H : \beta_{Pr} = 0$ against $K :\neq 0$

iii) In the model $IC = \beta_0 + \beta_{Pr}Pr + Z$ at level $\alpha = 0.05$ test $H : \beta_{Pr} = 0$ against $K :\neq 0$.

iv) Based on this output can you suggest a model which could be useful for predicting ice cream consumption? Briefly justify your answer.

To access the data use the same URL as in problem 1 but replace mnsales.txt with icecream.txt.

4. A firm has developed a new type of light bulb and is interested in investigating the drop in light output as a % of original output, y. It is know that y depends on the cleanliness of the surface, x_1, and the length of time the bulb has been in operation, x_2. There are two possible values for x_1 are 0 and 1. Zero denotes a clean bulb and one denotes a dirty bulb. The URL for the data is lightbulb.txt. Use the data to build a regression model that relates y to some function of x_1 and x_2. Include a residual analysis.

5. A supermarket chain is interested in exploring the relationship between sales of its store-brand canned vegetables, y, and the amount spent on advertising, x_1, and the amount of alloted shelf space, x_2. One of the chain's supermarkets was randomly selected, and over a 20-week period y, x_1 and x_2 were observed. The URL for the data is canveg.txt.

a. Fit the following model to the data;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + Z$$

b. Conduct an F-test to investigate the overall usefulness of this model at $\alpha = 0.05$.

c. Test $\beta_3 = 0$ at $\alpha = 0.05$.

d. Explain what it means to say that advertising expenditures and shelf space interact.

e. Explain how you could be misled by using a first-order model instead of an interaction model to explain how advertising expenditures and shelf space influence sales.