

Problem Set 6

Note: the data sets for the problems below are available in the file 'PS6 data sets.xls'.

Problem 1

A real estate appraiser is interested in developing a regression model to predict the fair market value of houses in a particular town. She collected the data shown in the table below. The appraiser wants to determine if the selling price of the houses can be accounted for by the total square footage of living area, the size of the garage and the number of bedrooms.

| Sq. Feet (in 000s) | Garage | Bedrooms | Price (in \$000s) |
|-----------------------|---------|----------|----------------------|
| 1,000 | none | 2 | 65 |
| 1,100 | none | 2 | 73 |
| 1,150 | one-car | 2 | 85 |
| 1,400 | none | 3 | 87 |
| 1,700 | one-car | 3 | 98 |
| 1,800 | one-car | 4 | 105 |
| 1,900 | none | 3 | 95 |
| 1,900 | one-car | 4 | 125 |
| 2,100 | two-car | 4 | 125 |
| 2,100 | two-car | 4 | 137 |
| 2,300 | two-car | 4 | 150 |

Problem 2

Roger Gallagher owns a used car lot that deals solely in used Corvettes. He would like to develop a regression model to help predict the price he can expect to receive for the cars he owns. He collected the following data concerning a number of cars he has sold in recent months.

| Mileage | Year | T-Top | Price |
|---------|------|-------|----------|
| 115 | 1968 | yes | \$13,875 |
| 95 | 1970 | no | \$11,000 |
| 125 | 1972 | no | \$8,000 |
| 85 | 1974 | yes | \$14,950 |
| 77 | 1976 | yes | \$15,625 |
| 105 | 1978 | no | \$11,300 |
| 88 | 1979 | no | \$13,250 |
| 73 | 1981 | yes | \$16,500 |
| 55 | 1983 | no | \$16,500 |
| 65 | 1987 | yes | \$19,500 |
| 45 | 1988 | no | \$22,300 |
| 15 | 1988 | no | \$25,500 |
| 23 | 1991 | yes | \$31,900 |

Let Y = selling price, X1 = mileage, X2 = model year, and X3 = whether or not the car had a t-top.

- a) If Roger wants to use a simple linear regression function to estimate the selling price of a car, which X variable do you recommend he use?
- b) Determine the parameter estimates for the regression function represented by:

$$Y = a + b_1 * X_1 + b_2 * X_2$$
 What is the estimated regression function? Does X2 help to explain the selling price of the cars if X1 is also in the model? What might be the reason for this?
- c) Set up a binary variable (X3) to indicate whether or not each car in the sample has a t-top. Determine the parameter estimates for the regression function represented by:

$$Y = a + b_1 * X_1 + b_2 * X_3$$
 Does X3 help to explain the selling price of the cars if X1 is also in the model? Explain.
- d) According to the previous model, on average, how much does a t-top add to the value of a car?
- e) Determine the parameter estimates for the regression function represented by:

$$Y = a + b_1 * X_1 + b_2 * X_2 + b_3 * X_3$$
- f) Of all the regression functions considered here, which do you recommend Roger use?

Problem 3

A marketing manager wishes to forecast corporate sales for the coming year and to better understand the factors that influence them. The marketing manager wants the forecast for at least four reasons:

- (1) to obtain estimates needed as part of the corporate planning activity;
- (2) to get an idea of the staffing requirements in sales and sales service to handle the company's increased sales;
- (3) to help in planning budget allocations for advertising, dealer discounts, and so on; and
- (4) to help make better policy decisions concerning price, advertising, and product development expenditures.

With this initial identification of the problem, the marketing manager sits down with the sales manager and others in her organization to determine the factors that might affect the company's sales. They come up with the following:

Sales depend upon

personal disposable income,
 dealers' allowances,
 prices,
 product development expenditures,
 capital investments,
 advertising,
 sales expenses,
 total industry advertising,
 and some random effects.

The marketing manager recognizes that some of these factors will have a more important effect than others on the company's sales; others may turn out to be unimportant. Since any one of them, however, may have an important impact, data are gathered on all of them at this early stage in the process.

The file 'PS6 data sets.xls' includes semiannual data, covering the period from 1980 through 1999, on nine variables (eight independent variables and the dependent variable SALES). A list of the nine variables is given below:

1. Personal Disposable Income in millions of dollars. (PDI)
 2. Dealers' Allowances in thousands of dollars. (DEALS)
 3. Price in dollars. (PRICE)
 4. Product Development in thousands of dollars. (R&D)
 5. Capital Investments in thousands of dollars. (INVEST)
 6. Advertising in thousands of dollars. (ADVERTIS)
 7. Sales Expenses in thousands of dollars. (EXPENSE)
 8. Total Industry Advertising in thousands of dollars. (TOTINDAD)
 9. Company Sales in thousands of dollars. (SALES)
-
- a) Generate and study the correlation matrix. What do the various correlation coefficients signify?
 - b) Find the most appropriate regression equation for SALES. How do you interpret this equation?