8.1 Property Valuation: Scientific mass appraisal is a technique in which linear regression methods applied to the problem of property valuation. The objective in scientific mass appraisal is to predict the sale price of a home from selected physical characteristics of the building and taxes (local, school, county) paid on the building. Twenty-four observations were obtained from *Multiple Listing* (Vol. 87) for Erie. PA, which is designated as Area 12 in the directory. These data (Table 8.17) were originally presented by Narula and Wellington (1977). The list of variables are given in Table 8.18

Table 8.17 Building Characteristics and Sales Price

Row	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	Y
1	4.918	1.000	3.472	0.998	1.0	7	4	42	0	25.90
2	5.021	1.000	3.531	1.500	2.0	7	4	62	0	29.50
3	4.543	1.000	2.275	1.175	1.0	6	3	40	0	27.90
4	4.557	1.000	4.050	1.232	1.0	6	3	54	0	25.90
5	5.060	1.000	4.455	1.121	1.0	6	3	42	0	29.90
6	3.891	1.000	4.455	0.988	1.0	6	3	56	0	29.90
7	5.898	1.000	5.850	1.240	1.0	7	3	51	1	30.90
8	5.604	1.000	9.520	1.501	0.0	6	3	32	0	28.90
9	5.828	1.000	6.435	1.225	2.0	6	3	32	0	35.90
10	5.300	1.000	4.988	1.552	1.0	6	3	30	0	31.50
11	6.271	1.000	5.520	0.975	1.0	5	2	30	0	31.00
12	5.959	1.000	6.666	1.121	2.0	6	3	32	0	30.90
13	5.050	1.000	5.000	1.020	0.0	5	2	46	1	30.00
14	8.246	1.500	5.150	1.664	2.0	8	4	50	0	36.90
15	6.697	1.500	6.902	1.488	1.5	7	3	22	1	41.90
16	7.784	1.500	7.102	1.376	1.0	6	3	17	0	40.50
17	9.038	1.000	7.800	1.500	1.5	7	3	23	0	43.90
18	5.989	1.000	5.520	1.256	2.0	6	3	40	1	37.90
19	7.542	1.500	5.000	1.690	1.0	6	3	22	0	37.90
20	8.795	1.500	9.890	1.820	2.0	8	4	50	1	44.50
21	6.083	1.500	6.727	1.652	1.0	6	3	44	0	37.90
22	8.361	1.500	9.150	1.777	2.0	8	4	48	1	38.90
23	8.140	1.000	8.000	1.504	2.0	7	3	3	0	36.90
24	9.142	1.500	7.326	1.831	1.5	8	4	31	0	45.80

Table 8.18 List of Variables for Data in Table 8.17

Variable	Definition
Y	Sale price of the house in thousands of dollars
X_1	Taxes (local, county, school) in thousands of dollars
X_2	Number of bathrooms
X_3	Lot size (in thousands of square feet)
X_4	Living space (in thousands of square feet)
X_5	Number of garage stalls
X_6	Number of rooms
X_7	Number of bedrooms
X_8	Age of of the home (years)
X_9	Number of fireplaces

Answer the following questions, in each case justifying your answer by appropriate analyses.

(a) In a fitted regression model that relates the sale price to taxes and building characteristics, would you include all the variables?

(b) A veteran real estate agent has suggested that local taxes, number of rooms, and age of the house would adequately describe the sale price. Do you agree?

(c) A real estate expert who was brought into the project reasoned as follows: The selling price of a home is determined by its desirability and this is certainly a function of the physical characteristic of the building. This overall assessment is reflected in the local taxes paid by the homeowner; consequently, the best predictor of sale price is the local taxes. The building characteristics are therefore redundant in a regression equation which includes local taxes. An equation that relates sale price solely to local taxes would be adequate. Examine this assertion by examining several models. Do you agree? Present what you consider to be the most adequate model or models for predicting sale price of homes in Erie, PA.

In chapter 7 practice exercise

8.2 Refer to the Gasoline Consumption data in Tables 7.16 and 7.17

(a) Would you include all the variables to predict the gasoline consumption of the cars? Explain, giving reasons.

(b) Six alternative models have been suggested:
(i) Regression of Y on X_1
(ii) Regression of Y on X_{10}
(iii) Regression of Y on X_1 and X_{10}
(iv) Regression of Y on X_2 and X_{10}
(v) Regression of Y on X_8 and X_{10}
(vi) Regression of Y on X_8 and X_5, and X_{10}
Among these regression models, which would you choose to predict the gasoline consumption of automobiles? Can you suggest a better model?

(c) Plot Y against X_1, X_2, X_8, and X_{10} (one at a time). Do the plots suggest that the relationship between Y and the 11 predictor variables may not be linear?

(d) The gasoline consumption was determined by driving each car with the same load over the same track (a road length of about 123 miles). Instead of using Y (miles per gallon), it was suggested that we consider a new variable, $W = 100/Y$ (gallons per hundred miles). Plot W against X_1, X_2, X_8, and X_{10} and examine if the relationship between W and the 11 predictor variables is more linear than that between Y and the 11 predictor variables.

(e) Repeat Part (b) using W in place of Y. What are your conclusions?

(f) Regress Y on X_{13}, where $X_{13} = X_8/X_{10}$.

(g) Write a brief report describing your findings. Make a recommendation on the model to be used for predicting gasoline consumption of cars.

In chapter 3 practice exercise

8.3 Cigarette Consumption Data: Consider the Cigarette Consumption data described in Exercise 3.15 and given in Table 3.17. The organization wanted to construct a regression equation that relates statewide cigarette consumption (per capita basis) to various socioeconomic and demographic variables, and to determine whether these variables were useful in predicting the consumption of cigarettes.

(a) Construct a linear regression model that explains the per capita sale of cigarettes in a given state. In your analysis, pay particular attention to outliers. See if the deletion of an outlier affects your findings. Look at residual plots before deciding on a final model. You need not include all the variables in the model if your analysis indicates otherwise. Your objective should be to find the smallest number of variables that describes the state sale of cigarettes meaningfully and adequately.

(b) Write a report describing your findings.