LECTURE 6 PROBLEM

experiment was devised in which a set of consumers were randomly shown one computer from an inventory and were asked to guess its price. Four computers were used in the study, with list prices of $1,800, $2,500, $3,000, and $4,500. The following table summarizes the findings. Each column represents a different computer, with its price in the first row. The rest of the column contains the various guesses by randomly selected customers. (Each guess came from a different customer.)

	C1	C2	C3	C4
Actual price	1,800	2,500	3,000	4,500
Guesses	1,400	2,500	3,600	3,600
	2,300	2,600	3,300	5,500
	2,200	2,400	3,600	4,900
	3,000	2,700	3,600	5,600
	1,000	2,500	3,700	4,600
	1,500	2,600		4,300
	1,400	1,800		6,100
		2,400		

a. Explore the possibility of a linear relation between a customer's guesses and actual prices. Does there seem to be a strong linear relationship, based on a scatter plot and the correlation?

b. Build a linear model to represent a customer's price guess for a given computer price. What guess would the model predict for a computer that has a list price of $3,600?

c. What is the probability that the intercept term in the model of (a) is actually zero? Build a linear model with an intercept of zero to represent a customer's price guess for a given computer price. What guess would the model predict for a computer that has a list price of $3,600?

d. What advantages or disadvantages does the model in (b) have?

4. The Centers for Disease Control has collected the following data on cigarette advertising and brand preferences among adolescents and adults.

Brand	Advertising ($ million)	Brand preference Adolescent (%)	Brand preference Adult (%)
Marlboro	75	60.0	23.5
Camel	43	13.3	6.7
Newport	35	12.7	4.8
Kool	21	1.2	3.9
Winston	17	1.2	3.9
Benson & Hedges	4	1.0	3.0
Salem	3	0.3	2.5

a. Explore the possibility of a linear relation between advertising and brand preference for adults and for adolescents separately. Do there seem to be strong linear relationships, based on a scatter plot and the correlation?

b. Build linear models to represent the relation between advertising and brand preference for these two groups. What preferences would the models predict for advertising of $60 million?

c. Build nonlinear models to represent the relation between advertising and brand preference for these two groups, using the power function $y = ax^b$. What preferences would the models predict for advertising of $60 million? What are the advantages and disadvantages of a nonlinear model for this relationship?

5. The Energy Conservation Committee at National Electronics Company is trying to understand energy use at their plant. As a first step, the committee wants to build a model that will predict monthly energy use as a function of production volumes (*Production*), daily outside temperature (*Temperature*), and number of workdays (*Days*). The following table summarizes the data they have been able to collect for the previous year.

Month	Energy use	Temperature	Days	Production
1	450	42	24	121
2	442	56	21	116
3	499	62	24	132
4	484	68	25	109
5	479	78	25	115
6	507	85	26	119
7	515	89	25	118
8	501	81	24	116
9	513	73	24	132
10	480	67	25	127
11	492	58	24	122
12	466	50	23	117

a. Build a linear model to predict energy use based on all three potential explanatory variables.

b. What level of energy use would the model in (a) predict for a month in which there was an average temperature of 44, monthly production of 120, and 25 days of work at the plant?

c. What percentage of the variability in energy use is accounted for by the model in (a)?

d. According to the model in (a), what are the marginal impacts on energy use of a one-unit increase in temperature and a one-unit increase in production?

e. Evaluate each of the four regression parameters in (a) to determine whether any are likely to be zero. Eliminate those parameters with a high probability of being zero from the model and estimate a new model. Compare the advantages and disadvantages of this model to the one in (a).

6. The National Transportation Safety Board collects data by state (including the District of Columbia) on traffic fatalities. Part of this data is shown in the following table,