Exercises

3.5.1 In an attempt to predict the final grade of students in an Introduction to Statistics class, the professor gives each student a 20-point pretest at the beginning of the year. The table below gives the final grade, pretest score, ACT score, and year (1 = freshman, 2 = sophomore, etc.) of 10 students.

Grade	84.5	82.3	69.2	65.1	80.1	85.9	88.1	90.7	87.2	92.7
Pretest	9	8	18	10	6	8	16	11	15	19
Year	1	2	2	4	3	3	1	4	4	3
ACT	25	20	18	17	20	22	30	28	27	31

a. Find the regression equation that predicts Grade in terms of Pretest Score. Repeat using Year and then ACT. Which of these single-variable predictors is "best" at predicting the final grade based on the R^2 values? Does the pretest score alone appear to be a good predictor of the final grade? Explain.

b. Consider all four different combinations of two or three predictor variables. Determine which combination is "best" at predicting the final grade using the methods described in this section. Based on your results, does it seem worthwhile to give the pretest as a way of predicting the final grade? Does the year of the student appear to affect the final grade? Explain.

c. Use the multiple regression equation that predicts Grade in terms of Pretest and ACT to predict the grade of a student who has a pretest score of 18 and and ACT score of 28.

3.5.2 Use a multiple regression approach to fit a 7th degree polynomial to the 8 data points shown in the table below. Create a graph of the resulting polynomial on top of the data points.

x	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
y	0.90	9.00	2.10	7.80	13.50	9.30	0.40	6.20

3.5.3 Use a multiple regression approach to fit a model of the form $y = a + bx + c\ln(3x+1) + d/\left(1-x^2\right) + e\cos(3x)$ to the data in the table below.

x	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y	14.22	-4.46	16.61	-1.07	19.34	4.72	24.2	7.77	25.67	11.67	29.28	13.96	34.53	17.08	37.38

3.5.4 Use the definition of R^2 in formula (3.1) to calculate the R^2 value for the multiple-regression equation for predicting selling price in terms of area and acres as found in Example 3.5.2. Compare this value to the R Square value given in Figure 3.44. Also verify that the Adjusted R Square value in the figure is calculated according to formula (3.5).

3.5.5 Table 3.7 gives the poverty level, unemployment rate, high school graduation rate, and divorce rate (all in percentages) of 10 randomly selected states in 2007 (data collected by Matthew Schranz, 2011). Determine which combination of predictor variable(s) is "best" at predicting poverty level. Based on this data, does divorce rate appear to be related to poverty level at all?

Poverty Level	Unemployment Rate	High School Graduation Rate	Divorce Rate
14.3	4.3	69.6	3.9
9.3	3.8	71.9	3.7
8.9	3.8	86.5	2.6
12.8	5.3	81.9	3.8
5.8	3.5	81.7	3.8
15.5	5	68.6	4
12.8	5.3	73.8	3.9
9.4	2.7	82.5	3.1
10	4	88.6	3.6
11	4.5	88.5	2.9

Table 3.7

3.6 Spline Models

Consider the data in Table 3.8 which gives the Liters of milk given by a dairy cow on each of several different days after she begins producing. A graph of this data is shown in Figure 3.49. Our goal is to model Liters in terms of Day so that we can predict how much milk was given on the days not listed in the table.

Day	5	13	25	42	50	62	75	90	100	120
Liters	10	19	30	25	22	35	50	55	40	35

Table 3.8

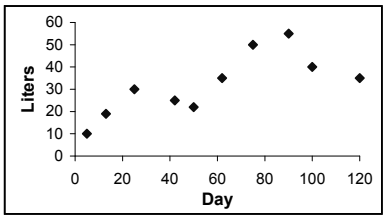

Figure 3.49

The graph of the data certainly does not resemble an exponential, logarithmic, or power curve, so a linearizable model does not seem appropriate. Low-order polynomials do not capture the trend of the data, and higher-order polynomials produce oscillations which do not seem appropriate. We will instead consider *spline* models where we simply "connect the dots" with either straight lines, forming a **linear spline** model, or with cubic polynomials,