Table 6.15 Advertising Pages (P), in Hundreds, and Advertising Revenue (R), in Millions of Dollars) for 41 Magazines in 1986

Magazine	P	R	Magazine	P	R
Cosmopolitan	25	50.0	Town and Country	1	7.0
Redbook	15	49.7	True Story	77	6.6
Glamour	20	34.0	Brides	13	6.2
Southern Living	17	30.7	Book Digest Magazine	5	5.8
Vogue	23	27.0	W	7	5.1
Sunset	17	26.3	Yankee	13	4.1
House and Garden	14	24.6	Playgirl	4	3.9
New York Magazine	22	16.9	Saturday Review	6	3.9
House Beautiful	12	16.7	New Woman	3	3.5
Mademoiselle	15	14.6	Ms.	6	3.3
Psychology Today	8	13.8	Cuisine	4	3.0
Life Magazine	7	13.2	Mother Earth News	3	2.5
Smithsonian	9	13.1	1001 Decorating Ideas	3	2.3
Rolling Stone	12	10.6	Self	5	2.3
Modern Bride	1	8.8	Decorating & Craft Ideas	4	1.8
Parents	6	8.7	Saturday Evening Post	4	1.5
Architectural Digest	12	8.5	McCall's Needlework and Craft	3	1.3
Harper's Bazaar	9	8.3	Weight Watchers	3	1.3
Apartment Life	7	8.2	High Times	4	1.0
Bon Appetit	9	8.2	Soap Opera Digest	2	0.3
Gourmet	7	7.3			

EXERCISES

6.1 Magazine Advertising: In a study of revenue from advertising, data were collected for 41 magazines in 1986 (Table 6.15). The variables observed are number of pages of advertising and advertising revenue. The names of the magazines are listed.

(a) Fit a linear regression equation relating advertising revenue to advertising pages. Verify that the fit is poor.

(b) Choose an appropriate transformation of the data and fit the model to the transformed data. Evaluate the fit.

(c) You should not be surprised by the presence of a large number of outliers because the magazines are highly heterogeneous and it is unrealistic to expect a single relationship to connect all of them. Delete the outliers and obtain an acceptable regression equation that relates advertising revenue to advertising pages.

6.2 Wind Chill Factor: Table 6.16 gives the effective temperatures (W), which are due to the wind chill effect, for various values of the actual temperatures (T) in still air and windspeed (V). The zero-wind condition is taken as the rate of chilling when one is walking through still air (an apparent wind of four

Table 6.16 Wind Chill Factor (°F) for Various Values of Windspeed, V, in Miles/Hour, and Temperature (°F)

	Actual Air Temperature (T)											
V	50	40	30	20	10	0	−10	−20	−30	−40	−50	−60
5	48	36	27	17	5	−5	−15	−25	−35	−46	−56	−66
10	40	29	18	5	−8	−20	−30	−43	−55	−68	−80	−93
15	35	23	10	−5	−18	−29	−42	−55	−70	−83	−97	−112
20	32	18	4	−10	−23	−34	−50	−64	−79	−94	−108	−121
25	30	15	−1	−15	−28	−38	−55	−72	−88	−105	−118	−130
30	28	13	−5	−18	−33	−44	−60	−76	−92	−109	−124	−134
35	27	11	−6	−20	−35	−48	−65	−80	−96	−113	−130	−137
40	26	10	−7	−21	−37	−52	−68	−83	−100	−117	−135	−140
45	25	9	−8	−22	−39	−54	−70	−86	−103	−120	−139	−143
50	25	8	−9	−23	−40	−55	−72	−88	−105	−123	−142	−145

miles per hour (mph)). The National Weather Service originally published the data; we have compiled it from a publication of the Museum of Science of Boston. The temperatures are measured in degrees Fahrenheit (°F), and the wind-speed in mph.

(a) The data in Table 6.16 are not given in a format suitable for direct application of regression programs. You may need to construct another table containing three columns, one column for each of the variables W, T, and V. This table can be found in the book's Web Site.[6]

(b) Fit a linear relationship between W, T, and V. The pattern of residuals should indicate the inadequacy of the linear model.

(c) After adjusting W for the effect of T (e.g., keeping T fixed), examine the relationship between W and V. Does the relationship between W and V appear linear?

(d) After adjusting W for the effect of V, examine the relationship between W and T. Does the relationship appear linear?

(e) Fit the model

$$W = \beta_0 + \beta_1\, T + \beta_2\, V + \beta_3\, \sqrt{V} + \varepsilon. \tag{6.18}$$

Does the fit of this model appear adequate? The W numbers were produced by the National Weather Service according to the formula (except for rounding errors)

$$W = 0.0817(3.71\sqrt{V} + 5.81 - 0.25V)(T - 91.4) + 91.4. \tag{6.19}$$

Does the formula above give an accurate numerical description of W?

[6]http://www.ilr.cornell.edu/~hadi/RABE4

Table 6.17 Annual World Crude Oil Production in Millions of Barrels (1880–1988)

Year	OIL	Year	OIL	Year	OIL
1880	30	1940	2,150	1972	18,584
1890	77	1945	2,595	1974	20,389
1900	149	1950	3,803	1976	20,188
1905	215	1955	5,626	1978	21,922
1910	328	1960	7,674	1980	21,722
1915	432	1962	8,882	1982	19,411
1920	689	1964	10.310	1984	19,837
1925	1,069	1966	12,016	1986	20,246
1930	1,412	1968	14,104	1988	21,338
1935	1,655	1970	16,690		

(f) Can you suggest a model better than those in (6.18) and (6.19)?

6.3 Refer to the Presidential Election Data in Table 5.17, where the response variable V is the proportion of votes obtained by a presidential candidate in United States. Since the response is a proportion, it has a value between 0 and 1. The transformation $Y = \log(V/(1-V))$ takes the variable V with values between 0 and 1 to a variable Y with values between $-\infty$ to $+\infty$. It is therefore more reasonable to expect that Y satisfies the normality assumption than does V.

(a) Consider fitting the model

$$\begin{aligned} Y &= \beta_0 + \beta_1 \cdot I + \beta_2 \cdot D + \beta_3 \cdot W + \beta_4 \cdot (G \cdot I) \\ &\quad + \beta_5 \cdot P + \beta_6 \cdot N + \varepsilon, \end{aligned} \tag{6.20}$$

which is the same model as in (5.11) but replacing V by Y.

(b) For each of the two models, examine the appropriate residual plots discussed in Chapter 4 to determine which model satisfies the standard assumptions more than the other, the original variable V or the transformed variable Y.

(c) What does the equation in (6.20) imply about the form of the model relating the original variables V in terms of the predictor variables? That is, find the form of the function

$$\begin{aligned} V &= f(\beta_0 + \beta_1 \cdot I + \beta_2 \cdot D + \beta_3 \cdot W + \beta_4 \cdot (G \cdot I) \\ &\quad + \beta_5 \cdot P + \beta_6 \cdot N + \varepsilon). \end{aligned} \tag{6.21}$$

[Hint: This is a nonlinear function referred to as the *logistic function*, which is discussed in Chapter 12.]

6.4 Oil Production Data: The data in Table 6.17 are the annual world crude oil production in millions of barrels for the period 1880–1988. The data are taken from Moore and McCabe (1993), p. 147.

Table 6.18 The Average Price Per Megabyte in Dollars From 1988–1998

Year	Price	Year	Price
1988	11.54	1994	0.705
1989	9.30	1995	0.333
1990	6.86	1996	0.179
1991	5.23	1997	0.101
1992	3.00	1998	0.068
1993	1.46		

Source: Kindly provided by Jim Porter, Disk/Trends in Wired April 1998.

(a) Construct a scatter plot of the oil production variable (OIL) versus Year and observe that the scatter of points on the graph is not linear. In order to fit a linear model to these data, OIL must be transformed.

(b) Construct a scatter plot of log(OIL) versus Year. The scatter of points now follows a straight line from 1880 to 1973. Political turmoil in the oil-producing regions of the Middle East affected patterns of oil production after 1973.

(c) Fit a linear regression of log(OIL) on Year. Assess the goodness of fit of the model.

(d) Construct the index plot of the standardized residuals. This graph shows clearly that one of the standard assumptions is violated. Which one?

6.5 One of the remarkable technological developments in computer industry has been the ability to store information densely on hard disk. The cost of storage has steadily declined. Table 6.18 shows the average price per megabyte in dollars from 1988–1998.

(a) Does a linear time trend describe the data? Define a new variable t by coding 1988 as 1, 1989 as 2, etc.

(b) Fit the model $P_t = P_0 e^{\beta\, t}$, where P_t is the price in period t. Does this model describe the data?

(c) Introduce an indicator variable which takes the value 0 for the years 1988–1991, and 1 for the remaining years. Fit a model to connecting $\log(P_t)$ with time t, the indicator variable, and the variable created by taking the product of time and the indicator variable. Interpret the coefficients of the fitted model.