EXERCISES

7.1 The eigenvalues of the correlation matrix of a set of predictor variables are

$$4.603, \quad 1.175, \quad 0.203, \quad 0.015, \quad 0.003, \quad 0.001$$

and the corresponding eigenvectors are given in Table **7.15**

Table 7.15 Six Eigenvectors of the Correlation Matrix of the Predictors.

	V_1	V_2	V_3	V_4	V_5	V_6
X_1	–0.462	0.058	–0.149	–0.793	0.338	–0.135
X_2	–0.462	0.053	–0.278	0.122	–0.150	0.818
X_3	–0.321	–0.596	0.728	–0.008	0.009	0.107
X_4	–0.202	0.798	0.562	0.077	0.024	0.018
X_5	–0.462	–0.046	–0.196	0.590	0.549	–0.312
X_6	–0.465	0.001	–0.128	0.052	–0.750	–0.450

(a) How many sets of collinearity are there in this data set. Explain.

(b) What are the variables involved in each set? Explain.

7.2 In the analysis of the Advertising data in Section **7.5** it is suggested that the regression of sales S_t against E_t and three of the remaining four variables $(A_t, P_t, A_{t-1}, S_{t-1})$ may resolve the collinearity problem. Run the four suggested regressions and, for each of them, examine the resulting VIF_j's to see if collinearity has been eliminated. data can be found in the folder R codes/Data on Canvas

7.3 Gasoline Consumption: To study the factors that determine the gasoline consumption of cars, data were collected on 30 models of cars. Besides the gasoline consumption (Y), measured in miles per gallon for each car, 11 other measurements representing physical and mechanical characteristics are made. Definitions of variables are given in Table 9.16. The source of the data in Table 9.17 is *Motor Trend* magazine for the year 1975. We wish to determine whether the data set is collinear.

(a) Compute the correlation matrix of the predictor variables $X_1, \cdots, X_{11}$ and the corresponding pairwise scatter plots. Identify any evidence of collinearity.

(b) Compute the eigenvalues, eigenvectors, and the condition number of the correlation matrix. Is collinearity present in the data?

(c) Identify the variables involved in collinearity by examining the eigenvectors corresponding to small eigenvalues.

(d) Regress Y on the 11 predictor variables and compute the VIF for each of the predictors. Which predictors are affected by the presence of collinearity?

data can be found in the folder R codes/Data on Canvas

Table 7.16 Variables for the Gasoline Consumption Data in Table 7.17

Variable	Definition
Y	Miles/gallon
X_1	Displacement (cubic inches)
X_2	Horsepower (feet/pound)
X_3	Torque (feet/pound)
X_4	Compression ratio
X_5	Rear axle ratio
X_6	Carburetor (barrels)
X_7	Number of transmission speeds
X_8	Overall length (inches)
X_9	Width (inches)
X_{10}	Weight (pounds)
X_{11}	Type of transmission (1 = automatic; 0 = manual)

Table 7.17 Gasoline Consumption and Automotive Variables

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
18.9	350.0	165	260	8.00	2.56	4	3	200.3	69.9	3910	1
17.0	350.0	170	275	8.50	2.56	4	3	199.6	72.9	3860	1
20.0	250.0	105	185	8.25	2.73	1	3	196.7	72.2	3510	1
18.3	351.0	143	255	8.00	3.00	2	3	199.9	74.0	3890	1
20.1	225.0	95	170	8.40	2.76	1	3	194.1	71.8	3365	0
11.2	440.0	215	330	8.20	2.88	4	3	184.5	69.0	4215	1
22.1	231.0	110	175	8.00	2.56	2	3	179.3	65.4	3020	1
21.5	262.0	110	200	8.50	2.56	2	3	179.3	65.4	3180	1
34.7	89.7	70	81	8.20	3.90	2	4	155.7	64.0	1905	0
30.4	96.9	75	83	9.00	4.30	2	5	165.2	65.0	2320	0
16.5	350.0	155	250	8.50	3.08	4	3	195.4	74.4	3885	1
36.5	85.3	80	83	8.50	3.89	2	4	160.6	62.2	2009	0
21.5	171.0	109	146	8.20	3.22	2	4	170.4	66.9	2655	0
19.7	258.0	110	195	8.00	3.08	1	3	171.5	77.0	3375	1
20.3	140.0	83	109	8.40	3.40	2	4	168.8	69.4	2700	0
17.8	302.0	129	220	8.00	3.00	2	3	199.9	74.0	3890	1
14.4	500.0	190	360	8.50	2.73	4	3	224.1	79.8	5290	1
14.9	440.0	215	330	8.20	2.71	4	3	231.0	79.7	5185	1
17.8	350.0	155	250	8.50	3.08	4	3	196.7	72.2	3910	1
16.4	318.0	145	255	8.50	2.45	2	3	197.6	71.0	3660	1
23.5	231.0	110	175	8.00	2.56	2	3	179.3	65.4	3050	1
21.5	360.0	180	290	8.40	2.45	2	3	214.2	76.3	4250	1
31.9	96.9	75	83	9.00	4.30	2	5	165.2	61.8	2275	0
13.3	460.0	223	366	8.00	3.00	4	3	228.0	79.8	5430	1
23.9	133.6	96	120	8.40	3.91	2	5	171.5	63.4	2535	0
19.7	318.0	140	255	8.50	2.71	2	3	215.3	76.3	4370	1
13.9	351.0	148	243	8.00	3.25	2	3	215.5	78.5	4540	1
13.3	351.0	148	243	8.00	3.26	2	3	216.1	78.5	4715	1
13.8	360.0	105	295	8.25	3.15	4	3	209.3	77.4	4215	1

data refers to Practice Exercise for Chapter 5, and can be found in the folder R codes/Data on Canvas

7.4 Refer to the Presidential Election Data in Table 5.19 and consider fitting a model relating V to all the variables (including a time trend representing year of election) plus as many interaction terms involving two or three variables as you possibly can.

(a) What is the maximum number of terms (coefficients) in a linear regression model that you can fit to these data? [*Hint*: Consider the number of observations in the data.]

(b) Examine the predictor variables in the above model for the presence of collinearity. (Compute the correlation matrix, the condition number, and the VIFs.)

(c) Identify the subsets of variables involved in collinearity. Attempt to solve the collinearity problem by deleting some of the variables involved in collinearity.

(d) Fit a model relating V to the set of predictors you found to be free from collinearity.

7.5 Refer to the Presidential Election Data in Table 5.19 and consider fitting the model in (5.12). See Practice Exercise for Chapter 5

(a) Examine the predictor variables in this model for the presence of collinearity. (Compute the correlation matrix, the condition number, and the VIFs.)

(b) Identify the subsets of variables involved in collinearity. Attempt to solve the collinearity problem by deleting some of the variables involved in collinearity.

(c) Fit a model relating V to the set of predictors you found to be free from collinearity.

(d) Compare the results with those obtained in Exercise **7.4**