(c) Is the bivariate normal distribution a viable population model? Explain with reference to Q–Q plots and a scatter diagram.

5.20. A wildlife ecologist measured x_1 = tail length (in millimeters) and x_2 = wing length (in millimeters) for a sample of n = 45 female hook-billed kites. These data are displayed in Table 5.12. Using the data in the table,

Table 5.12 Bird Data

x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)	x_1 (Tail length)	x_2 (Wing length)
191	284	186	266	173	271
197	285	197	285	194	280
208	288	201	295	198	300
180	273	190	282	180	272
180	275	209	305	190	292
188	280	187	285	191	286
210	283	207	297	196	285
196	288	178	268	207	286
191	271	202	271	209	303
179	257	205	285	179	261
208	289	190	280	186	262
202	285	189	277	174	245
200	272	211	310	181	250
192	282	216	305	189	262
199	280	189	274	188	258

Source: Data courtesy of S. Temple.

(a) Find and sketch the 95% confidence ellipse for the population means μ_1 and μ_2. Suppose it is known that μ_1 = 190 mm and μ_2 = 275 mm for *male* hook-billed kites. Are these plausible values for the mean tail length and mean wing length for the female birds? Explain.

(b) Construct the simultaneous 95% T^2-intervals for μ_1 and μ_2 and the 95% Bonferroni intervals for μ_1 and μ_2. Compare the two sets of intervals. What advantage, if any, do the T^2-intervals have over the Bonferroni intervals?

(c) Is the bivariate normal distribution a viable population model? Explain with reference to Q–Q plots and a scatter diagram.

5.21. Using the data on bone mineral content in Table 1.8, construct the 95% Bonferroni intervals for the individual means. Also, find the 95% simultaneous T^2-intervals. Compare the two sets of intervals.

5.22. A portion of the data contained in Table 6.10 in Chapter 6 is reproduced in Table 5.13. These data represent various costs associated with transporting milk from farms to dairy plants for gasoline trucks. Only the first 25 multivariate observations for gasoline trucks are given. Observations 9 and 21 have been identified as outliers from the full data set of 36 observations. (See [2].)

(c) Calculate the linear combinations of mean components most responsible for rejecting $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$ in Part b.

(d) Bond rating companies are interested in a company's ability to satisfy its outstanding debt obligations as they mature. Does it appear as if one or more of the foregoing financial ratios might be useful in helping to classify a bond as "high" or "medium" quality? Explain.

(e) Repeat part (b) assuming normal populations with unequal covariance matices (see (6-27), (6-28) and (6-29)). Does your conclusion change?

6.22. Researchers interested in assessing pulmonary function in nonpathological populations asked subjects to run on a treadmill until exhaustion. Samples of air were collected at definite intervals and the gas contents analyzed. The results on 4 measures of oxygen consumption for 25 males and 25 females are given in Table 6.12 on page 348. The variables were

$$X_1 = \text{resting volume } O_2 \text{ (L/min)}$$
$$X_2 = \text{resting volume } O_2 \text{ (mL/kg/min)}$$
$$X_3 = \text{maximum volume } O_2 \text{ (L/min)}$$
$$X_4 = \text{maximum volume } O_2 \text{ (mL/kg/min)}$$

(a) Look for gender differences by testing for equality of group means. Use $\alpha = .05$. If you reject $H_0: \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$, find the linear combination most responsible.

(b) Construct the 95% simultaneous confidence intervals for each $\mu_{1i} - \mu_{2i}, i = 1, 2, 3, 4$. Compare with the corresponding Bonferroni intervals.

(c) The data in Table 6.12 were collected from graduate-student volunteers, and thus they do not represent a random sample. Comment on the possible implications of this information.

6.23. Construct a one-way MANOVA using the width measurements from the iris data in Table 11.5. Construct 95% simultaneous confidence intervals for differences in mean components for the two responses for each pair of populations. Comment on the validity of the assumption that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3$.

6.24. Researchers have suggested that a change in skull size over time is evidence of the interbreeding of a resident population with immigrant populations. Four measurements were made of male Egyptian skulls for three different time periods: period 1 is 4000 B.C., period 2 is 3300 B.C., and period 3 is 1850 B.C. The data are shown in Table 6.13 on page 349 (see the skull data on the website www.prenhall.com/statistics). The measured variables are

$$X_1 = \text{maximum breadth of skull (mm)}$$
$$X_2 = \text{basibregmatic height of skull (mm)}$$
$$X_3 = \text{basialveolar length of skull (mm)}$$
$$X_4 = \text{nasal height of skull (mm)}$$

Construct a one-way MANOVA of the Egyptian skull data. Use $\alpha = .05$. Construct 95% simultaneous confidence intervals to determine which mean components differ among the populations represented by the three time periods. Are the usual MANOVA assumptions realistic for these data? Explain.

6.25. Construct a one-way MANOVA of the crude-oil data listed in Table 11.7 on page 662. Construct 95% simultaneous confidence intervals to determine which mean components differ among the populations. (You may want to consider transformations of the data to make them more closely conform to the usual MANOVA assumptions.)

8.17. Using the data on bone mineral content in Table 1.8, perform a principal component analysis of **S**.

8.18. The data on national track records for women are listed in Table 1.9.

(a) Obtain the sample correlation matrix **R** for these data, and determine its eigenvalues and eigenvectors.

(b) Determine the first two principal components for the standardized variables. Prepare a table showing the correlations of the standardized variables with the components, and the cumulative percentage of the total (standardized) sample variance explained by the two components.

(c) Interpret the two principal components obtained in Part b. (Note that the first component is essentially a normalized unit vector and might measure the athletic excellence of a given nation. The second component might measure the relative strength of a nation at the various running distances.)

(d) Rank the nations based on their score on the first principal component. Does this ranking correspond with your inituitive notion of athletic excellence for the various countries?

8.19. Refer to Exercise 8.18. Convert the national track records for women in Table 1.9 to speeds measured in meters per second. Notice that the records for 800 m, 1500 m, 3000 m, and the marathon are given in minutes. The marathon is 26.2 miles, or 42,195 meters, long. Perform a principal components analysis using the covariance matrix **S** of the speed data. Compare the results with the results in Exercise 8.18. Do your interpretations of the components differ? If the nations are ranked on the basis of their score on the first principal component, does the subsequent ranking differ from that in Exercise 8.18? Which analysis do you prefer? Why?

8.20. The data on national track records for men are listed in Table 8.6. (See also the data on national track records for men on the website www.prenhall.com/statistics) Repeat the principal component analysis outlined in Exercise 8.18 for the men. Are the results consistent with those obtained from the women's data?

8.21. Refer to Exercise 8.20. Convert the national track records for men in Table 8.6 to speeds measured in meters per second. Notice that the records for 800 m, 1500 m, 5000 m, 10,000 m and the marathon are given in minutes. The marathon is 26.2 miles, or 42,195 meters, long. Perform a principal component analysis using the covariance matrix **S** of the speed data. Compare the results with the results in Exercise 8.20. Which analysis do you prefer? Why?

8.22. Consider the data on bulls in Table 1.10. Utilizing the seven variables YrHgt, FtFrBody, PrctFFB, Frame, BkFat, SaleHt, and SaleWt, perform a principal component analysis using the covariance matrix **S** and the correlation matrix **R**. Your analysis should include the following:

(a) Determine the appropriate number of components to effectively summarize the sample variability. Construct a scree plot to aid your determination.

(b) Interpret the sample principal components.

(c) Do you think it is possible to develop a "body size" or "body configuration" index from the data on the seven variables above? Explain.

(d) Using the values for the first two principal components, plot the data in a two-dimensional space with $\hat{y}_1$ along the vertical axis and $\hat{y}_2$ along the horizontal axis. Can you distinguish groups representing the three breeds of cattle? Are there any outliers?

(e) Construct a Q–Q plot using the first principal component. Interpret the plot.

9.26. Consider the mice-weight data in Example 8.6. Start with the sample *covariance* matrix. (See Exercise 8.15 for $\sqrt{s_{ii}}$.)

(a) Obtain the principal component solution to the factor model with $m = 1$ and $m = 2$.

(b) Find the maximum likelihood estimates of the loadings and specific variances for $m = 1$ and $m = 2$.

(c) Perform a varimax rotation of the solutions in Parts a and b.

9.27. Repeat Exercise 9.26 by factoring **R** instead of the sample covariance matrix **S**. Also, for the mouse with standardized weights $[.8, -.2, -.6, 1.5]$, obtain the factor scores using the maximum likelihood estimates of the loadings and Equation (9-58).

9.28. Perform a factor analysis of the national track records for women given in Table 1.9. Use the sample covariance matrix **S** and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix **R**. Does it make a difference if **R**, rather than **S**, is factored? Explain.

9.29. Refer to Exercise 9.28. Convert the national track records for women to speeds measured in meters per second. (See Exercise 8.19.) Perform a factor analysis of the speed data. Use the sample covariance matrix **S** and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix **R**. Does it make a difference if **R**, rather than **S**, is factored? Explain. Compare your results with the results in Exercise 9.28. Which analysis do you prefer? Why?

9.30. Perform a factor analysis of the national track records for men given in Table 8.6. Repeat the steps given in Exercise 9.28. Is the appropriate factor model for the men's data different from the one for the women's data? If not, are the interpretations of the factors roughly the same? If the models are different, explain the differences.

9.31. Refer to Exercise 9.30. Convert the national track records for men to speeds measured in meters per second. (See Exercise 8.21.) Perform a factor analysis of the speed data. Use the sample covariance matrix **S** and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix **R**. Does it make a difference if **R**, rather than **S**, is factored? Explain. Compare your results with the results in Exercise 9.30. Which analysis do you prefer? Why?

9.32. Perform a factor analysis of the data on bulls given in Table 1.10. Use the seven variables YrHgt, FtFrBody, PrctFFB, Frame, BkFat, SaleHt, and SaleWt. Factor the sample covariance matrix **S** and interpret the factors. Compute factor scores, and check for outliers. Repeat the analysis with the sample correlation matrix **R**. Compare the results obtained from **S** with the results from **R**. Does it make a difference if **R**, rather than **S**, is factored? Explain.

9.33. Perform a factor analysis of the psychological profile data in Table 4.6. Use the sample correlation matrix **R** constructed from measurements on the five variables, Indep, Supp, Benev, Conform and Leader. Obtain both the principal component and maximum likelihood solutions for $m = 2$ and $m = 3$ factors. Can you interpret the factors? Your analysis should include factor rotation and the computation of factor scores.

Note: Be aware that a maximum likelihood solution may result in a Heywood case.

9.34. The pulp and paper properties data are given in Table 7.7. Perform a factor analysis using observations on the four paper property variables, BL, EM, SF, and BS and the sample correlation matrix **R**. Can the information in these data be summarized by a single factor? If so, can you interpret the factor? Try both the principal component and maximum likelihood solution methods. Repeat this analysis with the sample covariance matrix **S**. Does your interpretation of the factor(s) change if **S** rather than **R** is factored?

10.16. Andrews and Herzberg [1] give data obtained from a study of a comparison of nondiabetic and diabetic patients. Three primary variables,

$$X_1^{(1)} = \text{glucose intolerance}$$
$$X_2^{(1)} = \text{insulin response to oral glucose}$$
$$X_3^{(1)} = \text{insulin resistance}$$

and two secondary variables,

$$X_1^{(2)} = \text{relative weight}$$
$$X_2^{(2)} = \text{fasting plasma glucose}$$

were measured. The data for $n = 46$ nondiabetic patients yield the covariance matrix

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right] = \left[\begin{array}{rrr|rr} 1106.000 & 396.700 & 108.400 & .787 & 26.230 \\ 396.700 & 2382.000 & 1143.000 & -.214 & -23.960 \\ 108.400 & 1143.000 & 2136.000 & 2.189 & -20.840 \\ \hline .787 & -.214 & 2.189 & .016 & .216 \\ 26.230 & -23.960 & -20.840 & .216 & 70.560 \end{array}\right]$$

Determine the sample canonical variates and their correlations. Interpret these quantities. Are the first canonical variates good summary measures of their respective sets of variables? Explain. Test for the significance of the canonical relations with $\alpha = .05$.

10.17. Data concerning a person's desire to smoke and psychological and physical state were collected for $n = 110$ subjects. The data were responses, coded 1 to 5, to each of 12 questions (variables). The four standardized measurements related to the desire to smoke are defined as

$$z_1^{(1)} = \text{smoking 1 (first wording)}$$
$$z_2^{(1)} = \text{smoking 2 (second wording)}$$
$$z_3^{(1)} = \text{smoking 3 (third wording)}$$
$$z_4^{(1)} = \text{smoking 4 (fourth wording)}$$

The eight standardized measurements related to the psychological and physical state are given by

$$z_1^{(2)} = \text{concentration}$$
$$z_2^{(2)} = \text{annoyance}$$
$$z_3^{(2)} = \text{sleepiness}$$
$$z_4^{(2)} = \text{tenseness}$$
$$z_5^{(2)} = \text{alertness}$$
$$z_6^{(2)} = \text{irritability}$$
$$z_7^{(2)} = \text{tiredness}$$
$$z_8^{(2)} = \text{contentedness}$$

The correlation matrix constructed from the data is

$$\mathbf{R} = \left[\begin{array}{c|c} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \hline \mathbf{R}_{21} & \mathbf{R}_{22} \end{array}\right]$$