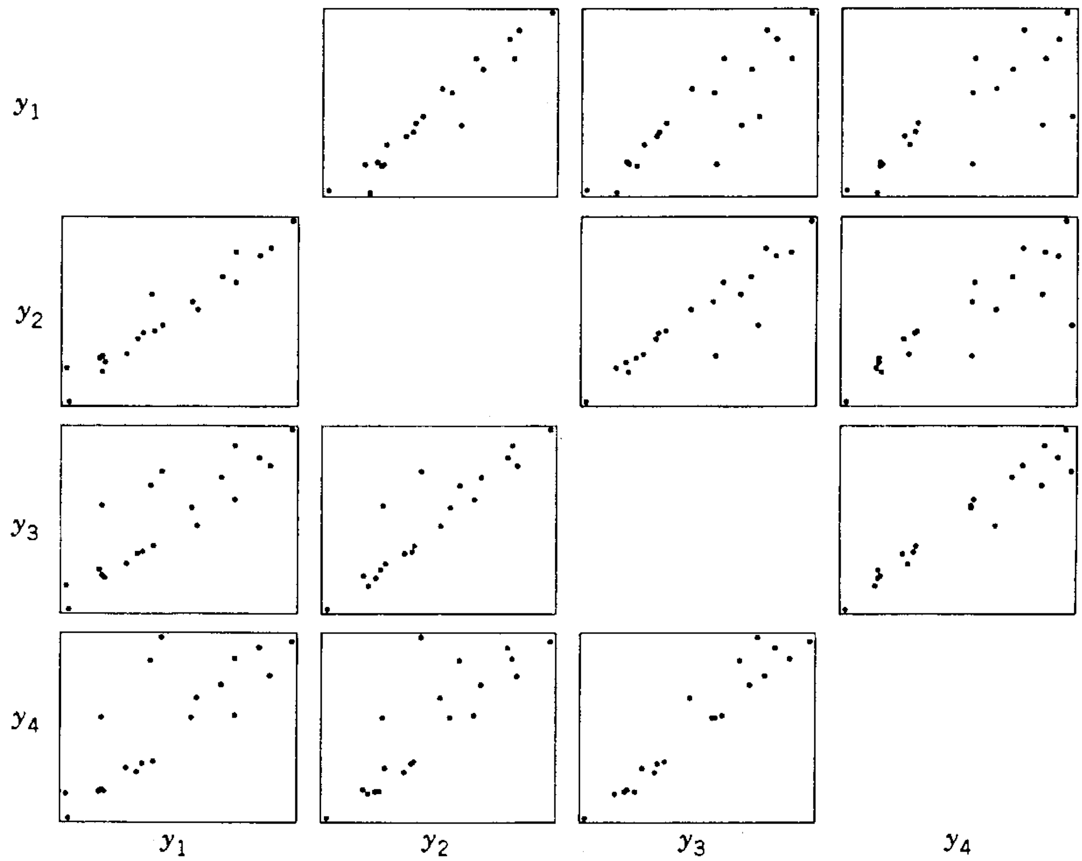

Figure 4.10. Scatter plots for the ramus bone data in Table 3.6.

The bivariate scatter plots are given in Figure 4.10. Three values are clearly separate from the other observations in the plot of y_1 versus y_4. In Table 3.6, the 9th, 12th, and 20th values of y_4 are not unusual, nor are the 9th, 12th, and 20th values of y_1. However, the increase from y_1 to y_4 is exceptional in each case. If these values are not due to errors in recording the data and if this sample is representative, then we appear to have a mixture of two populations. This should be taken into account in making inferences. □

PROBLEMS

4.1 Consider the two covariance matrices

$$\mathbf{\Sigma}_1 = \begin{pmatrix} 14 & 8 & 3 \\ 8 & 5 & 2 \\ 3 & 2 & 1 \end{pmatrix}, \qquad \mathbf{\Sigma}_2 = \begin{pmatrix} 6 & 6 & 1 \\ 6 & 8 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

Show that $|\mathbf{\Sigma}_2| > |\mathbf{\Sigma}_1|$ and that $\text{tr}(\mathbf{\Sigma}_2) < \text{tr}(\mathbf{\Sigma}_1)$. Thus the generalized variance of population 2 is greater than the generalized variance of population 1, even though the total variance is less. Comment on why this is true in terms of the variances and correlations.

4.2 For $\mathbf{z} = (\mathbf{T}')^{-1}(\mathbf{y} - \boldsymbol{\mu})$ in (4.4), show that $E(\mathbf{z}) = \mathbf{0}$ and $\text{cov}(\mathbf{z}) = \mathbf{I}$.

4.3 Show that the form of the likelihood function in (4.13) follows from the previous expression.

4.4 For $(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ in (4.3) and (4.6), show that $E[(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})] = p$. Assume $E(\mathbf{y}) = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$. Normality is not required.

4.5 Show that by adding and subtracting $\bar{\mathbf{y}}$, the exponent of (4.13) has the form given in (4.14), that is,

$$\frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}} + \bar{\mathbf{y}} - \boldsymbol{\mu}) = \frac{1}{2}\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \bar{\mathbf{y}}) + \frac{n}{2}(\bar{\mathbf{y}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{y}} - \boldsymbol{\mu}).$$

4.6 Show that $\sqrt{b_1}$ and b_2, as given in (4.18) and (4.19), are invariant to the transformation $z_i = ay_i + b$.

4.7 Show that if $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\beta_{2,p} = p(p+2)$ as in (4.34).

4.8 Show that $b_{1,p}$ and $b_{2,p}$, as given by (4.36) and (4.37), are invariant under the transformation $\mathbf{z}_i = \mathbf{A}\mathbf{y}_i + \mathbf{b}$, where $\mathbf{A}$ is nonsingular. Thus $b_{1,p}$ and $b_{2,p}$ do not depend on the units of measurement.

4.9 Show that $F_{(n)} = [(n-p-1)/p](1/w - 1)$ as in (4.45).

4.10 Suppose $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 6 & 1 & -2 \\ 1 & 13 & 4 \\ -2 & 4 & 4 \end{pmatrix}.$$

(a) Find the distribution of $z = 2y_1 - y_2 + 3y_3$.
(b) Find the joint distribution of $z_1 = y_1 + y_2 + y_3$ and $z_2 = y_1 - y_2 + 2y_3$.
(c) Find the distribution of y_2.
(d) Find the joint distribution of y_1 and y_3.
(e) Find the joint distribution of y_1, y_3, and $\frac{1}{2}(y_1 + y_2)$.

4.11 Suppose $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given in the previous problem.
(a) Find a vector $\mathbf{z}$ such that $\mathbf{z} = (\mathbf{T}')^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is $N_3(\mathbf{0}, \mathbf{I})$ as in (4.4).
(b) Find a vector $\mathbf{z}$ such that $\mathbf{z} = (\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is $N_3(\mathbf{0}, \mathbf{I})$ as in (4.5).
(c) What is the distribution of $(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$?

4.12 Suppose $\mathbf{y}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} -2 \\ 3 \\ -1 \\ 5 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 11 & -8 & 3 & 9 \\ -8 & 9 & -3 & 6 \\ 3 & -3 & 2 & 3 \\ 9 & 6 & 3 & 9 \end{pmatrix}.$$

(a) Find the distribution of $z = 4y_1 - 2y_2 + y_3 - 3y_4$.

(b) Find the joint distribution of $z_1 = y_1 + y_2 + y_3 + y_4$ and $z_2 = -2y_1 + 3y_2 + y_3 - 2y_4$.

(c) Find the joint distribution of $z_1 = 3y_1 + y_2 - 4y_3 - y_4$, $z_2 = -y_1 - 3y_2 + y_3 - 2y_4$, and $z_3 = 2y_1 + 2y_2 + 4y_3 - 5y_4$.

(d) What is the distribution of y_3?

(e) What is the joint distribution of y_2 and y_4?

(f) Find the joint distribution of y_1, $\frac{1}{2}(y_1 + y_2)$, $\frac{1}{3}(y_1 + y_2 + y_3)$, and $\frac{1}{4}(y_1 + y_2 + y_3 + y_4)$.

4.13 Suppose $\mathbf{y}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given in the previous problem.

(a) Find a vector $\mathbf{z}$ such that $\mathbf{z} = (\mathbf{T}')^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is $N_4(\mathbf{0}, \mathbf{I})$ as in (4.4).

(b) Find a vector $\mathbf{z}$ such that $\mathbf{z} = (\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{y} - \boldsymbol{\mu})$ is $N_4(\mathbf{0}, \mathbf{I})$ as in (4.5).

(c) What is the distribution of $(\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$?

4.14 Suppose $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Which of the following random variables are independent?

(a) y_1 and y_2

(b) y_1 and y_3

(c) y_2 and y_3

(d) (y_1, y_2) and y_3

(e) (y_1, y_3) and y_2

4.15 Suppose $\mathbf{y}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} -4 \\ 2 \\ 5 \\ -1 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 8 & 0 & -1 & 0 \\ 0 & 3 & 0 & 2 \\ -1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 7 \end{pmatrix}.$$

Which of the following random variables are independent?

(a) y_1 and y_2

(b) y_1 and y_3

(c) y_1 and y_4

(d) y_2 and y_3

(e) y_2 and y_4

(f) y_3 and y_4

(g) (y_1, y_2) and y_3

(h) (y_1, y_2) and y_4

(i) (y_1, y_3) and y_4

(j) y_1 and (y_2, y_4)

(k) y_1 and y_2 and y_3

(l) y_1 and y_2 and y_4

(m) (y_2, y_2) and (y_3, y_4)

(n) (y_1, y_3) and (y_2, y_4)

4.16 Assume $\mathbf{y}$ and $\mathbf{x}$ are subvectors, each 2×1, where $\binom{\mathbf{y}}{\mathbf{x}}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \left(\begin{array}{c} 2 \\ -1 \\ \hline 3 \\ 1 \end{array}\right), \qquad \boldsymbol{\Sigma} = \left(\begin{array}{cc|cc} 7 & 3 & -3 & 2 \\ 3 & 6 & 0 & 4 \\ \hline -3 & 0 & 5 & -2 \\ 2 & 4 & -2 & 4 \end{array}\right).$$

(a) Find $E(\mathbf{y}|\mathbf{x})$ by (4.7).

(b) Find $\text{cov}(\mathbf{y}|\mathbf{x})$ by (4.8).

4.17 Suppose $\mathbf{y}$ and $\mathbf{x}$ are subvectors, such that $\mathbf{y}$ is 2×1 and $\mathbf{x}$ is 3×1, with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ partitioned accordingly:

$$\boldsymbol{\mu} = \left(\begin{array}{c} 3 \\ -2 \\ \hline 4 \\ -3 \\ 5 \end{array}\right), \qquad \boldsymbol{\Sigma} = \left(\begin{array}{cc|ccc} 14 & -8 & 15 & 0 & 3 \\ -8 & 18 & 8 & 6 & -2 \\ \hline 15 & 8 & 50 & 8 & 5 \\ 0 & 6 & 8 & 4 & 0 \\ 3 & -2 & 5 & 0 & 1 \end{array}\right).$$

Assume that $\binom{\mathbf{y}}{\mathbf{x}}$ is distributed as $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(a) Find $E(\mathbf{y}|\mathbf{x})$ by (4.7).

(b) Find $\text{cov}(\mathbf{y}|\mathbf{x})$ by (4.8).

4.18 Suppose that $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ is a random sample from a nonnormal multivariate population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. If n is large, what is the approximate distribution of each of the following?

(a) $\sqrt{n}(\overline{\mathbf{y}} - \boldsymbol{\mu})$

(b) $\overline{\mathbf{y}}$

4.19 For the ramus bone data treated in Example 4.5.2, check each of the four variables for univariate normality using the following techniques:

(a) Q–Q plots;

(b) $\sqrt{b_1}$ and b_2 as given by (4.18) and (4.19);

(c) D'Agostino's test using D and Y given in (4.22) and (4.23);

(d) The test by Lin and Mudholkar using z defined in (4.24).

4.20 For the calcium data in Table 3.3, check for multivariate normality and outliers using the following tests:

(a) Calculate D_i^2 as in (4.27) for each observation.

(b) Compare the largest value of D_i^2 with the critical value in Table A.6.

(c) Compute u_i and v_i in (4.28) and (4.29) and plot them. Is there an indication of nonlinearity or outliers?

(d) Calculate $b_{1,p}$ and $b_{2,p}$ in (4.36) and (4.37) and compare them with critical values in Table A.5.

4.21 For the probe word data in Table 3.5, check each of the five variables for univariate normality and outliers using the following tests:

(a) Q–Q plots;

(b) $\sqrt{b_1}$ and b_2 as given by (4.18) and (4.19);

(c) D'Agostino's test using D and Y given in (4.22) and (4.23);

(d) The test by Lin and Mudholkar using z defined in (4.24).

4.22 For the probe word data in Table 3.5, check for multivariate normality and outliers using the following tests:

(a) Calculate D_i^2 as in (4.27) for each observation.

(b) Compare the largest value of D_i^2 with the critical value in Table A.6.

(c) Compute u_i and v_i in (4.28) and (4.29) and plot them. Is there an indication of nonlinearity or outliers?

(d) Calculate $b_{1,p}$ and $b_{2,p}$ in (4.36) and (4.37) and compare them with critical values in Table A.5.

4.23 Six hematology variables were measured on 51 workers (Royston 1983):

y_1 = hemoglobin concentration
y_2 = packed cell volume
y_3 = white blood cell count
y_4 = lymphocyte count
y_5 = neutrophil count
y_6 = serum lead concentration

The data are given in Table 4.3. Check each of the six variables for univariate normality using the following tests:

(a) Q–Q plots;

(b) $\sqrt{b_1}$ and b_2 as given by (4.18) and (4.19);

(c) D'Agostino's test using D and Y given in (4.22) and (4.23);

(d) The test by Lin and Mudholkar using z defined in (4.24).

Table 4.3. Hematology Data

Observation Number	y_1	y_2	y_3	y_4	y_5	y_6
1	13.4	39	4100	14	25	17
2	14.6	46	5000	15	30	20
3	13.5	42	4500	19	21	18
4	15.0	46	4600	23	16	18
5	14.6	44	5100	17	31	19
6	14.0	44	4900	20	24	19
7	16.4	49	4300	21	17	18
8	14.8	44	4400	16	26	29
9	15.2	46	4100	27	13	27
10	15.5	48	8400	34	42	36

(continued)

Table 4.3. (Continued)

Observation Number	y_1	y_2	y_3	y_4	y_5	y_6
11	15.2	47	5600	26	27	22
12	16.9	50	5100	28	17	23
13	14.8	44	4700	24	20	23
14	16.2	45	5600	26	25	19
15	14.7	43	4000	23	13	17
16	14.7	42	3400	9	22	13
17	16.5	45	5400	18	32	17
18	15.4	45	6900	28	36	24
19	15.1	45	4600	17	29	17
20	14.2	46	4200	14	25	28
21	15.9	46	5200	8	34	16
22	16.0	47	4700	25	14	18
23	17.4	50	8600	37	39	17
24	14.3	43	5500	20	31	19
25	14.8	44	4200	15	24	29
26	14.9	43	4300	9	32	17
27	15.5	45	5200	16	30	20
28	14.5	43	3900	18	18	25
29	14.4	45	6000	17	37	23
30	14.6	44	4700	23	21	27
31	15.3	45	7900	43	23	23
32	14.9	45	3400	17	15	24
33	15.8	47	6000	23	32	21
34	14.4	44	7700	31	39	23
35	14.7	46	3700	11	23	23
36	14.8	43	5200	25	19	22
37	15.4	45	6000	30	25	18
38	16.2	50	8100	32	38	18
39	15.0	45	4900	17	26	24
40	15.1	47	6000	22	33	16
41	16.0	46	4600	20	22	22
42	15.3	48	5500	20	23	23
43	14.5	41	6200	20	36	21
44	14.2	41	4900	26	20	20
45	15.0	45	7200	40	25	25
46	14.2	46	5800	22	31	22
47	14.9	45	8400	61	17	17
48	16.2	48	3100	12	15	18
49	14.5	45	4000	20	18	20
50	16.4	49	6900	35	22	24
51	14.7	44	7800	38	34	16

4.24 For the hematology data in Table 4.3, check for multivariate normality using the following techniques:

(a) Calculate D_i^2 as in (4.27) for each observation.

(b) Compare the largest value of D_i^2 with the critical value in Table A.6 (extrapolate).

(c) Compute u_i and υ_i in (4.28) and (4.29) and plot them. Is there an indication of nonlinearity or outliers?

(d) Calculate $b_{1,p}$ and $b_{2,p}$ in (4.36) and (4.37) and compare them with critical values in Table A.5.