Exercises

9.1. Show that the covariance matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & .63 & .45 \\ .63 & 1.0 & .35 \\ .45 & .35 & 1.0 \end{bmatrix}$$

for the $p = 3$ standardized random variables Z_1, Z_2, and Z_3 can be generated by the $m = 1$ factor model

$$\begin{aligned} Z_1 &= .9F_1 + \varepsilon_1 \\ Z_2 &= .7F_1 + \varepsilon_2 \\ Z_3 &= .5F_1 + \varepsilon_3 \end{aligned}$$

where $\text{Var}(F_1) = 1$, $\text{Cov}(\boldsymbol{\varepsilon}, F_1) = \mathbf{0}$, and

$$\boldsymbol{\Psi} = \text{Cov}(\boldsymbol{\varepsilon}) = \begin{bmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{bmatrix}$$

That is, write $\boldsymbol{\rho}$ in the form $\boldsymbol{\rho} = \mathbf{LL}' + \boldsymbol{\Psi}$.

9.2. Use the information in Exercise 9.1.

(a) Calculate communalities $h_i^2, i = 1, 2, 3$, and interpret these quantities.

(b) Calculate $\text{Corr}(Z_i, F_1)$ for $i = 1, 2, 3$. Which variable might carry the greatest weight in "naming" the common factor? Why?

9.3. The eigenvalues and eigenvectors of the correlation matrix $\boldsymbol{\rho}$ in Exercise 9.1 are

$$\begin{aligned} \lambda_1 &= 1.96, & \mathbf{e}_1' &= [.625, .593, .507] \\ \lambda_2 &= .68, & \mathbf{e}_2' &= [-.219, -.491, .843] \\ \lambda_3 &= .36, & \mathbf{e}_3' &= [.749, -.638, -.177] \end{aligned}$$

(a) Assuming an $m = 1$ factor model, calculate the loading matrix $\mathbf{L}$ and matrix of specific variances $\boldsymbol{\Psi}$ using the principal component solution method. Compare the results with those in Exercise 9.1.

(b) What proportion of the total population variance is explained by the first common factor?

9.4. Given $\boldsymbol{\rho}$ and $\boldsymbol{\Psi}$ in Exercise 9.1 and an $m = 1$ factor model, calculate the reduced correlation matrix $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho} - \boldsymbol{\Psi}$ and the principal factor solution for the loading matrix $\mathbf{L}$. Is the result consistent with the information in Exercise 9.1? Should it be?

9.5. Establish the inequality (9-19).

Hint: Since $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}}$ has zeros on the diagonal,

$$(\text{sum of squared entries of } \mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}}) \le (\text{sum of squared entries of } \mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}')$$

Now, $\mathbf{S}-\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' = \hat{\lambda}_{m+1}\hat{\mathbf{e}}_{m+1}\hat{\mathbf{e}}'_{m+1} + \cdots + \hat{\lambda}_p\hat{\mathbf{e}}_p\hat{\mathbf{e}}'_p = \hat{\mathbf{P}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{P}}'_{(2)}$, where $\hat{\mathbf{P}}_{(2)} = [\hat{\mathbf{e}}_{m+1}\vdots\cdots\vdots\hat{\mathbf{e}}_p]$ and $\hat{\mathbf{\Lambda}}_{(2)}$ is the diagonal matrix with elements $\hat{\lambda}_{m+1}, \ldots, \hat{\lambda}_p$.

Use (sum of squared entries of $\mathbf{A}$) $=$ tr $\mathbf{AA}'$ and tr $[\hat{\mathbf{P}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{P}}'_{(2)}] =$ tr $[\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}]$.

9.6. Verify the following matrix identities.

(a) $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{I} - (\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}$

Hint: Premultiply both sides by $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})$.

(b) $(\mathbf{LL}' + \mathbf{\Psi})^{-1} = \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}$

Hint: Postmultiply both sides by $(\mathbf{LL}' + \mathbf{\Psi})$ and use (a).

(c) $\mathbf{L}'(\mathbf{LL}' + \mathbf{\Psi})^{-1} = (\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}$

Hint: Postmultiply the result in (b) by $\mathbf{L}$, use (a), and take the transpose, noting that $(\mathbf{LL}' + \mathbf{\Psi})^{-1}$, $\mathbf{\Psi}^{-1}$, and $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}$ are symmetric matrices.

9.7. *(The factor model parameterization need not be unique.)* Let the factor model with $p = 2$ and $m = 1$ prevail. Show that

$$\sigma_{11} = \ell_{11}^2 + \psi_1, \qquad \sigma_{12} = \sigma_{21} = \ell_{11}\ell_{21}$$
$$\sigma_{22} = \ell_{21}^2 + \psi_2$$

and, for given σ_{11}, σ_{22}, and σ_{12}, there is an infinity of choices for $\mathbf{L}$ and $\mathbf{\Psi}$.

9.8. *(Unique but improper solution: Heywood case.)*
Consider an $m = 1$ factor model for the population with covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .4 & .9 \\ .4 & 1 & .7 \\ .9 & .7 & 1 \end{bmatrix}$$

Show that there is a unique choice of $\mathbf{L}$ and $\mathbf{\Psi}$ with $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, but that $\psi_3 < 0$, so the choice is not admissible.

9.9. In a study of liquor preference in France, Stoetzel [14] collected preference rankings of $p = 9$ liquor types from $n = 1442$ individuals. A factor analysis of the 9×9 sample correlation matrix of rank orderings gave the following estimated loadings:

	Estimated factor loadings		
Variable (X_1)	F_1	F_2	F_3
Liquors	.64	.02	.16
Kirsch	.50	−.06	−.10
Mirabelle	.46	−.24	−.19
Rum	.17	.74	.97*
Marc	−.29	.66	−.39
Whiskey	−.29	−.08	.09
Calvados	−.49	.20	−.04
Cognac	−.52	−.03	.42
Armagnac	−.60	−.17	.14

*This figure is too high. It exceeds the maximum value of .64, as a result of an approximation method for obtaining the estimated factor loadings used by Stoetzel.

Given these results, Stoetzel concluded the following: The major principle of liquor preference in France is the distinction between sweet and strong liquors. The second motivating element is price, which can be understood by remembering that liquor is both an expensive commodity and an item of conspicuous consumption. Except in the case of the two most popular and least expensive items (rum and marc), this second factor plays a much smaller role in producing preference judgments. The third factor concerns the sociological and primarily the regional, variability of the judgments. (See [14], p. 11.)

(a) Given what you know about the various liquors involved, does Stoetzel's interpretation seem reasonable?

(b) Plot the loading pairs for the first two factors. Conduct a graphical orthogonal rotation of the factor axes. Generate approximate rotated loadings. Interpret the rotated loadings for the first two factors. Does your interpretation agree with Stoetzel's interpretation of these factors from the unrotated loadings? Explain.

9.10. The correlation matrix for chicken-bone measurements (see Example 9.14) is

$$\begin{bmatrix} 1.000 & & & & & \\ .505 & 1.000 & & & & \\ .569 & .422 & 1.000 & & & \\ .602 & .467 & .926 & 1.000 & & \\ .621 & .482 & .877 & .874 & 1.000 & \\ .603 & .450 & .878 & .894 & .937 & 1.000 \end{bmatrix}$$

The following estimated factor loadings were extracted by the maximum likelihood procedure:

Variable	Estimated factor loadings F_1	F_2	Varimax rotated estimated factor loadings F_1^*	F_2^*
1. Skull length	.602	.200	.484	.411
2. Skull breadth	.467	.154	.375	.319
3. Femur length	.926	.143	.603	.717
4. Tibia length	1.000	.000	.519	.855
5. Humerus length	.874	.476	.861	.499
6. Ulna length	.894	.327	.744	.594

Using the *unrotated* estimated factor loadings, obtain the maximum likelihood estimates of the following.

(a) The specific variances.

(b) The communalities.

(c) The proportion of variance explained by each factor.

(d) The residual matrix $\mathbf{R} - \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' - \hat{\boldsymbol{\Psi}}_z$.

9.11. Refer to Exercise 9.10. Compute the value of the varimax criterion using both unrotated and rotated estimated factor loadings. Comment on the results.

9.12. The *covariance* matrix for the logarithms of turtle measurements (see Example 8.4) is

$$\mathbf{S} = 10^{-3}\begin{bmatrix} 11.072 & & \\ 8.019 & 6.417 & \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

The following maximum likelihood estimates of the factor loadings for an $m = 1$ model were obtained:

Variable	Estimated factor loadings F_1
1. ln(length)	.1022
2. ln(width)	.0752
3. ln(height)	.0765

Using the estimated factor loadings, obtain the maximum likelihood estimates of each of the following.

(a) Specific variances.

(b) Communalities.

(c) Proportion of variance explained by the factor.

(d) The residual matrix $\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}}$.
Hint: Convert $\mathbf{S}$ to $\mathbf{S}_n$.

9.13. Refer to Exercise 9.12. Compute the test statistic in (9-39). Indicate why a test of $H_0: \boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$ (with $m = 1$) versus $H_1: \boldsymbol{\Sigma}$ unrestricted cannot be carried out for this example. [See (9-40).]

9.14. The maximum likelihood factor loading estimates are given in (9A-6) by

$$\hat{\mathbf{L}} = \hat{\boldsymbol{\Psi}}^{1/2}\hat{\mathbf{E}}\hat{\boldsymbol{\Delta}}^{1/2}$$

Verify, for this choice, that

$$\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}} = \hat{\boldsymbol{\Delta}}$$

where $\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\Lambda}} - \mathbf{I}$ is a diagonal matrix.

9.15. Hirschey and Wichern [7] investigate the consistency, determinants, and uses of accounting and market-value measures of profitability. As part of their study, a factor analysis of accounting profit measures and market estimates of economic profits was conducted. The correlation matrix of accounting historical, accounting replacement, and market-value measures of profitability for a sample of firms operating in 1977 is as follows:

Variable	HRA	HRE	HRS	RRA	RRE	RRS	Q	REV
Historical return on assets, HRA	1.000							
Historical return on equity, HRE	.738	1.000						
Historical return on sales, HRS	.731	.520	1.000					
Replacement return on assets, RRA	.828	.688	.652	1.000				
Replacement return on equity, RRE	.681	.831	.513	.887	1.000			
Replacement return on sales, RRS	.712	.543	.826	.867	.692	1.000		
Market Q ratio, Q	.625	.322	.579	.639	.419	.608	1.000	
Market relative excess value, REV	.604	.303	.617	.563	.352	.610	.937	1.000

The following rotated principal component estimates of factor loadings for an $m = 3$ factor model were obtained:

Variable	Estimated factor loadings F_1	F_2	F_3
Historical return on assets	.433	.612	.499
Historical return on equity	.125	.892	.234
Historical return on sales	.296	.238	.887
Replacement return on assets	.406	.708	.483
Replacement return on equity	.198	.895	.283
Replacement return on sales	.331	.414	.789
Market Q ratio	.928	.160	.294
Market relative excess value	.910	.079	.355
Cumulative proportion of total variance explained	.287	.628	.908

(a) Using the estimated factor loadings, determine the specific variances and communalities.

(b) Determine the residual matrix, $\mathbf{R} - \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' - \hat{\mathbf{\Psi}}_z$. Given this information and the cumulative proportion of total variance explained in the preceding table, does an $m = 3$ factor model appear appropriate for these data?

(c) Assuming that estimated loadings less than .4 are small, interpret the three factors. Does it appear, for example, that market-value measures provide evidence of profitability distinct from that provided by accounting measures? Can you separate accounting historical measures of profitability from accounting replacement measures?

9.16. Verify that factor scores constructed according to (9-50) have sample mean vector $\mathbf{0}$ and zero sample covariances.

9.17. Refer to Example 9.12. Using the information in this example, evaluate $(\hat{\mathbf{L}}_z'\hat{\mathbf{\Psi}}_z^{-1}\hat{\mathbf{L}}_z)^{-1}$.

Note: Set the fourth diagonal element of $\hat{\mathbf{\Psi}}_z$ to .01 so that $\hat{\mathbf{\Psi}}_z^{-1}$ can be determined. Will the regression and generalized least squares methods for constructing factors scores for standardized stock price observations give nearly the same results? *Hint:* See equation (9-57) and the discussion following it.

The following exercises require the use of a computer.

9.18. Refer to Exercise 8.16 concerning the numbers of fish caught.

(a) Using only the measurements $x_1 - x_4$, obtain the principal component solution for factor models with $m = 1$ and $m = 2$.

(b) Using only the measurements $x_1 - x_4$, obtain the maximum likelihood solution for factor models with $m = 1$ and $m = 2$.

(c) Rotate your solutions in Parts (a) and (b). Compare the solutions and comment on them. Interpret each factor.

(d) Perform a factor analysis using the measurements $x_1 - x_6$. Determine a reasonable number of factors m, and compare the principal component and maximum likelihood solutions after rotation. Interpret the factors.

9.19. A firm is attempting to evaluate the quality of its sales staff and is trying to find an examination or series of tests that may reveal the potential for good performance in sales

The firm has selected a random sample of 50 sales people and has evaluated each on 3 measures of performance: growth of sales, profitability of sales, and new-account sales. These measures have been converted to a scale, on which 100 indicates "average" performance. Each of the 50 individuals took each of 4 tests, which purported to measure creativity, mechanical reasoning, abstract reasoning, and mathematical ability, respectively. The $n = 50$ observations on $p = 7$ variables are listed in Table 9.12 on page 536.

(a) Assume an orthogonal factor model for the *standardized variables* $Z_i = (X_i - \mu_i)/\sqrt{\sigma_{ii}}$, $i = 1, 2, \ldots, 7$. Obtain either the principal component solution or the maximum likelihood solution for $m = 2$ and $m = 3$ common factors.

(b) Given your solution in (a), obtain the rotated loadings for $m = 2$ and $m = 3$. Compare the two sets of rotated loadings. Interpret the $m = 2$ and $m = 3$ factor solutions.

(c) List the estimated communalities, specific variances, and $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ for the $m = 2$ and $m = 3$ solutions. Compare the results. Which choice of m do you prefer at this point? Why?

(d) Conduct a test of $H_0: \mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$ versus $H_1: \mathbf{\Sigma} \neq \mathbf{LL}' + \mathbf{\Psi}$ for both $m = 2$ and $m = 3$ at the $\alpha = .01$ level. With these results and those in Parts b and c, which choice of m appears to be the best?

(e) Suppose a new salesperson, selected at random, obtains the test scores $\mathbf{x}' = [x_1, x_2, \ldots, x_7] = [110, 98, 105, 15, 18, 12, 35]$. Calculate the salesperson's factor score using the weighted least squares method and the regression method.

Note: The components of $\mathbf{x}$ must be standardized using the sample means and variances calculated from the original data.

9.20. Using the air-pollution variables X_1, X_2, X_5, and X_6 given in Table 1.5, generate the sample *covariance* matrix.

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

(b) Find the maximum likelihood estimates of $\mathbf{L}$ and $\mathbf{\Psi}$ for $m = 1$ and $m = 2$.

(c) Compare the factorization obtained by the principal component and maximum likelihood methods.

9.21. Perform a varimax rotation of both $m = 2$ solutions in Exercise 9.20. Interpret the results. Are the principal component and maximum likelihood solutions consistent with each other?

9.22. Refer to Exercise 9.20.

(a) Calculate the factor scores from the $m = 2$ maximum likelihood estimates by (i) weighted least squares in (9-50) and (ii) the regression approach of (9-58).

(b) Find the factor scores from the principal component solution, using (9-51).

(c) Compare the three sets of factor scores.

9.23. Repeat Exercise 9.20, starting from the sample *correlation* matrix. Interpret the factors for the $m = 1$ and $m = 2$ solutions. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain.

9.24. Perform a factor analysis of the census-tract data in Table 8.5. Start with $\mathbf{R}$ and obtain both the maximum likelihood and principal component solutions. Comment on your choice of m. Your analysis should include factor rotation and the computation of factor scores.

9.25. Perform a factor analysis of the "stiffness" measurements given in Table 4.3 and discussed in Example 4.14. Compute factor scores, and check for outliers in the data. Use the sample covariance matrix $\mathbf{S}$.

Table 9.12 Salespeople Data

	Index of:			Score on:			
Salesperson	Sales growth (x_1)	Sales profit-ability (x_2)	New-account sales (x_3)	Creativity test (x_4)	Mechanical reasoning test (x_5)	Abstract reasoning test (x_6)	Mathe-matics test (x_7)
1	93.0	96.0	97.8	09	12	09	20
2	88.8	91.8	96.8	07	10	10	15
3	95.0	100.3	99.0	08	12	09	26
4	101.3	103.8	106.8	13	14	12	29
5	102.0	107.8	103.0	10	15	12	32
6	95.8	97.5	99.3	10	14	11	21
7	95.5	99.5	99.0	09	12	09	25
8	110.8	122.0	115.3	18	20	15	51
9	102.8	108.3	103.8	10	17	13	31
10	106.8	120.5	102.0	14	18	11	39
11	103.3	109.8	104.0	12	17	12	32
12	99.5	111.8	100.3	10	18	08	31
13	103.5	112.5	107.0	16	17	11	34
14	99.5	105.5	102.3	08	10	11	34
15	100.0	107.0	102.8	13	10	08	34
16	81.5	93.5	95.0	07	09	05	16
17	101.3	105.3	102.8	11	12	11	32
18	103.3	110.8	103.5	11	14	11	35
19	95.3	104.3	103.0	05	14	13	30
20	99.5	105.3	106.3	17	17	11	27
21	88.5	95.3	95.8	10	12	07	15
22	99.3	115.0	104.3	05	11	11	42
23	87.5	92.5	95.8	09	09	07	16
24	105.3	114.0	105.3	12	15	12	37
25	107.0	121.0	109.0	16	19	12	39
26	93.3	102.0	97.8	10	15	07	23
27	106.8	118.0	107.3	14	16	12	39
28	106.8	120.0	104.8	10	16	11	49
29	92.3	90.8	99.8	08	10	13	17
30	106.3	121.0	104.5	09	17	11	44
31	106.0	119.5	110.5	18	15	10	43
32	88.3	92.8	96.8	13	11	08	10
33	96.0	103.3	100.5	07	15	11	27
34	94.3	94.5	99.0	10	12	11	19
35	106.5	121.5	110.5	18	17	10	42
36	106.5	115.5	107.0	08	13	14	47
37	92.0	99.5	103.5	18	16	08	18
38	102.0	99.8	103.3	13	12	14	28
39	108.3	122.3	108.5	15	19	12	41
40	106.8	119.0	106.8	14	20	12	37
41	102.5	109.3	103.8	09	17	13	32
42	92.5	102.5	99.3	13	15	06	23
43	102.8	113.8	106.8	17	20	10	32
44	83.3	87.3	96.3	01	05	09	15
45	94.8	101.8	99.8	07	16	11	24
46	103.5	112.0	110.8	18	13	12	37
47	89.5	96.0	97.3	07	15	11	14
48	84.3	89.8	94.3	08	08	08	09
49	104.3	109.5	106.5	14	12	12	36
50	106.0	118.5	105.0	12	16	11	39

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?

9.35. Repeat Exercise 9.34 using observations on the pulp fiber characteristic variables AFL, LFF, FFF, and ZST. Can these data be summarized by a single factor? Explain.

9.36. Factor analyze the Mali family farm data in Table 8.7. Use the sample correlation matrix $\mathbf{R}$. Try both the principal component and maximum likelihood solution methods for $m = 3$, 4, and 5 factors. Can you interpret the factors? Justify your choice of m. Your analysis should include factor rotation and the computation of factor scores. Can you identify any outliers in these data?

References

1. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.
2. Bartlett, M. S. "The Statistical Conception of Mental Factors." *British Journal of Psychology*, **28** (1937), 97–104.
3. Bartlett, M. S. "A Note on Multiplying Factors for Various Chi-Squared Approximations." *Journal of the Royal Statistical Society (B)* **16** (1954), 296–298.
4. Dixon, W. S. *Statistical Software Manual to Accompany BMDP Release 7/version 7.0* (paperback). Berkeley, CA: University of California Press, 1992.
5. Dunn, L. C. "The Effect of Inbreeding on the Bones of the Fowl." *Storrs Agricultural Experimental Station Bulletin*, **52** (1928), 1–112.
6. Harmon, H. H. *Modern Factor Analysis* (3rd ed.). Chicago: The University of Chicago Press, 1976.
7. Hirschey, M., and D. W. Wichern. "Accounting and Market-Value Measures of Profitability: Consistency, Determinants and Uses." *Journal of Business and Economic Statistics*, **2**, no. 4 (1984), 375–383.
8. Joreskog, K. G. "Factor Analysis by Least Squares and Maximum Likelihood." In *Statistical Methods for Digital Computers*, edited by K. Enslein, A. Ralston, and H. S. Wilf. New York: John Wiley, 1975.
9. Kaiser, H.F. "The Varimax Criterion for Analytic Rotation in Factor Analysis." *Psychometrika*, **23** (1958), 187–200.
10. Lawley, D. N., and A. E. Maxwell. *Factor Analysis as a Statistical Method* (2nd ed.). New York: American Elsevier Publishing Co., 1971.
11. Linden, M. "A Factor Analytic Study of Olympic Decathlon Data." *Research Quarterly*, **48**, no. 3 (1977), 562–568.
12. Maxwell, A. E. *Multivariate Analysis in Behavioral Research*. London: Chapman and Hall, 1977.
13. Morrison, D. F. *Multivariate Statistical Methods* (4th ed.). Belmont, CA: Brooks/Cole Thompson Learning, 2005.
14. Stoetzel, J. "A Factor Analysis of Liquor Preference." *Journal of Advertising Research*, **1** (1960), 7–11.
15. Wright, S. "The Interpretation of Multivariate Systems." In *Statistics and Mathematics in Biology*, edited by O. Kempthorne and others. Ames, IA: Iowa State University Press, 1954, 11–33.