Chapter 9

9.1

$$L' = [.9 \quad .7 \quad .5]; \qquad LL' = \begin{bmatrix} .81 & .63 & .45 \\ .63 & .49 & .35 \\ .45 & .35 & .25 \end{bmatrix}$$

so $\boldsymbol{\rho} = LL' + \Psi$

9.2 a) For $m=1$

$$h_1^2 = \ell_{11}^2 = .81$$
$$h_2^2 = \ell_{21}^2 = .49$$
$$h_3^2 = \ell_{31}^2 = .25$$

The communalities are those parts of the variances of the variables explained by the single factor.

b) $\mathrm{Corr}(Z_i,F_1) = \mathrm{Cov}(Z_i,F_1)$, $i = 1,2,3$. By (9-5) $\mathrm{Cov}(Z_i,F_1) = \ell_{i1}$. Thus $\mathrm{Corr}(Z_1,F_1) = \ell_{11} = .9$; $\mathrm{Corr}(Z_2,F_1) = \ell_{21} = .7$; $\mathrm{Corr}(Z_3,F_1) = \ell_{31} = .5$. The first variable, Z_1, has the largest correlation with the factor and therefore will probably carry the most weight in naming the factor.

9.3 a) $L = \sqrt{\lambda_1}\,\underset{\sim}{e}_1 = \sqrt{1.96}\begin{bmatrix} .625 \\ .593 \\ .507 \end{bmatrix} = \begin{bmatrix} .876 \\ .831 \\ .711 \end{bmatrix}$. Slightly different from result in Exercise 9.1.

b) Proportion of total variance explained $= \dfrac{\lambda_1}{p} = \dfrac{1.96}{3} = .65$

9.4

$$\boldsymbol{\rho} = \boldsymbol{\rho} - \Psi = LL' = \begin{bmatrix} .81 & .63 & .45 \\ .63 & .49 & .35 \\ .45 & .35 & .25 \end{bmatrix}$$

$$L = \sqrt{\lambda_1}\,\underset{\sim}{e}_1 = \sqrt{1.55}\begin{bmatrix} .7229 \\ .5623 \\ .4016 \end{bmatrix} = \begin{bmatrix} .9 \\ .7 \\ .5 \end{bmatrix}$$

Result is consistent with results in Exercise 9.1. It should be since $m = 1$ common factor completely determines $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho} - \Psi$.

9.5

Since $\tilde{\Psi}$ is diagonal and $S - \tilde{L}\tilde{L}' - \tilde{\Psi}$ has zeros on the diagonal, (sum of squared entries of $S - \tilde{L}\tilde{L}' - \tilde{\Psi}$) $\le$ (sum of squared entries of $S - \tilde{L}\tilde{L}$). By the hint, $S - \tilde{L}\tilde{L} = \hat{P}_{(2)}\hat{\Lambda}_{(2)}\hat{P}'_{(3)}$ which has sum of squared entries

$$\mathrm{tr}[\hat{P}_{(2)}\hat{\Lambda}_{(2)}\hat{P}'_{(2)}(\hat{P}_{(2)}\hat{\Lambda}_{(2)}\hat{P}'_{(2)})'] = \mathrm{tr}[\hat{P}_{(2)}\hat{\Lambda}_{(2)}\hat{\Lambda}'_{(2)}\hat{P}'_{(2)}]$$

$$= \mathrm{tr}[\hat{\Lambda}_{(2)}\hat{\Lambda}'_{(2)}\hat{P}'_{(2)}\hat{P}_{(2)}] = \mathrm{tr}[\hat{\Lambda}_{(2)}\hat{\Lambda}'_{(2)}]$$

$$= \hat{\lambda}^2_{m+1} + \hat{\lambda}^2_{m+2} + \cdots + \hat{\lambda}^2_p$$

Therefore,

$$(\text{sum of squared entries of } S - \tilde{L}\tilde{L}' - \tilde{\Psi}) \le \hat{\lambda}^2_{m+1} + \hat{\lambda}^2_{m+2} + \cdots + \hat{\lambda}^2_p$$

9.6

a) Follows directly from hint.

b) Using the hint, we post multiply by $(LL' + \Psi)$ to get

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

$$= \Psi^{-1}(LL' + \Psi) - \Psi^{-1}\underbrace{L(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}(L}_{\text{(use part (a))}}L' + \Psi)$$

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

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

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

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

Note all these multiplication steps are reversible.

c) Multiplying the result in (b) by L we get

$$(LL' + \Psi)^{-1}L = \Psi^{-1}L - \Psi^{-1}\underbrace{L(I + L'\Psi^{-1}L)^{-1}L'\Psi^{-1}L}_{\text{(use part (a))}}$$

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

Result follows by taking the transpose of both sides of the final equality.

9.7 From the equation $\Sigma = LL' + \Psi$, $m = 1$, we have

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \ell_{11}^2 + \psi_1 & \ell_{11}\ell_{21} \\ \ell_{11}\ell_{21} & \ell_{21}^2 + \psi_2 \end{bmatrix}$$

so $\sigma_{11} = \ell_{11}^2 + \psi_1$, $\sigma_{22} = \ell_{21}^2 + \psi_2$ and $\sigma_{12} = \ell_{11}\ell_{21}$.
Let $\rho = \sigma_{12}/\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}$. Then, for any choice $|\rho|\sqrt{\sigma_{22}} \le \ell_{21} \le \sqrt{\sigma_{22}}$, set $\ell_{11} = \sigma_{12}/\ell_{21}$ and check $\sigma_{12} = \ell_{11}\ell_{21}$. We obtain $\psi_1 = \sigma_{11} - \ell_{11}^2 = \sigma_{11} - \frac{\sigma_{12}}{\ell_{21}^2} \ge \sigma_{11} - \frac{\sigma_{12}^2}{\rho^2\sigma_{22}} = \sigma_{11} - \sigma_{11} = 0$ and $\psi_2 = \sigma_{22} - \ell_{21}^2 \ge \sigma_{22} - \sigma_{22} = 0$. Since ℓ_{21} was arbitrary within a suitable interval, there are an infinite number of solutions to the factorization.

9.8 $\Sigma = LL' + \Psi$ for $m = 1$ implies

$$\begin{pmatrix} 1 = \ell_{11}^2 + \psi_1 & .4 = \ell_{11}\ell_{21} & .9 = \ell_{11}\ell_{31} \\ & 1 = \ell_{21}^2 + \psi_2 & .7 = \ell_{21}\ell_{31} \\ & & 1 = \ell_{31}^2 + \psi_3 \end{pmatrix}$$

Now $\frac{\ell_{11}}{\ell_{21}} = \frac{.9}{.7}$ and $\ell_{11}\ell_{21} = .4$, so $\ell_{11}^2 = (\frac{.9}{.7})(.4)$ and $\ell_{11} = \pm\ .717$. Thus $\ell_{21} = \pm\ .558$. Finally, from $.9 = \ell_{11}\ell_{31}$, we have $\ell_{31} = \pm\ .9/.717 = \pm 1.255$.

Note all the loadings must be of the same sign because all the covariances are positive. We have

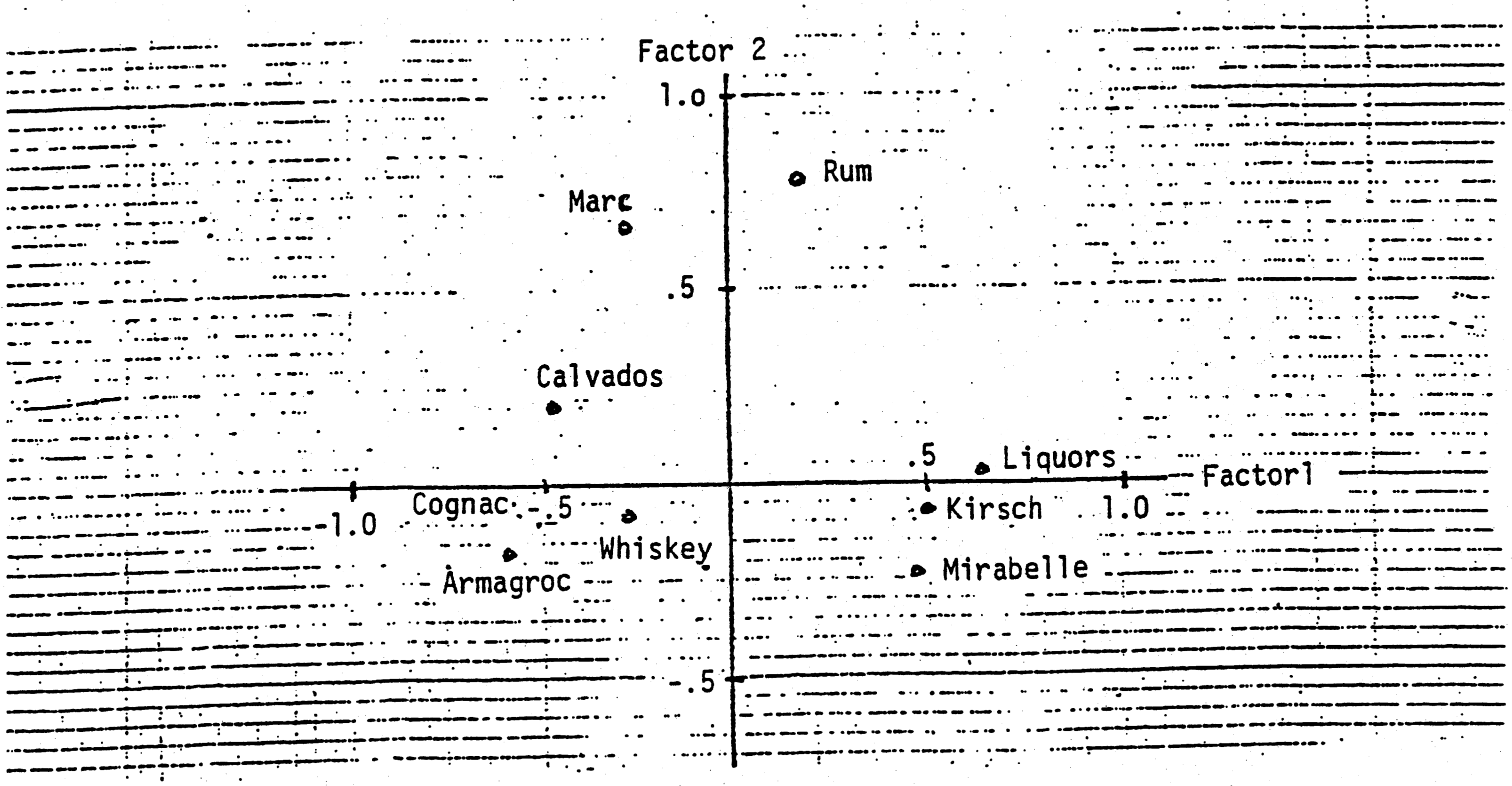
$$LL' = \begin{bmatrix} .717 \\ .558 \\ 1.255 \end{bmatrix} [.717\ .558\ 1.255] = \begin{bmatrix} .514 & .4 & .9 \\ .4 & .3111 & .7 \\ .9 & .7 & 1.575 \end{bmatrix}$$

so $\psi_3 = 1 - 1.575 = -.575$, which is inadmissible as a variance.

9.9

(a) Stoetzel's interpretation seems reasonable. The first factor seems to contrast sweet with strong liquors.

(b)



It doesn't appear as if rotation of the factor axes is necessary.

(a) & (b)

The specific variances and communalities based on the unrotated factors, are given in the following table:

Variable	Specific Variance	Communality
Skull length	.5976	.4024
Skull breadth	.7582	.2418
Femur length	.1221	.8779
Tibia length	.0000	1.0000
Humerus length	.0095	.9905
Ulna length	.0938	.9062

(c) The proportion of variance explained by each factor is:

$$\text{Factor 1}: \quad \frac{1}{6}\sum_{i=1}^{p} \ell_{1i}^2 = \frac{4.0001}{6} \quad \text{or} \quad 66.7\%$$

$$\text{Factor 2}: \quad \frac{1}{6}\sum_{i=1}^{p} \ell_{2i}^2 = \frac{.4177}{6} \quad \text{or} \quad 6.7\%$$

(d) $R - \hat{L}_z \hat{L}_z' - \hat{\Psi} =$

$$\begin{bmatrix} 0 & & & & & \\ .193 & 0 & & & & \\ -.017 & -.032 & 0 & & & \\ .000 & .000 & .000 & 0 & & \\ -.000 & .001 & .000 & .000 & 0 & \\ -.001 & -.018 & .003 & .000 & .000 & 0 \end{bmatrix}$$

9.11 Substituting the factor loadings given in the table (Exercise 9.10) into equation (9-45) gives.

$$V\ (\text{unrotated}) = .01087$$

$$V\ (\text{rotated}) = .04692$$

Although the rotated loadings are to be preferred by the varimax ("simple structure") criterion, interpretation of the factors seems clearer with the unrotated loadings.

9.12

The covariance matrix for the logarithms of turtle measurements is:

$$\mathbf{S} = 10^{-3} \times \begin{bmatrix} 11.0720040 & 8.0191419 & 8.1596480 \\ 8.0191419 & 6.4167255 & 6.0052707 \\ 8.1596480 & 6.0052707 & 6.7727585 \end{bmatrix}$$

The maximum likelihood estimates of the factor loadings for an m=1 model are

Variable	Estimated factor loadings F_1
1. ln(length)	0.1021632
2. ln(width)	0.0752017
3. ln(height)	0.0765267

Therefore,

$$\hat{L} = \begin{bmatrix} 0.1021632 \\ 0.0752017 \\ 0.0765267 \end{bmatrix}, \quad \hat{L}\hat{L}' = 10^{-3} \times \begin{bmatrix} 10.4373 & 7.6828 & 7.8182 \\ 7.6828 & 5.6553 & 5.7549 \\ 7.8182 & 5.7549 & 5.8563 \end{bmatrix}$$

(b) Since $\hat{h}_i^2 = \hat{l}_{i1}^2$ for an m=1 model, the communalities are

$$\hat{h}_1^2 = 0.0104373, \quad \hat{h}_2^2 = 0.0056553, \quad \hat{h}_3^2 = 0.0058563$$

(a) To find specific variances ψ_i's, we use the equation

$$\hat{\psi}_i = s_{ii} - \hat{h}_i^2$$

Note that in this case, we should use S_n to get s_{ii}, not S because the maximum likelihood estimation method is used.

$$S_n = \frac{n-1}{n} S = \frac{23}{24} S = 10^{-3} \times \begin{bmatrix} 10.6107 & 7.685 & 7.8197 \\ 7.685 & 6.1494 & 5.7551 \\ 7.8197 & 5.7551 & 6.4906 \end{bmatrix}$$

Thus we get

$$\hat{\psi}_1 = 0.0001734, \quad \hat{\psi}_2 = 0.0004941, \quad \hat{\psi}_3 = 0.0006342$$

(c) The proportion explained by the factor is

$$\frac{\hat{h}_1^2 + \hat{h}_2^2 + \hat{h}_3^2}{s_{11} + s_{22} + s_{33}} = \frac{0.0219489}{0.0232507} = .9440$$

(d) From (a)-(c), the residual matrix is:

$$S_n - \hat{L}\hat{L}' - \hat{\Psi} = 10^{-6} \times \begin{bmatrix} 0 & 2.1673 & 1.4474 \\ 2.1673 & 0 & 0.112497 \\ 1.4474 & 0.112497 & 0 \end{bmatrix}.$$

9.13

Equation (9-40) requires $m < \frac{1}{2}(2p+1 - \sqrt{8p+1})$. Here we have $m = 1$, $p = 3$ and the strict inequality does not hold.

9.14 Since

$$\hat{\Psi}^{\frac{1}{2}}\hat{\Psi}^{-1}\hat{\Psi}^{\frac{1}{2}} = I, \quad \hat{\Delta}^{\frac{1}{2}}\hat{\Delta}^{\frac{1}{2}} = \hat{\Delta} \quad \text{and} \quad \hat{E}'\hat{E} = I,$$

$$\hat{L}'\hat{\Psi}^{-1}\hat{L} = \hat{\Delta}^{\frac{1}{2}}\hat{E}'\hat{\Psi}^{\frac{1}{2}}\hat{\Psi}^{-1}\hat{\Psi}^{\frac{1}{2}}\hat{E}\hat{\Delta}^{\frac{1}{2}} = \hat{\Delta}^{\frac{1}{2}}\hat{E}'\hat{E}\hat{\Delta}^{\frac{1}{2}} = \hat{\Delta}^{\frac{1}{2}}\hat{\Delta}^{\frac{1}{2}} = \hat{\Delta}.$$

9.15

(a)

variable	variance	communality
HRA	0.188966	0.811034
HRE	0.133955	0.866045
HRS	0.068971	0.931029
RRA	0.100611	0.899389
RRE	0.079682	0.920318
RRS	0.096522	0.903478
Q	0.02678	0.97322
REV	0.039634	0.960366

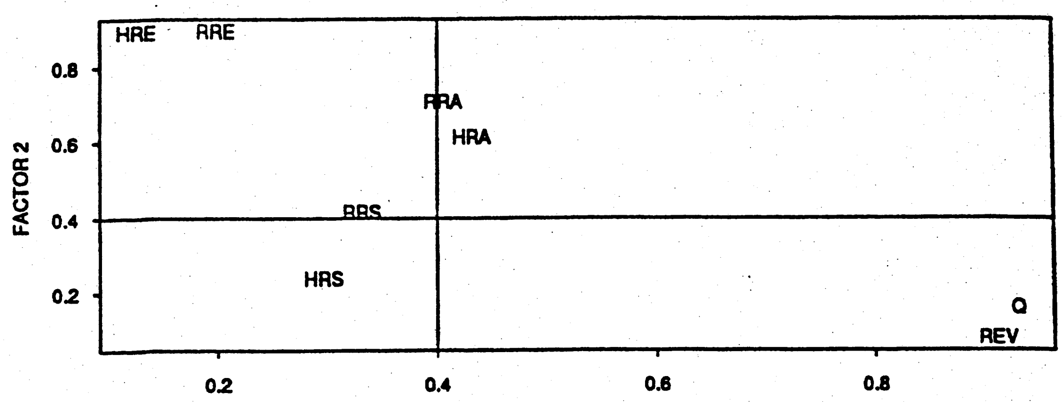
(b) **Residual Matrix**

0	0.021205	0.014563	-0.022111	-0.093691	-0.078402	-0.02145	-0.015523
0.021205	0	0.063146	-0.107308	-0.058312	-0.052289	-0.005516	0.035712
0.014563	0.063146	0	-0.065101	-0.009639	-0.070351	0.005454	0.013953
-0.022111	-0.107308	-0.065101	0	0.036263	0.058415	0.00695	-0.033857
-0.093691	-0.058312	-0.009639	0.036263	0	0.032645	0.008854	0.00065
-0.078402	-0.052289	-0.070351	0.058415	0.032645	0	0.002626	-0.004011
-0.02145	-0.005516	0.005454	0.00695	0.008854	0.002626	0	-0.02449
-0.015523	0.035712	0.013953	-0.033857	0.00065	-0.004011	-0.02449	0

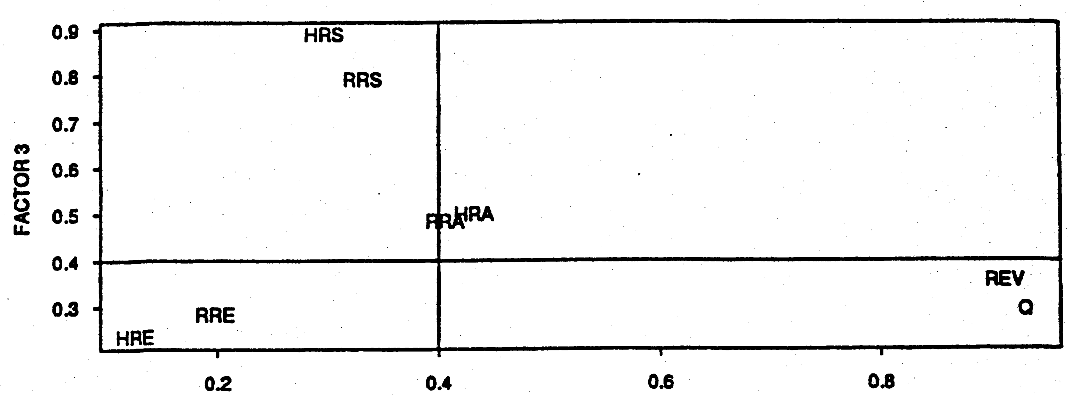
The m=3 factor model appears appropriate.

(c) The first factor is related to market-value measures (Q, REV). The second factor is related to accounting historical measures on equity (HRE, RRE). The third factor is related to accounting historical measures on sales (HRS, RRS). Accounting historical measures on assets (HRA,RRA) are weakly related to all factors. Therefore, market-value measures provide evidence of profitability distinct from that provided by the accounting measures. However, we cannot separate accounting historical measures of profitability from accounting replacement measures.

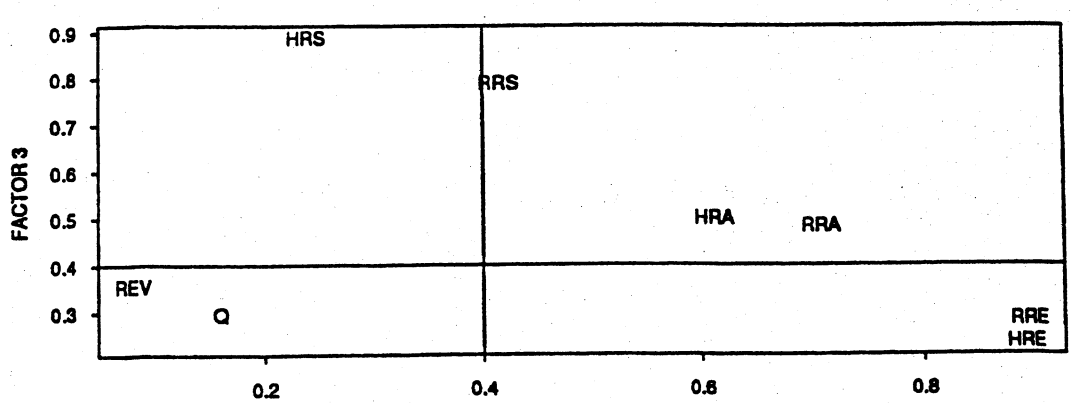
PROBLEM 9.15



FACTOR 1
Rotated Factor Pattern



FACTOR 1
Rotated Factor Pattern



FACTOR 2
Rotated Factor Pattern

9.16 From (9-50) $\hat{\mathbf{f}}_j = \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$ and

$$\sum_{j=1}^{n} \mathbf{f}_j = \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}}) = \mathbf{0} .$$

Since $$\hat{\mathbf{f}}_j\hat{\mathbf{f}}_j' = \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\hat{\Psi}^{-1}\hat{L}\hat{\Delta}^{-1} ,$$

$$\sum_{j=1}^{n} \hat{\mathbf{f}}_j\hat{\mathbf{f}}_j' = \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\hat{\Psi}^{-1}\hat{L}\hat{\Delta}^{-1}$$

$$= n\, \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-1}S_n\hat{\Psi}^{-1}\hat{L}\hat{\Delta}^{-1} .$$

Using (9A-1),

$$\sum_{j=1}^{n} \mathbf{f}_j\mathbf{f}_j' = n\, \hat{\Delta}^{-1}\hat{L}'\hat{\Psi}^{-\frac{1}{2}}\hat{\Psi}^{-\frac{1}{2}}\hat{L}(I + \hat{\Delta})\hat{\Delta}^{-1}$$

$$= n\, \hat{\Delta}^{-1}\hat{\Delta}(I + \hat{\Delta})\hat{\Delta}^{-1} = n(I + \hat{\Delta}^{-1}),$$

a diagonal matrix. Consequently, the factor scores have sample mean vector $\mathbf{0}$ and zero sample covariances.

9.17 Using the information in Example 9.12, we have

$(\hat{\mathbf{L}}_z'\hat{\Psi}_z^{-1}\hat{\mathbf{L}}_z)^{-1} = \begin{pmatrix} .2220 & -.0283 \\ -.0283 & .0137 \end{pmatrix}$ which, apart from rounding error, is a diagonal matrix. Since the number in the (1,1) position, .2220, is appreciably different from 0, and the observations have been standardized, equation (9-57) suggests the regression and generalized least squares methods for computing factor scores could give somewhat different results.

9.18. Factor analysis of Wisconsin fish data

(a) Principal component solution using $x_1 - x_4$

Initial Factor Method: Principal Components

	1	2	3	4
Eigenvalue	2.1539	0.7876	0.6157	0.4429
Difference	1.3663	0.1719	0.1728	
Proportion	0.5385	0.1969	0.1539	0.1107
Cumulative	0.5385	0.7354	0.8893	1.0000

Factor Pattern (m = 1)

	FACTOR1
BLUEGILL	0.77273
BCRAPPIE	0.73867
SBASS	0.64983
LBASS	0.76738

Factor Pattern (m = 2)

	FACTOR1	FACTOR2
BLUEGILL	0.77273	-0.40581
BCRAPPIE	0.73867	-0.36549
SBASS	0.64983	0.67309
LBASS	0.76738	0.19047

(b) Maximum likelihood solution using $x_1 - x_4$

Initial Factor Method: Maximum Likelihood

Factor Pattern (m = 1)

	FACTOR1
BLUEGILL	0.70812
BCRAPPIE	0.63002
SBASS	0.48544
LBASS	0.65312

Factor Pattern (m = 2)

	FACTOR1	FACTOR2
BLUEGILL	0.98748	-0.02251
BCRAPPIE	0.50404	0.25907
SBASS	0.28186	0.65863
LBASS	0.48073	0.41799

(c) Varimax rotation. Note that rotation is not possible with 1 factor.

Principal Components
Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2
BLUEGILL	0.85703	0.16518
BCRAPPIE	0.80526	0.17543
SBASS	0.08767	0.93147
LBASS	0.48072	0.62774

Maximum Likelihood
Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2
BLUEGILL	0.96841	0.19445
BCRAPPIE	0.43501	0.36324
SBASS	0.13066	0.70439
LBASS	0.37743	0.51319

For both solutions, Bluegill and Crappie load heavily on the first factor, while large-mouth and smallmouth bass load heavily on the second factor.

(d) Factor analysis using $x_1 - x_6$

Initial Factor Method: Principal Components

	1	2	3	4	5	6
Eigenvalue	2.3549	1.0719	0.9843	0.6644	0.5004	0.4242
Difference	1.2830	0.0876	0.3199	0.1640	0.0762	
Proportion	0.3925	0.1786	0.1640	0.1107	0.0834	0.0707
Cumulative	0.3925	0.5711	0.7352	0.8459	0.9293	1.0000

Factor Pattern (m = 3)

	FACTOR1	FACTOR2	FACTOR3
BLUEGILL	0.72944	-0.02285	-0.47611
BCRAPPIE	0.72422	0.01989	-0.20739
SBASS	0.60333	0.58051	0.26232
LBASS	0.76170	0.07998	-0.03199
WALLEYE	-0.39334	0.83342	-0.01286
NPIKE	0.44657	-0.18156	0.80285

Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3
BLUEGILL	0.85090	-0.12720	-0.13806
BCRAPPIE	0.74189	0.11256	-0.06957
SBASS	0.51192	0.46222	0.54231
LBASS	0.71176	0.28458	0.00311
WALLEYE	-0.24459	-0.21480	0.86227
NPIKE	0.05282	0.92348	-0.14613

Initial Factor Method: Maximum Likelihood

Factor Pattern

	FACTOR1	FACTOR2	FACTOR3
BLUEGILL	0.00000	1.00000	0.00000
BCRAPPIE	0.18979	0.49190	0.23481
SBASS	0.96466	0.26350	0.00000
LBASS	0.29875	0.46530	0.29435
WALLEYE	0.12927	-0.22770	-0.49746
NPIKE	0.24062	0.06520	0.46665

Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3
BLUEGILL	0.99637	0.06257	0.05767
BCRAPPIE	0.46485	0.21097	0.26931
SBASS	0.20017	0.97853	0.04905
LBASS	0.42801	0.31567	0.33099
WALLEYE	-0.20771	0.13392	-0.50492
NPIKE	0.02359	0.22600	0.47779

The first principal component factor influences the Bluegill, Crappie and the Bass. The Northern Pike alone loads heavily on the second factor, and the Walleye and smallmouth bass on the third factor. The MLE solution is different.

9.19 (a), (b) and (c) Maximum Likelihood (m = 3)

UNROTATED FACTOR LOADINGS (PATTERN)
FOR MAXIMUM LIKELIHOOD CANONICAL FACTORS

		Factor 1	Factor 2	Factor 3
Growth	1	0.772	0.295	0.527
Profits	2	0.570	0.347	0.721
Newaccts	3	0.774	0.433	0.355
Creative	4	0.389	0.921	0.000
Mechanic	5	0.509	0.426	0.334
Abstract	6	0.968	-0.250	0.000
Math	7	0.632	0.181	0.729
	VP	3.267	1.520	1.566

ROTATED FACTOR LOADINGS (PATTERN)

		Factor 1	Factor 2	Factor 3
Growth	1	0.794	0.374	0.437
Profits	2	0.912	0.316	0.184
Newaccts	3	0.653	0.544	0.437
Creative	4	0.255	0.967	0.019
Mechanic	5	0.541	0.464	0.208
Abstract	6	0.390	0.054	0.953
Math	7	0.919	0.179	0.295
	VP	3.180	1.720	1.454

		Communalities	Specific Variances
1	Growth	0.9615	.0385
2	Profits	0.9648	.0352
3	Newaccts	0.9124	.0876
4	Creative	1.0000	.0000
5	Mechanic	0.5519	.4481
6	Abstract	1.0000	.0000
7	Math	0.9631	.0369

$$R = \begin{bmatrix} 1.0 & .926 & .884 & .572 & .708 & .674 & .927 \\ & 1.0 & .843 & .542 & .746 & .465 & .944 \\ & & 1.0 & .700 & .637 & .641 & .853 \\ & & & 1.0 & .591 & .147 & .413 \\ & & & & 1.0 & .386 & .575 \\ \text{(Symmetric)} & & & & & 1.0 & .566 \\ & & & & & & 1.0 \end{bmatrix}$$

$$\hat{L}\hat{L}' + \hat{\Psi} = \begin{bmatrix} 1.0 & .923 & .912 & .572 & .694 & .674 & .925 \\ & 1.0 & .848 & .542 & .679 & .465 & .948 \\ & & 1.0 & .700 & .696 & .641 & .826 \\ & & & 1.0 & .591 & .147 & .413 \\ & & & & 1.0 & .386 & .646 \\ & & & & & 1.0 & .566 \\ \text{(Symmetric)} & & & & & & 1.0 \end{bmatrix}$$

It is clear from an examination of the residual matrix $R - (\hat{L}\hat{L}' + \hat{\Psi})$ that an $m = 3$ factor solution represents the observed correlations quite well. However, it is difficult to provide interpretations for the factors. If we consider the rotated loadings, we see that the last two factors are dominated by the single variables "creative" and "abstract" respectively. The first factor links the salespeople performance variables with math ability.

(d) Using (9-39) with $n = 50$, $p = 7$, $m = 3$ we have

$$43.833 \ln\left(\frac{.000075933}{.000018427}\right) = 62.1 > \chi_3^2(.01) = 11.3$$

so we reject $H_0: \Sigma = LL' + \Psi$ for $m = 3$. Neither of the $m = 2$, $m = 3$ factor models appear to fit by the χ^2 criterion. We note that the matrices R, $\hat{L}\hat{L}' + \hat{\Psi}$ have small determinants and rounding error could affect the calculation of the test statistic. Again, the residual matrix above indicates a good fit for $m = 3$.

(e) $\underset{\sim}{z}' = [1.522, -.852, .465, .957, 1.129, .673, .497]$

Using the regression method for computing factor scores, we have; with $\underset{\sim}{f} = \hat{L}_z R_z^{-1}$:

Principal components $(m=3)$	Maximum likelihood $(m=3)$
$\underset{\sim}{f}' = [.686, .271, 1.395]$	$\underset{\sim}{f}' = [-.702, .679, -.751]$

Factor scores using weighted least squares can only be computed for the principal component solutions since $\hat{\Psi}^{-1}$ cannot be computed for the maximum likelihood solutions. ($\hat{\Psi}$ has zeros on the main diagonal for the maximum likelihood solutions). Using (9-50),

Principal components $(m = 3)$

$\underset{\sim}{f}' = [.344, .233, 1.805]$

9.20

$$
\begin{array}{cc}
 & \begin{array}{cccc} X_1 & X_2 & X_5 & X_6 \end{array} \\
S = & \begin{bmatrix} 2.50 & -2.77 & -.59 & -2.23 \\ & 300.52 & 6.78 & 30.78 \\ & & 11.36 & 3.13 \\ \text{(symmetric)} & & & 31.98 \end{bmatrix}
\end{array}
$$

(a) Principal components (m = 2)

	Factor 1 loadings	Factor 2 loadings
X_1 (wind)	-.17	-.37
X_2 (solar rad.)	17.32	-.61
X_5 (NO_2)	.42	.74
X_6 (O_3)	1.96	5.19

(b) Maximum likelihood estimates of the loadings are obtained from $\hat{L} = \hat{V}^{1/2}\hat{L}_z$ where $\hat{L}_z$ are the loadings obtained from the sample correlation matrix R. (For $\hat{L}_z$ see problem 9.23). Note: Maximum likelihood estimates of the loadings for $m = 2$ may be difficult to obtain for some computer packages without good estimates of the communalities. One choice for initial estimates of the communalities are the communalities from the $m = 2$ principal components solution.

(c) Maximum likelihood estimation (with $m = 2$) does a better job of accounting for the covariances in S than the $m = 2$ principal component solution. On the other hand, the principal component solution generally produces uniformly smaller estimates of the specific variances. For the unrotated $m = 2$ solution, the first factor is dominated by X_2 = solar radiation and X_6 = O_3. The second factor seems to be a contrast between the pair X_1 = wind; X_2 = solar radiation and the pair X_5 = NO_2 and X_6 = O_3.

9.21 Principal components ($m = 2$)

	Rotated loadings	
	Factor 1	Factor 2
X_1 (wind)	.10	-.46
X_2 (solar rad.)	2.00	.05
X_5 (NO_2)	.05	.87
X_6 (O_3)	.71	5.49

Again the first factor is dominated by solar radiation and, to some extent, ozone. The second factor might be interpreted as a contrast between wind and the pair of pollutants NO_2 and O_3. Recall solar radiation and ozone have the largest sample variances. This will affect the estimated loadings obtained by the principal component method.

9.22 (a) Since, for maximum likelihood estimates, $\hat{L} = D^{\frac{1}{2}}\hat{L}_z$ and $S = D^{\frac{1}{2}}RD^{\frac{1}{2}}$, the factor scores generated by the equations for $\underset{\sim}{f}_j$ in (9-58) will be identical. Similarly, the factor scores generated by the weighted least squares formulas in (9-50) will be identical.

The factor scores generated by the regression method with maximum likelihood estimates ($m = 2$; see problem 9.23.) are given below for the first 10 cases.

Case	$\hat{f}_1$	$\hat{f}_2$
1	0.316	-0.544
2	0.252	-0.546
3	0.129	-0.509
4	0.332	-0.790
5	0.492	-0.012
6	0.515	-0.370
7	0.530	-0.456
8	1.070	0.724
9	0.384	-0.023
10	-0.179	0.105

(b) Factor scores using principal component estimates (m = 2) and (9-51) for the first 10 cases are given below:

Case	$\hat{f}_1$	$\hat{f}_2$
1	1.203	-0.368
2	1.646	-1.029
3	1.447	-0.937
4	0.717	0.795
5	0.856	-0.049
6	0.811	0.394
7	0.518	0.950
8	-0.083	1.168
9	0.410	0.259
10	-0.492	0.072

(c) The sets of factor scores are quite different. Factor scores depend heavily on the method used to estimate loadings and specific variances as well as the method used to generate them.

9.23

Principal components (m = 2)

	Factor 1 loadings	Factor 2 loadings	Rotated loadings Factor 1	Rotated loadings Factor 2
X_1 (wind)	-.56	-.24	-.31	$\boxed{-.53}$
X_2 (solar rad.)	.65	-.52	$\boxed{.83}$	-.04
X_5 (NO_2)	.48	.74	-.05	$\boxed{.88}$
X_6 (O_3)	.77	-.20	$\boxed{.74}$	.30

Maximum likelihood (m = 2)

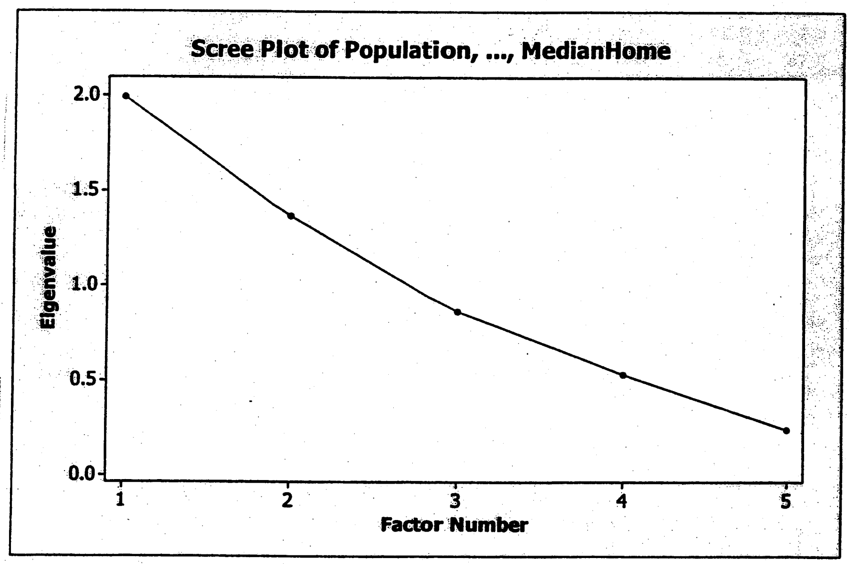
	Factor 1 loadings	Factor 2 loadings	Rotated loadings Factor 1	Rotated loadings Factor 2
X_1 (wind)	-.38	.32	-.09	$\boxed{.49}$
X_2 (solar rad.)	.50	.27	$\boxed{.56}$	-.10
X_5 (NO_2)	.25	-.04	.17	-.19
X_6 (O_3)	.65	-.03	$\boxed{.49}$	$\boxed{-.43}$

Examining the rotated loadings, we see that both solution methods yield similar estimated loadings for the first factor. It might be called a "ozone pollution factor". There are some differences for the second factor. However, the second factor appears to compare one of the pollutants with wind. It might be called a "pollutant transport" factor. We note that the interpretations of the factors might differ depending upon the choice of R or S (see problems 9.20 and 9.21) for analysis. Also the two solution methods give somewhat different results indicating the solution is not very stable. Some of the observed correlations between the variables are very small implying that a $m = 1$ or $m = 2$ factor model for these four variables will not be a completely satisfactory description of the underlying structure. We may need about as many factors as variables. If this is the case, there is nothing to be gained by proposing a factor model.

9.24

$$\mathbf{R} = \begin{pmatrix} 1.0 & -.192 & .313 & -.119 & .026 \\ -.192 & 1.0 & -.065 & .373 & .685 \\ .313 & -.065 & 1.0 & -.411 & -.010 \\ -.119 & .373 & -.411 & 1.0 & .180 \\ .026 & .685 & -.010 & .180 & 1.0 \end{pmatrix}$$

The correlations are relatively small with the possible exception of .685, the correlation between Percent Professional Degree and Median Home Value. Consequently, a factor analysis with fewer than 4 or 5 factors may be problematic. The scree plot, shown below, reinforces this conjecture. The scree plot falls off almost linearly, there is no sharp elbow. However, we present a factor analysis with $m = 3$ factors for both the principal components and maximum likelihood solutions.



Principal Component Factor Analysis ($m = 3$)

Unrotated Factor Loadings and Communalities

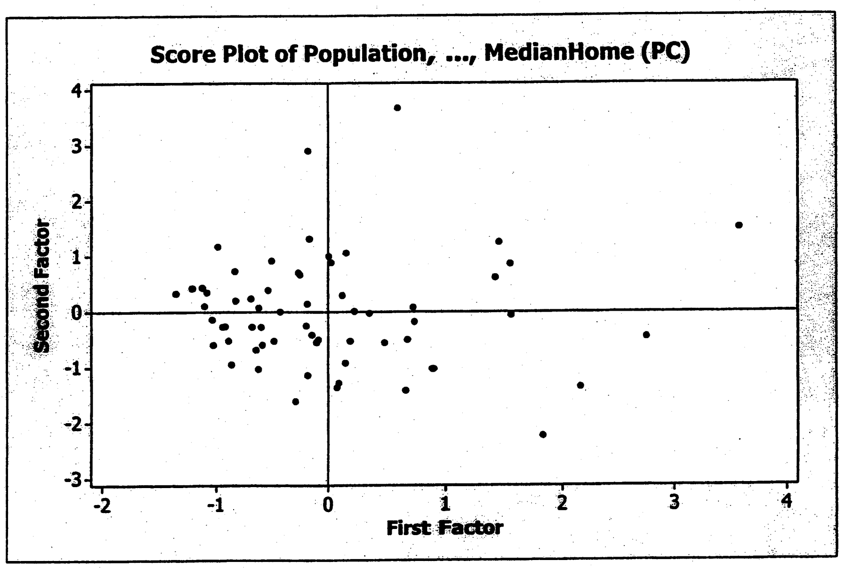
Variable	Factor1	Factor2	Factor3	Communality
Population	-0.371	-0.541	-0.729	0.962
PerCentProDeg	0.837	-0.381	0.153	0.870
PerCentEmp>16	-0.460	-0.708	0.209	0.756
PerCentGovEmp	0.676	0.295	-0.512	0.807
MedianHome	0.696	-0.584	0.064	0.830
Variance	1.9919	1.3675	0.8642	4.2236
% Var	0.398	0.274	0.173	0.845

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Communality
Population	-0.059	-0.118	-0.972	0.962
PerCentProDeg	0.907	0.160	0.147	0.870
PerCentEmp>16	0.102	-0.801	-0.321	0.756
PerCentGovEmp	0.277	0.850	-0.082	0.807
MedianHome	0.908	0.009	-0.068	0.830
Variance	1.7382	1.4050	1.0803	4.2236
% Var	0.348	0.281	0.216	0.845

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3
Population	-0.019	0.138	-0.940
PerCentProDeg	0.522	-0.028	0.109
PerCentEmp>16	0.169	-0.577	-0.135
PerCentGovEmp	0.052	0.658	-0.278
MedianHome	0.544	-0.099	-0.070



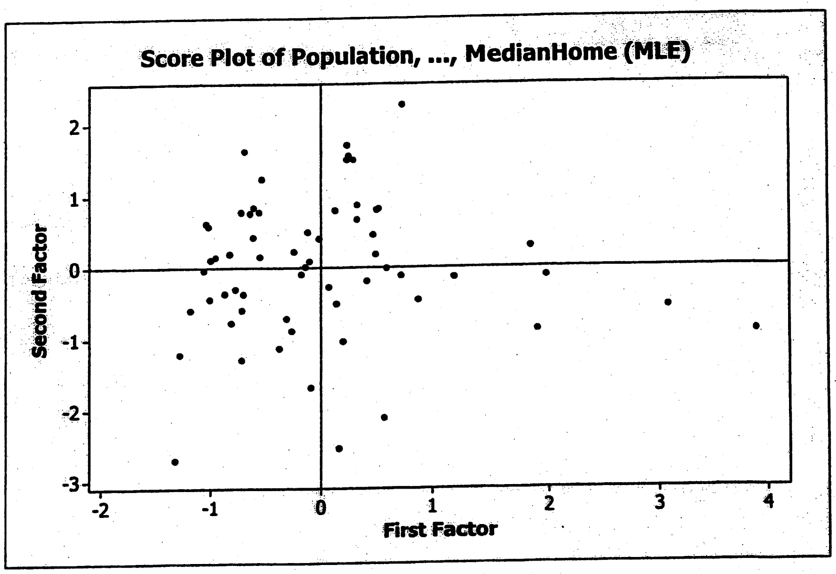
Maximum Likelihood Factor Analysis ($m = 3$)

* NOTE * Heywood case

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Population	-0.047	-0.999	-0.000	1.000
PerCentProDeg	0.989	0.146	-0.000	1.000
PerCentEmp>16	-0.020	-0.313	0.941	0.984
PerCentGovEmp	0.362	0.103	-0.395	0.298
MedianHome	0.701	-0.059	-0.015	0.496
Variance	1.6043	1.1310	1.0419	3.7772
% Var	0.321	0.226	0.208	0.755

```
Rotated Factor Loadings and Communalities
Varimax Rotation

Variable         Factor1  Factor2  Factor3  Communality
Population        -0.036    0.155   -0.987        1.000
PerCentProDeg      0.985   -0.090    0.145        1.000
PerCentEmp>16      0.047    0.977   -0.165        0.984
PerCentGovEmp      0.333   -0.430    0.041        0.298
MedianHome         0.699   -0.054   -0.061        0.496

Variance          1.5750   1.1740   1.0282       3.7772
% Var              0.315    0.235    0.206        0.755


Factor Score Coefficients

Variable         Factor1  Factor2  Factor3
Population         0.137   -0.177   -1.046
PerCentProDeg      1.017   -0.053   -0.046
PerCentEmp>16      0.070    1.025    0.159
PerCentGovEmp     -0.001   -0.010   -0.002
MedianHome        -0.000   -0.001   -0.000
```



A $m = 3$ factor solution explains from 75% to 85% of the variance depending on the solution method. Using the rotated loadings, the first factor in both methods has large loadings on Percent Professional Degree and Median Home Value. It is difficult to label this factor but since income is probably somewhere in this mix, it might be labeled an "affluence" or "white collar" factor. The second and third factors from the two solutions are similar as well. The second factor is a bipolar factor with large loadings (in absolute value) on Percent Employed over 16 and Percent Government Employment. We call this factor an "employment" factor. The third factor is clearly a "population" factor. Factor scores for the first two factors from the two solutions methods are similar.

9.25

$$S = \begin{bmatrix} 105,625 & 94,734 & 87,242 & 94,280 \\ & 101,761 & 76,186 & 81,204 \\ & & 91,809 & 90,343 \\ \text{(Symmetric)} & & & 104,329 \end{bmatrix}$$

A $m = 1$ factor model appears to represent these data quite well.

	Principal Components Factor 1 loadings	Maximum Likelihood Factor 1 loadings
Shock wave	317.	320.
Vibration	293.	291.
Static test 1	287.	275.
Static test 2	307.	297.
Proportion Variance Explained	90.1%	86.9%

Factor scores ($m = 1$) using the regression method for the first few cases are:

Principal Components	Maximum Likelihood
-.009	-.033
1.530	1.524
.808	.719
-.804	-.802

The factor scores produced from the two solution methods are very similar. The correlation between the two sets of scores is .992.

The outliers, specimens 9 and 16, were identified in Example 4.15.

9.26

a) Principal Components

	m = 1		m = 2		
	Factor 1 loadings	$\tilde{\Psi}_i$	Factor 1 loadings	Factor 2 loadings	$\tilde{\Psi}_i$
Litter 1	27.9	309.0	27.9	-6.2	271.2
Litter 2	30.4	205.7	30.4	-4.9	182.2
Litter 3	31.5	344.3	31.5	18.5	1.7
Litter 4	32.9	310.0	32.9	-8.0	245.8
Percentage Variance Explained	76.4%		76.4%	9.4%	

b) Maximum Likelihood

	m = 1	
	Factor loadings	$\hat{\Psi}_i$
Litter 1	26.8	370.2
Litter 2	30.5	198.2
Litter 3	28.4	529.6
Litter 4	30.4	471.0
Percentage Variance Explained	68.8%	

The maximum likelihood estimates of the factor loadings for $m = 2$ were not obtained due to convergence difficulties in the computer program.

c) It is only necessary to rotate the $m = 2$ solution.

Principal Components (m = 2)

	Rotated loadings Factor 1	Factor 2
Litter 1	26.2	11.4
Litter 2	27.5	13.8
Litter 3	14.7	33.4
Litter 4	31.4	12.8
Percentage Variance Explained	53.5%	32.4%

9.27 Principal Components (m = 2)

	Factor 1 loadings	Factor 2 loadings	$\tilde{\Psi}_i$	Rotated loadings Factor 1	Factor 2
Litter 1	.86	.44	.06	.33	.91
Litter 2	.91	.12	.15	.59	.71
Litter 3	.85	-.36	.14	.87	.32
Litter 4	.87	-.21	.20	.78	.44
Percentage Variance Explained	76.5%	9.5%		45.4%	40.6%

Maximum Likelihood (m = 1)

	Factor 1 loadings	$\hat{\Psi}_i$
Litter 1	.81	.34
Litter 2	.91	.17
Litter 3	.78	.39
Litter 4	.81	.34
Percentage Variance Explained	68.8%	

$$\hat{f} = \hat{L}_z R^{-1} \underset{\sim}{z} = .297$$

9.28 The covariance matrix **S** (see below) is dominated by the marathon since the marathon times are given in minutes. It is unlikely that a factor analysis will be useful; however, the principal component solution with m = 2 is given below. Using the unrotated loadings, the first factor explains about 98% of the variance and the largest factor loading is associated with the marathon. Using the rotated loadings, the first factor explains about 87% of the variance and again the largest loading is associated with the marathon. The second factor, with either unrotated or rotated loadings, explains relatively little of the remaining variance and can be ignored. The first factor might be labeled a "running endurance" factor but this factor provides us with little insight into the nature of the running events. It is better to factor analyze the correlation matrix **R** in this case.

Covariances: 100m(s), 200m(s), 400m(s), 800m, 1500m, 3000m, Marathon

	100m(s)	200m(s)	400m(s)	800m	1500m	3000m
100m(s)	0.15532					
200m(s)	0.34456	0.86309				
400m(s)	0.89130	2.19284	6.74546			
800m	0.02770	0.06617	0.18181	0.00755		
1500m	0.08389	0.20276	0.50918	0.02141	0.07418	
3000m	0.23388	0.55435	1.42682	0.06138	0.21616	0.66476
Marathon	4.33418	10.38499	28.90373	1.21965	3.53984	10.70609

	Marathon
Marathon	270.27015

Principal Component Factor Analysis of **S** ($m = 2$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
100m(s)	0.267	-0.230	0.124
200m(s)	0.640	-0.582	0.749
400m(s)	1.785	-1.881	6.725
800m	0.075	-0.027	0.006
1500m	0.217	-0.073	0.052
3000m	0.654	-0.158	0.453
Marathon	16.438	0.238	270.270
Variance	274.36	4.02	278.38
% Var	0.984	0.014	0.999

Rotated Factor Loadings and Communalities
Varimax Rotation

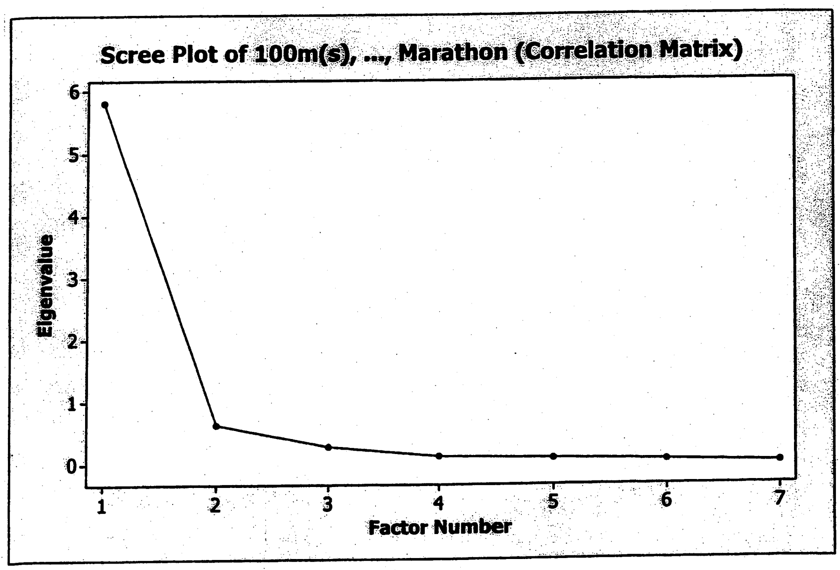
Variable	Factor1	Factor2	Communality
100m(s)	0.172	-0.308	0.124
200m(s)	0.401	-0.767	0.749
400m(s)	1.030	-2.380	6.725
800m	0.061	-0.051	0.006
1500m	0.178	-0.143	0.052
3000m	0.560	-0.373	0.453
Marathon	15.517	-5.431	270.270
Variance	242.38	36.00	278.38
% Var	0.869	0.129	0.999

The correlation matrix **R** for the women's track records follows.

Correlations: 100m(s), 200m(s), 400m(s), 800m, 1500m, 3000m, Marathon

	100m(s)	200m(s)	400m(s)	800m	1500m	3000m
200m(s)	0.941					
400m(s)	0.871	0.909				
800m	0.809	0.820	0.806			
1500m	0.782	0.801	0.720	0.905		
3000m	0.728	0.732	0.674	0.867	0.973	
Marathon	0.669	0.680	0.677	0.854	0.791	0.799

The scree plot below suggests at most a $m = 2$ factor solution.



Principal Component Factor Analysis of **R** (m =2)

Unrotated Factor Loadings and Communalities

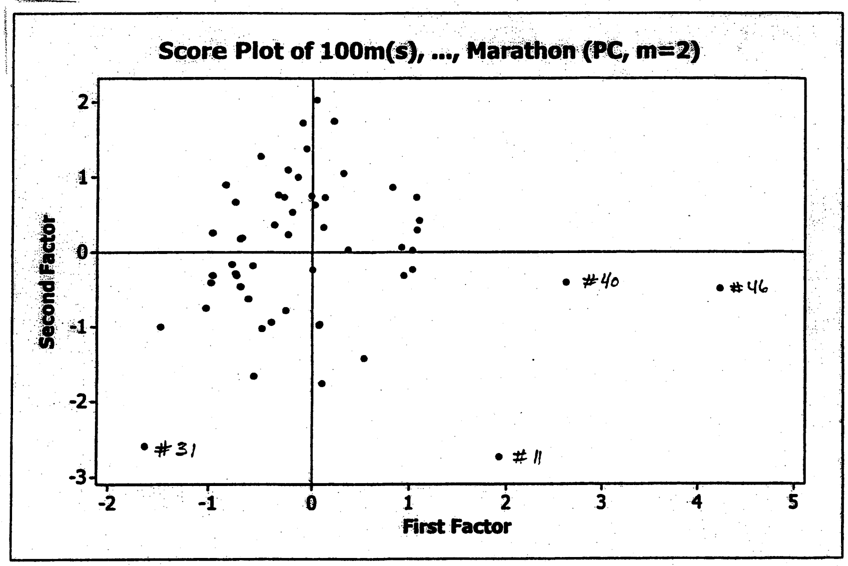
Variable	Factor1	Factor2	Communality
100m(s)	0.910	-0.323	0.933
200m(s)	0.923	-0.328	0.960
400m(s)	0.887	-0.364	0.919
800m	0.951	0.128	0.921
1500m	0.938	0.245	0.940
3000m	0.906	0.336	0.934
Marathon	0.856	0.309	0.828
Variance	5.8076	0.6287	6.4363
% Var	0.830	0.090	0.919

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m(s)	0.438	-0.861	0.933
200m(s)	0.444	-0.874	0.960
400m(s)	0.393	-0.875	0.919
800m	0.778	-0.562	0.921
1500m	0.849	-0.468	0.940
3000m	0.888	-0.381	0.934
Marathon	0.833	-0.365	0.828
Variance	3.3530	3.0833	6.4363
% Var	0.479	0.440	0.919

Factor Score Coefficients

Variable	Factor1	Factor2
100m(s)	-0.240	-0.480
200m(s)	-0.244	-0.488
400m(s)	-0.288	-0.525
800m	0.259	0.035
1500m	0.386	0.172
3000m	0.481	0.280
Marathon	0.445	0.255



Maximum Likelihood Factor Analysis of R ($m = 2$)

Unrotated Factor Loadings and Communalities

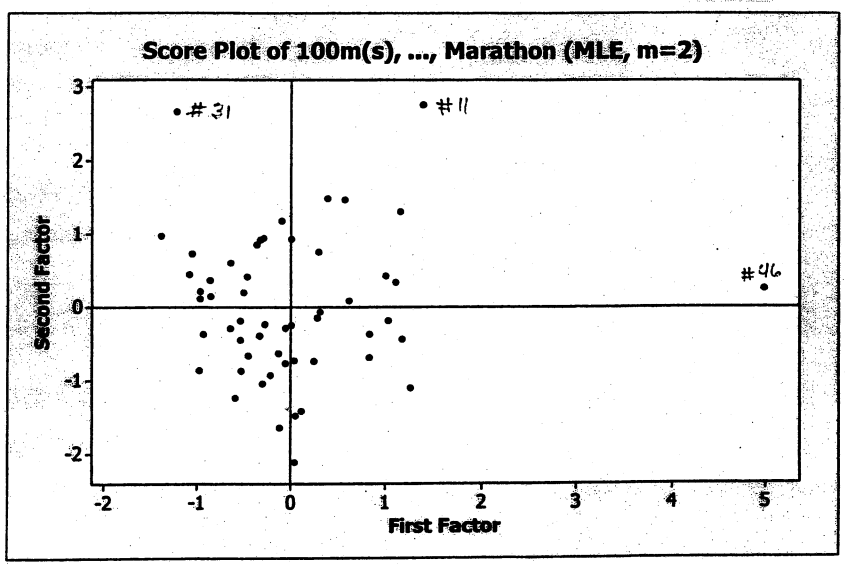
Variable	Factor1	Factor2	Communality
100m(s)	0.876	0.371	0.906
200m(s)	0.899	0.410	0.976
400m(s)	0.827	0.405	0.848
800m	0.925	-0.006	0.856
1500m	0.974	-0.187	0.984
3000m	0.945	-0.282	0.972
Marathon	0.809	-0.091	0.662
Variance	5.6104	0.5927	6.2032
% Var	0.801	0.085	0.886

```
Rotated Factor Loadings and Communalities
Varimax Rotation
```

Variable	Factor1	Factor2	Communality
100m(s)	0.455	0.836	0.906
200m(s)	0.449	0.880	0.976
400m(s)	0.395	0.832	0.848
800m	0.728	0.571	0.856
1500m	0.879	0.460	0.984
3000m	0.915	0.367	0.972
Marathon	0.690	0.432	0.662
Variance	3.1806	3.0225	6.2032
% Var	0.454	0.432	0.886

```
Factor Score Coefficients
```

Variable	Factor1	Factor2
100m(s)	-0.107	0.237
200m(s)	-0.481	1.019
400m(s)	-0.077	0.157
800m	0.036	0.025
1500m	0.772	-0.317
3000m	0.595	-0.369
Marathon	0.024	-0.003



The results from the two solution methods are very similar. Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events load highly on this factor. The second factor appears to contrast the shorter running events (100m, 200m, 400m) with the longer events (800m, 1500m, 3000m, marathon). This bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 800m through the marathon load highly on this factor. The second factor might be classified as a "running speed" factor. Note, for both factors, the remaining running events in each case have moderately large loadings on the factor. The two factor solution accounts for 89%-92% (depending on solution method) of the total variance. The plots of the factor scores indicate that observations #46 (Samoa), #11 (Cook Islands) and #31 (North Korea) are outliers.

9.29 The covariance matrix **S** for the running events measured in meters/second is given below. Since all the running event variables are now on a commensurate measurement scale, it is likely a factor analysis of **S** will produce nearly the same results as a factor analysis of the correlation matrix **R**. The results for a $m = 2$ factor analysis of **S** using the principal component method are shown below. A factor analysis of **R** follows.

Covariances: 100m/s, 200m/s, 400m/s, 800m/s, 1500m/s, 3000m/s, Marm/s

	100m/s	200m/s	400m/s	800m/s	1500m/s	3000m/s
100m/s	0.0905383					
200m/s	0.0956063	0.1146714				
400m/s	0.0966724	0.1138699	0.1377889			
800m/s	0.0650640	0.0749249	0.0809409	0.0735228		
1500m/s	0.0822198	0.0960189	0.0954430	0.0864542	0.1238405	
3000m/s	0.0921422	0.1054364	0.1083164	0.0997547	0.1437148	0.1765843
Marm/s	0.0810999	0.0933103	0.1018807	0.0943056	0.1184578	0.1465604

	Marm/s
Marm/s	0.1667141

Principal Component Factor Analysis of S ($m = 2$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
100m/s	0.265	-0.110	0.083
200m/s	0.306	-0.127	0.110
400m/s	0.324	-0.152	0.128
800m/s	0.256	0.016	0.066
1500m/s	0.335	0.062	0.116
3000m/s	0.393	0.116	0.168
Marm/s	0.362	0.130	0.148
Variance	0.73215	0.08607	0.81822
% Var	0.829	0.097	0.926

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.128	-0.257	0.083
200m/s	0.147	-0.297	0.110
400m/s	0.145	-0.327	0.128
800m/s	0.204	-0.156	0.066
1500m/s	0.293	-0.173	0.116
3000m/s	0.373	-0.170	0.168
Marm/s	0.359	-0.139	0.148
Variance	0.45423	0.36399	0.81822
% Var	0.514	0.412	0.926

Factor Score Coefficients

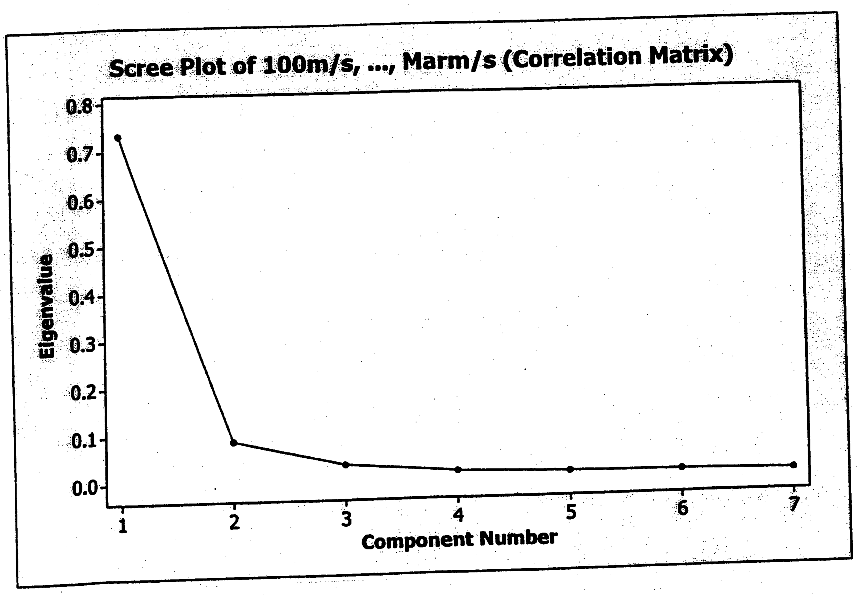
Variable	Factor1	Factor2
100m/s	-0.171	-0.363
200m/s	-0.222	-0.471
400m/s	-0.306	-0.603
800m/s	0.104	-0.025
1500m/s	0.287	0.085
3000m/s	0.542	0.280
Marm/s	0.558	0.335

Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events have roughly the same size loadings on this factor. The second factor appears to contrast the shorter running events (100m, 200m, 400m) with the longer events (800m, 1500m, 3000m, marathon). This bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 800m through the marathon have higher loadings on this factor. The second factor might be classified as a "running speed" factor. Note, for both factors, the remaining running events in each case have moderate and roughly equal loadings on the factor. The two factor solution accounts for 93% of the variance.

The correlation matrix **R** is shown below along with the scree plot. A two factor solution seems warranted.

Correlations: 100m/s, 200m/s, 400m/s, 800m/s, 1500m/s, 3000m/s, Marm/s

	100m/s	200m/s	400m/s	800m/s	1500m/s	3000m/s
200m/s	0.938					
400m/s	0.866	0.906				
800m/s	0.797	0.816	0.804			
1500m/s	0.776	0.806	0.731	0.906		
3000m/s	0.729	0.741	0.694	0.875	0.972	
Marm/s	0.660	0.675	0.672	0.852	0.824	0.854



Principal Component Factor Analysis of **R** ($m = 2$)

Unrotated Factor Loadings and Communalities

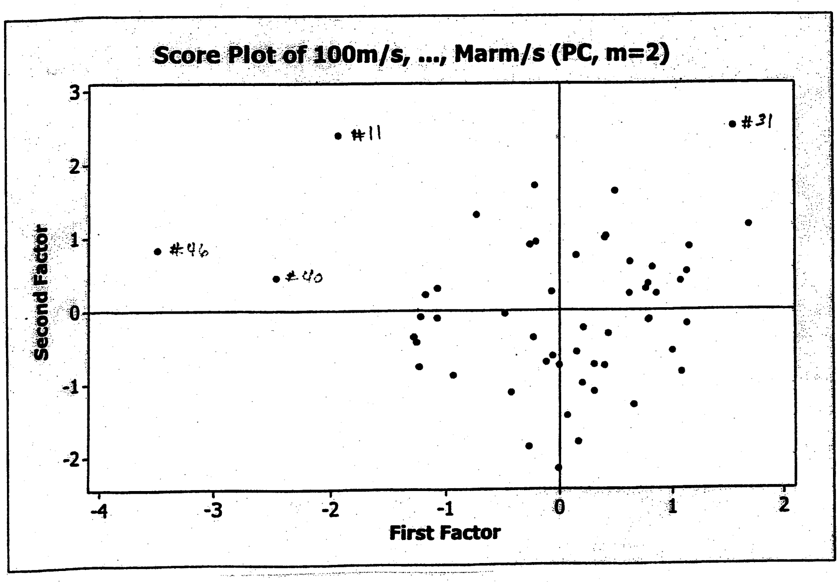
Variable	Factor1	Factor2	Communality
100m/s	0.903	-0.342	0.932
200m/s	0.921	-0.335	0.960
400m/s	0.887	-0.352	0.911
800m/s	0.948	0.123	0.914
1500m/s	0.943	0.227	0.941
3000m/s	0.919	0.320	0.947
Marm/s	0.866	0.354	0.875
Variance	5.8323	0.6477	6.4799
% Var	0.833	0.093	0.926

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.418	-0.870	0.932
200m/s	0.436	-0.878	0.960
400m/s	0.400	-0.867	0.911
800m/s	0.771	-0.565	0.914
1500m/s	0.839	-0.486	0.941
3000m/s	0.886	-0.402	0.947
Marm/s	0.871	-0.341	0.875
Variance	3.3675	3.1125	6.4799
% Var	0.481	0.445	0.926

Factor Score Coefficients

Variable	Factor1	Factor2
100m/s	-0.252	-0.489
200m/s	-0.243	-0.484
400m/s	-0.265	-0.499
800m/s	0.248	0.025
1500m/s	0.358	0.142
3000m/s	0.455	0.249
Marm/s	0.484	0.293



Maximum Likelihood Factor Analysis of **R** ($m = 2$)

Unrotated Factor Loadings and Communalities

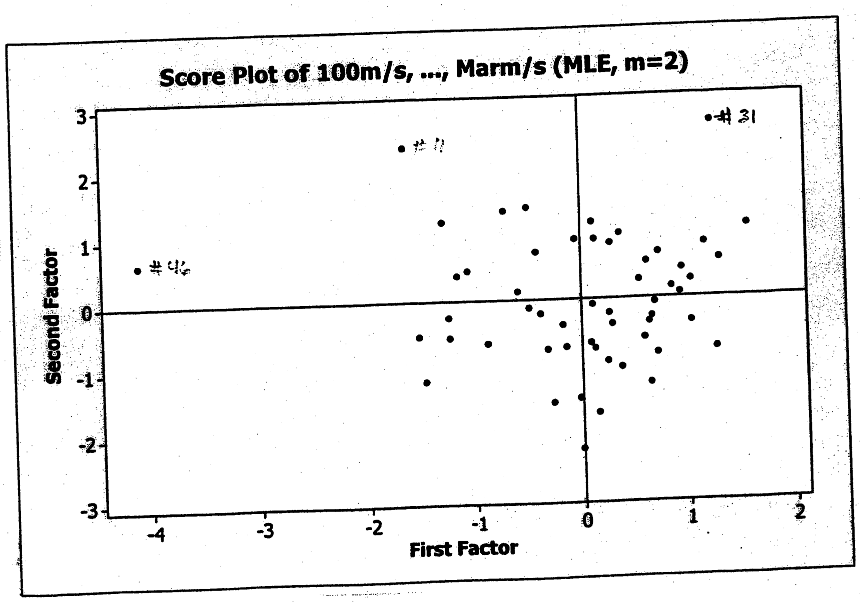
Variable	Factor1	Factor2	Communality
100m/s	0.880	-0.349	0.896
200m/s	0.910	-0.393	0.983
400m/s	0.844	-0.352	0.836
800m/s	0.921	0.042	0.850
1500m/s	0.966	0.193	0.971
3000m/s	0.945	0.302	0.984
Marm/s	0.834	0.203	0.737
Variance	5.6844	0.5716	6.2560
% Var	0.812	0.082	0.894

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.441	-0.838	0.896
200m/s	0.435	-0.891	0.983
400m/s	0.412	-0.816	0.836
800m/s	0.726	-0.568	0.850
1500m/s	0.859	-0.482	0.971
3000m/s	0.914	-0.386	0.984
Marm/s	0.765	-0.389	0.737
Variance	3.2395	3.0165	6.2560
% Var	0.463	0.431	0.894

Factor Score Coefficients

Variable	Factor1	Factor2
100m/s	-0.073	-0.167
200m/s	-0.521	-1.122
400m/s	-0.048	-0.106
800m/s	0.039	-0.014
1500m/s	0.379	0.124
3000m/s	0.949	0.518
Marm/s	0.041	0.017



The results from the two solution methods are very similar and very similar to the principal component factor analysis of the covariance matrix S. Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events load highly on this factor. The second factor appears to contrast the shorter running events (100m, 200m, 400m) with the longer events (800m, 1500m, 3000m, marathon). This bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 800m through the marathon load highly on this factor. The second factor might be classified as a "running speed" factor. Note, for both factors, the remaining running events in each case have moderately large loadings on the factor. The two factor solution accounts for 89%-93% (depending on solution method) of the total variance. The plots of the factor scores indicate that observations #46 (Samoa), #11 (Cook Islands) and #31 (North Korea) are outliers.

The results of the $m = 2$ factor analysis of women's track records when time is measured in meters per second are very much the same as the results for the $m = 2$ factor analysis of **R** presented in Exercise 9.28. If the correlation matrix **R** is factor analyzed, it makes little difference whether running event time is measured in seconds (or minutes) as in Exercise 9.28 or in meters per second. It does make a difference if the covariance matrix **S** is factor analyzed, since the measurement scales in Exercise 9.28 are quite different from the meters/second scale.

9.30 The covariance matrix **S** (see below) is dominated by the marathon since the marathon times are given in minutes. It is unlikely that a factor analysis will be useful; however, the principal component solution with m = 2 is given below. Using the unrotated loadings, the first factor explains about 98% of the variance and the largest factor loading is associated with the marathon. Using the rotated loadings, the first factor explains about 83% of the variance and again the largest loading is associated with the marathon. The second factor, with either unrotated or rotated loadings, explains relatively little of the remaining variance and can be ignored. The first factor might be labeled a "running endurance" factor but this factor provides us with little insight into the nature of the running events. It is better to factor analyze the correlation matrix **R** in this case.

Covariances: 100m, 200m, 400m, 800m, 1500m, 5000m, 10,000m, Marathon

	100m	200m	400m	800m	1500m	5000m
100m	0.048973					
200m	0.111044	0.300903				
400m	0.256022	0.666818	2.069956			
800m	0.008264	0.022929	0.057938	0.002751		
1500m	0.025720	0.066193	0.168473	0.007131	0.023034	
5000m	0.124575	0.317734	0.853486	0.034348	0.105833	0.578875
10,000m	0.265613	0.688936	1.849941	0.074257	0.229701	1.262533
Marathon	1.340139	3.541038	9.178857	0.378905	1.192564	6.430489

	10,000m	Marathon
10,000m	2.819569	
Marathon	14.342538	80.135356

Principal Component Factor Analysis of **S** ($m = 2$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
100m	0.152	-0.107	0.034
200m	0.401	-0.270	0.234
400m	1.044	-0.979	2.049
800m	0.043	-0.015	0.002
1500m	0.134	-0.033	0.019
5000m	0.722	-0.125	0.537
10,000m	1.610	-0.223	2.643
Marathon	8.950	0.179	80.130
Variance	84.507	1.141	85.649
% Var	0.983	0.013	0.996

Rotated Factor Loadings and Communalities
Varimax Rotation

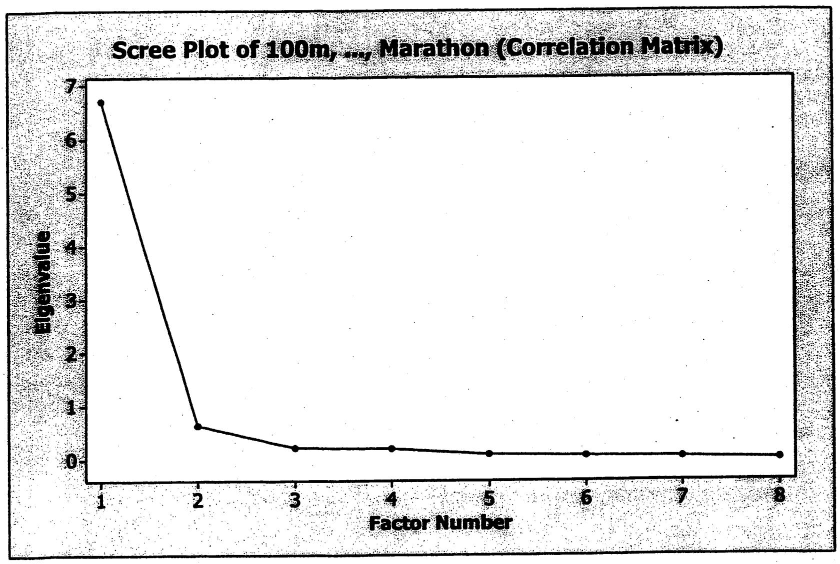
Variable	Factor1	Factor2	Communality
100m	0.097	-0.158	0.034
200m	0.262	-0.406	0.234
400m	0.573	-1.312	2.049
800m	0.033	-0.031	0.002
1500m	0.110	-0.083	0.019
5000m	0.615	-0.399	0.537
10,000m	1.392	-0.841	2.643
Marathon	8.294	-3.367	80.130
Variance	71.529	14.119	85.649
% Var	0.832	0.164	0.996

The correlation matrix **R** for the men's track records follows.

Correlations: 100m, 200m, 400m, 800m, 1500m, 5000m, 10,000m, Marathon

	100m	200m	400m	800m	1500m	5000m	10,000m
200m	0.915						
400m	0.804	0.845					
800m	0.712	0.797	0.768				
1500m	0.766	0.795	0.772	0.896			
5000m	0.740	0.761	0.780	0.861	0.917		
10,000m	0.715	0.748	0.766	0.843	0.901	0.988	
Marathon	0.676	0.721	0.713	0.807	0.878	0.944	0.954

The scree plot below suggests at most a $m = 2$ factor solution.



Principal Component Factor Analysis of **R** (m =2)

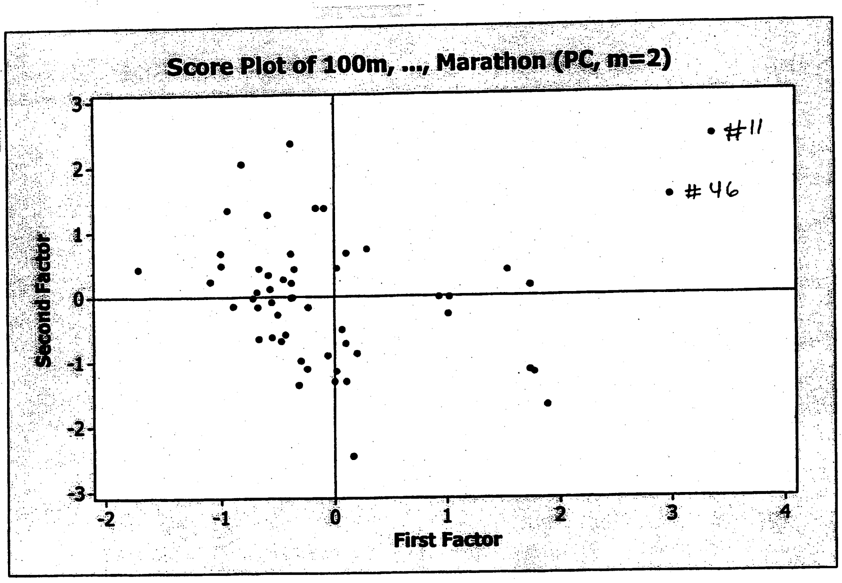
```
Unrotated Factor Loadings and Communalities

Variable   Factor1  Factor2  Communality
100m        0.861    0.423      0.920
200m        0.896    0.376      0.944
400m        0.878    0.276      0.847
800m        0.914   -0.071      0.840
1500m       0.948   -0.123      0.913
5000m       0.957   -0.236      0.972
10,000m     0.947   -0.267      0.969
Marathon    0.917   -0.309      0.937

Variance   6.7033   0.6384     7.3417
% Var       0.838    0.080      0.918
```

```
Rotated Factor Loadings and Communalities
Varimax Rotation
```

Variable	Factor1	Factor2	Communality
100m	0.375	0.882	0.920
200m	0.433	0.870	0.944
400m	0.485	0.782	0.847
800m	0.739	0.543	0.840
1500m	0.798	0.526	0.913
5000m	0.879	0.447	0.972
10,000m	0.892	0.417	0.969
Marathon	0.896	0.365	0.937
Variance	4.1168	3.2249	7.3417
% Var	0.515	0.403	0.918

```
Factor Score Coefficients
```

Variable	Factor1	Factor2
100m	-0.335	0.586
200m	-0.283	0.533
400m	-0.183	0.413
800m	0.176	0.004
1500m	0.233	-0.053
5000m	0.349	-0.186
10,000m	0.380	-0.224
Marathon	0.420	-0.277



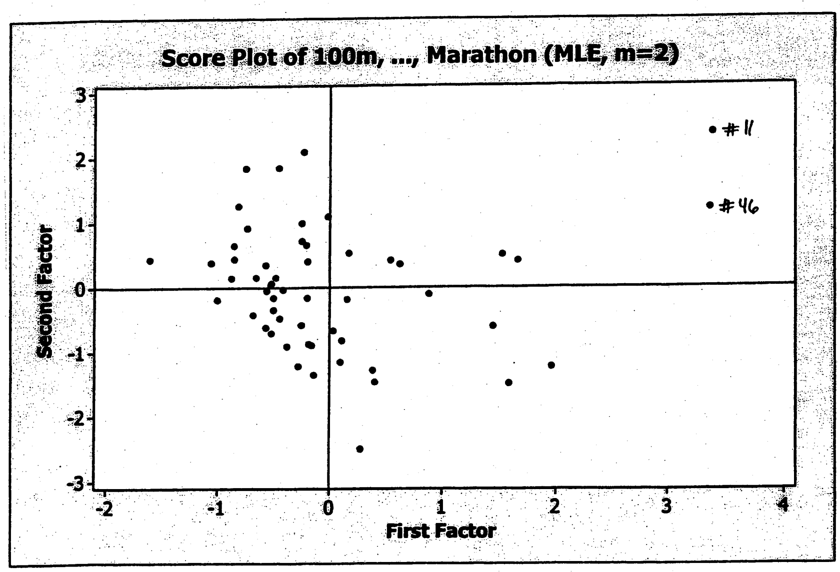
Maximum Likelihood Factor Analysis of **R** ($m = 2$)

```
Unrotated Factor Loadings and Communalities
```

Variable	Factor1	Factor2	Communality
100m	0.780	0.507	0.866
200m	0.814	0.548	0.963
400m	0.810	0.339	0.772
800m	0.875	0.147	0.788
1500m	0.927	0.083	0.866
5000m	0.991	-0.077	0.988
10,000m	0.989	-0.106	0.989
Marathon	0.949	-0.104	0.912
Variance	6.4134	0.7299	7.1432
% Var	0.802	0.091	0.893

```
Rotated Factor Loadings and Communalities
Varimax Rotation

Variable   Factor1  Factor2  Communality
100m         0.401    0.839        0.866
200m         0.409    0.892        0.963
400m         0.515    0.712        0.772
800m         0.671    0.581        0.788
1500m        0.748    0.554        0.866
5000m        0.886    0.450        0.988
10,000m      0.900    0.424        0.989
Marathon     0.865    0.405        0.912

Variance    3.9446   3.1986       7.1432
% Var        0.493    0.400        0.893
```

```
Factor Score Coefficients

Variable   Factor1  Factor2
100m        -0.125    0.256
200m        -0.490    0.994
400m        -0.044    0.104
800m        -0.011    0.054
1500m        0.003    0.056
5000m        0.558   -0.209
10,000m      0.761   -0.423
Marathon     0.089   -0.051
```



The results from the two solution methods are very similar. Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events load highly on this factor. The second factor appears to contrast the shorter running events with the longer events although the nature of the contrast is a bit different for the two methods. For the principal component method, the 100m, 200m and 400m events have positive loadings and the 800m, 1500m, 5000m, 10,000m and marathon events have negative loadings. For the maximum likelihood method, the 100m, 200m, 400m, 800m and 1500m events are in one group (positive loadings) and the 5000, 10,000m and marathon are in the other group (negative loadings). Nevertheless, this bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 800m through the marathon load highly on this factor. The second factor might be classified as a "running speed" factor. Note, for both factors, the remaining running events in each case have moderately large loadings on the factor. The two factor solution accounts for 89%-92% (depending on solution method) of the total variance. The plots of the factor scores indicate that observations #46 (Samoa) and #11 (Cook Islands) are outliers. The factor analysis of the men's track records is very much the same as that for the women's track records in Exercise 9.28.

9.31 The covariance matrix **S** for the running events measured in meters/second is given below. Since all the running event variables are now on a commensurate measurement scale, it is likely a factor analysis of **S** will produce nearly the same results as a factor analysis of the correlation matrix **R**. The results for a $m = 2$ factor analysis of **S** using the principal component method are shown below. A factor analysis of **R** follows.

Covariances: 100m/s, 200m/s, 400m/s, 800m/s, 1500m/s, 5000m/s, 10,000m/s, ...

	100m/s	200m/s	400m/s	800m/s	1500m/s
100m/s	0.0434979				
200m/s	0.0482772	0.0648452			
400m/s	0.0434632	0.0558678	0.0688217		
800m/s	0.0314951	0.0432334	0.0428221	0.0468840	
1500m/s	0.0425034	0.0535265	0.0537207	0.0523058	0.0729140
5000m/s	0.0469252	0.0587731	0.0617664	0.0571560	0.0766388
10,000m/s	0.0448325	0.0572512	0.0599354	0.0553945	0.0745719
Marathonm/s	0.0431256	0.0562945	0.0567342	0.0541911	0.0736518

	5000m/s	10,000m/s	Marathonm/s
5000m/s	0.0959398		
10,000m/s	0.0937357	0.0942894	
Marathonm/s	0.0905819	0.0909952	0.0979276

Principal Component Factor Analysis of **S** ($m = 2$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Communality
100m/s	0.171	-0.093	0.038
200m/s	0.219	-0.113	0.061
400m/s	0.223	-0.101	0.060
800m/s	0.195	-0.007	0.038
1500m/s	0.256	0.014	0.066
5000m/s	0.301	0.056	0.094
10,000m/s	0.296	0.067	0.092
Marathonm/s	0.293	0.083	0.093
Variance	0.49405	0.04622	0.54027
% Var	0.844	0.079	0.923

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.080	-0.178	0.038
200m/s	0.105	-0.222	0.061
400m/s	0.116	-0.215	0.060
800m/s	0.151	-0.124	0.038
1500m/s	0.212	-0.145	0.066
5000m/s	0.273	-0.138	0.094
10,000m/s	0.275	-0.127	0.092
Marathonm/s	0.283	-0.112	0.093
Variance	0.32860	0.21168	0.54027
% Var	0.562	0.362	0.923

Factor Score Coefficients

Variable	Factor1	Factor2
100m/s	-0.197	-0.377
200m/s	-0.287	-0.561
400m/s	-0.254	-0.526
800m/s	0.048	-0.078
1500m/s	0.159	-0.022
5000m/s	0.379	0.184
10,000m/s	0.415	0.240
Marathonm/s	0.489	0.334

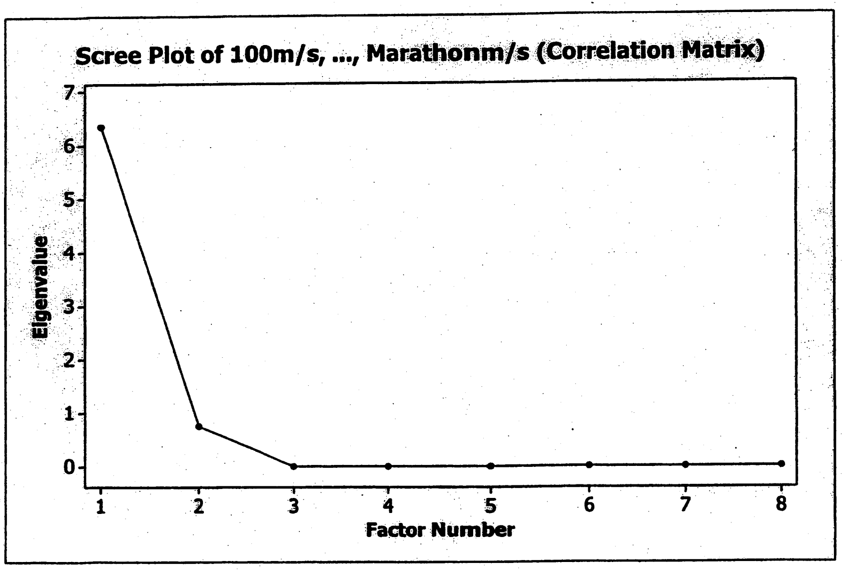
Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events have roughly the same size loadings on this factor. The second factor appears to contrast the shorter running events (100m, 200m, 400m, 800m) with the longer events (1500m, 5000m, 10,000, marathon). This bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 1500m through the marathon have higher loadings on this factor. The second factor might be classified as a "running speed" factor. Note, the 800m run has about equal (in absolute value) loadings on both factors and the remaining running events in each case have moderate and roughly equal loadings on the factor. The two factor solution accounts for 92% of the variance.

The correlation matrix **R** is shown next along with the scree plot. A two factor solution seems warranted.

Correlations: 100m/s, 200m/s, 400m/s, 800m/s, 1500m/s, 5000m/s, 10,000m/s, ...

	100m/s	200m/s	400m/s	800m/s	1500m/s
200m/s	0.909				
400m/s	0.794	0.836			
800m/s	0.697	0.784	0.754		
1500m/s	0.755	0.778	0.758	0.895	
5000m/s	0.726	0.745	0.760	0.852	0.916
10,000m/s	0.700	0.732	0.744	0.833	0.899
Marathonm/s	0.661	0.706	0.691	0.800	0.872

	5000m/s	10,000m/s
10,000m/s	0.986	
Marathonm/s	0.935	0.947



Principal Component Factor Analysis of **R** $(m = 2)$

Unrotated Factor Loadings and Communalities

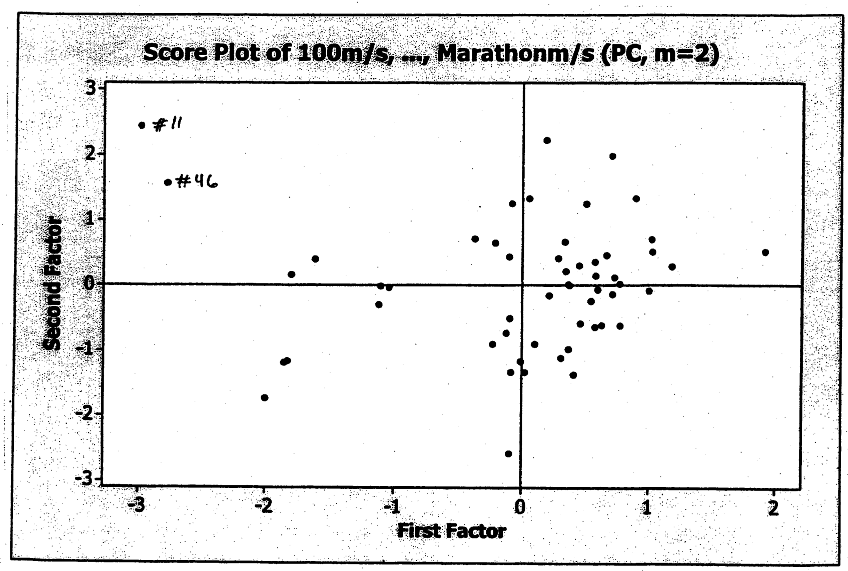
Variable	Factor1	Factor2	Communality
100m/s	0.854	-0.430	0.913
200m/s	0.888	-0.387	0.939
400m/s	0.868	-0.297	0.841
800m/s	0.910	0.076	0.834
1500m/s	0.947	0.133	0.914
5000m/s	0.954	0.242	0.968
10,000m/s	0.943	0.274	0.965
Marathonm/s	0.912	0.312	0.929
Variance	6.6258	0.6765	7.3023
% Var	0.828	0.085	0.913

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.369	-0.881	0.913
200m/s	0.423	-0.872	0.939
400m/s	0.466	-0.790	0.841
800m/s	0.741	-0.534	0.834
1500m/s	0.805	-0.515	0.914
5000m/s	0.882	-0.437	0.968
10,000m/s	0.895	-0.405	0.965
Marathonm/s	0.896	-0.355	0.929
Variance	4.1116	3.1907	7.3023
% Var	0.514	0.399	0.913

Factor Score Coefficients

Variable	Factor1	Factor2
100m/s	-0.315	-0.566
200m/s	-0.270	-0.522
400m/s	-0.186	-0.418
800m/s	0.178	-0.004
1500m/s	0.236	0.056
5000m/s	0.341	0.178
10,000m/s	0.371	0.215
Marathonm/s	0.405	0.261



Maximum Likelihood Factor Analysis of R $(m = 2)$

Unrotated Factor Loadings and Communalities

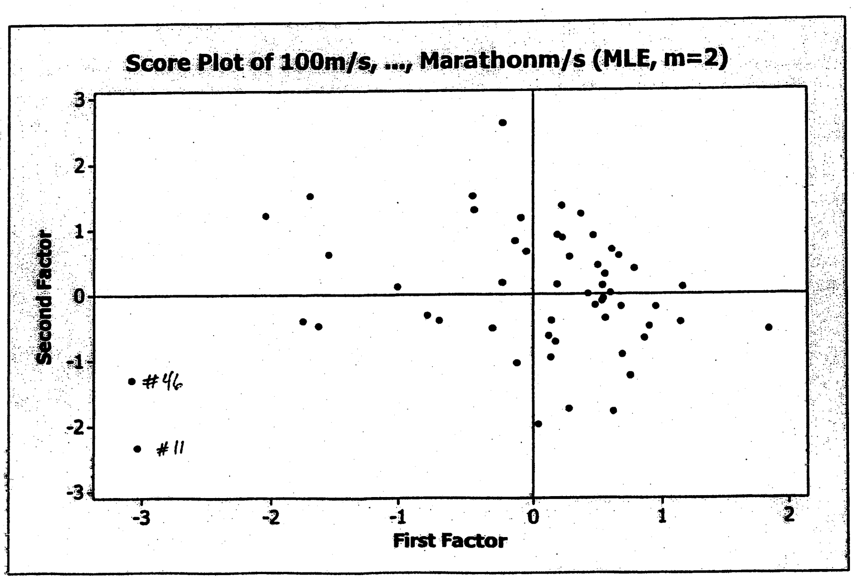
Variable	Factor1	Factor2	Communality
100m/s	0.773	0.511	0.859
200m/s	0.806	0.554	0.957
400m/s	0.797	0.351	0.758
800m/s	0.870	0.140	0.777
1500m/s	0.928	0.067	0.865
5000m/s	0.989	-0.088	0.985
10,000m/s	0.986	-0.117	0.986
Marathonm/s	0.942	-0.108	0.899
Variance	6.3380	0.7485	7.0865
% Var	0.792	0.094	0.886

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
100m/s	0.394	0.839	0.859
200m/s	0.400	0.893	0.957
400m/s	0.497	0.715	0.758
800m/s	0.670	0.573	0.777
1500m/s	0.757	0.540	0.865
5000m/s	0.890	0.440	0.985
10,000m/s	0.903	0.413	0.986
Marathonm/s	0.860	0.398	0.899
Variance	3.9325	3.1540	7.0865
% Var	0.492	0.394	0.886

Factor Score Coefficients

Variable	Factor1	Factor2
100m/s	-0.128	0.268
200m/s	-0.457	0.951
400m/s	-0.046	0.111
800m/s	-0.008	0.055
1500m/s	0.012	0.055
5000m/s	0.570	-0.219
10,000m/s	0.711	-0.388
Marathonm/s	0.089	-0.047



The results from the two solution methods are very similar and very similar to the principal component factor analysis of the covariance matrix S. Using the unrotated loadings, the first factor might be identified as a "running excellence" factor. All the running events load highly on this factor. The second factor appears to contrast the shorter running events with the longer events although there is some difference in the groupings depending on the solution method. The 800m and 1500m runs are in the longer group for the principal component method and in the shorter group for the maximum likelihood method. Nevertheless, this bipolar factor might be called a "running speed-running endurance" factor. After rotation the overall excellence factor disappears and the first factor appears to represent "running endurance" since the running events 800m through the marathon load highly on this factor. The second factor might be classified as a "running speed" factor. Note, for both factors, the remaining running events in each case have moderately large loadings on the factor. The two factor solution accounts for 89%-91% (depending on solution method) of the total variance. The plots of the factor scores indicate that observations #46 (Samoa) and #11 (Cook Islands) are outliers.

The results of the $m = 2$ factor analysis of men's track records when time is measured in meters per second are very much the same as the results for the $m = 2$ factor analysis of **R** presented in Exercise 9.30. If the correlation matrix **R** is factor analyzed, it makes little difference whether running event time is measured in seconds (or minutes) as in Exercise 9.30 or in meters per second. It does make a difference if the covariance matrix **S** is factor analyzed, since the measurement scales in Exercise 9.30 are quite different from the meters/second scale.

9.32. Factor analysis of data on bulls

Factor analysis using sample covariance matrix S

Initial Factor Method: Principal Components

	1	2	3	4	5	6	7
Eigenvalue	20579.6126	4874.6748	5.4292	3.3163	0.4688	0.0741	0.0045
Difference	15704.9378	4869.2456	2.1129	2.8475	0.3948	0.0695	
Proportion	0.8082	0.1914	0.0002	0.0001	0.0000	0.0000	0.0000
Cumulative	0.8082	0.9996	0.9998	1.0000	1.0000	1.0000	1.0000

Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.48777	0.39033	0.38532	YrHgt
X4	0.75367	0.65725	-0.00086	FtFrBody
X5	0.37408	0.62342	0.64446	PrctFFB
X6	0.48170	0.36809	0.33505	Frame
X7	0.11083	-0.38394	-0.49074	BkFat
X8	0.66769	0.29875	0.33038	SaleHt
X9	0.96506	-0.26204	0.00009	SaleWt

SAS scales the loadings obtained from a covariance matrix and then rotates the scaled loadings. The scaling is $\hat{l}_{ij} / \sqrt{s_{ii}}$.

Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.50195	0.42460	0.32637	YrHgt
X4	0.25853	0.90600	0.33514	FtFrBody
X5	0.83816	0.45576	0.18354	PrctFFB
X6	0.44716	0.42166	0.31943	Frame
X7	-0.60974	-0.06913	0.15478	BkFat
X8	0.40890	0.46689	0.50894	SaleHt
X9	-0.13508	0.30219	0.94363	SaleWt

Initial Factor Method: Maximum Likelihood

Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.00000	1.00000	0.00000	YrHgt
X4	0.42819	0.62380	0.39838	FtFrBody
X5	0.85244	0.52282	0.00000	PrctFFB
X6	-0.01180	0.94025	0.03120	Frame
X7	-0.36162	-0.34428	0.39308	BkFat
X8	0.08393	0.85951	0.28992	SaleHt
X9	0.00598	0.36843	0.83599	SaleWt

Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.94438	0.28442	0.16509	YrHgt
X4	0.41219	0.50159	0.55648	FtFrBody
X5	0.23003	0.94883	0.21635	PrctFFB
X6	0.88812	0.25026	0.18382	Frame
X7	-0.25711	-0.51405	0.27102	BkFat
X8	0.75340	0.26667	0.43720	SaleHt
X9	0.25282	-0.05273	0.87634	SaleWt

Factor analysis using sample correlation matrix R

Initial Factor Method: Principal Components

	1	2	3	4	5	6	7
Eigenvalue	4.1207	1.3371	0.7414	0.4214	0.1858	0.1465	0.0471
Difference	2.7836	0.5957	0.3200	0.2356	0.0393	0.0994	
Proportion	0.5887	0.1910	0.1059	0.0602	0.0265	0.0209	0.0067
Cumulative	0.5887	0.7797	0.8856	0.9458	0.9723	0.9933	1.0000

Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.91334	-0.04948	-0.35794	YrHgt
X4	0.83700	0.15014	0.38772	FtFrBody
X5	0.72177	-0.36484	0.48930	PrctFFB
X6	0.88091	0.00894	-0.38949	Frame
X7	-0.37900	0.82646	-0.03335	BkFat
X8	0.91927	0.11715	-0.15210	SaleHt
X9	0.54798	0.69440	0.21811	SaleWt

Varimax Rotated Factor Pattern

	FACTOR1	FACTOR2	FACTOR3	
X3	0.94188	0.27085	-0.06532	YrHgt
X4	0.44792	0.78354	0.24262	FtFrBody
X5	0.26505	0.87071	-0.25513	PrctFFB
X6	0.93812	0.21799	-0.01382	Frame
X7	-0.23541	-0.37460	0.79502	BkFat
X8	0.83365	0.41206	0.13094	SaleHt
X9	0.34932	0.39692	0.74194	SaleWt

Initial Factor Method: Maximum Likelihood

Factor Pattern

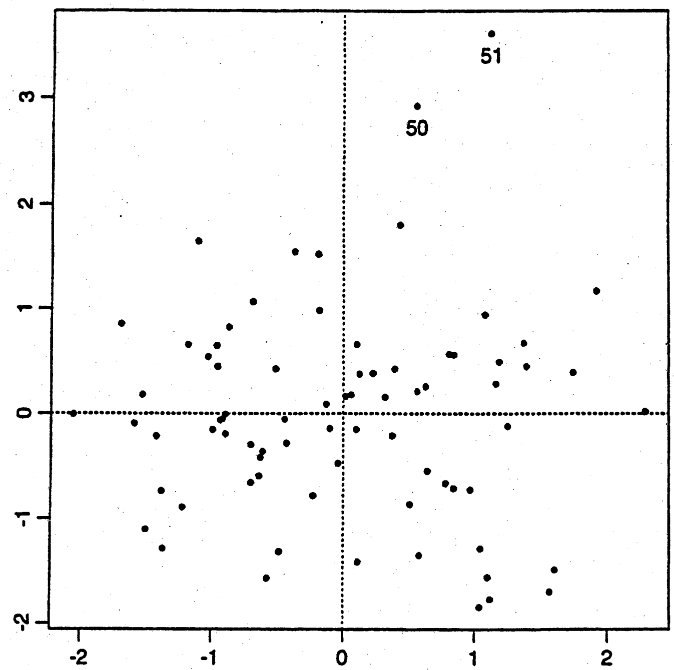
	FACTOR1	FACTOR2	FACTOR3	
X3	0.00000	1.00000	0.00000	YrHgt
X4	0.42819	0.62380	0.39838	FtFrBody
X5	0.85244	0.52282	0.00000	PrctFFB
X6	-0.01180	0.94025	0.03120	Frame
X7	-0.36162	-0.34428	0.39308	BkFat
X8	0.08393	0.85951	0.28992	SaleHt
X9	0.00598	0.36843	0.83599	SaleWt

Varimax Rotated Factor Pattern

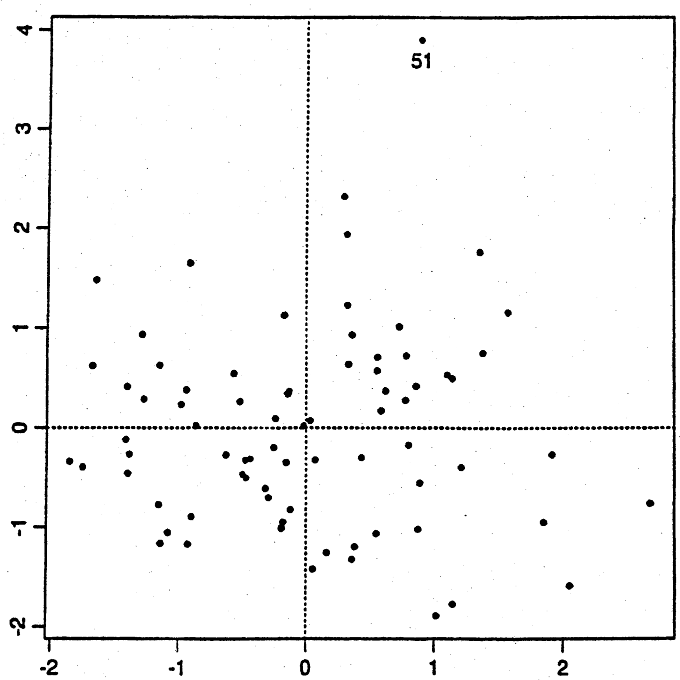
	FACTOR1	FACTOR2	FACTOR3	
X3	0.94438	0.28442	0.16509	YrHgt
X4	0.41219	0.50159	0.55648	FtFrBody
X5	0.23003	0.94883	0.21635	PrctFFB
X6	0.88812	0.25026	0.18382	Frame
X7	-0.25711	-0.51405	0.27102	BkFat
X8	0.75340	0.26667	0.43720	SaleHt
X9	0.25282	-0.05273	0.87634	SaleWt

The interpretation of factors from R is different of the interpretation of factors from S.

Factor scores for the first two factors using S
and varimax rotated PC estimates of factor loadings



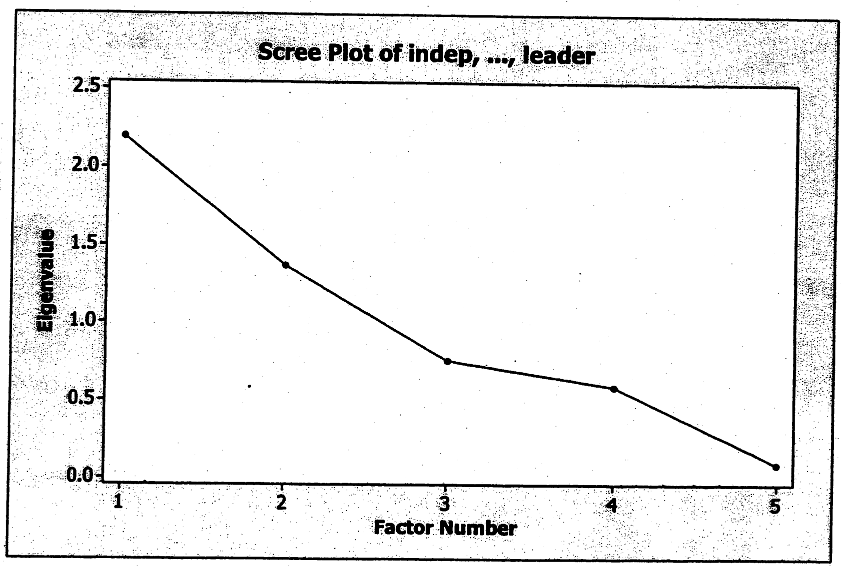
Factor scores for the first two factors using R
and varimax rotated PC estimates of factor loadings



9.33 The correlation matrix **R** and the scree plot follow. The correlations are relatively modest. These correlations and the scree plot suggest $m = 2$ factors is probably too few. An initial factor analysis with $m = 2$ confirms this conjecture. Consequently, we give a $m = 3$ factor solution.

Correlations: indep, supp, benev, conform, leader

	indep	supp	benev	conform
supp	-0.173			
benev	-0.561	0.018		
conform	-0.471	-0.327	0.298	
leader	0.187	-0.401	-0.492	-0.333



Principal Component Factor Analysis of **R** ($m = 3$)

Unrotated Factor Loadings and Communalities

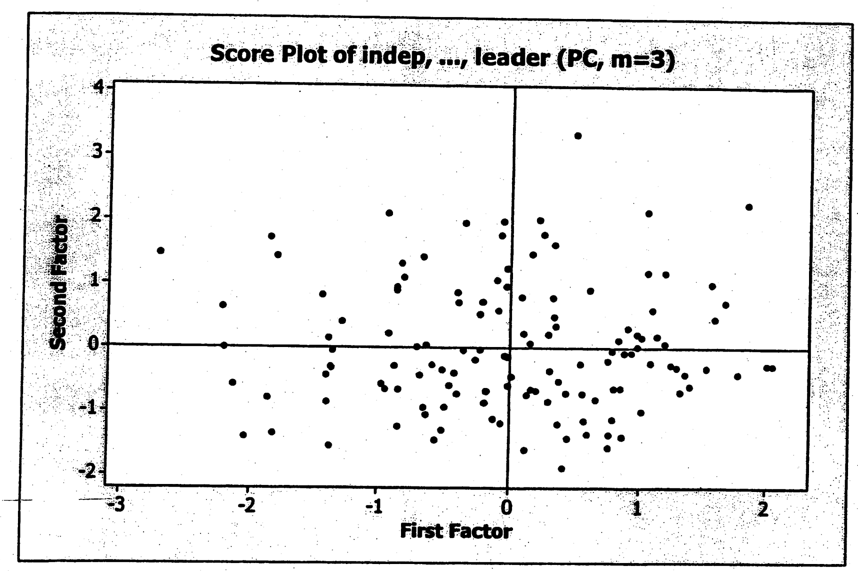
Variable	Factor1	Factor2	Factor3	Communality
indep	-0.772	0.101	-0.580	0.943
supp	0.180	0.922	0.163	0.909
benev	0.813	-0.009	0.100	0.670
conform	0.651	-0.574	-0.256	0.819
leader	-0.696	-0.422	0.563	0.979
Variance	2.1966	1.3682	0.7559	4.3207
% Var	0.439	0.274	0.151	0.864

```
Rotated Factor Loadings and Communalities
Varimax Rotation
```

Variable	Factor1	Factor2	Factor3	Communality
indep	-0.971	0.018	-0.003	0.943
supp	0.136	-0.312	0.890	0.909
benev	0.700	-0.418	-0.081	0.670
conform	0.419	-0.379	-0.707	0.819
leader	-0.155	0.971	-0.111	0.979
Variance	1.6506	1.3587	1.3114	4.3207
% Var	0.330	0.272	0.262	0.864

```
Factor Score Coefficients
```

Variable	Factor1	Factor2	Factor3
indep	-0.752	-0.362	-0.147
supp	0.119	-0.129	0.690
benev	0.372	-0.127	-0.010
conform	0.073	-0.277	-0.545
leader	0.240	0.832	0.008

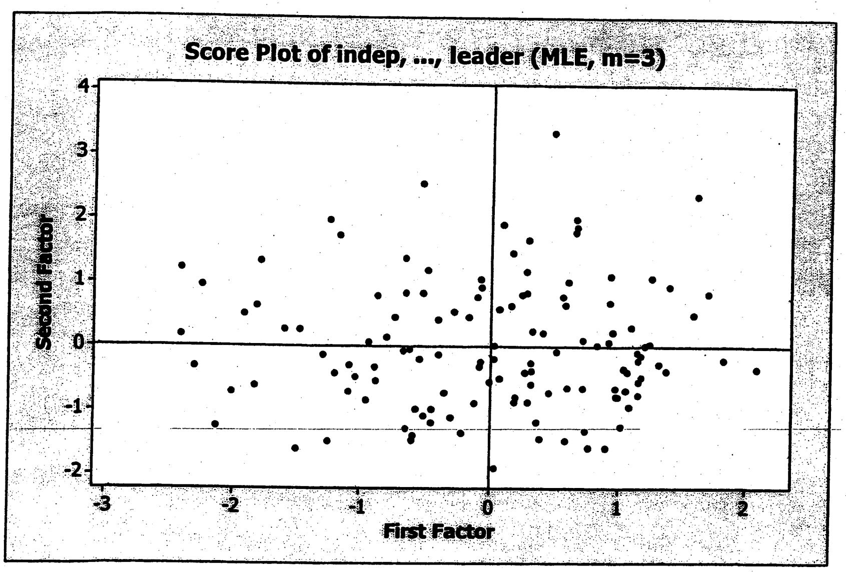


Maximum Likelihood Factor Analysis of **R** ($m = 3$)

```
* NOTE * Heywood case

Unrotated Factor Loadings and Communalities
```

Variable	Factor1	Factor2	Factor3	Communality
indep	-0.788	0.187	0.587	1.000
supp	-0.464	-0.401	-0.790	1.000
benev	0.532	-0.492	-0.086	0.532
conform	0.664	-0.333	0.194	0.589
leader	0.000	1.000	0.000	1.000
Variance	1.5591	1.5486	1.0133	4.1211
% Var	0.312	0.310	0.203	0.824

```
Rotated Factor Loadings and Communalities
Varimax Rotation

Variable   Factor1  Factor2  Factor3  Communality
indep       -0.992    0.034    0.122        1.000
supp         0.048   -0.192   -0.980        1.000
benev        0.562   -0.454    0.098        0.532
conform      0.515   -0.371    0.432        0.589
leader      -0.129    0.968    0.213        1.000

Variance    1.5842   1.3199   1.2170       4.1211
% Var        0.317    0.264    0.243        0.824


Factor Score Coefficients

Variable   Factor1  Factor2  Factor3
indep       -1.016   -0.130   -0.024
supp        -0.123    0.219   -1.069
benev       -0.000    0.000    0.000
conform     -0.000    0.000   -0.000
leader       0.011    1.081   -0.211
```



Using the unrotated loadings and including moderate loadings of magnitudes .4—.5, the factors are all bipolar and appear to be difficult to interpret. Moreover, the arrangement of relatively large loadings on each factor is quite different for the two solution methods. The rotated loadings are consistent with one another for the two solution methods and, although all the factors are bipolar, may be easier to interpret. The first factor is a contrast between Independence and the pair Benevolence and Conformity. Perhaps this factor could be classified as a "conforming—not conforming" factor. The second factor is essentially a "leadership" factor although if moderate loadings are included, this factor is a contrast between Leadership and Benevolence. Teenagers with above average scores on Leadership tend to be above average on this factor, while those with above average scores on Benevolence tend to be below average on this factor. Perhaps we could label this factor a "lead—follow" factor. The third factor is essentially a "support" factor although, again, if moderate loadings are used, this factor is a contrast between Support and Conformity. To our minds however, the latter is difficult to interpret. The factor scores for the first two factors are similar for the two solutions methods. No outliers are immediately evident.

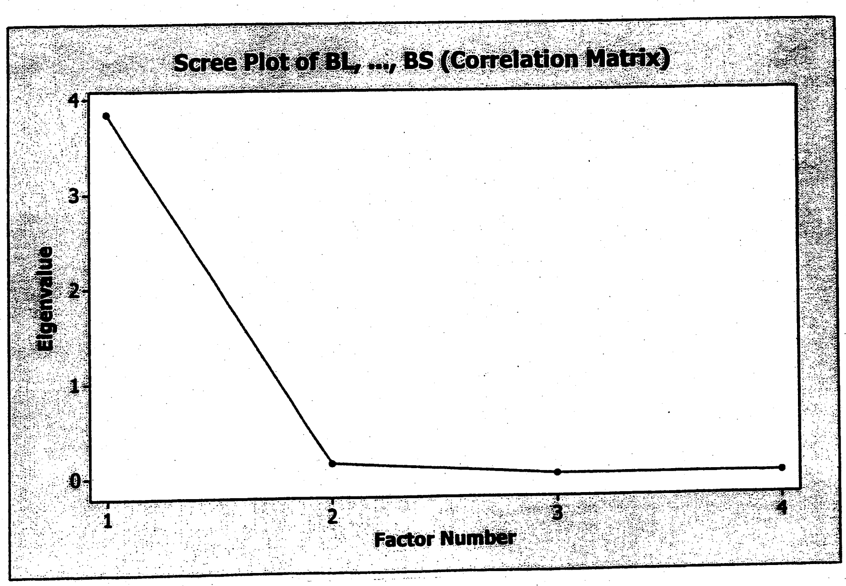
9.34 A factor analysis of the paper property variables with either **S** or **R** suggests a $m = 1$ factor solution is reasonable. All variables load highly on a single factor. The covariance matrix **S** and correlation matrix **R** follow along with a scree plot using **R**. For completeness, the results for a $m = 2$ factor solution using both solution methods is also given. Plots of factor scores from the two factor model suggest that observations 58, 59, 60 and 61 may be outliers.

Covariances: BL, EM, SF, BS

	BL	EM	SF	BS
BL	8.302871			
EM	1.886636	0.513359		
SF	4.147318	0.987585	2.140046	
BS	1.972056	0.434307	0.987966	0.480272

Correlations: BL, EM, SF, BS

	BL	EM	SF
EM	0.914		
SF	0.984	0.942	
BS	0.988	0.875	0.975



Principal Component Factor Analysis of **S** ($m = 1$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Communality
BL	2.878	8.285
EM	0.664	0.441
SF	1.449	2.101
BS	0.684	0.468
Variance	11.295	11.295
% Var	0.988	0.988

Factor Score Coefficients

Variable	Factor1
BL	0.734
EM	0.042
SF	0.188
BS	0.042

The first factor explains 99% of the total variance. All variables, given their measurement scales, load highly on this factor. Note: There is no factor rotation with one factor.

Principal Component Factor Analysis of **R** ($m = 1$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Communality
BL	0.992	0.984
EM	0.951	0.905
SF	0.996	0.991
BS	0.980	0.960
Variance	3.8395	3.8395
% Var	0.960	0.960

Factor Score Coefficients

Variable	Factor1
BL	0.258
EM	0.248
SF	0.259
BS	0.255

The first factor explains 96% of the variance. All variables load highly and about equally on this factor. This factor might be called a "paper properties index."

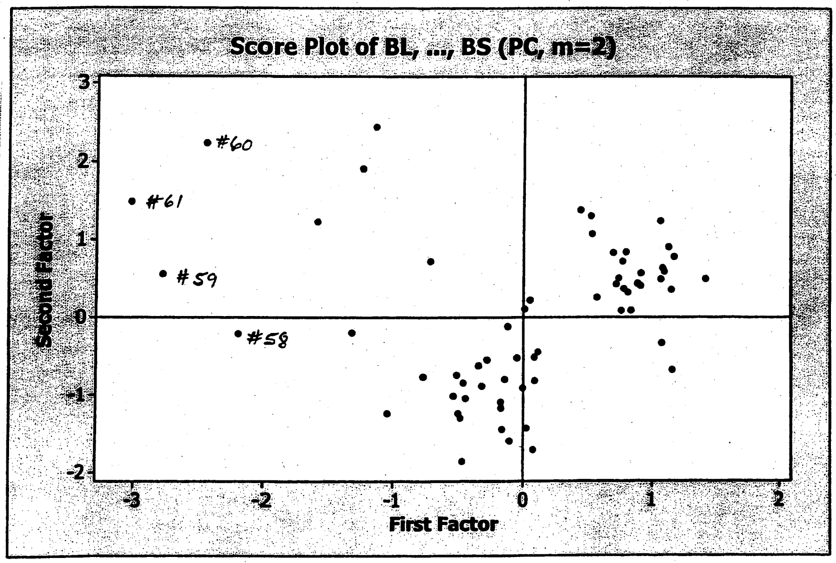
Maximum Likelihood Factor Analysis of **R** ($m = 1$)

```
* NOTE * Heywood case

Unrotated Factor Loadings and Communalities

Variable  Factor1  Communality
BL          1.000        1.000
EM          0.914        0.835
SF          0.984        0.968
BS          0.988        0.975

Variance   3.7784       3.7784
% Var       0.945        0.945


Factor Score Coefficients

Variable  Factor1
BL          1.000
EM          0.000
SF          0.000
BS          0.000
```

The results are similar to the results for the principal component method. The first factor explains about 95% of the variance and all variables load highly and about equally on this factor. Again, the factor might be called a "paper properties index."

Principal Component Factor Analysis of **R** ($m = 2$)

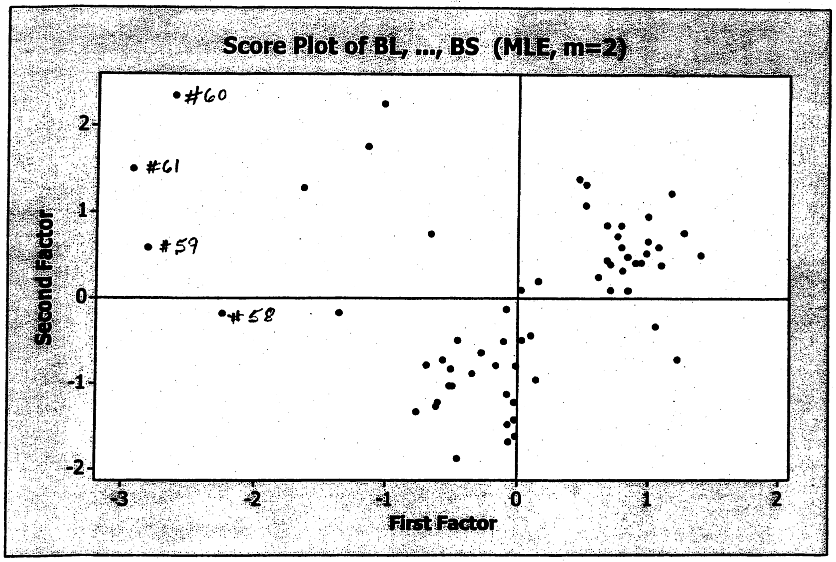
```
Unrotated Factor Loadings and Communalities

Variable  Factor1  Factor2  Communality
BL          0.992   -0.098        0.993
EM          0.951    0.307        0.999
SF          0.996   -0.008        0.991
BS          0.980   -0.191        0.996

Variance   3.8395   0.1403       3.9798
% Var       0.960    0.035        0.995


Rotated Factor Loadings and Communalities
Varimax Rotation

Variable  Factor1  Factor2  Communality
BL          0.817    0.571        0.993
EM          0.522    0.852        0.999
SF          0.761    0.642        0.991
BS          0.868    0.493        0.996

Variance   2.2717   1.7082       3.9798
% Var       0.568    0.427        0.995


Factor Score Coefficients

Variable  Factor1  Factor2
BL          0.650   -0.361
EM         -1.235    1.821
SF          0.232    0.128
BS          1.081   -0.868
```



Using the unrotated loadings, the second factor explains very little of the variance beyond that of the first factor and is not needed. Since the unrotated loadings provide a clear interpretation of the first factor there is no need to consider the rotated loadings. The potential outliers are evident in the plot of factor scores.

Maximum Likelihood Factor Analysis of **R** ($m = 2$)

```
* NOTE * Heywood case

Unrotated Factor Loadings and Communalities

Variable  Factor1  Factor2  Communality
BL          0.988    0.103        0.986
EM          0.875    0.485        1.000
SF          0.975    0.185        0.984
BS          1.000    0.000        1.000

Variance   3.6900   0.2800       3.9700
% Var       0.922    0.070        0.992


Rotated Factor Loadings and Communalities
Varimax Rotation

Variable  Factor1  Factor2  Communality
BL          0.809    0.576        0.986
EM          0.523    0.853        1.000
SF          0.757    0.641        0.984
BS          0.870    0.492        1.000

Variance   2.2572   1.7128       3.9700
% Var       0.564    0.428        0.992


Factor Score Coefficients

Variable  Factor1  Factor2
BL         -0.000   -0.000
EM         -1.016    1.795
SF         -0.000   -0.000
BS          1.759   -1.078
```



The results are similar to the results for the principal component method. Using the unrotated loadings, the first factor explains 92% of the total variance and the second factor explains very little of the remaining variance. Since the unrotated loadings provide a clear interpretation of the first factor (paper properties index) there is no need to consider the rotated loadings. The same potential outliers are evident in the plot of factor scores.

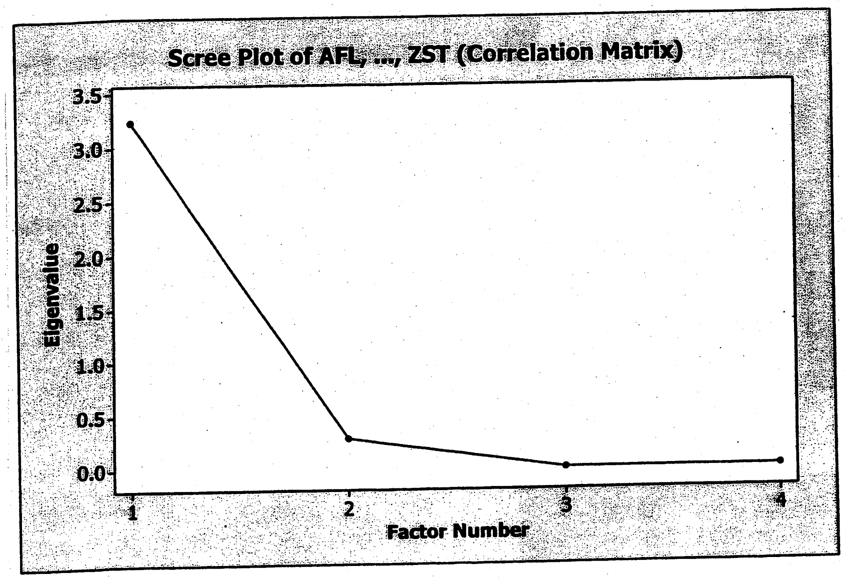
9.35 A factor analysis of the pulp fiber characteristic variables with **S** and **R** for $m = 1$ and $m = 2$ factors is summarized below. The covariance matrix **S** and correlation matrix **R** follow along with a scree plot using **R**. Plots of factor scores from the two factor model suggest that observations 60 and 61 and possibly observations 57, 58 and 59 may be outliers. A $m = 1$ factor solution using **R** appears to be the best choice.

Covariances: AFL, LFF, FFF, ZST

	AFL	LFF	FFF	ZST
AFL	0.06227			
LFF	3.35980	221.05161		
FFF	-3.21404	-185.63707	308.39989	
ZST	0.00577	0.34760	-0.40633	0.00087

Correlations: AFL, LFF, FFF, ZST

	AFL	LFF	FFF
LFF	0.906		
FFF	-0.733	-0.711	
ZST	0.784	0.793	-0.785



Principal Component Factor Analysis of **S** ($m = 1$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Communality
AFL	0.216	0.047
LFF	13.250	175.573
FFF	-16.729	279.858
ZST	0.025	0.001
Variance	455.48	455.48
% Var	0.860	0.860

Factor Score Coefficients

Variable	Factor1
AFL	0.000
LFF	0.433
FFF	-0.645
ZST	0.000

The first factor explains 86% of the total variance and represents a contrast between FFF (with a negative loading) and the AFL, LFF and ZST group, all with positive loadings. AFL (average fiber length), LFF (long fiber fraction) and ZST (zero span tensile strength) may all have to do with paper strength while FFF (fine fiber fraction) may have something to do with paper quality. Perhaps we could label this factor a "strength—quality" factor.

Principal Component Factor Analysis of **R** ($m = 1$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Communality
AFL	0.936	0.877
LFF	0.933	0.870
FFF	-0.878	0.770
ZST	0.917	0.841
Variance	3.3577	3.3577
% Var	0.839	0.839

Factor Score Coefficients

Variable	Factor1
AFL	0.279
LFF	0.278
FFF	-0.261
ZST	0.273

The first factor explains 84% of the variance and the pattern of loadings is consistent with that of the $m = 1$ factor analysis of the covariance matrix **S**. Again, we might label this bi polar factor a "strength—quality" factor.

Maximum Likelihood Factor Analysis of **R** ($m = 1$)

Unrotated Factor Loadings and Communalities

Variable	Factor1	Communality
AFL	0.949	0.900
LFF	0.945	0.894
FFF	-0.784	0.614
ZST	0.846	0.717
Variance	3.1245	3.1245
% Var	0.781	0.781

Factor Score Coefficients

Variable	Factor1
AFL	0.422
LFF	0.394
FFF	-0.090
ZST	0.132

The first factor explains 78% of the variance and the pattern of loadings is consistent with that of the $m = 1$ factor analysis of the covariance matrix **R** using the principal component method. Again, we might label this bi polar factor a "strength—quality" factor.

Because the different measurement scales make the factor loadings obtained from the covariance matrix difficult to interpret, we continue with a factor analysis of the correlation matrix **R** with $m = 2$.

Principal Component Factor Analysis of R ($m = 2$)

Unrotated Factor Loadings and Communalities

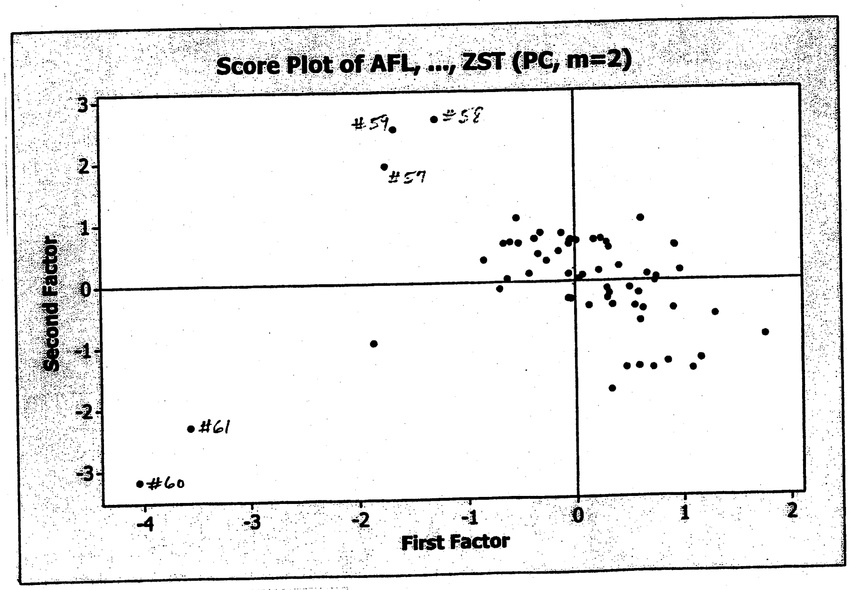
Variable	Factor1	Factor2	Communality
AFL	0.936	0.256	0.942
LFF	0.933	0.288	0.953
FFF	-0.878	0.423	0.949
ZST	0.917	-0.150	0.863
Variance	3.3577	0.3493	3.7070
% Var	0.839	0.087	0.927

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
AFL	0.868	-0.434	0.942
LFF	0.887	-0.408	0.953
FFF	-0.372	0.900	0.949
ZST	0.583	-0.723	0.863
Variance	2.0176	1.6893	3.7070
% Var	0.504	0.422	0.927

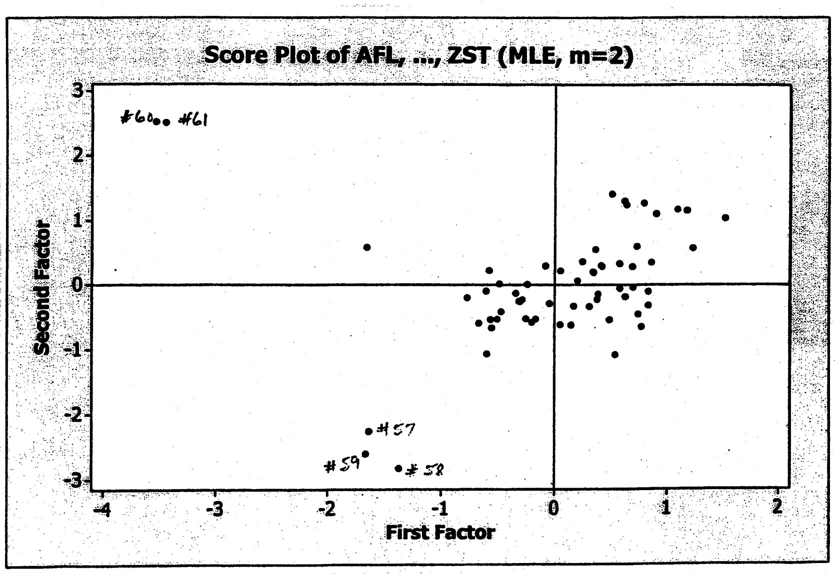
Factor Score Coefficients

Variable	Factor1	Factor2
AFL	0.696	0.359
LFF	0.757	0.429
FFF	0.613	1.075
ZST	-0.082	-0.501



Maximum Likelihood Factor Analysis of **R** $(m = 2)$

```
Unrotated Factor Loadings and Communalities

Variable   Factor1   Factor2   Communality
AFL          0.913    -0.205         0.876
LFF          0.926    -0.292         0.943
FFF         -0.890    -0.388         0.944
ZST          0.866     0.033         0.752

Variance    3.2351    0.2796        3.5146
% Var        0.809     0.070         0.879


Rotated Factor Loadings and Communalities
Varimax Rotation

Variable   Factor1   Factor2   Communality
AFL          0.819     0.454         0.876
LFF          0.886     0.397         0.943
FFF         -0.407    -0.882         0.944
ZST          0.625     0.601         0.752

Variance    2.0124    1.5023        3.5146
% Var        0.503     0.376         0.879


Factor Score Coefficients

Variable   Factor1   Factor2
AFL          0.336    -0.101
LFF          0.922    -0.423
FFF          0.534    -1.197
ZST          0.049     0.076
```

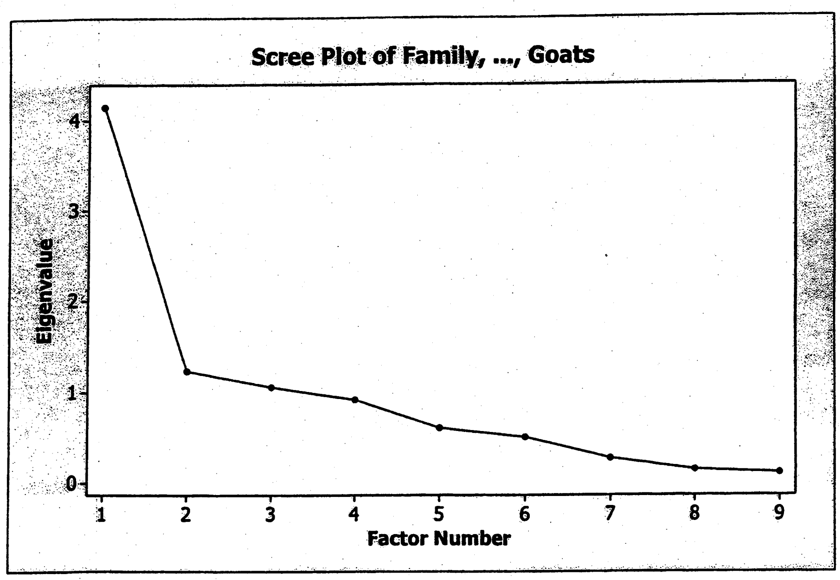


Examining the unrotated loadings for both solution methods, we see that the second factor explains little (about 7%-8%) of the remaining variance. Also, this factor has moderate to very small loadings on all the variables with the possible exception of variable FFF. If retained, this factor might be called a "fine fiber" of "quality" factor. Using the rotated loadings, the second factor looks much like the first factor for both solution methods. That is, this factor appears to be a contrast between variable FFF and the group of variables AFL, LFF and ZST. To summarize, there seems to be no gain in understanding from adding a second factor to the model. A one factor model appears be sufficient in this case. However, plots of the factor scores for m = 2 suggest observations 60, 61 and, perhaps, observations 57, 58 and 59 may be outliers.

9.36 The correlation matrix **R** and the scree plot is shown below. After $m = 2$ there is no sharp elbow in the scree plot and the plot falls off almost linearly. Potential choices for m are 2, 3, 4 and 5. We give the results for $m = 4$ but, to our minds, here is a case where a factor model is not particularly well defined.

Correlations: Family, DistRd, Cotton, Maze, Sorg, Millet, Bull, Cattle, Goats

	Family	DistRd	Cotton	Maze	Sorg	Millet	Bull	Cattle
DistRd	-0.084							
Cotton	0.724	0.028						
Maze	0.679	-0.054	0.730					
Sorg	0.568	-0.071	0.383	0.109				
Millet	0.506	0.022	0.389	0.217	0.382			
Bull	0.727	-0.088	0.765	0.623	0.443	0.353		
Cattle	0.336	-0.063	0.175	0.197	0.404	0.081	0.520	
Goats	0.484	0.031	0.399	0.136	0.424	0.305	0.560	0.357



Principal Component Factor Analysis of **R** ($m = 4$)

Unrotated Factor Loadings and Communalities

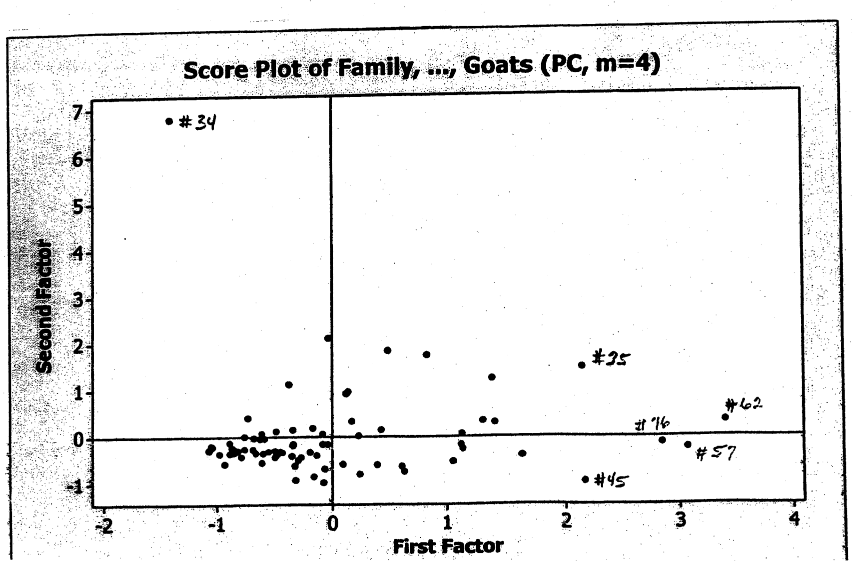
Variable	Factor1	Factor2	Factor3	Factor4	Communality
Family	0.903	-0.111	-0.002	-0.118	0.842
DistRd	-0.068	-0.080	-0.855	0.482	0.974
Cotton	0.837	-0.380	-0.070	0.028	0.851
Maze	0.687	-0.616	0.175	0.158	0.907
Sorg	0.633	0.503	-0.070	-0.219	0.706
Millet	0.547	0.048	-0.396	-0.582	0.798
Bull	0.896	-0.033	0.125	0.189	0.856
Cattle	0.502	0.510	0.286	0.466	0.811
Goats	0.629	0.421	-0.178	0.096	0.614
Variance	4.1443	1.2364	1.0581	0.9205	7.3593
% Var	0.460	0.137	0.118	0.102	0.818

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Family	0.714	0.320	-0.473	0.080	0.842
DistRd	-0.026	-0.022	0.006	-0.986	0.974
Cotton	0.856	0.150	-0.301	-0.076	0.851
Maze	0.951	0.008	0.032	0.047	0.907
Sorg	0.092	0.564	-0.606	0.112	0.706
Millet	0.226	-0.026	-0.863	-0.029	0.798
Bull	0.724	0.535	-0.210	0.043	0.856
Cattle	0.148	0.879	0.108	0.074	0.811
Goats	0.180	0.629	-0.406	-0.145	0.614
Variance	2.7840	1.8985	1.6476	1.0291	7.3593
% Var	0.309	0.211	0.183	0.114	0.818

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3	Factor4
Family	0.197	-0.013	-0.171	0.063
DistRd	0.014	0.042	0.030	-0.963
Cotton	0.344	-0.115	-0.024	-0.090
Maze	0.494	-0.165	0.247	0.023
Sorg	-0.199	0.246	-0.374	0.100
Millet	-0.078	-0.260	-0.697	-0.001
Bull	0.224	0.204	0.110	0.005
Cattle	-0.063	0.633	0.329	0.019
Goats	-0.114	0.338	-0.156	-0.164



Maximum Likelihood Factor Analysis of R $(m = 4)$

Unrotated Factor Loadings and Communalities

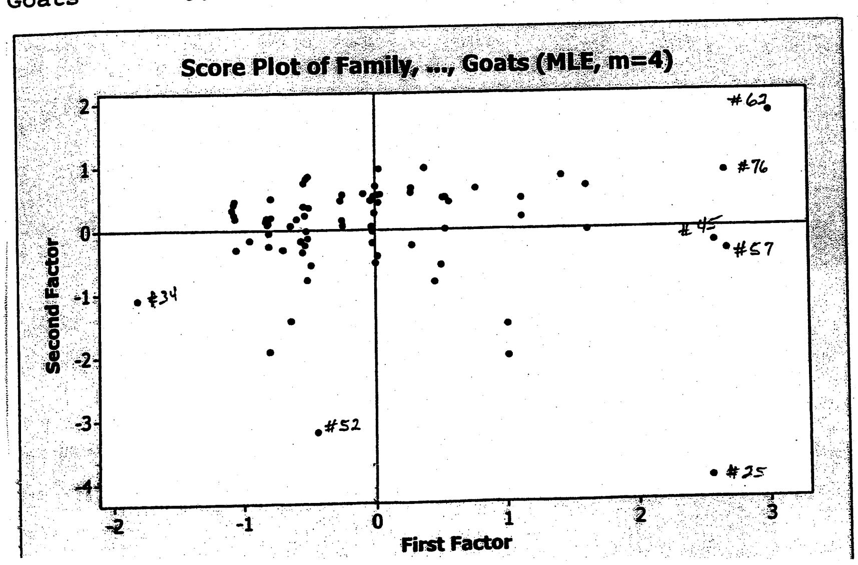
Variable	Factor1	Factor2	Factor3	Factor4	Communality
Family	0.752	-0.324	-0.162	-0.374	0.837
DistRd	-0.064	0.056	-0.044	-0.003	0.009
Cotton	0.794	-0.238	-0.307	-0.044	0.782
Maze	0.980	0.170	0.025	-0.002	0.990
Sorg	0.211	-0.567	-0.071	-0.527	0.649
Millet	0.276	-0.269	-0.301	-0.361	0.369
Bull	0.746	-0.616	-0.096	0.131	0.962
Cattle	0.290	-0.608	0.640	-0.074	0.869
Goats	0.249	-0.607	-0.151	-0.109	0.465
Variance	2.9824	1.7047	0.6610	0.5841	5.9322
% Var	0.331	0.189	0.073	0.065	0.659

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Family	0.630	-0.605	0.229	-0.148	0.837
DistRd	-0.040	0.017	-0.081	0.025	0.009
Cotton	0.713	-0.362	0.075	-0.370	0.782
Maze	0.980	-0.034	0.166	-0.016	0.990
Sorg	0.034	-0.740	0.303	-0.089	0.649
Millet	0.206	-0.558	-0.028	-0.120	0.369
Bull	0.540	-0.324	0.437	-0.612	0.962
Cattle	0.039	-0.154	0.915	-0.079	0.869
Goats	0.072	-0.466	0.268	-0.414	0.465
Variance	2.2098	1.7035	1.2850	0.7340	5.9322
% Var	0.246	0.189	0.143	0.082	0.659

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3	Factor4
Family	0.013	-0.606	-0.078	0.247
DistRd	0.001	-0.002	-0.009	-0.002
Cotton	0.033	-0.161	-0.162	-0.113
Maze	0.995	0.440	0.109	0.681
Sorg	-0.023	-0.404	0.017	0.206
Millet	0.003	-0.185	-0.062	0.052
Bull	-0.026	0.215	0.103	-1.426
Cattle	-0.141	0.091	0.896	0.385
Goats	-0.009	-0.093	-0.010	-0.023



The two solution methods for $m = 4$ factors produce somewhat different results. The patterns for unrotated loadings on the first two factors are similar but not identical. The patterns of loadings for the two solution methods on the third and fourth factors are quite different. Notice that DistRd does not load on any factor in the maximum likelihood solution. The factor loading patterns are more alike for the two solution methods using the rotated loadings, although factors 2 & 3 in the principal component solution appear to be reversed in the maximum likelihood solution. The rotated loadings on factor 4 for the two methods are quite different. Again, DistRd does not load on any factor in the maximum likelihood solution, it appears to define factor 4 in the principal component solution. (From **R** we see that DistRd is not correlated with any of the other variables.) Variables Family, Cotton, Maze, and Bullocks load highly on the first factor. The variables Family, Sorghum, Millet and Goats load highly on the second factor (maximum likelihood solution) or the third factor (principal component solution). Growing cotton and maze is labor intensive and bullocks are helpful. The first factor might be called a "family farm—row crop" factor. Millet and sorghum are grasses and may provide feed for goats. Consequently, the second (or third in the case of the principal component solution) factor might be called a "family farm—grass crop" factor. The third factor in the maximum likelihood solution (second factor in the principal component solution) may have different interpretations depending on the solution method but in both cases, Bullocks and Cattle load highly on this factor. Perhaps this factor could be called a "livestock" factor. The rotated loadings are considerably different on the fourth factor. This factor is clearly a "distance to the road" factor in the principal component solution. The interpretation is not clear in the maximum likelihood solution. The fact that the two solution methods produce somewhat different results and explain quite different proportions of the total variation (82% for principal components, 66% for maximum likelihood) reinforces the notion that a linear factor model is not particularly well defined for this problem. Plots of factor scores for the first two factors indicated there are several potential outliers. If these observations are removed, the results could change.