- The contribution to the *total* variance,

$$tr(\Sigma) = \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} Var(X_i)$$

from the ith common factor is

$$\ell_{1i}^2 + \ell_{2i}^2 + \cdots + \ell_{pi}^2$$

- For the principal component solution, say the 1st common factor

$$\tilde{\ell}_{11}^2 + \tilde{\ell}_{21}^2 + \cdots + \tilde{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right)'\left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right) = \hat{\lambda}_1$$

and

$$\begin{pmatrix}\text{Proportion of total} \\ \text{sample variance} \\ \text{due to } j\text{th factor}\end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\[2ex] \dfrac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}$$

is frequently used as a heuristic device for determining the appropriate number of common factor

- **Example 9.3 (Factor analysis of consumer-preference data)** In a consumer-preference study, a random sample of customers were asked to rate several attributes of a new product. The responses, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next:

$$\begin{array}{lc} \textit{Attribute (Variable)} & \\ \text{Taste} & 1 \\ \text{Good buy for money} & 2 \\ \text{Flavor} & 3 \\ \text{Suitable for snack} & 4 \\ \text{Provides lots of energy} & 5 \end{array}
\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1.00 & .02 & .96 & .42 & .01 \\ .02 & 1.00 & .13 & .71 & .85 \\ .96 & .13 & 1.00 & .50 & .11 \\ .42 & .71 & .50 & 1.00 & .79 \\ .01 & .85 & .11 & .79 & 1.00 \end{bmatrix} \end{array}$$

Variable	Estimated factor loadings $\tilde{\ell}_{ij} = \sqrt{\hat{\lambda}_i}\,\hat{e}_{ij}$ F_1	F_2	Communalities $\tilde{h}_i^2$	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Taste	.56	.82	.98	.02
2. Good buy for money	.78	−.53	.88	.12
3. Flavor	.65	.75	.98	.02
4. Suitable for snack	.94	−.10	.89	.11
5. Provides lots of energy	.80	−.54	.93	.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	.571	.932		

- comparison of PCA and FA

- Modified Principal Component Approach --- Principal Factor Solution
  - Idea: the common factor should account for the *off-diagonal* elements of Σ, as well as the *communality portion* of the diagonal elements

$$h_i^2 = \sigma_{ii} - \psi_i^2$$

  - algorithm (<u>Note</u>: $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$)

    1. Guess $\tilde{\Psi}$
    2. Set $\tilde{L}$ = the largest m eigenvectors of the eigendecomposition of $S - \tilde{\Psi}$
    3. Set $\tilde{\Psi} = \text{diag}(S - \tilde{L}\tilde{L}')$

    Repeat steps 2 and 3 until convergence

    - A popular choice for initial estimates of specific variances is the inverse of the diagonal elements of $\mathbf{S}^{-1}$ (for factoring the correlation matrix $\mathbf{R}$, one minus the initial estimates is equal to the square of the multiple correlation coefficient between X_i and other p–1 variables)

  - some of the eigenvalues of $S - \tilde{\Psi}$ may be negative
  - In the iteration, some communality estimates may exceeds Var(X_i), which results in a negative estimate of the specific variance. This is referred to as *Heywood case*.

- Maximum Likelihood Method
  - Assume the common factors and the specific factors are jointly normally distributed

  - Then, the distribution of

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{L}\ \mathbf{F} + \boldsymbol{\varepsilon}$$

  is multivariate normal and its likelihood is

$$L(\boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\mathbf{\Sigma}|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right)\text{tr}\left[\mathbf{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})' + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]}$$
$$= (2\pi)^{\frac{-(n-1)p}{2}} |\mathbf{\Sigma}|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right)\text{tr}\left[\mathbf{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]}$$
$$\times (2\pi)^{-\frac{p}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right)(\bar{\mathbf{x}}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})}$$

  where $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$

    - To make $\mathbf{L}$ well defined, we can impose the computationally convenient uniqueness condition

$$\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{\Delta} \qquad \text{a diagonal matrix}$$

  (Q: dimension reduction of the parameter space caused by the condition =)

  - Maximizing the likelihood is equivalent to minimizing

$$\ln|\mathbf{\Sigma}| - \ln|\mathbf{S}_n| + \text{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}_n) - p \qquad (\text{❖})$$

  subject to $\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{\Delta}$, a diagonal matrix.

    - <u>Note</u>: the function take the value zero if $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ equals $\mathbf{S}_n$
    - The MLE of $\mathbf{L}$ and $\mathbf{\Psi}$ can be obtained by numerical maximization

- **Result 9.1.** Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ is the covariance matrix for the m common factor model
The maximum likelihood estimators $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ maximize (❖) subject to $\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$ being diagonal.

The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2 \qquad \text{for } i = 1, 2, \ldots, p$$

so

$$\begin{pmatrix}\text{Proportion of total sample} \\ \text{variance due to } j\text{th factor}\end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}}$$

- If the variables are standardized so that $\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, then the covariance matrix $\boldsymbol{\rho}$ of $\mathbf{Z}$ has the representation

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$$

Thus, $\boldsymbol{\rho}$ has a factorization with loading matrix $\mathbf{L}_\mathbf{z} = \mathbf{V}^{-1/2}\mathbf{L}$ and specific variance matrix $\boldsymbol{\Psi}_\mathbf{z} = \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$

- By the invariance property of MLE, the MLE of $\boldsymbol{\rho}$ is

$$\begin{aligned}\hat{\boldsymbol{\rho}} &= (\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2} \\ &= \hat{\mathbf{L}}_\mathbf{z}\hat{\mathbf{L}}_\mathbf{z}' + \hat{\boldsymbol{\Psi}}_\mathbf{z}\end{aligned}$$

where $\hat{\mathbf{V}}^{-1/2}$ and $\hat{\mathbf{L}}$ are the maximum likelihood estimators of $\mathbf{V}^{-1/2}$ and $\mathbf{L}$.

- Note: There is usually no relationship between the principal components of the sample correlation matrix and the sample covariance matrix. For maximum likelihood factor analysis, however, the results of analyzing either matrix are essentially equivalent (this is not true of principal factor analysis)

- The MLE method could produce very different results when $m \to m+1$
- The MLE approach can also experience difficulties with Heywood cases
- comparison of the PC and MLE approaches (Example: stock-price data)

	Maximum likelihood			Principal components		
	Estimated factor loadings		Specific variances	Estimated factor loadings		Specific variances
Variable	F_1	F_2	$\hat{\psi}_i = 1 - \hat{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. J P Morgan	.115	.755	.42	.732	−.437	.27
2. Citibank	.322	.788	.27	.831	−.280	.23
3. Wells Fargo	.182	.652	.54	.726	−.374	.33
4. Royal Dutch Shell	1.000	−.000	.00	.605	.694	.15
5. Texaco	.683	−.032	.53	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.323	.647		.487	.769	

- residual matrix

MLE

$$\begin{bmatrix} 0 & .001 & -.002 & .000 & .052 \\ .001 & 0 & .002 & .000 & -.033 \\ -.002 & .002 & 0 & .000 & .001 \\ .000 & .000 & .000 & 0 & .000 \\ .052 & -.033 & .001 & .000 & 0 \end{bmatrix}$$

PC

$$\begin{bmatrix} 0 & -.099 & -.185 & -.025 & .056 \\ -.099 & 0 & -.134 & .014 & -.054 \\ -.185 & -.134 & 0 & .003 & .006 \\ -.025 & .014 & .003 & 0 & -.156 \\ .056 & -.054 & .006 & -.156 & 0 \end{bmatrix}$$

- A large sample test for the number of common factors (i.e., m)
  - Recall: choosing m might be done by examining the total variance explained
  - Assume the common factors and the specific factors are jointly normally distributed
  - null and alternative hypotheses

$$H_0: \underset{(p\times p)}{\mathbf{\Sigma}} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{\Psi}}$$

H_1: $\mathbf{\Sigma}$ any other positive definite matrix.

- under alternative, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\mathbf{\Sigma}} = ((n-1)/n)\mathbf{S} = \mathbf{S}_n$, the maximized likelihood is proportional to $|\mathbf{S}_n|^{-n/2}e^{-np/2}$
- under null, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$, where $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$ are the MLE of $\mathbf{L}$ and $\mathbf{\Psi}$, the maximized likelihood is proportional to

$$|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|^{-n/2} \exp\left(-\tfrac{1}{2}n\,\mathrm{tr}[(\hat{\mathbf{L}}\hat{\mathbf{L}} + \hat{\mathbf{\Psi}})^{-1}\mathbf{S}_n]\right)$$

- likelihood ratio statistic

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right] = n\ln\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

with degrees of freedom,

$$\begin{aligned} v - v_0 &= \tfrac{1}{2}p(p+1) - [p(m+1) - \tfrac{1}{2}m(m-1)] \\ &= \tfrac{1}{2}[(p-m)^2 - p - m] \end{aligned}$$

- Bartlett (1954) has shown that the chi-square approximation can be improved by replacing n with $(n - 1 - (2p + 4m + 5)/6)$
- Using Bartlett's correction, we reject H_0 at the α significant level if

$$(n - 1 - (2p + 4m + 5)/6)\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{[(p-m)^2-p-m]/2}(\alpha)$$

- Note: because the number of degrees of freedom must be positive, it follows that

$$m < \tfrac{1}{2}(2p + 1 - \sqrt{8p+1})$$

in order to apply the test

❖ **Reading**: Textbook, 9.3, Supplement 9A