Factor Analysis:

Objective: Describe covariance relationships among a large set of measured traits with a few linear combinations of underlying but unobservable traits.

\[ X_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix} \]
denotes the set of \( p \) measurements recorded for the \( i \)-th experimental unit.

\[ \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \]
and \( \text{Var}(X) = \Sigma_{pp} \)

The big idea is to construct a linear model:

\[
\begin{bmatrix}
X_{i1} - \mu_1 \\
X_{i2} - \mu_2 \\
\vdots \\
X_{ip} - \mu_p
\end{bmatrix} =
\begin{bmatrix}
l_{11} & l_{12} & \cdots & l_{1m} \\
l_{21} & l_{22} & \cdots & l_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
l_{p1} & l_{p2} & \cdots & l_{pm}
\end{bmatrix}
\begin{bmatrix}
F_{i1} \\
F_{i2} \\
\vdots \\
F_{im}
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_{i1} \\
\varepsilon_{i2} \\
\vdots \\
\varepsilon_{im}
\end{bmatrix}
\]

\( p \) measurements, matrix of factor loadings, \( m \) factors which describe major features of members of the population.

Random "errors" corresponding to measurement error and variation not accounted for by the common factors (variation in specific factors).
Note that we want
\[ m = \text{number of factors} \]
to be much smaller than
\[ p = \text{number of measured attributes}. \]

Matrix of factor loadings
\[
L = \begin{bmatrix}
    l_{11} & l_{12} & \cdots & l_{1 j} & \cdots & l_{1 m} \\
    l_{21} & l_{22} & \cdots & l_{2 j} & \cdots & l_{2 m} \\
    \vdots & \vdots & & \vdots & & \vdots \\
    l_{p1} & l_{p2} & \cdots & l_{pj} & \cdots & l_{pm}
\end{bmatrix}
\]
\[ p \times m \]

j-th column is the vector of loadings for the j-th factor

\[ l_{ij} \] is called the loading of the i-th variable on the j-th factor

\[
X_i - \mu = L \xi_i + \epsilon_i
\]

Restrictions:
\[
E(\epsilon_i) = 0
\]
\[
\text{Var}(\epsilon_i) = \Psi = \begin{bmatrix}
    \psi_1 & \psi_2 & \cdots & \psi_p
\end{bmatrix}
\]
\[ \epsilon_i \text{'s and } \xi_i \text{'s are independent.} \]

Orthogonal Factor model:
\[
E(\xi_i) = 0
\]
\[
\text{Var}(\xi_i) = I_{m \times m} = \begin{bmatrix}
    1 & \cdots & 0 \\
    \vdots & \ddots & \vdots \\
    0 & \cdots & 1
\end{bmatrix}
\]

The real restriction is that the factors are uncorrelated. Variances can be made equal to 1 by properly scaling the factor loadings.

[an alternative is the Oblique factor model]
This model imposes a covariance structure on \( X \)

1) \[
\Sigma = \text{Var}(X) = \text{Var}(L\xi + \epsilon)
\]
\[
= \text{Var}(L\xi) + \text{Var}(\epsilon)
\]
\[
= LL' + \Psi
\]

The off diagonal elements of \( LL' \) are \( \sigma_{ij} \), the covariances in \( \Sigma \).

\[
\text{Var}(x_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i
\]

\[
\text{Cov}(x_i, x_j) = \sum_{k=1}^{m} l_{ik} l_{jk}
\]

2) \[
\text{Cov}(x, \epsilon) = \text{Cov}(L\xi, \epsilon) = L\psi
\]
so \[
\text{Cov}(x_i, \epsilon_i) = l_{i.}
\]

The proportion of the variance of the \( i \)-th measurement \( X_i \) contributed by the \( m \) factors \( F_1, F_2, \ldots, F_m \) is called the \( i \)-th communality.

The remaining proportion of the variance of the \( i \)-th measurement associated with \( \epsilon_i \), is called the uniqueness or specific variance.

\[
\sigma_{ii} = \frac{\sum_{k=1}^{m} l_{ik}^2}{\text{Var}(x_i)} = \frac{\sum_{k=1}^{m} l_{ik}^2 + \psi_i}{\text{Var}(x_i)}
\]

Call it \( h_i^2 \), communality, specific variance

Then

\[
\sigma_{ii} = h_i^2 + \psi_i
\]

8.5

8.6
limitations of the orthogonal factor model:

1. Linearity:
   For a non-linear model, say
   \[(X_i - \mu_i) = \left[\alpha_1 F_1 F_3 + \alpha_2 \ln(F_3)\right]^2 + \varepsilon_i,\]
   the covariance approximation
   \[LL' + \Psi\]
   may be quite inadequate.

   The amount of observed data is not adequate to check this assumption of linearity.

   Linear combinations of the factors may provide a good approximation for non-linear relationships over a small range of factor values.

2. The \(\frac{P(P+1)}{2}\) elements of \(\Sigma\) are described with \(mp\) factor loadings in \(L\) and \(p\) specific variances \(\{\Psi_i\}\).

   The factor model is most useful when \(m\) is small, but in many cases \(mp+p\) parameters are not adequate and \(\Sigma\) is not close to \(LL' + \Psi\) when \(p = 12\), \(\frac{P(P+1)}{2} = 78\), but for \(m = 2\) factors, \(mp+p = 36\).

   Look at example 9.2 in T+W for a 3x3 covariance matrix which cannot be exactly described by a one factor model.
\[ S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

where \( \bar{x} \) is the sample mean.

Measure attributes of each sampled member, \( x_1, x_2, \ldots, x_n \) where \( \bar{x} \) is the sample mean of members of the population. Obtain a random sample of \( p \) parameters for the unobservable factors. We will also estimate scores.

Consider matrix of factor loadings \( L \) and specific variances \( \Phi \).

Methods of Estimation:

\[ [\Phi] = [L] ^T \]

Let \( \mathbf{L} \) be an orthogonal transformation of the factors, uniquely up to an orthogonal rotation. Then the factor model is determined.
If standardized measurements

\[
\xi_j = \begin{bmatrix}
\frac{(x_{1j} - \bar{x}_1)}{\sqrt{s_{11}}}
\end{bmatrix}
\]

are used, replace \( S \) with \( R \), the matrix of sample correlations.

**Methods of estimation:**

1. **Principal Component method**
2. **Principal Factor method**
3. **Maximum Likelihood method**

---

**Principal Component Method:**

**Spectral decomposition of \( \Sigma_{p \times p} \)**

\[
\Sigma = \lambda_1 \xi_1 \xi_1' + \lambda_2 \xi_2 \xi_2' + \cdots + \lambda_p \xi_p \xi_p'
\]

where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 \)

are the eigenvalues for \( \Sigma \).

Suppose \( \lambda_1, \lambda_2, \ldots, \lambda_m \) are "large" and \( \lambda_{k+1} \) is "small." Then

\[
\Sigma \approx \sum_{k=1}^{m} \lambda_k \xi_k \xi_k'
\]

\[
= \begin{bmatrix}
\lambda_1 \xi_1 \xi_1' & \lambda_2 \xi_2 \xi_2' & \cdots & \lambda_m \xi_m \xi_m'
\end{bmatrix}
\begin{bmatrix}
\sqrt{\lambda_1} \xi_1
\sqrt{\lambda_2} \xi_2
\vdots
\sqrt{\lambda_m} \xi_m
\end{bmatrix}
\]

\[
= LL'
\]
Then from the factor model
\[ \Sigma = LL' + \Psi \]

we have
\[
\begin{bmatrix}
\psi_1 \\
\vdots \\
\psi_m \\
\end{bmatrix} = \Sigma - \begin{bmatrix}
\lambda_1 e_1 \\
\vdots \\
\lambda_m e_m \\
\end{bmatrix} \begin{bmatrix}
\lambda_1 e_1' \\
\vdots \\
\lambda_m e_m' \\
\end{bmatrix}
\]

so
\[
\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \lambda_j^2 e_{ij} = \sigma_{ii} - \sum_{j=1}^{m} \lambda_j e_{ij}^2
\]
\[
= \sigma_{ii} - h_i^2
\]

Estimate \( L \) and \( \Psi \) by substituting estimated eigenvectors and eigenvalues from \( S \) or \( R \).

\[
\tilde{L} = \begin{bmatrix}
\sqrt{\lambda_1} \tilde{e}_1 \\
\sqrt{\lambda_2} \tilde{e}_2 \\
\vdots \\
\sqrt{\lambda_m} \tilde{e}_m \\
\end{bmatrix}
\]

\[
\tilde{\Psi} = \begin{bmatrix}
\tilde{\psi}_1 \\
\tilde{\psi}_2 \\
\vdots \\
\tilde{\psi}_m \\
\end{bmatrix} = \text{diag}(S) - \text{diag}(\tilde{L} \tilde{L}')
\]

Estimated specific variance:
\[
\tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \lambda_j \tilde{e}_{ij}^2
\]

Estimated communalities:
\[
\tilde{h}_i^2 = \sum_{j=1}^{m} \tilde{\lambda}_j^2 \tilde{e}_{ij}^2 = \sum_{j=1}^{m} \hat{\lambda}_j e_{ij}^2
\]
\[
= s_{ii} - \tilde{\psi}_i
\]
Example 9.4 (Page 489 in J+W)

Stock price data consisting of \( n = 100 \) weekly rates of returns on \( p = 5 \) stocks.

The data were standardized and the factor analysis was performed on the sample correlation matrix \( R \).

\[
\begin{bmatrix}
1 & .58 & .51 & .39 & .46 \\
.60 & 1 & .39 & .32 \\
.44 & .42 & 1 & .52 \\
.52 & 1 & & & \\
\end{bmatrix}
\]

Two factor (m=2) solution:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Specific Variances ( \hat{\psi}_i ) = 1 - ( \hat{\eta}_i^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allied Chem</td>
<td>.783</td>
<td>-.217</td>
<td>.34</td>
</tr>
<tr>
<td>Dupont</td>
<td>.773</td>
<td>-.458</td>
<td>.19</td>
</tr>
<tr>
<td>Union Carbide</td>
<td>.794</td>
<td>-.234</td>
<td>.31</td>
</tr>
<tr>
<td>Exxon</td>
<td>.713</td>
<td>.473</td>
<td>.27</td>
</tr>
<tr>
<td>Texaco</td>
<td>.712</td>
<td>.524</td>
<td>.22</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
.571 & .733 \\
\end{bmatrix}
\]

Cumulative proportion of total (standardized) sample variance:

Residual Matrix:

\[
R - (\bar{\Lambda} \bar{\Lambda}' + \bar{\Psi}) =
\begin{bmatrix}
0 & -.13 & -.16 & -.07 & .02 \\
0 & .12 & .06 & .01 & \\
0 & .02 & .02 & \\
0 & .23 & & \\
\end{bmatrix}
\]
Comments:

1. Estimated loadings on a factor do not change as the number of factors is increased.

2. Diagonal elements of $S$ (or $R$) are exactly equal to the diagonal elements of $\tilde{L}\tilde{L}' + \breve{\Sigma}$, but sample covariances may not be exactly reproduced.

   Select the number of factors $m$ to make off-diagonal elements small for the residual matrix

   $$S - [\tilde{L}\tilde{L}' + \breve{\Sigma}]$$

Note that the sum of the squared elements of

$$S - (\tilde{L}\tilde{L}' + \breve{\Sigma})$$

is

$$\sum_{k=m+1}^{p} \lambda_k^2$$

Choose $m$ big enough to make this sum very small.

3. Contribution of the $k$-th factor to the total variance $\text{tr}(S) = \sum_{i=1}^{p} \hat{\lambda}_i$ is

   $$\hat{L}_k' I \hat{L}_k = [\sqrt{\hat{\lambda}_k} \hat{e}_k]' [\sqrt{\hat{\lambda}_k} \hat{e}_k] = \hat{\lambda}_k$$

   \[
   \left\{ \begin{array}{l}
   \text{Proportion of total sample variance due to $k$-th factor} \\
   \frac{\hat{\lambda}_k}{\text{tr}(S)} \quad \text{using $S$} \\
   \frac{\hat{\lambda}_k}{p} \quad \text{using $R$}
   \end{array} \right. 
   \]
Principal Factor Method:

Consider the model for the correlation matrix

\[ R = LL' + \Psi \]

Then

\[ LL' = R - \Psi = \begin{bmatrix} h_1^2 & r_{12} & \cdots & r_{1p} \\ r_{21} & h_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & h_p^2 \end{bmatrix} \]

where \( h_i^2 = 1 - \psi_i \).

Suppose initial estimates are available for the communalities

\[ (h_1^*)^2 (h_2^*)^2 \cdots (h_p^*)^2 \]

Then

\[ R_r = \begin{bmatrix} (h_1^*)^2 & r_{12}^* & \cdots & r_{1p}^* \\ (h_2^*)^2 & 1 & \cdots & r_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ (h_p^*)^2 & r_{p1}^* & \cdots & 1 \end{bmatrix} = L_r^* L_r'^* \]

where

\[ L_r^* = \begin{bmatrix} \sqrt{\hat{\lambda}_1^*} & \hat{e}_1^* & \cdots & \hat{e}_m^* \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\hat{\lambda}_m^*} & \hat{e}_1^* & \cdots & \hat{e}_m^* \end{bmatrix} \]

\[ \psi_i^* = 1 - \sum_{j=1}^{m} \hat{\lambda}_j^* [\hat{e}_{ij}^*]^2 \]

and \( \hat{\lambda}_1^* \geq \cdots \geq \hat{\lambda}_m^* \)

are the largest eigenvalues for \( R_r \)

and \( \hat{e}_1^*, \ldots, \hat{e}_m^* \) are the corresponding eigenvectors.

\[ (\hat{h}_i^*)^2 = \sum_{j=1}^{m} \hat{\lambda}_j^* [\hat{e}_{ij}^*]^2 = 1 - \psi_i^* \]
Apply this procedure iteratively

1) Start with
\[
(h_i^*)^2 = R^2 \text{ value for the regression of } X_i \\
\text{on the other variables}
\]

2) Compute factor loadings from eigenvalues and eigenvectors of \( R_p \)

3) Compute new \((h_i^*)^2\) values

[Repeat steps (2) and (3) until]

algorithm converges

Problems:
- some eigenvalues of \( R_p \) can be negative
- choice of \( m \)

If \( m \) is too large some communalities may become larger than one, and the iterations will terminate unless you use one of the following options:

**HEYWOOD**
Fix any communality that is larger than one equal to one and continue iterations with respect to the remaining variables.

**ULTRA HEYWOOD**
Continue iterations with respect to all variables, regardless of the size of the communalities
Maximum Likelihood Method

A likelihood function is needed so additional assumptions are made

$$X_1, X_2, \ldots, X_n \sim NID(\mu, \Sigma_{pxp})$$

and

$$X_j = L F_j + \xi_j$$

where

$$\Sigma = L L' + \Psi$$

$$F_1, F_2, \ldots, F_n \sim NID(0, I_{mxm})$$

$$\xi_1, \xi_2, \ldots, \xi_n \sim NID(0, \Psi_{pxp})$$

the $\xi_i$'s and $F_i$'s are independent.
For $m$ factors:

- Estimated communalities:
  \[ \hat{h}_i^2 = \frac{\sum_{k=1}^{p} \hat{\lambda}_{ik}^2}{\text{trace}(S_n)} \]
  for $i = 1, 2, \ldots, p$

- Proportion of total sample variance due to the $k$-th factor:
  \[ \frac{\sum_{i=1}^{p} \hat{\lambda}_{ik}^2}{p} \]

A further restriction which specifies a unique solution for \( \Lambda \) within a diagonal matrix \( \Lambda \) is:

\( L' \Lambda L = A \)

See supplement for details of maximizing the likelihood function.

The m.a.e.\(_2\) and:

\[ \hat{X} = \hat{X} \begin{bmatrix} \hat{a}_1 \hat{a}_2 \end{bmatrix} \]

\[ \hat{L}_{p \times m} = \begin{bmatrix} \hat{a}_{i1} & \hat{a}_{i2} \end{bmatrix} \]
Maximum likelihood estimates for the example with 5 stocks.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Loadings for Factor 1</th>
<th>Loadings for Factor 2</th>
<th>Specific Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allied Chem.</td>
<td>.684</td>
<td>.189</td>
<td>.50</td>
</tr>
<tr>
<td>Du Pont</td>
<td>.694</td>
<td>.517</td>
<td>.35</td>
</tr>
<tr>
<td>Union Carbide</td>
<td>.681</td>
<td>.248</td>
<td>.47</td>
</tr>
<tr>
<td>Exxon</td>
<td>.621</td>
<td>-.073</td>
<td>.61</td>
</tr>
<tr>
<td>Texaco</td>
<td>.792</td>
<td>-.442</td>
<td>.18</td>
</tr>
<tr>
<td>Cumulative proportion of variance</td>
<td>.485</td>
<td>.598</td>
<td></td>
</tr>
</tbody>
</table>

Residual Matrix

\[ R = \hat{L} \hat{L}' - \hat{\Phi} \]

\[
\begin{bmatrix}
0 & .005 & -.004 & -.024 & -.004 \\
0 & -.003 & -.004 & .000 \\
0 & .031 & -.004 \\
0 & .000 & \\
0 &
\end{bmatrix}
\]

Note:
The elements of the residual matrix are smaller for the m.l.e method.
A large sample chi-squared criterion for deciding if the number of factors $m$ is sufficient:

\[ H_0: \Sigma_{p\times p} = L_{p \times m} \Lambda_{m \times m} L_{m \times p} + \Psi_{p \times p} \]

\[ H_A: \Sigma \text{ is any positive definite covariance matrix} \]

Likelihood ratio test:

Under $H_0$: \[ \hat{\Sigma} = \hat{L} \hat{L}' + \hat{\Psi} \]
\[ \hat{\Lambda} = \hat{\Sigma}^{-1} \]

Under $H_A$: \[ \hat{\Sigma} = \hat{S}_m = \frac{1}{n} \hat{A} \]
\[ \hat{\Lambda} = \hat{S}_m^{-1} \]

\[ -2 \ln(\Lambda) = n \left[ \ln |\hat{L} \hat{L}' + \hat{\Psi}| - \ln |\frac{1}{n} \hat{S}_m| \right] \]
\[ + n \left[ \frac{1}{2} \left( (\hat{L} \hat{L}' + \hat{\Psi})^{-1} \Sigma_{\hat{S}_m} \right) - p \right] \]

\[ \text{this is zero} \]

\[ = n \left[ \ln |\hat{L} \hat{L}' + \hat{\Psi}| - \ln |\frac{1}{n} \hat{S}_m| \right] \]

\[ \text{d.f.} = \frac{1}{2} \left[ (p-m)^2 - p - m \right] \]

Bartlett correction: Reject $H_0$ if

\[ \left( n-1 - \frac{3p+4m-5}{6} \right) \ln \left( \frac{|\hat{L} \hat{L}' + \hat{\Psi}|}{|\frac{1}{n} \hat{S}_m|} \right) \]

\[ > \chi^2_{\frac{1}{2} \left[ (p-m)^2 - p - m \right], \alpha} \]

Must have

\[ m < \frac{1}{2} \left( 2p+1 - \sqrt{8p+1} \right) \text{ for positive d.f.} \]
Determination of d.f.

Under $H_A \cup H_0$:

Estimate $p$ means

$$\frac{P(P+1)}{2} \text{ elements of } \Sigma$$

Under $H_0$:

Estimate $p$ means

$$\Sigma = LL' + \Psi$$

$p$ diagonal elements

$p(m-1)$ elements in $L$ with constraint $L'\Psi L = \text{diagonal matrix}$

df = \left[ p + \frac{P(P+1)}{2} \right] - \left[ p + mp + p - \frac{m(m-1)}{2} \right]

\begin{align*}
&= \frac{1}{2} \left[ (p-m)^2 - p - m \right]
\end{align*}

Likelihood ratio test for $H_0$:

$$\Sigma_{p \times p} = L_{p \times m} L_{p \times m}' + \Psi_{p \times p}$$

diagonal matrix

tends to suggest too many factors ($m$ values that are too large)

- non-normality
- outliers
- non-linearity
- correlated measurement errors

Consider AIC or SBC values
Kaiser Measure of Sampling Adequacy: (Psychometrika, 1970, 401-415)
(Kaiser + Rice, Ed. Psych Meas. 1974, 24, 111-119)

\[
MSA = 1 - \frac{\sum_{j \neq k} g_{jk}^2}{\sum_{j \neq k} r_{jk}^2}
\]

\(r_{jk}\) = correlation between the j-th and k-th variables

\(g_{jk}\) = partial correlation between the j-th and k-th variables controlling for all other variables in the analysis

Guidelines:

\[.9 < MSA\] excellent data
\[.8 < MSA \leq .9\] very good
\[.7 < MSA \leq .8\] good
\[.6 < MSA \leq .7\] mediocre
\[.5 < MSA \leq .6\] miserable
\[MSA < .5\] unacceptable

Tucker and Lewis
Reliability Coefficient
(Psychometrika, 1973, 38, 1-10)

From m.l.e. method, compute

\[\hat{\Lambda}_m\] estimate of factor loadings

\[\hat{\Sigma} = \text{diag}\left( R - \hat{\Lambda}_m \hat{\Lambda}_m' \right)\]

\[G_m = \hat{\Sigma}_{m}^{-\frac{1}{2}} \left( R - \hat{\Lambda}_m \hat{\Lambda}_m' \right) \hat{\Sigma}_{m}^{-\frac{1}{2}}\]

for the model with \(m\) factors

\(g_{mij}\) is called a "partial correlation" between variables \(X_i\) and \(X_j\) controlling for the \(m\) common factors.
Sum of squared "partial correlations" between variables controlling for m common factors:

\[ F_m = \sum_{i<j} g_{mij}^2 \]

Mean square:

\[ M_m = \frac{F_m}{df_m} \]

where

\[ df_m = \frac{1}{2} \left[ (p-m)^2 - p - m \right] \]

are the d.f. for the log-likelihood ratio test of

\( H_0: m \) factors are sufficient

"Mean square" for model with zero common factors:

Then \( G_0 = R \) and

\[ M_0 = \left( \sum_{i<j} r_{ij}^2 \right) / df_0 \]

where

\[ df_0 = \frac{p(p-1)}{2} \]

are d.f. for testing that all correlations are zero.

Reliability coefficient

\[ \hat{\rho}_m = \frac{M_0 - M_m}{M_0 - \frac{1}{n_m}} \]

where

\[ n_m = (n-1) - \frac{2p+5}{6} - \frac{2m}{3} \]

is a Bartlett correction factor.
Cornbach’s Alpha

A set of observe items \[ X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \]
that “measure” the same latent trait should all have high positive correlations.

Define

\[
\bar{\gamma} = \frac{1}{\frac{p(p-1)}{2}} \sum_{i<j} \text{Cov}(X_i, X_j)
\]

\[
= \frac{1}{p} \sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\text{Var}(X_i)}
\]

\[
= \frac{2}{p-1} \sum_{i=1}^{p} \frac{S_{ii}}{\sum_{i} S_{ii}}
\]

Cornbach’s Alpha:

\[
\alpha = \frac{p \bar{\gamma}}{1 + (p-1) \bar{\gamma}}
\]

- \( \alpha = 0 \) if \( \bar{\gamma} = 0 \)
- \( \alpha = 1 \) if \( \bar{\gamma} = 1 \)
- When the analysis is done, be sure that all scales are orientated in the same direction, so all correlations are positive.
Rational:

 Measure $p$ items $\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_p \end{bmatrix}$

 Suppose each of the $p$ items has a high positive correlation with a latent trait

 Then each trait will have a high positive correlation with $\sum_{i=1}^{p} X_i$

 $\text{Var}(\sum_{i=1}^{p} X_i) = \sum_{i=1}^{p} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)$

 $= \sum_{i=1}^{p} \text{Var}(X_i) + \sum_{i<j} \rho_{ij} \sqrt{\text{Var}(X_i) \text{Var}(X_j)}$

 If $\rho_{ij} = 1$ for all $i \neq j$

 then

 $\text{Var}(\sum_{i=1}^{p} X_i) = \sum_{i=1}^{p} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j)$

 $= \sum_{i=1}^{p} \text{Var}(X_i) \sqrt{\text{Var}(X_i) \text{Var}(X_j)}$

 $= \left[ \sum_{i=1}^{p} \rho_{ii} \text{Var}(X_i) \right]^{2}$

 $= p^2$ for standardized variables, $\text{Var}(X_i) = 1$

 if $\rho_{ij} = 1$ for all $i \neq j$.

 This is the maximum value for $\text{Var}(\sum_{i=1}^{p} X_i)$

 Also

 $\sum_{i=1}^{p} \text{Var}(X_i) = p$ for standardized variables
In the extreme case where
\[ p_{ij} = 1 \text{ for all } i \neq j \]

\[ \frac{\sum \text{Var}(x_i)}{\text{Var}(\varepsilon x_i)} = \frac{p}{p^2} = \frac{1}{p} \]

and

\[ \alpha = \frac{p}{p-1} \left[ 1 - \frac{\sum \text{Var}(x_i)}{\text{Var}(\varepsilon x_i)} \right] \]

\[ = \frac{p}{p-1} \left[ 1 - \frac{p}{p^2} \right] = 1 \]

Note that:

\[ \alpha = \frac{p}{p-1} \left[ 1 - \frac{\sum \text{Var}(x_i)}{\text{Var}(\varepsilon x_i)} \right] \]

\[ = \frac{p \bar{r}}{1 + (p-1) \bar{r}} \quad \text{where} \quad \bar{r} = \frac{\text{Cov}}{\text{Var}} \]

\[ \text{Factor Rotation:} \]

Let \( T_{m \times m} \) be an orthogonal matrix \( (TT' = T'T = I_{m \times m}) \).

Then, an orthogonal transformation of the factor loading matrix is

\[ L_{p \times m} T_{m \times m} = \hat{L}^*_{p \times m} \]

and the decomposition of the covariance matrix is

\[ \hat{L} L' + \hat{\varphi} = \hat{L} T T' \hat{L}' + \hat{\varphi} \]

\[ = \hat{L}^* (\hat{L}^*)' + \hat{\varphi} \]

Estimated specific variances and communalities are not altered by orthogonal transformations of \( \hat{L} \).
The varimax criterion:

Define \( \hat{\xi}_{ij}^* = \hat{\xi}_{ij} / \hat{h}_i \)

to be the “scaled” loading of the
\( i \)-th variable on the \( j \)-th factor
after rotation.

The varimax procedure
selects the orthogonal transformation
that maximizes

\[
V = \frac{1}{P} \sum_{j=1}^{m} \left[ \sum_{i=1}^{p} (\hat{\xi}_{ij}^*)^4 - \frac{1}{p} \left( \sum_{i=1}^{p} \hat{\xi}_{ij}^* \right)^2 \right]
\]

\( \propto \sum_{j=1}^{m} (\text{variance of squares of \( \hat{\xi}_{ij}^* \)}) \)

Scaling gives variables with smaller
communalities more influence.

The stock market example

(\( m = 2 \) factors)

(maximum likelihood)

\( \Rightarrow F_1^* \)

\( \Rightarrow F_2^* \)

After rotation each of the \( p \)
variables (or measured traits) should
have a high loading on only
one factor. (This is not always
possible).
Varimax rotation of \( m = 2 \) factors for the stock market example:
(maximum likelihood method)

<table>
<thead>
<tr>
<th>Variable</th>
<th>m.e.e.'s</th>
<th></th>
<th>Varimax rotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F_1 )</td>
<td>( F_2 )</td>
<td>( F_1^* )</td>
</tr>
<tr>
<td>Allied Chem.</td>
<td>.68</td>
<td>.19</td>
<td>.60</td>
</tr>
<tr>
<td>Du Pont</td>
<td>.69</td>
<td>.52</td>
<td>.85</td>
</tr>
<tr>
<td>Union Carbide</td>
<td>.68</td>
<td>.25</td>
<td>.64</td>
</tr>
<tr>
<td>Exxon</td>
<td>.62</td>
<td>-.09</td>
<td>.36</td>
</tr>
<tr>
<td>Texaco</td>
<td>.79</td>
<td>-.44</td>
<td>.21</td>
</tr>
</tbody>
</table>

Quartimax criterion:

Scaled loadings:

\[
\hat{\ell}_{ij} = \frac{\hat{\ell}_{ij}}{\hat{h}_i}
\]

where \( \hat{h}_i^2 = \sum_{j=1}^{m} \hat{\ell}_{ij}^2 \)

Select an orthogonal transformation (rotation) of the loadings to maximize

\[
Q = \frac{1}{m} \sum_{i=1}^{p} \left[ \sum_{j=1}^{m} (\hat{\ell}_{ij}^*)^4 - \frac{1}{m} \left( \sum_{j=1}^{m} \hat{\ell}_{ij}^2 \right)^2 \right]
\]

— the general market factor was destroyed by the rotation (it is possible to keep some factors fixed while rotating)

— relationship to pattern of correlations
Objectives:

1. Each of the $p$ variables (or traits) should have a fairly high loading on the same factor.

2. Each variable (or trait) should have a high loading on at most one other factor and near zero loadings on remaining factors.

Quartimax rotation of $m=2$ factors for the stock returns data (maximum likelihood method)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_1^*$</th>
<th>$F_2^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allied Chem.</td>
<td>.68</td>
<td>.19</td>
<td>.71</td>
<td>.05</td>
</tr>
<tr>
<td>Dupont</td>
<td>.69</td>
<td>.52</td>
<td>.83</td>
<td>-.26</td>
</tr>
<tr>
<td>Union Carbide</td>
<td>.68</td>
<td>.25</td>
<td>.72</td>
<td>-.01</td>
</tr>
<tr>
<td>Exxon</td>
<td>.62</td>
<td>-.07</td>
<td>.56</td>
<td>.27</td>
</tr>
<tr>
<td>Texaco</td>
<td>.79</td>
<td>-.44</td>
<td>.60</td>
<td>.68</td>
</tr>
</tbody>
</table>
PROMAX METHOD:

An "oblique" or non-orthogonal transformation

(step 1) First do the varimax rotation to obtain the loadings $L^{*}_{p \times m}$

(step 2) Construct another $p \times m$ matrix $Q$ where

$$q_{ij} = |l_{ij}^{*}|^{k-1} l_{ij}^{*} \quad \text{and} \quad \sum_j q_{ij} = 1 \quad (k \geq 1 \text{ is chosen by trial and error, usually } k < 4)$$

(step 3) Find a matrix $U$ such that each column of $L^{*}U$ is close to the corresponding column of $Q$.

Choose the $j$-th column of $U$ to minimize

$$(q_{ij} - L^{*}y_j)^{'}(q_{ij} - L^{*}y_j)$$

This yields

$$U = (L^{*}L^{*})^{-1} L^{*}Q$$

(step 4) Rescale $U$ so transformed factors have unit variance

$$D^2 = \text{diag}((U^{'}U)^{-1})$$

$$M = UD$$
PROMAX transformation yields factors with loadings

\[ L^\circ = L^* M \]

Also,

\[ \Phi = (M'M)^{-1} \]

is the correlation matrix for the new factors

\[ L^* E = L^* MM^{-1} E = L^\circ E^\circ \]

\[ \Rightarrow \]

\[ \text{Var}(E^\circ) = \text{Var}(M^{-1}E) = M^{-1} \text{Var}(E)(M^{-1})' = M^{-1} \left[ I (M^{-1})' \right] = (M'M)^{-1} \]
Estimation of Factor Scores:

\[(x_j - \mu)_{p \times 1} = L_{p \times m} \xi_j + \xi_j\]

If this model is correct

\[\text{Var}(\xi_j) = \Psi = \begin{bmatrix} \psi_1 & \cdots & \psi_p \end{bmatrix}\]

Weighted least squares estimation of \(\xi_j\):

\[\hat{\xi}_j = (L' \Psi^{-1} L)^{-1} L' \Psi^{-1} (x_j - \mu)\]

\[\hat{\xi}_j = (\hat{\xi}' \hat{\Psi} \hat{\xi})^{-1} \hat{\xi}' \hat{\Psi}^{-1} (x_j - \bar{x})\]
Ordinary (unweighted) least squares estimation is sometimes used when factor loadings are obtained from the principal component method (since specific variances tend to be more nearly equal, i.e. $\hat{\gamma}_1 = \ldots = \hat{\gamma}_p$)

$$
\tilde{F}_j = (\tilde{L}'\tilde{L})^{-1}\tilde{L}'(x_j - \bar{x})
$$

$$
= \begin{bmatrix}
\frac{1}{\sqrt{\lambda_1}} \hat{\epsilon}_1' (x_j - \bar{x}) \\
\vdots \\
\frac{1}{\sqrt{\lambda_m}} \hat{\epsilon}_m' (x_j - \bar{x})
\end{bmatrix}
$$

$$
\tilde{L} = \begin{bmatrix}
\sqrt{\lambda_1} \hat{\epsilon}_1 \\
\vdots \\
\sqrt{\lambda_m} \hat{\epsilon}_m
\end{bmatrix}
$$

**Regression Method:**

Consider the joint distribution of $(x_j - \mu)$ and $F_j$.

Assume multivariate normality as in the maximum likelihood approach to factor analysis.

$$
\begin{bmatrix}
x_j - \mu \\
F_j
\end{bmatrix} \sim N_{p+m}(\mu, \Sigma)
$$

where

$$
\Sigma = \begin{bmatrix}
LL' + 4 & L \\
L' & I_{m \times m}
\end{bmatrix}
$$

if the $m$ factor model is correct.
Obtain the conditional mean of $\mathbf{f}_j$ given $\mathbf{x}_j - \mu = (x_j - \mu)$

$$E(\mathbf{f}_j | \mathbf{x}_j - \mu) = \mathbf{L}' (\mathbf{L} \mathbf{L}' + \Phi)^{-1} (x_j - \mu)$$

Use the estimated conditional mean vector as the estimate of the factor scores

$$\hat{\mathbf{f}}_j = \mathbf{L}' (\mathbf{L} \mathbf{L}' + \Phi)^{-1} (x_j - \bar{x})$$

To reduce bad effects of an incorrect choice of $m$ replace $\mathbf{L} \mathbf{L}' + \Phi$ with $S = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})'$ to obtain

$$\hat{\mathbf{f}}_j = \mathbf{L}' S^{-1} (x_j - \bar{x})$$

Estimated factor scores for stock market example (maximum likelihood solution, $m = 2$, varimax rotation) (Regression estimates)

Recall $\mathbf{f}_j \sim NID_m(\mathbf{0}, \mathbf{I}_{m \times m})$

- check for outliers
- multivariate normality (circular contours)
- univariate tests of normality for the factor scores