Exploratory and Confirmatory Factor Analysis

Michael Friendly

Feb. 25, Mar. 3, 10, 2008
SCS Short Course
web notes: http://www.math.yorku.ca/SCS/Courses/factor/

Part 2: EFA Outline

1. Basic ideas of factor analysis
   - Linear regression on common factors
   - Partial linear independence
   - Partial linear independence: demonstration
   - Common variance vs. unique variance

2. Factor estimation methods
   - Example: Spearman's 'Two-factor' theory
   - Example: Holzinger & Swineford 9 abilities data

3. Factor and component rotation
   - Thurstone's Postulates of Simple Structure
   - Rotation methods: Overview
   - Oblique rotations
   - Procrustes rotations

4. Factor Scores

Basic Ideas of Factor Analysis

Overview & goals
- Goal of factor analysis: Parsimony—account for a set of observed variables in terms of a small number of latent, underlying constructs (common factors).
  - Fewer common factors than PCA components
  - Unlike PCA, does not assume that variables are measured without error
- Observed variables can be modeled as regressions on common factors
- Common factors can “account for” or explain the correlations among observed variables
- How many different underlying constructs (common factors) are needed to account for correlations among a set of observed variables?
  - Rank of correlation matrix = number of linearly independent variables.
  - Factors of a matrix: \( R = \Lambda \Lambda^T \) (“square root” of a matrix)
- Variance of each variable can be decomposed into common variance (communality) and unique variance (uniqueness)

A set of observed variables, \( x_1, x_2, \ldots, x_p \) is considered to arise as a set of linear combinations of some unobserved, latent variables called common factors, \( \xi_1, \xi_2, \ldots, \xi_k \).

That is, each variable can be expressed as a regression on the common factors. For two variables and one common factor, \( \xi \), the model is:

\[
\begin{align*}
  x_1 &= \lambda_1 \xi + z_1 \\
  x_2 &= \lambda_2 \xi + z_2
\end{align*}
\]

The common factors are shared among two or more variables. The unique factor, \( z_i \), associated with each variable represents the unique component of that variable.
Basic ideas: 1. Linear regression on common factors

**Assumptions:**
- Common and unique factors are uncorrelated:
  \[ r(\xi, z_1) = r(\xi, z_2) = 0 \]
- Unique factors are all uncorrelated and centered:
  \[ r(z_1, z_2) = 0 \quad E(z_i) = 0 \]
- This is a critical difference between factor analysis and component analysis: in PCA, the residuals are correlated.
- Another critical difference—more important—is that factor analysis only attempts to account for common variance, not total variance.

For \( k \) common factors, the common factor model can be expressed as

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_p
\end{bmatrix} = \begin{bmatrix}
  \lambda_{11} & \cdots & \lambda_{1k} \\
  \lambda_{21} & \cdots & \lambda_{2k} \\
  \vdots & \ddots & \vdots \\
  \lambda_{p1} & \cdots & \lambda_{pk}
\end{bmatrix} \begin{bmatrix}
  \xi_1 \\
  \xi_2 \\
  \vdots \\
  \xi_k
\end{bmatrix} + \begin{bmatrix}
  z_1 \\
  z_2 \\
  \vdots \\
  z_p
\end{bmatrix}
\]

or, in matrix terms:

\[ x = \Lambda \xi + z \]  \hspace{1cm} (2)

This model is not testable, since the factors are unobserved variables. However, the model (2) implies a particular form for the variance-covariance matrix, \( \Sigma \), of the observed variables, which is testable:

\[ \Sigma = \Lambda \Phi \Lambda^T + \Psi \]  \hspace{1cm} (3)

where:
- \( \Lambda_{p \times k} = \) factor pattern (“loadings”)
- \( \Phi_{k \times k} = \) matrix of correlations among factors.
- \( \Psi = \) diagonal matrix of unique variances of observed variables.

It is usually assumed initially that the factors are uncorrelated \( (\Phi = I) \), but this assumption may be relaxed if oblique rotation is used.

Basic ideas: 2. Partial linear independence

- The factors “account for” the correlations among the variables, since the variables may be correlated only through the factors.
- If the common factor model holds, the **partial correlations of the observed variables with the common factor(s) partialled out are all zero**:
  \[ r(x_i, x_j|\xi) = r(z_i, z_j) = 0 \]
- With one common factor, this has strong implications for the observed correlations:
  \[
  \begin{align*}
  r_{12} &= E(x_1, x_2) = E[(\lambda_1 \xi + z_1)(\lambda_2 \xi + z_2)] \\
          &= \lambda_1 \lambda_2 \\
  r_{13} &= \lambda_1 \lambda_3 \\
  \text{i.e. } r_{ij} &= \lambda_i \lambda_j
  \end{align*}
  \]

That is, the correlations in any pair of rows/cols of the correlation matrix are proportional **if the one factor model holds**. The correlation matrix has the structure:

\[
R_{(p \times p)} = \begin{bmatrix}
\lambda_1 & & \\
& \lambda_2 & \\
& & \ddots \\
& & & \lambda_p
\end{bmatrix} + \begin{bmatrix}
\mathbf{u}_1^2 & & \\
& \mathbf{u}_2^2 & \\
& & \ddots \\
& & & \mathbf{u}_p^2
\end{bmatrix}
\]

Similarly, if the common factor model holds with \( k \) factors, the pattern of correlations can be reproduced by the product of the matrix of factor loadings, \( \Lambda \) and its transpose:
### Simple example

Consider the following correlation matrix of 5 measures of “mental ability”

<table>
<thead>
<tr>
<th></th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>x5</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1.00</td>
<td>.72</td>
<td>.63</td>
<td>.54</td>
<td>.45</td>
</tr>
<tr>
<td>x2</td>
<td>.72</td>
<td>1.00</td>
<td>.56</td>
<td>.48</td>
<td>.40</td>
</tr>
<tr>
<td>x3</td>
<td>.63</td>
<td>.56</td>
<td>1.00</td>
<td>.42</td>
<td>.35</td>
</tr>
<tr>
<td>x4</td>
<td>.54</td>
<td>.48</td>
<td>.42</td>
<td>1.00</td>
<td>.30</td>
</tr>
<tr>
<td>x5</td>
<td>.45</td>
<td>.40</td>
<td>.35</td>
<td>.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

These correlations are exactly consistent with the idea of a single common factor ($g$).

The factor loadings, or correlations of the variables with $g$ are .9 .8 .7 .6 .5, e.g., $r_{12} = .9 \times .8 = .72$; $r_{13} = .9 \times .7 = .63$; etc.

Thus, the correlation matrix can be expressed exactly as

$$R_{(5 \times 5)} = \begin{bmatrix} .9 & .8 & .7 & .6 & .5 \\ .7 & .6 & .5 & & \\ & & & & \end{bmatrix} \begin{bmatrix} .19 & .36 & & & \\ & .51 & & & \\ & & .64 & & \\ & & & .75 & \end{bmatrix}$$

### Implications

The implications of this are:

- The matrix $(R - \Psi)$, i.e., the correlation matrix with communalitites on the diagonal is of rank $k \ll p$. [PCA: rank($R$) = $p$]
- Thus, FA should produce fewer factors than PCA, which “factors” the matrix $R$ with 1s on the diagonal.
- The matrix of correlations among the variables with the factors partialled out is:

$$R - \Lambda \Lambda^T = \Psi = \begin{bmatrix} u_1^2 & & & \\ & u_2^2 & & \\ & & \ddots & \\ & & & u_p^2 \end{bmatrix}$$

Thus, if the $k$-factor model fits, there remain no correlations among the observed variables when the factors have been taken into account.

### Partial linear independence: demonstration

- Generate two factors, MATH and VERBAL.
- Then construct some observed variables as linear combinations of these.

```sas
data scores; drop n;
do N = 1 to 800; *-- 800 observations;
  MATH = normal(13579);
  VERBAL = normal(13579);
  mat_test = normal(76543) + 1.*MATH - .2*VERBAL;
  eng_test = normal(76543) + 1.*MATH + 1.*VERBAL;
  sci_test = normal(76543) + .7*MATH - .3*VERBAL;
  his_test = normal(76543) - .2*MATH + .5*VERBAL;
  output;
end;
label MATH = ‘Math Ability Factor’
  VERBAL = ‘Verbal Ability Factor’
  mat_test = ‘Mathematics test’
  eng_test = ‘English test’
  sci_test = ‘Science test’
  his_test = ‘History test’;
```

```sas
proc corr nosimple noprob;
  var mat_test eng_test sci_test his_test;
  title2 ‘Simple Correlations among TESTS’;
```

<table>
<thead>
<tr>
<th></th>
<th>mat_test</th>
<th>eng_test</th>
<th>sci_test</th>
<th>his_test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics test</td>
<td>1.000</td>
<td>-0.069</td>
<td>0.419</td>
<td>-0.144</td>
</tr>
<tr>
<td>English test</td>
<td>-0.069</td>
<td>1.000</td>
<td>-0.097</td>
<td>0.254</td>
</tr>
<tr>
<td>Science test</td>
<td>0.419</td>
<td>-0.097</td>
<td>1.000</td>
<td>-0.227</td>
</tr>
<tr>
<td>History test</td>
<td>-0.144</td>
<td>0.254</td>
<td>-0.227</td>
<td>1.000</td>
</tr>
</tbody>
</table>

```sas
proc corr nosimple noprob;
  var mat_test eng_test sci_test his_test;
  partial MATH VERBAL;
  title2 ‘Partial Correlations, partialling Factors’;
```

<table>
<thead>
<tr>
<th></th>
<th>mat_test</th>
<th>eng_test</th>
<th>sci_test</th>
<th>his_test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics test</td>
<td>1.000</td>
<td>-0.048</td>
<td>-0.015</td>
<td>0.035</td>
</tr>
<tr>
<td>English test</td>
<td>-0.048</td>
<td>1.000</td>
<td>0.028</td>
<td>-0.072</td>
</tr>
<tr>
<td>Science test</td>
<td>-0.015</td>
<td>0.028</td>
<td>1.000</td>
<td>-0.064</td>
</tr>
<tr>
<td>History test</td>
<td>0.035</td>
<td>-0.072</td>
<td>-0.064</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Basic ideas: 3. Common variance vs. unique variance

- Factor analysis provides an account of the variance of each variable as common variance (communality) and unique variance (uniqueness).
- From the factor model (with uncorrelated factors, $\Phi = I$),

\[
x = \Lambda \xi + z
\]

(4)

it can be shown that the common variance of each variable is the sum of squared loadings:

\[
\begin{align*}
\text{var}(x_i) &= \sum \lambda_{ik}^2 + \text{var}(z_i) \\
&= h_i^2 \text{(communality)} + u_i^2 \text{(uniqueness)}
\end{align*}
\]

If a measure of reliability is available, the unique variance can be further divided into error variance, $e_i^2$, and specific variance, $s_i^2$. Using standardized variables:

\[
\text{reliability} \quad \begin{array}{c}
\text{var}(x_i) = 1 = h_i^2 \\
\text{(communality)} + s_i^2 \\
\text{(uniqueness)} + e_i^2
\end{array}
\]

Correlations or covariances?

Correlations or covariances?

As we saw in PCA, factors can be extracted from either the covariance matrix ($\Sigma$) of the observed variables, with the common factor model:

\[
\Sigma = \Lambda \Phi \Lambda^T + \Psi
\]

or the correlation matrix ($R$), with the model

\[
R = \Lambda \Phi \Lambda^T + \Psi
\]

- If the variables are standardized, these are the same: $R = \Sigma$
- If the units of the variables are important & meaningful, analyze $\Sigma$
- Some methods of factor extraction are scale free— you get equivalent results whether you analyse $R$ or $\Sigma$.
- Below, I’ll describe things in terms of $\Sigma$. 

E.g., for two tests, each with reliability $r_{x_i} = .80$, and

\[
\begin{align*}
x_1 &= .8 \xi + .6 z_1 \\
x_2 &= .6 \xi + .8 z_1
\end{align*}
\]

we can break down the variance of each variable as:

\[
\begin{array}{c|cc|c|c}
\text{var} &= \text{common} + \text{unique} & \rightarrow \text{(specific} + \text{error)} \\
x_1: & .64 & .36 & .16 & .20 \\
x_1: & .36 & .64 & .44 & .20 \\
\end{array}
\]
Factor Estimation Methods: Basic ideas

Common characteristics

Many methods of factor extraction for EFA have been proposed, but they have some common characteristics:

- **Initial solution with uncorrelated factors** ($\Phi = I$)
  - The model becomes
  \[ \Sigma = \Lambda \Lambda^T + \Psi \]
  - If we know (or can estimate) the communalities ($= 1 - \text{uniqueness} = 1 - \psi_{ii}$), we can factor the "reduced covariance (correlation) matrix", $\Sigma - \Psi$
  \[ \Sigma - \Psi = \Lambda \Lambda^T = (UD^{1/2})(D^{1/2}U^T) \tag{5} \]
  - In (5), $U$ is the matrix of eigenvectors of $\Sigma - \Psi$ and $D$ is the diagonal matrix of eigenvalues.

- **Initial estimates of communalities**: A good prior estimate of the communality of a variable is its’ $R^2$ (SMC) with all other variables.
  \[ \text{SMC}_i \equiv R^2_{x_i | \text{others}} \leq h_i^2 = \text{communality} = 1 - \psi_{ii} \]

Factor Estimation Methods: Fit functions

Given $S_{(p \times p)}$, an observed variance-covariance matrix of $x_{(p \times 1)}$, the computational problem is to estimate $\hat{\Lambda}$ and $\hat{\Psi}$ such that:

\[ \hat{\Sigma} = \hat{\Lambda} \hat{\Lambda}^T + \hat{\Psi} \approx S \]

Let $F(S, \hat{\Sigma}) = \text{measure of distance between } S \text{ and } \hat{\Sigma}$. Factoring methods differ in the measure $F$ used to assess badness of fit:

- **Iterated PFA (ULS, PRINIT) [NOT Scale Free]** Minimizes the sum of squares of differences between $S$ and $\hat{\Sigma}$.
  \[ F_{ULS} = \text{tr}(S - \hat{\Sigma})^2 \]

- **Generalized Least Squares (GLS) [Scale Free]** Minimizes the sum of squares of differences between $S$ and $\hat{\Sigma}$, weighted inversely by the variances of the observed variables.
  \[ F_{GLS} = \text{tr}(I - S^{-1}\hat{\Sigma})^2 \]

Factor Estimation Methods

- **Maximum likelihood [Scale Free]** Finds the parameters that maximize the likelihood ("probability") of observing the data ($S$) given that the FA model fits for the population $\Sigma$.
  \[ F_{ML} = \text{tr}(S\hat{\Sigma}^{-1}) - \log |\hat{\Sigma}^{-1}S| - p \]
  - In large samples, $(N - 1)F_{\text{min}} \sim \chi^2$
  - The hypothesis tested is
    \[ H_0 : k \text{ factors are sufficient} \]
    vs.
    \[ H_1 : k > k \text{ factors are required} \]
  - Good news: This is the only EFA method that gives a significance test for the number of common factors.
  - Bad news: This $\chi^2$ test is extremely sensitive to sample size.
Example: Spearman’s ‘two-factor’ theory

Spearman used this data on 5 tests to argue for a ‘two-factor’ theory of ability:

- general ability factor—accounts for all correlations
- unique factors for each test

```sas
data spear5 (TYPE=CORR);
  input _TYPE_ $ _NAME_ $ test1 - test5;
  label test1='Mathematical judgement'
           test2='Controlled association'
           test3='Literary interpretation'
           test4='Selective judgement'
           test5='Spelling';
  datalines;
  CORR test1 1.00 . . . .
  CORR test2 .485 1.00 . . .
  CORR test3 .400 .397 1.00 . .
  CORR test4 .397 .397 .335 1.00 .
  CORR test5 .295 .247 .275 .195 1.00
  N 100 100 100 100 100;
```

NB: The _TYPE_ = ‘N’ observation is necessary for a proper $\chi^2$ test.

Use `METHOD=ML` to test 1 common factor model
```
proc factor data=spear5
  method=ml /* use maximum likelihood */
  residuals /* print residual correlations */
  nfact=1; /* estimate one factor */
title2 'Test of hypothesis of one general factor';
```

Initial output:
```
Initial Factor Method: Maximum Likelihood
Prior Communality Estimates: SMC

TEST1    TEST2    TEST3    TEST4    TEST5
  0.334390  0.320497  0.249282  0.232207  0.123625

1 factors will be retained by the NFACTOR criterion.

Iter   Criterion   Ridge   Change   Communalities
  1  0.007610   0.000000  0.160632  0.495000  0.463500  0.348200  0.317900  0.158300
  2  0.007591   0.000000  0.004291  0.495299  0.466200  0.343900  0.320300  0.158900
  3  0.007591   0.000000  0.000201  0.495400  0.466200  0.343900  0.320300  0.158700
```

Hypothesis tests & fit statistics:
```
Significance tests based on 100 observations:

Test of H0: No common factors.
  vs HA: At least one common factor.
  Chi-square = 87.205  df = 10  Prob>chi**2 = 0.0001

Test of H0: 1 Factors are sufficient.
  vs HA: More factors are needed.
  Chi-square = 0.727  df = 5  Prob>chi**2 = 0.9815

Chi-square without Bartlett’s correction = 0.7510547937
Akaike’s Information Criterion = -9.248945206
Schwarz’s Bayesian Criterion = -22.27479614
Tucker and Lewis’s Reliability Coefficient = 1.1106900868
```

NB: The 1-factor model fits exceptionally well—too well? (like Mendel’s peas)

Factor pattern (“loadings”):
```
Factor Pattern

FACTOR1

TEST1  0.70386  Mathematical judgement
TEST2  0.68282  Controlled association
TEST3  0.58643  Literary interpretation
TEST4  0.56594  Selective judgement
TEST5  0.39837  Spelling
```

NB: For uncorrelated factors, the factor coefficients are also correlations of the variables with the factors.

- Mathematical judgment is the ‘best’ measure of the g factor (general intelligence)
- Spelling is the worst measure
Example: Spearman’s 'two-factor' theory

Common and unique variance:

<table>
<thead>
<tr>
<th>FACTOR1</th>
<th>Common</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEST1</td>
<td>0.70386</td>
<td>.495</td>
</tr>
<tr>
<td>TEST2</td>
<td>0.68282</td>
<td>.466</td>
</tr>
<tr>
<td>TEST3</td>
<td>0.58643</td>
<td>.344</td>
</tr>
<tr>
<td>TEST4</td>
<td>0.56594</td>
<td>.320</td>
</tr>
<tr>
<td>TEST5</td>
<td>0.39837</td>
<td>.159</td>
</tr>
</tbody>
</table>

- Mathematical judgement is the 'best' measure of the g factor
- Spelling is the worst measure

Example: Holzinger & Swineford 9 abilities data

Nine tests from a battery of 24 ability tests given to junior high school students at two Chicago schools in 1939.

```
title 'Holzinger & Swineford 9 Ability Variables';
data psych9(type=CORR);
  Input _NAME_ $1-3 _TYPE_ $5-9 X1 X2 X4 X6 X7 X9 X10 X12 X13;
  label X1='Visual Perception' X2='Cubes' X4='Lozenges'
    X6='Paragraph Comprehension' X7='Sentence Completion'
    X9='Word Meaning' X10='Addition' X12='Counting Dots'
    X13='Straight-curved Caps';
datalines;
  X1 CORR 1. . . . . . . . . .
  X2 CORR .318 1. . . . . . . .
  X4 CORR .436 .419 1. . . . . . .
  X6 CORR .335 .234 .323 1. . . . . . .
  X7 CORR .304 .157 .283 .722 1. . . . .
  X9 CORR .326 .195 .350 .714 .685 1. . . . .
  X10 CORR .116 .057 .056 .203 .246 .170 1. . . . .
  X12 CORR .314 .145 .229 .095 .181 .113 .585 1. . . . .
  X13 CORR .489 .239 .361 .309 .345 .280 .408 .512 1. . . . .
N   145 145 145 145 145 145 145 145 145
MEAN 29.60 24.80 15.97 9.95 18.85 90.18 68.59 109.75 191.8
STD  6.89  4.43  8.29  3.36  4.63  7.92  23.70  20.92  36.91
RELI .7563 .5677 .9365 .7499 .7536 .8701 .9518 .9374 .8889
run;
```

“Little Jiffy:” Principal factor analysis using SMC, Varimax rotation

- The 9 tests were believed to tap 3 factors: Visual, Verbal & Speed
- The default analysis is METHOD=PRINCIPAL, PRIORS=ONE \leftrightarrow PCA!
- The results are misleading, about both the number of factors and their interpretation.

```
title2 'Principal factor solution';
proc Factor data=psych9
  Method=PRINCIPAL
  Priors=SMC
  Round flag=.3
  Scree
  Rotate=VARIMAX;
run;
```

- method=PRINCIPAL is non-iterative; method=PRINIT uses iterated PFA
- ROUND option prints coefficients \times 100, rounded; FLAG option prints a * next to larger values

```
Eigenvalues of the Reduced Correlation Matrix:
Total = 4.05855691  Average = 0.45095077

<table>
<thead>
<tr>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.07328008</td>
<td>1.99393040</td>
<td>0.7572</td>
</tr>
<tr>
<td>2</td>
<td>1.07934969</td>
<td>0.45916492</td>
<td>0.2659</td>
</tr>
<tr>
<td>3</td>
<td>0.62018476</td>
<td>0.1528</td>
<td>0.1528</td>
</tr>
<tr>
<td>4</td>
<td>0.03035486</td>
<td>0.0075</td>
<td>0.0075</td>
</tr>
<tr>
<td>5</td>
<td>-0.07788705</td>
<td>-0.03243783</td>
<td>-0.03243783</td>
</tr>
<tr>
<td>6</td>
<td>-0.11032489</td>
<td>-0.0572</td>
<td>-0.0572</td>
</tr>
<tr>
<td>7</td>
<td>-0.14897447</td>
<td>-0.10824191</td>
<td>-0.10824191</td>
</tr>
<tr>
<td>8</td>
<td>-0.17546086</td>
<td>-0.0572</td>
<td>-0.0572</td>
</tr>
<tr>
<td>9</td>
<td>-0.23196521</td>
<td>-0.0572</td>
<td>-0.0572</td>
</tr>
</tbody>
</table>

2 factors will be retained by the PROPORTION criterion.

- NB: The default criteria (PROPORTION=1.0 or MINEIGEN=0) are seriously misleading.
Example: Holzinger & Swineford 9 abilities data

Scree plot

Holzinger & Swineford 9 Ability Variables
Principal factor solution (SMC)

Initial (unrotated) factor pattern:

<table>
<thead>
<tr>
<th>Factor Pattern</th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1 Visual Perception</td>
<td>57 *</td>
<td>13 *</td>
</tr>
<tr>
<td>X2 Cubes</td>
<td>37 *</td>
<td>4 *</td>
</tr>
<tr>
<td>X4 Lozenges</td>
<td>53 *</td>
<td>2 *</td>
</tr>
<tr>
<td>X6 Paragraph Comprehen</td>
<td>74 *</td>
<td>-39 *</td>
</tr>
<tr>
<td>X7 Sentence Completion</td>
<td>72 *</td>
<td>-31 *</td>
</tr>
<tr>
<td>X9 Word Meaning</td>
<td>71 *</td>
<td>-38 *</td>
</tr>
<tr>
<td>X10 Addition</td>
<td>41 *</td>
<td>44 *</td>
</tr>
<tr>
<td>X12 Counting Dots</td>
<td>46 *</td>
<td>59 *</td>
</tr>
<tr>
<td>X13 Straight-curved Caps</td>
<td>62 *</td>
<td>36 *</td>
</tr>
</tbody>
</table>

Interpretation ??

Varimax rotated factor pattern:

<table>
<thead>
<tr>
<th>Rotated Factor Pattern</th>
<th>Factor1</th>
<th>Factor2</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1 Visual Perception</td>
<td>39 *</td>
<td>43 *</td>
</tr>
<tr>
<td>X2 Cubes</td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>X4 Lozenges</td>
<td>42 *</td>
<td>32 *</td>
</tr>
<tr>
<td>X6 Paragraph Comprehen</td>
<td>83 *</td>
<td>11</td>
</tr>
<tr>
<td>X7 Sentence Completion</td>
<td>77 *</td>
<td>17</td>
</tr>
<tr>
<td>X9 Word Meaning</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>X10 Addition</td>
<td>8</td>
<td>59 *</td>
</tr>
<tr>
<td>X12 Counting Dots</td>
<td>3</td>
<td>75 *</td>
</tr>
<tr>
<td>X13 Straight-curved Caps</td>
<td>30</td>
<td>65 *</td>
</tr>
</tbody>
</table>

Interpretation ??

Maximum likelihood solutions

```
title2 'Maximum likelihood solution, k=2';
proc Factor data=psych9
   Method=ML
   NFact=2;
run;
```

Interpretation ??

In PCA, you can obtain the solution for all components, and just delete the ones you don’t want.

In iterative EFA methods, you have to obtain separate solutions for different numbers of common factors.

Here, we just want to get the $\chi^2$ test, and other fit statistics for the $k = 2$ factor ML solution.
Example: Holzinger & Swineford 9 abilities data
Maximum likelihood solution, k=2

<table>
<thead>
<tr>
<th>Test</th>
<th>Pr &gt; Chi-Square</th>
<th>DF</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: No common factors</td>
<td>&lt;.0001</td>
<td>36</td>
<td>483.4478</td>
</tr>
<tr>
<td>HA: At least one common factor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H0: 2 Factors are sufficient</td>
<td>&lt;.0001</td>
<td>19</td>
<td>61.1405</td>
</tr>
<tr>
<td>HA: More factors are needed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chi-Square without Bartlett’s Correction 63.415857
Akaike’s Information Criterion 25.415857
Schwarz’s Bayesian Criterion -31.142084
Tucker and Lewis’s Reliability Coefficient 0.821554

The sample size was supplied with the _TYPE_=N observations in the correlation matrix. Otherwise, use the option NOBS=n on the PROC FACTOR statement. (If you don’t, the default is NOBS=10000!)

Test of H0: No common factors → H0: R = I: all variables uncorrelated
H0: k = 2 is rejected here

Example: Holzinger & Swineford 9 abilities data
Maximum likelihood solution, k=3

proc Factor data=psych9
   Outstat=FACTORS /* Output data set */
   Method=ML
   NFact=3
   Round flag=.3
   Rotate=VARIMAX;

Specify k = 3 factors
Obtain an OUTSTAT= data set— I’ll use this to give a breakdown of the variance of each variable
A VARIMAX rotation will be more interpretable than the initial solution

Unrotated factor solution:

<table>
<thead>
<tr>
<th>Factor Pattern</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1 Visual Perception</td>
<td>51 *</td>
<td>18</td>
<td>43 *</td>
</tr>
<tr>
<td>X2 Cubes</td>
<td>32 *</td>
<td>7</td>
<td>39 *</td>
</tr>
<tr>
<td>X4 Lozenges</td>
<td>48 *</td>
<td>8</td>
<td>49 *</td>
</tr>
<tr>
<td>X6 Paragraph Comprehension</td>
<td>81 *</td>
<td>-30 *</td>
<td>-8</td>
</tr>
<tr>
<td>X7 Sentence Completion</td>
<td>80 *</td>
<td>-21</td>
<td>-16</td>
</tr>
<tr>
<td>X9 Word Meaning</td>
<td>77 *</td>
<td>-28</td>
<td>-4</td>
</tr>
<tr>
<td>X10 Addition</td>
<td>40 *</td>
<td>55 *</td>
<td>-37 *</td>
</tr>
<tr>
<td>X12 Counting Dots</td>
<td>40 *</td>
<td>72 *</td>
<td>-6</td>
</tr>
<tr>
<td>X13 Straight-curved Caps</td>
<td>56 *</td>
<td>43 *</td>
<td>16</td>
</tr>
</tbody>
</table>

H0: k = 3 is not rejected here
The \( \chi^2 \) test is highly dependent on sample size; other fit measures (later)
Interlude: Significance tests & fit statistics for EFA I

- As we have seen, ML solution \( \chi^2 = (N - 1)F_{\text{min}} \) (large sample test)
- Adding another factor always reduces \( \chi^2 \), but also reduces \( df \).
  - \( \chi^2 / df \) gives a rough measure of goodness-of-fit, taking \# factors into account. Values of \( \chi^2 / df \leq 2 \) are considered “good.”
  - Test \( \Delta \chi^2 = \chi^2_{m} - \chi^2_{m-1} \) on \( \Delta df = df_m - df_{m-1} \) degrees of freedom
  - \( \text{Pr}(\Delta \chi^2, \Delta df) \) tests if there is a significant improvement in adding one more factor.
- **Akaike Information Criterion (AIC):** penalizes model fit by \( 2 \times \# \) free parameters
  \[
  \text{AIC} = \chi^2 + 2(\# \text{ free parameters}) = \chi^2 + [p(p - 1) - 2df]
  \]
- **Bayesian Information Criterion (BIC):** greater penalty with larger \( N \)
  \[
  \text{BIC} = \chi^2 + \log N(\# \text{ free parameters})
  \]
- AIC and BIC: choose model with the smallest values

Interlude: Significance tests & fit statistics for EFA II

- **Tucker-Lewis Index (TLI):** Compares the \( \chi^2 / df \) for the null model \( (k = 0) \) to the \( \chi^2 / df \) for a proposed model with \( k = m \) factors
  \[
  TLI = \frac{(\chi^2_{0} / df_0) - (\chi^2_{m} / df_m)}{(\chi^2_{0} / df_0) - 1}
  \]
  - Theoretically, \( 0 \leq TLI \leq 1 \). “Acceptable” models should have at least \( TLI > .90 \); “good” models: \( TLI > .95 \)
- In CFA, there are many more fit indices. Among these, the **Root Mean Square Error of Approximation (RMSEA)** is popular now.
  \[
  \text{RMSEA} = \sqrt{\frac{(\chi^2 / df) - 1}{N - 1}}
  \]
  - “Adequate” models have \( \text{RMSEA} \leq .08 \); “good” models: \( \text{RMSEA} \leq .05 \).
Factor estimation methods

Example: Holzinger & Swineford 9 abilities data

Comparing solutions

Collect the test statistics in tables for comparison...

<table>
<thead>
<tr>
<th>k</th>
<th>H0: No common factors</th>
<th>Prob</th>
<th>ChiSq</th>
<th>DF</th>
<th>ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>483.4478</td>
<td>36</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>1</td>
<td>1 Factor is sufficient</td>
<td></td>
<td>172.2485</td>
<td>27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>2</td>
<td>2 Factors are sufficient</td>
<td></td>
<td>61.1405</td>
<td>19</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>3</td>
<td>3 Factors are sufficient</td>
<td></td>
<td>9.5453</td>
<td>12</td>
<td>0.6558</td>
</tr>
</tbody>
</table>

From these, various fit indices can be calculated...

<table>
<thead>
<tr>
<th>k</th>
<th>Chi2/df</th>
<th>Chi2</th>
<th>DF</th>
<th>diff</th>
<th>Pr &gt;</th>
<th>AIC</th>
<th>BIC</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>13.4291</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>6.3796</td>
<td>311.199</td>
<td>9</td>
<td>0</td>
<td>123.805</td>
<td>43.433</td>
<td>0.5672</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3.2179</td>
<td>111.108</td>
<td>8</td>
<td>0</td>
<td>25.416</td>
<td>-31.142</td>
<td>0.8216</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.7954</td>
<td>51.595</td>
<td>7</td>
<td>&lt;.0001</td>
<td>-14.052</td>
<td>-49.772</td>
<td>1.0165</td>
<td></td>
</tr>
</tbody>
</table>

All measures agree on $k = 3$ factors!

Factor and component rotation

In multiple regression, you can replace the $p$ regressors with a set of $p$ linear combinations of them without changing the $R^2$.

data demo;  
do i=1 to 20;  
  x1 = normal(0); x2 = normal(0); ^= random data;  
y = x1 + x2 + normal(0);  
x3 = x1 + x2; x4 = x1 - x2; ^= rotate 45 deg;  
output;  
end;  
proc reg data=demo;  
  model y = x1 x2;  
  model y = x3 x4;  

The models using $(x_1, x_2)$ and $(x_3, x_4)$ both have the same $R^2$:

Root MSE 1.36765 R-square [0.6233]

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T for HO: Parameter=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>-0.234933</td>
<td>0.30603261</td>
<td>-0.768</td>
<td></td>
</tr>
<tr>
<td>X1</td>
<td>1</td>
<td>1.151320</td>
<td>0.37796755</td>
<td>3.046</td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>1.112546</td>
<td>0.29270456</td>
<td>3.801</td>
<td></td>
</tr>
</tbody>
</table>

Root MSE 1.36765 R-square [0.6233]

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T for HO: Parameter=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>-0.234933</td>
<td>0.30603261</td>
<td>-0.768</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>1</td>
<td>1.131933</td>
<td>0.21980594</td>
<td>5.150</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>1</td>
<td>0.019387</td>
<td>0.25681328</td>
<td>0.075</td>
<td></td>
</tr>
</tbody>
</table>

Similarly, in component (or factor) analysis, you can replace a set of components by any (non-singular) set of linear combinations of them without changing the variation accounted for.

This process is called rotation

Rotating Factor Solutions

Rotation does not affect the overall goodness of fit; communalities are identical.

The need for rotation arises because factor solutions are interpreted based on the size of loadings.

Rotated and unrotated solutions may differ greatly in interpretation

Ex: Political attitudes toward government policies:

Unrotated | Rotated
---|---
X1: spend more on schools .766 -.232 .783 .163
X2: reduce unemployment .670 -.203 .685 .143
X3: control big business .574 -.174 .587 .123
X4: relax immigration .454 .533 .143 .685
X5: minority job programs .389 .457 .123 .587
X6: expand childcare .324 .381 .102 .489
Simple structure
To make the interpretation of factors as simple as possible:
- Each variable should have non-zero loadings on a small number of factors – preferably 1.
- Each factor should have major loadings on only a few variables – the rest near 0.

Rotation methods
- Purpose:
  - Make the pattern (loadings) more interpretable
  - Increase number of loadings near 1, 0, or -1
  - → simple structure
  - Only for EFA — in CFA, we specify (and test) a hypothesized factor structure directly.

- Orthogonal rotation — factors remain uncorrelated
  - Varimax tries to clean up the columns of the pattern matrix
  - Quartimax tries to clean up the rows of the pattern matrix
  - Equamax tries to do both

- Oblique rotation — factors become correlated, pattern may be simpler
  - Promax — uses result of an orthogonal method and tries to make it better, allowing factors to become correlated.
  - Crawford-Ferguson — a family of methods, allowing weights for row parsimony and column parsimony.

Before CFA, Procrustes (target) rotation was used to test how close you could come to a hypothesized factor pattern.

Analytic rotation methods
These all attempt to reduce ideas of “simple structure” to mathematical functions which can be optimized.
- **Varimax** — Minimize complexity of each factor (# non-zero loadings) → maximize variance of each column of squared loadings.
  - \( \sigma_i^2 = \frac{\sum_j (\lambda_{ij}^2)^2 - (\sum_j \lambda_{ij}^2)^2}{p} \) = variance of col \( j \) of squared loadings
  - Rotate pairs of cols. \( j, j' \) to find angle to make \( \sigma_j^2 + \sigma_{j'}^2 \) large
  - Repeat for all pairs of columns.

- **Orthomax** — Minimize complexity of each variable.
  - Communality = \( h_i^2 = \sum_k \lambda_{ik}^2 \) = constant (unchanged by rotation)
  - → minimize complexity by maximizing variance of squared loadings in each row.
  - \( (h_i^2)^2 = (\sum_j \lambda_{ij}^2)^2 = \sum_j \lambda_{ij}^4 + 2\sum_{m<n} \lambda_{im} \lambda_{jn} \) = constant

- **Equamax** — Tries to achieve simple structure in both rows (variables) and columns (factors).

Example: Holzinger & Swineford 9 abilities data
Maximum likelihood solution, \( k=3 \)

```plaintext
proc factor data=psych9
  Method=ML NFact=3
  round flag=.3
  outstat=FACT / * output data set for rotations * /
  stderr / * get standard errors * /
  rotate=varimax; / * varimax rotation * /
run;
```

Varimax rotated factor solution:

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Visual Perception</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>X2</td>
<td>Cubes</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>X4</td>
<td>Lozenges</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>X6</td>
<td>Paragraph Comprehension</td>
<td>84 *</td>
<td>7</td>
</tr>
<tr>
<td>X7</td>
<td>Sentence Completion</td>
<td>80 *</td>
<td>18</td>
</tr>
<tr>
<td>X9</td>
<td>Word Meaning</td>
<td>78 *</td>
<td>6</td>
</tr>
<tr>
<td>X10</td>
<td>Addition</td>
<td>17</td>
<td>76 *</td>
</tr>
<tr>
<td>X12</td>
<td>Counting Dots</td>
<td>-1</td>
<td>79 *</td>
</tr>
<tr>
<td>X13</td>
<td>Straight-curved Caps</td>
<td>20</td>
<td>52 *</td>
</tr>
</tbody>
</table>
Oblique rotations

- Orthogonal factors are often unnecessarily restrictive; they arise purely from mathematical convenience.
- One can sometimes achieve a simpler structure in the factor loadings by allowing the factors to be correlated.
- For latent variables in a given domain (intelligence, personality, depression), correlated factor often make more sense.

When $\Phi \neq I$, there are two matrices which can be interpreted:

- **Pattern loading ($\Lambda$)**
  $\lambda_{ij} = \text{regression coefficient for } x_i \text{ from factor } \xi_j$

- **Structure loading ($\Gamma = \Lambda\Phi$)**
  $\gamma_{ij} = \text{correlation of } x_i \text{ with factor } \xi_j$

Oblique rotation methods

- **Promax** is the most widely used oblique rotation method
  - Does an initial varimax rotation
  - Transform $\lambda_{ij} \rightarrow \lambda^3_{ij}$: makes loadings closer to 0/1
  - Oblique, least squares rotation to $\Lambda^3$ as target
  - Other oblique rotation methods include the Crawford-Ferguson family, minimizing
    
    $f_{CF} = c_1 \times \text{row parsimony} + c_2 \times \text{col parsimony}$

  - Many people try several rotation methods to see which gives most interpretable result.

Example: Holzinger & Swineford 9 abilities data

**Promax rotation**

For other rotations, use the `OUTSTAT=` data set from a prior run:

```
proc factor data=FACT
  Method=ML NFact=3
  Round flag=.3
  rotate=promax;
run;
```
**Example: Holzinger & Swineford 9 abilities data**

**Promax rotation**

Target matrix defined from initial Varimax:

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>3</td>
<td>2</td>
<td>83 *</td>
</tr>
<tr>
<td>X2</td>
<td>1</td>
<td>0</td>
<td>100 *</td>
</tr>
<tr>
<td>X4</td>
<td>3</td>
<td>0</td>
<td>92 *</td>
</tr>
<tr>
<td>X6</td>
<td>100 *</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>X7</td>
<td>98 *</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>X9</td>
<td>97 *</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>X10</td>
<td>1</td>
<td>100 *</td>
<td>0</td>
</tr>
<tr>
<td>X12</td>
<td>0</td>
<td>93 *</td>
<td>3</td>
</tr>
<tr>
<td>X13</td>
<td>2</td>
<td>40 *</td>
<td>29</td>
</tr>
</tbody>
</table>

**Rotated Factor Pattern (Standardized Regression Coefficients)**

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Visual Perception</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>X2</td>
<td>Cubes</td>
<td>0</td>
<td>-6</td>
</tr>
<tr>
<td>X4</td>
<td>Lozenges</td>
<td>7</td>
<td>-7</td>
</tr>
<tr>
<td>X6</td>
<td>Paragraph Comprehension</td>
<td>86 *</td>
<td>-3</td>
</tr>
<tr>
<td>X7</td>
<td>Sentence Completion</td>
<td>82 *</td>
<td>9</td>
</tr>
<tr>
<td>X9</td>
<td>Word Meaning</td>
<td>80 *</td>
<td>-4</td>
</tr>
<tr>
<td>X10</td>
<td>Addition</td>
<td>13</td>
<td>80 *</td>
</tr>
<tr>
<td>X12</td>
<td>Counting Dots</td>
<td>-14</td>
<td>79 *</td>
</tr>
<tr>
<td>X13</td>
<td>Straight-curved Caps</td>
<td>6</td>
<td>45 *</td>
</tr>
</tbody>
</table>

**Inter-Factor Correlations**

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor1</td>
<td>100 *</td>
<td>27</td>
<td>45 *</td>
</tr>
<tr>
<td>Factor2</td>
<td>27</td>
<td>100 *</td>
<td>38 *</td>
</tr>
<tr>
<td>Factor3</td>
<td>45 *</td>
<td>38 *</td>
<td>100 *</td>
</tr>
</tbody>
</table>

**Procrustes (target) rotations**

- Before CFA, the way to “test” a specific hypothesis for the factor pattern was by rotation to a “target matrix.”
- We can specify a hypothesis by a matrix of 1s and 0s, e.g.,

\[
B = \begin{bmatrix}
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1
\end{bmatrix}
\]

- **Procrustes rotation**: Find a transformation matrix \( T_{k \times k} \) such that \( \Delta T \approx B \) (least squares fit)

  - If \( T \) is orthogonal (\( TT^T = I \)), this is an orthogonal Procrustes rotation
  - Usually \( TT^T \neq I \) → oblique Procrustes rotation
  - Goodness of fit = sum of squares of differences, \( \text{tr}(\Delta T - B)^T(\Delta T - B) \)
Example: Holzinger & Swineford 9 abilities data

Procrustes rotation

Enter the hypothesized target as a matrix of 0/1 (transposed):

title2 'Procrustes rotation: 3 non-overlapping factors';
data hypothesis;
   input _name_ X1 X2 X4 X6 X7 X9 X10 X12 X13;
   list;
   datalines;
   FACTOR1 1 1 1 0 0 0 0 0 0
   FACTOR2 0 0 0 1 1 1 0 0 0
   FACTOR3 0 0 0 0 0 0 1 1 1;
proc factor data=FACT
   rotate=procrustes target=hypothesis
   round flag=.3 PLOT;
run;

Target matrix: Factor pattern:

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Visual Perception</td>
<td>100 *</td>
<td>0</td>
</tr>
<tr>
<td>X2</td>
<td>Cubes</td>
<td>100 *</td>
<td>0</td>
</tr>
<tr>
<td>X4</td>
<td>Lozenges</td>
<td>100 *</td>
<td>0</td>
</tr>
<tr>
<td>X6</td>
<td>Paragraph Comprehension</td>
<td>0</td>
<td>100 *</td>
</tr>
<tr>
<td>X7</td>
<td>Sentence Completion</td>
<td>0</td>
<td>100 *</td>
</tr>
<tr>
<td>X9</td>
<td>Word Meaning</td>
<td>0</td>
<td>100 *</td>
</tr>
<tr>
<td>X10</td>
<td>Addition</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X12</td>
<td>Counting Dots</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>X13</td>
<td>Straight-curved Caps</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Rotated Factor Pattern (Standardized Regression Coefficients)

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Visual Perception</td>
<td>61 *</td>
<td>3</td>
</tr>
<tr>
<td>X2</td>
<td>Cubes</td>
<td>52 *</td>
<td>-2</td>
</tr>
<tr>
<td>X4</td>
<td>Lozenges</td>
<td>66 *</td>
<td>5</td>
</tr>
<tr>
<td>X6</td>
<td>Paragraph Comprehension</td>
<td>3</td>
<td>87 *</td>
</tr>
<tr>
<td>X7</td>
<td>Sentence Completion</td>
<td>-5</td>
<td>83 *</td>
</tr>
<tr>
<td>X9</td>
<td>Word Meaning</td>
<td>7</td>
<td>80 *</td>
</tr>
<tr>
<td>X10</td>
<td>Addition</td>
<td>-29</td>
<td>13</td>
</tr>
<tr>
<td>X12</td>
<td>Counting Dots</td>
<td>9</td>
<td>-16</td>
</tr>
<tr>
<td>X13</td>
<td>Straight-curved Caps</td>
<td>34 *</td>
<td>4</td>
</tr>
</tbody>
</table>

Factor correlations:

<table>
<thead>
<tr>
<th></th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor1</td>
<td>100 *</td>
<td>48 *</td>
<td>34 *</td>
</tr>
<tr>
<td>Factor2</td>
<td>48 *</td>
<td>100 *</td>
<td>31 *</td>
</tr>
<tr>
<td>Factor3</td>
<td>34 *</td>
<td>31 *</td>
<td>100 *</td>
</tr>
</tbody>
</table>

Factors are slightly more correlated here than in Promax
The actual factor scores are obtained by applying the factor score coefficients to the standardized scores, $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$.

$$W_{n \times k} = Z_{n \times p}B_{p \times k}$$

In SAS, use PROC SCORE:

```sas
PROC FACTOR DATA=mydata
SCOREx /* produce factor scores */
OUTSTAT=fact;

PROC SCORE DATA=mydata
SCORE=fact /* uses _TYPE_='SCORE' obs */
OUT=myscores;
```