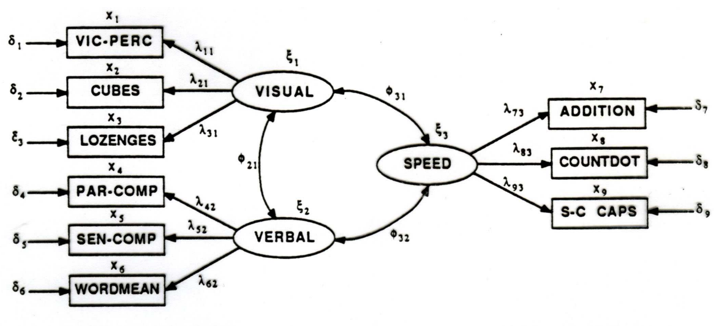
Exploratory and Confirmatory Factor Analysis

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

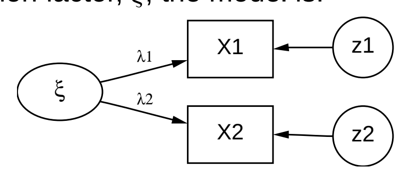
Overview & goals

- 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: $\boldsymbol{R} = \Lambda\Lambda^{\mathsf{T}}$ ("square root" of a matrix)
- Variance of each variable can be decomposed into common variance (communality) and unique variance (uniqueness)

Basic ideas: 1. Linear regression on common factors

- 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, ξ, the model is:

$$\begin{aligned} x_1 &= \lambda_1 \xi + z_1 \\ x_2 &= \lambda_2 \xi + z_2 \end{aligned}$$



- 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:

$$r(z_1, z_2) = 0 \qquad E(z_i) = 0$$

- 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

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

or, in matrix terms:

$$\boldsymbol{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{z} \tag{2}$$

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

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Psi} \tag{3}$$

where:

- $\boldsymbol{\Lambda}_{p \times k}$ = factor pattern (“loadings”)
- $\boldsymbol{\Phi}_{k \times k}$ = matrix of correlations among factors.
- $\boldsymbol{\Psi}$ = diagonal matrix of unique variances of observed variables.

It is usually assumed initially that the factors are uncorrelated ($\boldsymbol{\Phi} = \boldsymbol{I}$), 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***:

$$r(x_i, x_j | \xi) = r(z_i, z_j) = 0$$

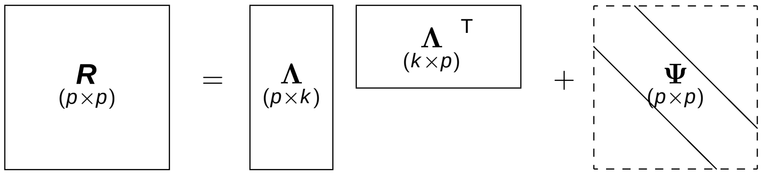
- With one common factor, this has strong implications for the observed correlations:

$$\begin{aligned} r_{12} &= E(x_1, x_2) = E[(\lambda_1 \xi + z_1)(\lambda_2 \xi + z_2)] \\ &= \lambda_1 \lambda_2 \\ r_{13} &= \lambda_1 \lambda_3 \\ \text{ie } r_{ij} &= \lambda_i \lambda_j \end{aligned}$$

- 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:

$$R_{(p \times p)} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_p \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_p \end{bmatrix} + \begin{bmatrix} u_1^2 & & & \\ & u_2^2 & & \\ & & \ddots & \\ & & & u_p^2 \end{bmatrix}$$

- 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, $\boldsymbol{\Lambda}$ and its transpose:



Simple example

Consider the following correlation matrix of 5 measures of "mental ability"

```
x1   1.00  .72  .63  .54  .45
x2    .72 1.00  .56  .48  .40
x3    .63  .56 1.00  .42  .35
x4    .54  .48  .42 1.00  .30
x5    .45  .40  .35  .30 1.00
```

- 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

$$R_{(5\times5)} = \begin{bmatrix} .9 \\ .8 \\ .7 \\ .6 \\ .5 \end{bmatrix} \begin{bmatrix} .9 & .8 & .7 & .6 & .5 \end{bmatrix} + \begin{bmatrix} .19 & & & & \\ & .36 & & & \\ & & .51 & & \\ & & & .64 & \\ & & & & .75 \end{bmatrix}$$

Implications

The implications of this are:

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

$$(\boldsymbol{R} - \Lambda\Lambda^{\mathsf{T}}) = \Psi = \begin{bmatrix} u_1^2 & & \\ & \ddots & \\ & & u_p^2 \end{bmatrix} = \text{a diagonal matrix}$$

- 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.

```
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';
```


Partial linear independence: demonstration

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

	mat_test	eng_test	sci_test	his_test
Mathematics test	1.000	-0.069	0.419	-0.144
English test	-0.069	1.000	-0.097	0.254
Science test	0.419	-0.097	1.000	-0.227
History test	-0.144	0.254	-0.227	1.000

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

	mat_test	eng_test	sci_test	his_test
Mathematics test	1.000	-0.048	-0.015	0.035
English test	-0.048	1.000	0.028	-0.072
Science test	-0.015	0.028	1.000	-0.064
History test	0.035	-0.072	-0.064	1.000

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, $\boldsymbol{\Phi} = \boldsymbol{I}$),

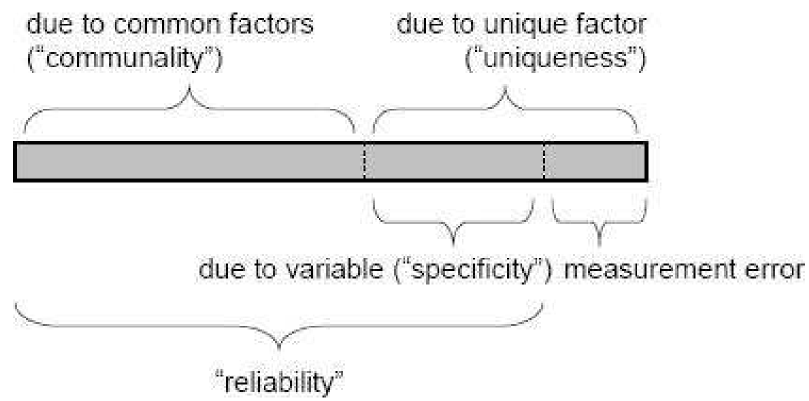
$$\boldsymbol{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{z} \tag{4}$$

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

$$\begin{aligned} var(x_i) &= \underbrace{\lambda_{i1}^2 + \cdots + \lambda_{ik}^2} + var(z_i) \\ &= h_i^2 \text{(communality)} + u_i^2 \text{(uniqueness)} \end{aligned}$$

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:

$$var(x_i) = 1 = \overbrace{h_i^2 + s_i^2}^{\text{reliability}} + e_i^2, \qquad \underbrace{s_i^2 + e_i^2}_{\text{uniqueness}}$$



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

$$\begin{aligned} x_1 &= .8\xi + .6z_1 \\ x_2 &= .6\xi + .8z_1 \end{aligned}$$

we can break down the variance of each variable as:

	var	=	common	+	unique	→(specific	+	error)
x_1:	1	=	.64	+	.36	→ .16	+	.20
x_1:	1	=	.36	+	.64	→ .44	+	.20

Factor Estimation Methods: Basic ideas

Correlations or covariances?

Correlations or covariances?

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

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Psi}$$

or the correlation matrix ($\boldsymbol{R}$), with the model

$$\boldsymbol{R} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{\mathsf{T}} + \boldsymbol{\Psi}$$

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

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 ($\mathbf{\Phi} = \mathbf{I}$)
  - The model becomes
    $$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^{\mathsf{T}} + \mathbf{\Psi}$$
  - If we know (or can estimate) the communalities (= 1 - uniqueness = $1 - \psi_{ii}$), we can factor the "reduced covariance (correlation) matrix", $\mathbf{\Sigma} - \mathbf{\Psi}$
    $$\mathbf{\Sigma} - \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}^{\mathsf{T}} = (\mathbf{U}\mathbf{D}^{1/2})(\mathbf{D}^{1/2}\mathbf{U}^{\mathsf{T}}) \qquad (5)$$
  - In (5), $\mathbf{U}$ is the matrix of eigenvectors of $(\mathbf{\Sigma} - \mathbf{\Psi})$ and $\mathbf{D}$ 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.

$$\text{SMC}_i \equiv R^2_{x_i \mid others} \leq h_i^2 = \text{communality} = 1 - \psi_{ii}$$

Factor Estimation Methods: Basic ideas

Common characteristics

- Most iterative methods cycle between estimating factor loadings (given communality estimates) and estimating the communalities (given factor loadings). The process stops when things don't change too much.
  1. Obtain initial estimate of $\widehat{\mathbf{\Psi}}$
  2. Estimate $\widehat{\mathbf{\Lambda}}$ from eigenvectors/values of $(\mathbf{\Sigma} - \widehat{\mathbf{\Psi}})$
  3. Update estimate of $\widehat{\mathbf{\Psi}}$, return to step 2 if max $|\widehat{\mathbf{\Psi}} - \widehat{\mathbf{\Psi}}_{\text{last}}| < \epsilon$

Factor Estimation Methods: Fit functions

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

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^{\mathsf{T}} + \widehat{\Psi} \quad \approx \quad \mathbf{S}$$

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

Factor Estimation Methods

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

```
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 χ^2 test.

Example: Spearman's ’two-factor’ theory

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    0.00761    0.000  0.16063    0.4950 0.4635 0.3482 0.3179 0.1583
  2    0.00759    0.000  0.00429    0.4953 0.4662 0.3439 0.3203 0.1589
  3    0.00759    0.000  0.00020    0.4954 0.4662 0.3439 0.3203 0.1587
```


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 = 1.1106908068
```

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

Example: Spearman's ’two-factor’ theory

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:

```
            FACTOR1    Common  Unique

TEST1       0.70386      .495    .505     Mathematical judgement
TEST2       0.68282      .466    .534     Controlled association
TEST3       0.58643      .344    .656     Literary interpretation
TEST4       0.56594      .320    .680     Selective judgement
TEST5       0.39837      .159    .841     Spelling
```

- Mathematical judgment 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 Comprehen' 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;
```


Example: Holzinger & Swineford 9 abilities data

"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 ↔ 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 ×100, rounded; FLAG option prints a * next to larger values

Example: Holzinger & Swineford 9 abilities data

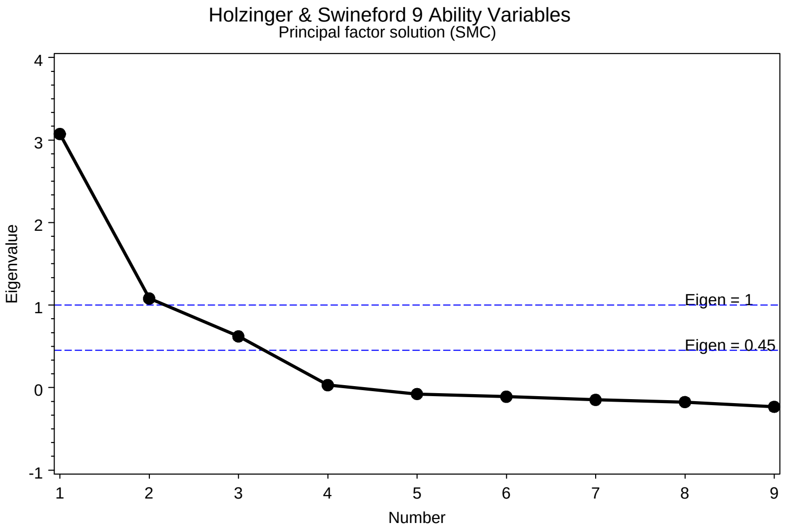
Output: Eigenvalues

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

      Eigenvalue    Difference    Proportion    Cumulative

   1  3.07328008    1.99393040        0.7572        0.7572
   2  1.07934969    0.45916492        0.2659        1.0232
   3  0.62018476    0.58982990        0.1528        1.1760
   4  0.03035486    0.10824191        0.0075        1.1835
   5  -.07788705    0.03243783       -0.0192        1.1643
   6  -.11032489    0.03864959       -0.0272        1.1371
   7  -.14897447    0.02648639       -0.0367        1.1004
   8  -.17546086    0.05650435       -0.0432        1.0572
   9  -.23196521                    -0.0572        1.0000

 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



Example: Holzinger & Swineford 9 abilities data

Initial (unrotated) factor pattern:

Factor Pattern

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

- Interpretation ??

Example: Holzinger & Swineford 9 abilities data

Varimax rotated factor pattern:

Rotated Factor Pattern

		Factor1	Factor2
X1	Visual Perception	39 *	43 *
X2	Cubes	28	25
X4	Lozenges	42 *	32 *
X6	Paragraph Comprehen	83 *	11
X7	Sentence Completion	77 *	17
X9	Word Meaning	80 *	10
X10	Addition	8	59 *
X12	Counting Dots	3	75 *
X13	Straight-curved Caps	30	65 *

- Interpretation ??

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solutions

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

- 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 χ^2 test, and other fit statistics for the $k = 2$ factor ML solution.

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solution, k=2

```
           Significance Tests Based on 145 Observations
                                                           Pr >
            Test                       DF     Chi-Square   ChiSq

H0: No common factors                  36       483.4478   <.0001
HA: At least one common factor
H0: 2 Factors are sufficient           19        61.1405   <.0001
HA: More factors are needed

    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 H_0 : No common factors $\rightarrow H_0 : \boldsymbol{R} = \boldsymbol{I}$: all variables uncorrelated
- $H_0 : k = 2$ 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

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solution, k=3

```
                                                         Pr >
            Test                     DF     Chi-Square   ChiSq

H0: No common factors                36       483.4478   <.0001
HA: At least one common factor
H0: 3 Factors are sufficient         12         9.5453   0.6558
HA: More factors are needed


    Chi-Square without Bartlett's Correction         9.948300
    Akaike's Information Criterion                 -14.051700
    Schwarz's Bayesian Criterion                   -49.772505
    Tucker and Lewis's Reliability Coefficient       1.016458
```

- $H_0 : k = 3$ is *not* rejected here
- The χ^2 test is highly dependent on sample size; other fit measures (later)

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solution, k=3

Unrotated factor solution:

```
                        Factor Pattern

                                 Factor1     Factor2     Factor3

X1     Visual Perception          51 *        18          43 *
X2     Cubes                      32 *         7          39 *
X4     Lozenges                   48 *         8          49 *
X6     Paragraph Comprehen        81 *       -30 *        -8
X7     Sentence Completion        80 *       -21         -16
X9     Word Meaning               77 *       -28          -4
X10    Addition                   40 *        55 *       -37 *
X12    Counting Dots              40 *        72 *        -6
X13    Straight-curved Caps       56 *        43 *        16
```

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solution, k=3

Varimax rotated factor solution:

Rotated Factor Pattern

		Factor1	Factor2	Factor3
X1	Visual Perception	20	19	64 *
X2	Cubes	11	4	50 *
X4	Lozenges	21	7	65 *
X6	Paragraph Comprehen	84 *	7	23
X7	Sentence Completion	80 *	18	17
X9	Word Meaning	78 *	6	25
X10	Addition	17	76 *	-5
X12	Counting Dots	-1	79 *	26
X13	Straight-curved Caps	20	52 *	47 *

Example: Holzinger & Swineford 9 abilities data

Decomposing the variance of each variable

Using the OUTSTAT= data set (communalities) and the reliabilities in the PSYCH9 data set, we can decompose the variance of each variable...

Name	Reliability	Common Variance	Unique Variance	Specific Variance	Error Variance
Visual Perception	0.756	0.482	0.518	0.275	0.244
Cubes	0.568	0.264	0.736	0.304	0.432
Lozenges	0.937	0.475	0.525	0.462	0.064
Paragraph Comprehen	0.750	0.760	0.240	-0.010	0.250
Sentence Completion	0.754	0.702	0.298	0.052	0.246
Word Meaning	0.870	0.677	0.323	0.193	0.130
Addition	0.952	0.607	0.393	0.345	0.048
Counting Dots	0.937	0.682	0.318	0.256	0.063
Straight-curved Caps	0.889	0.525	0.475	0.364	0.111

Assuming $k = 3$ factors: Verbal, Speed, Visual—

- Paragraph comprehension and Sentence completion are better measures of the Verbal factor, even though Word meaning is more reliable.
- Addition and Counting Dots are better measures of Speed; S-C Caps also loads on the Visual factor
- Visual factor: Lozenges most reliable, but Visual Perception has greatest common variance. Cubes has large specific variance and error variance.

Interlude: Significance tests & fit statistics for EFA I

- As we have seen, ML solution $\rightarrow \chi^2 = (N-1)F_{min}$ (large sample test)
- Adding another factor always reduces χ^2, but also reduces df.
  - χ^2/df gives a rough measure of goodness-of-fit, taking # factors into account. Values of $\chi^2/df <= 2$ 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
  - $\Pr(\Delta\chi^2, \Delta df)$ tests if there is a significant improvement in adding one more factor.
- Akaike Information Criterion (AIC): penalizes model fit by $2 \times$ # free parameters

$$AIC = \chi^2 + 2(\#\text{ free parameters}) = \chi^2 + [p(p-1) - 2df]$$

- Bayesian Information Criterion (BIC): greater penalty with larger N

$$BIC = \chi^2 + \log N(\#\text{ free parameters})$$

- AIC and BIC: choose model with the smallest values

Interlude: Significance tests & fit statistics for EFA II

- Tucker-Lewis Index (TLI) : Compares the χ^2/df for the null model ($k = 0$) to the χ^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.

$$RMSEA = \sqrt{\frac{(\chi^2/df) - 1}{N - 1}}$$

  - "Adequate" models have RMSEA $\leq .08$; "good' models: RMSEA $\leq .05$.

Example: Holzinger & Swineford 9 abilities data

Comparing solutions

Collect the test statistics in tables for comparison...

k	Test	ChiSq	DF	Prob ChiSq
0	H0: No common factors	483.4478	36	<.0001
1	H0: 1 Factor is sufficient	172.2485	27	<.0001
2	H0: 2 Factors are sufficient	61.1405	19	<.0001
3	H0: 3 Factors are sufficient	9.5453	12	0.6558

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

k	Chi2/df	diff Chi2	diff DF	Pr > diff	AIC	BIC	TLI
0	13.4291	.	.	.	.	.	.
1	6.3796	311.199	9	0	123.805	43.433	0.5672
2	3.2179	111.108	8	0	25.416	-31.142	0.8216
3	0.7954	51.595	7	<.0001	-14.052	-49.772	1.0165

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;
```


Factor and Component Rotation

- The models using (x1, x2) and (x3, x4) both have the same R^2:

```
    Root MSE       1.36765     R-square      0.6233

                   Parameter      Standard    T for H0:
Variable   DF       Estimate         Error   Parameter=0

INTERCEP    1      -0.234933    0.30603261        -0.768
X1          1       1.151320    0.37796755         3.046
X2          1       1.112546    0.29270456         3.801
  -----------------------------------------------------
    Root MSE       1.36765     R-square      0.6233

                   Parameter      Standard    T for H0:
Variable   DF       Estimate         Error   Parameter=0

INTERCEP    1      -0.234933    0.30603261        -0.768
X3          1       1.131933    0.21980594         5.150
X4          1       0.019387    0.25681328         0.075
```

- 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

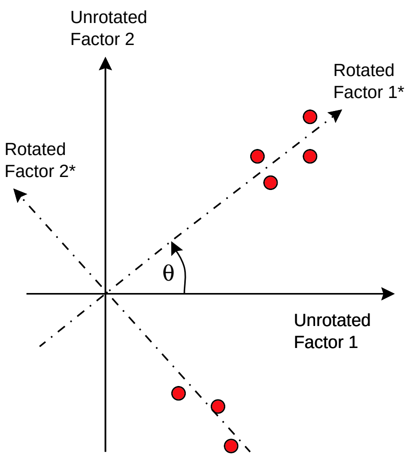
Ex: Political attitudes toward government policies:

	Unrotated F1	Unrotated F2	Rotated F1′	Rotated F2′
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.



$$\Lambda^* = \Lambda \overset{T}{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}$$

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 rotatation — 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.
  - $\sigma_j^2 = [\Sigma_i (\lambda_{ij}^2)^2 - (\Sigma_i \lambda_{ij}^2)/p]/p$ = variance of col j of squared loadings
  - Rotate pairs of cols. $j, j\prime$ to find angle to make $\sigma_j^2 + \sigma_{j\prime}^2$ large
  - Repeat for all pairs of columns.
- ***Orthomax*** — Minimize complexity of each variable.
  - Communality = $h_i^2 = \Sigma_{j=1}^k \lambda_{ij}^2$ = constant (unchanged by rotation)
  - → minimize complexity by maximizing variance of squared loadings in each *row*.

$$(h_i^2)^2 = (\Sigma_j \lambda_{ij}^2)^2 = \underbrace{\Sigma_j \lambda_{ij}^4}_{\text{max}} + 2(\Sigma_{m<n} \lambda_{im}^2 \lambda_{in}^2) = \text{constant}$$

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

Example: Holzinger & Swineford 9 abilities data

Maximum likelihood solution, k=3

```
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;
```

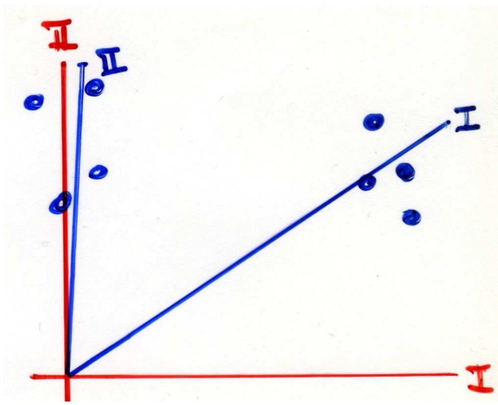
Varimax rotated factor solution:

Rotated Factor Pattern

		Factor1	Factor2	Factor3
X1	Visual Perception	20	19	64 *
X2	Cubes	11	4	50 *
X4	Lozenges	21	7	65 *
X6	Paragraph Comprehen	84 *	7	23
X7	Sentence Completion	80 *	18	17
X9	Word Meaning	78 *	6	25
X10	Addition	17	76 *	-5
X12	Counting Dots	-1	79 *	26
X13	Straight-curved Caps	20	52 *	47 *

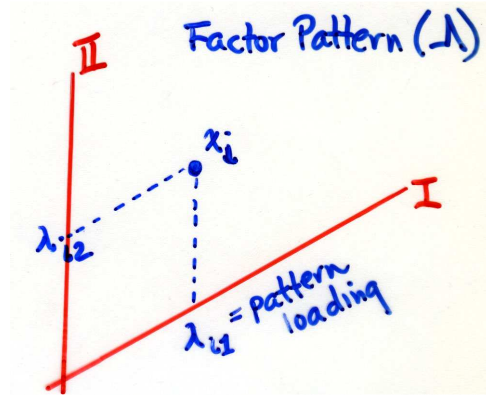
Oblique rotations

- 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.

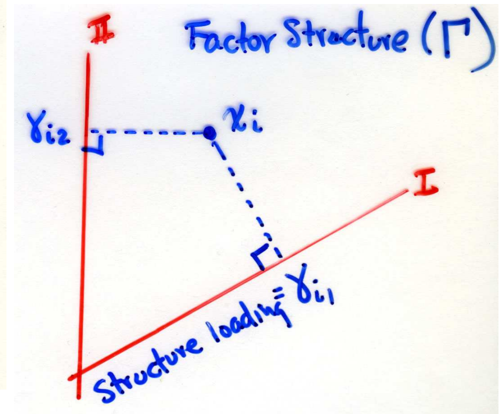


Oblique rotations

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



Pattern loading (Λ)
λ_{ij} = regression coefficient for x_i from factor ξ_j



Structure loading $(\Gamma = \Lambda\Phi)$
γ_{ij} = correlation of x_i with factor ξ_j

Oblique rotation methods

- Promax is the most widely used oblique rotation method
  - Does an initial varimax rotation
  - Transform $\lambda_{ij} \rightarrow \lambda_{ij}^3$: makes loadings closer to 0/1
  - Oblique, least squares rotation to Λ^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:

```
title2 'Promax rotation';
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:

The FACTOR Procedure
Rotation Method: Promax (power = 3)

Target Matrix for Procrustean Transformation

		Factor1	Factor2	Factor3
X1	Visual Perception	3	2	83 *
X2	Cubes	1	0	100 *
X4	Lozenges	3	0	92 *
X6	Paragraph Comprehen	100 *	0	2
X7	Sentence Completion	98 *	1	1
X9	Word Meaning	97 *	0	3
X10	Addition	1	100 *	0
X12	Counting Dots	0	93 *	3
X13	Straight-curved Caps	2	40 *	29

Factor pattern:

Rotated Factor Pattern (Standardized Regression Coefficients)

		Factor1	Factor2	Factor3
X1	Visual Perception	5	7	64 *
X2	Cubes	0	-6	53 *
X4	Lozenges	7	-7	68 *
X6	Paragraph Comprehen	86 *	-3	5
X7	Sentence Completion	82 *	9	-3
X9	Word Meaning	80 *	-4	8
X10	Addition	13	80 *	-23
X12	Counting Dots	-14	79 *	15
X13	Straight-curved Caps	6	45 *	39 *

Factor correlations:

Inter-Factor Correlations

	Factor1	Factor2	Factor3
Factor1	100 *	27	45 *
Factor2	27	100 *	38 *
Factor3	45 *	38 *	100 *

Example: Holzinger & Swineford 9 abilities data

Promax rotation

Factor structure:

Factor Structure (Correlations)

		Factor1	Factor2	Factor3
X1	Visual Perception	36 *	32 *	69 *
X2	Cubes	22	14	51 *
X4	Lozenges	36 *	21	68 *
X6	Paragraph Comprehen	87 *	21	42 *
X7	Sentence Completion	83 *	30	38 *
X9	Word Meaning	82 *	20	42 *
X10	Addition	24	75 *	13
X12	Counting Dots	14	81 *	38 *
X13	Straight-curved Caps	35 *	61 *	59 *

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.,

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

- Procrustes rotation: Find a transformation matrix $\boldsymbol{T}_{k \times k}$ such that $\boldsymbol{\Lambda T} \approx \boldsymbol{B}$ (least squares fit)
  - If $\boldsymbol{T}$ is orthogonal ($\boldsymbol{T}\boldsymbol{T}^{\mathsf{T}} = \boldsymbol{I}$), this is an orthogonal Procrustes rotation
  - Usually $\boldsymbol{T}\boldsymbol{T}^{\mathsf{T}} \neq \boldsymbol{I} \rightarrow$ oblique Procrustes rotation
  - Goodness of fit = sum of squares of differences, $\mathrm{tr}(\boldsymbol{\Lambda T} - \boldsymbol{B})^{\mathsf{T}}(\boldsymbol{\Lambda T} - \boldsymbol{B})$

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;
```


Example: Holzinger & Swineford 9 abilities data

Procrustes rotation

Target matrix: Factor pattern:

Target Matrix for Procrustean Transformation

		Factor1	Factor2	Factor3
X1	Visual Perception	100 *	0	0
X2	Cubes	100 *	0	0
X4	Lozenges	100 *	0	0
X6	Paragraph Comprehen	0	100 *	0
X7	Sentence Completion	0	100 *	0
X9	Word Meaning	0	100 *	0
X10	Addition	0	0	100 *
X12	Counting Dots	0	0	100 *
X13	Straight-curved Caps	0	0	100 *

Factor pattern:

Rotated Factor Pattern (Standardized Regression Coefficients)

		Factor1	Factor2	Factor3
X1	Visual Perception	61 *	3	15
X2	Cubes	52 *	-2	0
X4	Lozenges	66 *	5	1
X6	Paragraph Comprehen	3	87 *	-3
X7	Sentence Completion	-5	83 *	9
X9	Word Meaning	7	80 *	-4
X10	Addition	-29	13	80 *
X12	Counting Dots	9	-16	83 *
X13	Straight-curved Caps	34 *	4	51 *

Factor correlations:

Inter-Factor Correlations

	Factor1	Factor2	Factor3
Factor1	100 *	48 *	34 *
Factor2	48 *	100 *	31 *
Factor3	34 *	31 *	100 *

Factors are slightly more correlated here than in Promax

Factor Scores

- Factor scores represent the values of individual cases on the latent factor variables.
- Uses: classification, cluster analysis, regression, etc. based on results of factor analysis.
- Factor scores (unlike component scores) cannot be computed exactly, but must be estimated.
  - Reason: The unique factors (by definition) are uncorrelated with everything else.
  - Therefore a linear combination of the variables cannot be perfectly correlated with any common factor.
- Most factor analysis programs (`PROC FACTOR`, LISREL, EQS) estimate factor score coefficients by multiple regression, using the usual formula for standardized regression coefficients:

$$\boldsymbol{B}_{p\times k} = (\boldsymbol{R}_{xx})^{-1}\boldsymbol{R}_{x\xi} = \boldsymbol{R}_{xx}^{-1}\widehat{\Lambda}\widehat{\Phi}$$

Factor Scores

- The actual factor scores are obtained by applying the factor score coefficients to the standardized scores, $z_{ij} = (x_{ij} - \bar{x}_j)/s_j$.

$$\boldsymbol{W}_{n \times k} = \boldsymbol{Z}_{n \times p} \boldsymbol{B}_{p \times k}$$

- In SAS, use `PROC SCORE`:

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

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