Factor Analysis

Factor Analysis

Factor analysis is a statistical method used to study the dimensionality of a set of variables. In factor analysis, latent variables represent unobserved constructs and are referred to as factors or dimensions.

- Exploratory Factor Analysis (EFA)
  Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and on the linear relationships between the observed variables and the factors but only specifies the number of latent variables

Factor Analysis (Continued)

- Confirmatory Factor Analysis (CFA)

  Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population – characterized by allowing restrictions on the parameters of the model

Applications Of Factor Analysis

- Personality and cognition in psychology
  - Child Behavior Checklist (CBCL)
  - MMPI
- Attitudes in sociology, political science, etc.
- Achievement in education
- Diagnostic criteria in mental health

The Factor Analysis Model

The factor analysis model expresses the variation and covariation in a set of observed continuous variables y (j = 1 to p) as a function of factors η (k = 1 to m) and residuals ε (j = 1 to p). For person i,

$$y_{i1} = \nu_1 + \lambda_{11}\,\eta_{i1} + \lambda_{12}\,\eta_{i2} + \ldots + \lambda_{1k}\,\eta_{ik} + \ldots + \lambda_{1m}\,\eta_{im} + \varepsilon_{i1}$$

.

.

.

$$y_{ij} = \nu_j + \lambda_{j1}\,\eta_{i1} + \lambda_{j2}\,\eta_{i2} + \ldots + \lambda_{jk}\,\eta_{ik} + \ldots + \lambda_{jm}\,\eta_{im} + \varepsilon_{ij}$$

.

.

.

$$y_{ip} = \nu_p + \lambda_{p1}\,\eta_{i1} + \lambda_{p2}\,\eta_{i2} + \ldots + \lambda_{pk}\,\eta_{ik} + \ldots + \lambda_{pm}\,\eta_{im} + \varepsilon_{ip}$$

The Factor Analysis Model (Continued)

where

v_j are intercepts

λ_{jk} are factor loadings

η_{ik} are factor values

ε_{ij} are residuals with zero means and correlations of zero with the factors

The Factor Analysis Model (Continued)

In matrix form,

$$\boldsymbol{y}_\mathrm{i} = \boldsymbol{v} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_\mathrm{i} + \boldsymbol{\varepsilon}_\mathrm{i}\,,$$

where

$\boldsymbol{v}$	is the vector of intercepts v_j,
$\boldsymbol{\Lambda}$	is the matrix of factor loadings λ_jk,
$\boldsymbol{\Psi}$	is the matrix of factor variances/covariances, and
$\boldsymbol{\Theta}$	is the matrix of residual variances/covariances

with the population covariance matrix of observed variables $\boldsymbol{\Sigma}$,

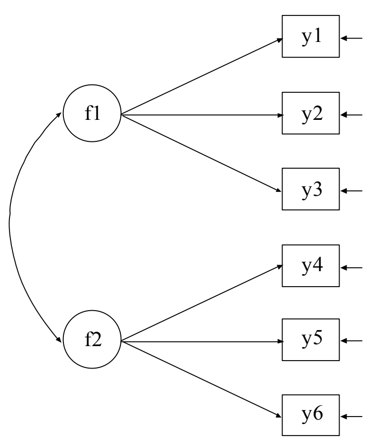
$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\, \boldsymbol{\Psi}\, \boldsymbol{\Lambda}' \;+\; \boldsymbol{\Theta}.$$

Factor Analysis Terminology

- Factor pattern: Λ
- Factor structure: $\Lambda * \Psi$, correlations between items and factors
- Heywood case: $\theta_{jj} < 0$
- Factor scores: $\hat{\eta}_i$
- Factor determinacy: quality of factor scores; correlation between η_i and $\hat{\eta}_i$

53

A Two-Factor Model



- Squares or rectangles represent observed variables
- Circles or ovals represent factors or latent variables
- Uni-directional arrows represent regressions or residuals
- Bi-directional arrows represent correlations/covariances

54

Formulas For The Path Diagram

$y_{i1} = v_1 + \lambda_{11} f_{i1} + 0\, f_{i2} + \varepsilon_{i1}$
$y_{i2} = v_2 + \lambda_{21} f_{i1} + 0\, f_{i2} + \varepsilon_{i2}$
$y_{i3} = v_3 + \lambda_{31} f_{i1} + 0\, f_{i2} + \varepsilon_{i3}$
$y_{i4} = v_4 + 0\, f_{i1} + \lambda_{42} f_{i2} + \varepsilon_{i4}$
$y_{i5} = v_5 + 0\, f_{i1} + \lambda_{52} f_{i2} + \varepsilon_{i5}$
$y_{i6} = v_6 + 0\, f_{i1} + \lambda_{62} f_{i2} + \varepsilon_{i6}$

Elements of $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\, \boldsymbol{\Psi}\, \boldsymbol{\Lambda}' + \boldsymbol{\Theta}$:

Variance of $y_1 = \sigma_{11} = \lambda_{11}^2\, \psi_{11} + \theta_{11}$

Covariance of $y_1, y_2 = \sigma_{21} = \lambda_{11}\, \psi_{11}\, \lambda_{21}$

Covariance of $y_1, y_4 = \sigma_{41} = \lambda_{11}\, \psi_{21}\, \lambda_{42}$

55

Recommendations For Using Factor Analysis In Practice

Issues

- History of EFA versus CFA
- Can hypothesized dimensions be found?
  - Validity of measurements

A Possible Research Strategy For Instrument Development

1. Pilot study 1
   - Small n, EFA
   - Revise, delete, add items

56

Recommendations For Using Factor Analysis In Practice (Continued)

2. Pilot study 2
   - Small n, EFA
   - Formulate tentative CFA model
3. Pilot study 3
   - Larger n, CFA
   - Test model from Pilot study 2 using random half of the sample
   - Revise into new CFA model
   - Cross-validate new CFA model using other half of data
4. Large scale study, CFA
5. Investigate other populations

57

Exploratory Factor Analysis

58

Exploratory Factor Analysis (EFA)

Used to explore the dimensionality of a measurement instrument by finding the smallest number of interpretable factors needed to explain the correlations among a set of variables – exploratory in the sense that it places no structure on the linear relationships between the observed variables and the factors but only specifies the number of latent variables

- Find the number of factors
- Determine the quality of a measurement instrument
  - Identify variables that are poor factor indicators
  - Identify factors that are poorly measured

59

Holzinger-Swineford Data

The data are taken from the classic 1939 study by Karl J. Holzinger and Frances Swineford. Twenty-six tests intended to measure a general factor and five specific factors were administered to seventh and eighth grade students in two schools, the Grant-White School (n = 145) and Pasteur School (n = 156). Students from the Grant-White School came from homes where the parents were American-born. Students from the Pasteur School came from the homes of workers in factories who were foreign-born.

Data for the analysis include nineteen test intended to measure four domains: spatial ability, verbal ability, speed, and memory. Data from the 145 students from the Grant-White School are used.

60

Holzinger-Swineford Variables

- SPATIAL TESTS
  - Visual perception test
  - Cubes
  - Paper form board
  - Lozenges
- VERBAL TESTS
  - General information
  - Paragraph comprehension
  - Sentence completion
  - Word classification
  - Word meaning

61

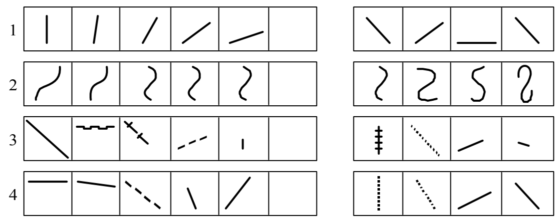
Holzinger-Swineford Variables (Continued)

- SPEED TESTS
  - Add
  - Code
  - Counting groups of dots
  - Straight and curved capitals
- MEMORY
  - Word recognition
  - Number recognition
  - Figure recognition
  - Object-number
  - Number-figure
  - Figure-word

62

Examples Of Holzinger-Swineford Variables

Test 1 Visual-Perception Test



Test 5 General Information

In each sentence below you have four choices for the last word, but only one is right. From the last four words of each sentence, select the right one and underline it.
EXAMPLE: Men see with their ears, nose, <u>eyes</u>, mouths.

1. Pumpkins grow on bushes, trees, vines, shrubs.
2. Coral comes from reefs, mines, trees, tusks.
3. Sugar cane grows mostly in Montana, Texas, Illinois, New York

63

Examples Of Holzinger-Swineford Variables (Continued)

Test 17 Object-Number

Name______________

Date______________

Here is a list of objects. Each one has a number. Study the list so that you can remember the number of each object.

Object	Number
apple	29
brush	71
candy	58
chair	44
cloud	53
dress	67
flour	15
grass	32
heart	86

After each object, write the number that belongs to it.

Object	Number
pupil	
chair	
house	
sugar	
flour	
river	
apple	
match	
train	

64

Sample Correlations For Holzinger-Swineford Data

	VISUAL	CUBES	PAPER	LOZENGES	GENERAL
VISUAL					
CUBES	.326				
PAPER	.372	.190			
LOZENGES	.449	.417	.366		
GENERAL	.328	.275	.309	.381	
PARAGRAP	.342	.228	.260	.328	.622
SENTENCE	.309	.159	.266	.287	.654
WORDC	.326	.156	.334	.380	.574
WORDM	.317	.195	.260	.347	.720
ADDITION	.104	.066	.128	.075	.314
CODE	.306	.151	.248	.181	.342
COUNTING	.308	.168	.198	.239	.210
STRAIGHT	.487	.248	.389	.373	.343
WORDR	.130	.082	.250	.161	.261
NUMBERR	.223	.135	.186	.205	.219
FIGURER	.419	.289	.307	.289	.177
OBJECT	.169	.011	.128	.139	.213
NUMBERF	.364	.264	.259	.353	.259
FIGUREW	.267	.110	.155	.180	.196

65

Sample Correlations For Holzinger-Swineford Data (Continued)

	PARAGRAP	SENTENCE	WORDC	WORDM	ADDITION
SENTENCE	.719				
WORDC	.520	.633			
WORDM	.714	.685	.537		
ADDITION	.209	.254	.297	.179	
CODE	.360	.248	.294	.287	.468
COUNTING	.104	.198	.290	.121	.587
STRAIGHT	.314	.356	.405	.272	.418
WORDR	.286	.233	.243	.250	.157
NUMBERR	.249	.157	.170	.213	.150
FIGURER	.288	.201	.299	..236	.137
OBJECT	.276	.251	.271	.285	.301
NUMBERF	.167	.176	.258	.213	.320
FIGUREW	.251	.241	.261	.277	.199

66

Sample Correlations For Holzinger-Swineford Data (Continued)

	CODE	COUNTING	STRAIGHT	WORDR	NUMBERR
COUNTING	.422				
STRAIGHT	.527	.528			
WORDR	.324	.130	.193		
NUMBERR	.238	.163	.138	.387	
FIGURER	.314	.128	.277	.382	.313
OBJECT	.357	.278	.191	.372	.346
NUMBERF	.346	.347	.325	.199	.318
FIGUREW	.290	.108	.252	.219	.183

	FIGURER	OBJECT	NUMBERF	FIGUREW
OBJECT	.339			
NUMBERF	.355	.452		
FIGUREW	.254	.327	.358	

EFA Model Estimation

Estimators

In EFA, a correlation matrix is analyzed.

- ULS – minimizes the residuals, observed minus estimated correlations
  - Fast
  - Not fully efficient
- ML – minimizes the differences between matrix summaries (determinant and trace) of observed and estimated correlations
  - Computationally more demanding
  - Efficient

EFA Model Indeterminacies And Rotations

A model that is identified has only one set of parameter values. To be identified, an EFA model must have m^2 restrictions on factor loadings, variances, and covariances. There are an infinite number of possible ways to place the restrictions. In software, restrictions are placed in two steps.

Step 1 – Mathematically convenient restrictions

- m(m+1)/2 come from fixing the factor variances to one and the factor covariances to zero
- m(m-1)/2 come from fixing (functions of) factor loadings to zero
  - ULS – $\Lambda'\Lambda$ diagonal
  - ML – $\Lambda' \Theta^{-1} \Lambda$ diagonal
  - General approach – fill the upper right hand corner of lambda with zeros

69

EFA Model Indeterminacies And Rotations (Continued)

Step 2 – Rotation to interpretable factors

Starting with a solution based on mathematically convenient restrictions, a more interpretable solution can be found using a rotation. There are two major types of rotations: orthogonal (uncorrelated factors) and oblique (correlated factors).

- Do an orthogonal rotation to maximize the number of factor loadings close to one and close to zero
- Do an oblique rotation of the orthogonal solution to obtain factor loadings closer to one and closer to zero

70

New EFA Features In Mplus Version 5

- Several new rotations including Quartimin and Geomin
- Standard errors for rotated loadings and factor correlations
- Non-normality robust standard errors and chi-square tests of model fit
- Modification indices for residual correlations
- Maximum likelihood estimation with censored, categorical, and count variables
- Exploratory factor analysis for complex survey data (stratification, clustering, and weights)

  TYPE = COMPLEX EFA # #;
- Exploratory factor mixture analysis with class-specific rotations

  TYPE = MIXTURE EFA # #;
- Two-level exploratory factor analysis for continuous and categorical variables with new rotations and standard errors, including unrestricted model for either level

  TYPE = TWOLEVEL EFA # # UW # # UB;

71

Determining The Number Of Factors That Explain The Correlations Among Variables

Descriptive Values

- Eigenvalues
- Residual Variances

Tests Of Model Fit

- RMSR – average residuals for the correlation matrix – recommend to be less than .05

72

Determining The Number Of Factors That Explain The Correlations Among Variables (Continued)

- Chi-Square – tests that the model does not fit significantly worse than a model where the variables correlate freely – p-values greater than .05 indicate good fit

  H_0: Factor model
  H_1: Unrestricted correlations model
  If $p < .05$, H_0 is rejected
  Note: We want large p

- RMSEA – function of chi-square – test of close fit – value less than .05 recommended

$$RMSEA = \sqrt{\max[(\chi^2 / n\,d - 1/n), 0]}\sqrt{G}$$

where d is the number of degrees of freedom of the model and G is the number of groups.

73

Steps In EFA

- Carefully develop or use a carefully developed set of variables that measure specific domains

- Determine the number of factors
  - Descriptive values
    - Eigenvalues
    - Residual variances
  - Tests of model fit
    - RMSR
    - Chi-square
    - RMSEA

74

Steps In EFA (Continued)

- Interpret the factors
- Determine the quality of the variables measuring the factors
  - Size loadings
  - Cross loadings
- Determine the quality of the factors
  - Number of variables that load on the factor
  - Factor determinacy – correlation between the estimated factor score and the factor
- Eliminate poor variables and factors and repeat EFA steps

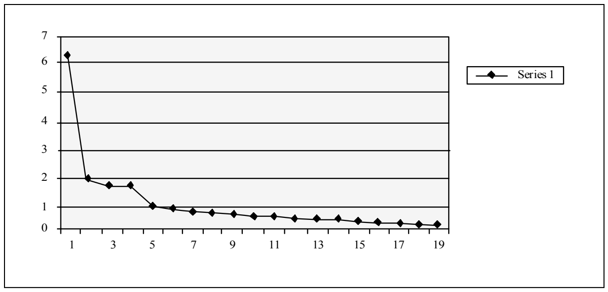
Input For Holzinger-Swineford EFA

```
TITLE:      EFA on 19 variables from Holzinger and Swineford
            (1939)

DATA:       FILE IS holzall.dat;
            FORMAT IS f3,2f2,f3,2f2/3x,13(1x,f3)/3x,11(1x,f3);

VARIABLE:   NAMES ARE id female grade agey agem school visual
            cubes paper lozenges general paragrap sentence wordc
            wordm addition code counting straight wordr numberr
            figurer object numberf figurew deduct numeric
            problemr series arithmet;

            USEV ARE visual cubes paper lozenges general
            paragrap sentence wordc wordm addition code counting
            straight wordr numberr figurer object numberf
            figurew;

            USEOBS IS school EQ 0;

ANALYSIS:   TYPE=EFA 1 8; ESTIMATOR = ML;
```

Determine The Number Of Factors

Examine The Eigenvalues

- Number greater than one
- Scree plot



77

Determine The Number Of Factors (Continued)

Examine The Fit Measures And Residual Variances
(ML, n = 145)

Factors	Chi-Square x^2 df p	RMSEA	RMSR	Negative Res. Var.
1	469.81 (152) .000	.120	.1130	no
2	276.44 (134) .000	.086	.0762	no
3	188.75 (117) .000	.065	.0585	no
4	110.34 (101) .248	.025	.0339	no
5	82.69 (86) .581	.000	.0280	no
6	no. conv.			
7	no. conv.			
8	no. conv.			

78

Interpret The Factors

- Examine factor loadings for the set of possible solutions
- Determine if factors support theory

Output Excerpts Holzinger-Swineford EFA Using 19 Variables

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL/ MEMORY 1	SPEED 2	VERBAL 3
VISUAL	**.740**	-.087	-.002
CUBES	**.522**	-.118	-.008
PAPER	**.508**	-.028	.058
LOZENGES	**.650**	-.153	.092
GENERAL	.043	.084	**.745**
PARAGRAP	.090	-.066	**.803**
SENTENCE	-.052	.046	**.851**
WORDC	.144	.136	**.547**
WORDM	.061	-.092	**.853**
ADDITION	-.257	**.923**	.073
CODE	.223	**.482**	.054
COUNTING	.112	**.728**	-.149
STRAIGHT	.389	.405	.013

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL/ MEMORY 1	SPEED 2	VERBAL 3
WORDR	**.284**	.063	.128
NUMBERR	**.374**	.038	.022
FIGURER	**.666**	-.072	-.063
OBJECT	.214	.270	.086
NUMBERF	**.534**	.237	-.146
FIGUREW	**.302**	.090	.094

Promax Factor Correlations

	1	2	3
1	1.000		
2	.536	1.000	
3	.539	.379	1.000

81

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
VISUAL	**.713**	.027	.008	.005
CUBES	**.541**	-.051	.007	-.050
PAPER	**.466**	.047	.070	.022
LOZENGES	**.650**	-.028	.106	-.062
GENERAL	.094	-.043	**.749**	.083
PARAGRAP	.040	.107	**.791**	-.092
SENTENCE	.002	-.050	**.846**	.052
WORDC	.155	.014	**.550**	.146
WORDM	.022	.078	**.840**	-.107
ADDITION	-.203	.108	.081	**.785**
CODE	.087	.289	.055	.419
COUNTING	.179	-.024	-.132	**.760**
STRAIGHT	.479	-.094	.033	.486

82

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
WORDR	-.037	**.551**	.098	-.052
NUMBERR	.062	**.532**	-.006	-.064
FIGURER	.368	.504	-.086	-.141
OBJECT	-.205	**.736**	.042	.119
NUMBERF	.275	.446	-.154	.178
FIGUREW	.082	**.376**	.080	.019

Promax Factor Correlations

	1	2	3	4
1	1.000			
2	.468	1.000		
3	.468	.421	1.000	
4	.360	.388	.325	1.000

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Varimax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
VISUAL	**.666**	.194	.183	.143
CUBES	**.487**	.072	.117	.042
PAPER	**.455**	.170	.191	.126
LOZENGES	**.608**	.135	.241	.068
GENERAL	.230	.133	**.743**	.183
PARAGRAP	.195	.244	**.772**	.038
SENTENCE	.158	.119	**.808**	.146
WORDC	.267	.174	**.589**	.242
WORDM	.180	.219	**.806**	.021
ADDITION	-.062	.189	.177	**.754**
CODE	.191	.367	.197	.486
COUNTING	.224	.110	.034	**.748**
STRAIGHT	.489	.103	.206	.545

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Varimax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3	SPEED 4
WORDR	.077	**.522**	.184	.064
NUMBERR	.144	**.506**	.103	.054
FIGURER	.398	.524	.081	.021
OBJECT	-.036	**.673**	.155	.229
NUMBERF	.326	.484	.034	.293
FIGUREW	.160	**.392**	.173	.118

85

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 5 Factor Solution

	SPATIAL 1	2	VERBAL 3	MEMORY 4	SPEED 5
VISUAL	**.613**	.211	.006	.050	.011
CUBES	**.552**	-.044	.029	-.028	-.044
PAPER	**.399**	.187	.058	.057	.021
LOZENGES	**.696**	-.070	.129	-.018	-.051
GENERAL	.137	-.094	**.771**	-.042	.096
PARAGRAP	-.006	.131	**.772**	.110	-.100
SENTENCE	-.010	.083	**.826**	-.049	.049
WORDC	.149	.061	**.543**	.018	.145
WORDM	.050	-.075	**.845**	.081	-.097
ADDITION	-.185	.032	.095	.113	**.765**
CODE	-.009	.291	.028	.295	.413
COUNTING	.167	.098	-.112	-.014	**.744**
STRAIGHT	.374	.474	-.013	-.124	.497

86

Output Excerpts Holzinger-Swineford EFA Using 19 Variables (Continued)

Promax Rotated Loadings – 5 Factor Solution

	SPATIAL 1	2	VERBAL 3	MEMORY 4	SPEED 5
WORDR	-.085	.098	.082	**.552**	-.060
NUMBERR	.071	-.094	.010	**.543**	-.059
FIGURER	.286	.144	-.101	.533	-.150
OBJECT	-.160	-.163	.056	**.745**	.124
NUMBERF	.358	-.256	-.126	.502	.195
FIGUREW	.074	.004	.075	**.386**	.026

Promax Factor Correlations

	1	2	3	4	5
1	1.000				
2	.206	1.000			
3	.415	.287	1.000		
4	.425	.335	.424	1.000	
5	.305	.035	.275	.343	1.000

87

Output Excerpts Using 19 Variables Quartimin Rotated Loadings

	SPATIAL	MEMORY	VERBAL	SPEED
VISUAL	**0.646**	0.076	0.092	0.050
CUBES	**0.488**	-0.010	0.064	-0.018
PAPER	**0.422**	0.077	0.128	0.053
FLAGS	**0.585**	0.017	0.178	-0.021
GENERAL	0.058	-0.049	**0.773**	0.093
PARAGRAP	0.019	0.088	**0.810**	-0.079
SENTENCE	-0.028	-0.064	**0.860**	0.056
WORDC	0.121	0.014	**0.584**	0.159
WORDM	0.000	0.058	**0.855**	-0.095
ADDITION	-0.196	0.093	0.100	**0.769**
CODE	0.084	0.283	0.100	0.431
COUNTING	0.149	-0.001	-0.081	**0.761**
STRAIGHT	0.418	-0.051	0.105	0.507

88

Output Excerpts Using 19 Variables Quartimin Rotated Loadings (Continued)

	SPATIAL	MEMORY	VERBAL	SPEED
WORDR	-0.006	**0.517**	0.124	-0.034
NUMBERR	0.086	**0.509**	0.028	0.041
FIGURER	0.366	0.505	-0.023	-0.100
OBJECT	-0.150	**0.683**	0.065	0.131
NUMBERF	0.274	0.447	-0.092	0.207
FIGUREW	0.091	**0.361**	0.113	0.037

QUARTIMIN FACTOR CORRELATIONS

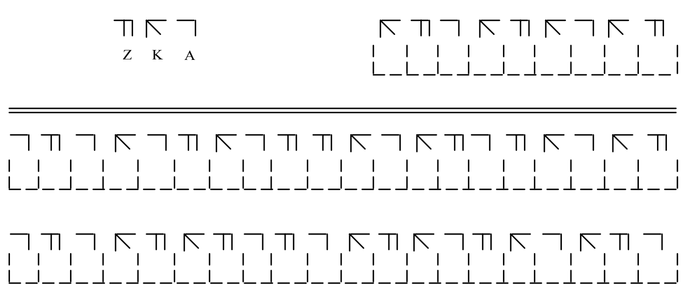
SPATIAL	1.000			
MEMORY	0.289	1.000		
VERBAL	0.371	0.377	1.000	
SPEED	0.266	0.323	0.290	1.000

Determine The Quality Of The Variables

Examine Cross Loadings

Four variables have cross loadings:

- Code (Speed) – loads on Memory and Speed factors
    - Requires matching letters to a set of figures



Determine The Quality Of The Variables (Continued)

- Straight (Speed) – loads on Spatial and Speed factors
  - Requires deciding if a letter consists of entirely straight lines or has curved lines

T	F	G	O	E	Y	Q	D	N	L	P	V	M	J	O

T	Z	F	S	H	J	K	L	D	T	F	U	C	T	U

E	H	K	B	L	S	J	R	R	O	P	T	K	E	M

N	T	D	C	U	S	Z	Y	H	O	T	L	F	C	N

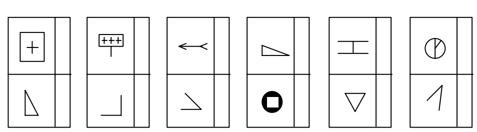
U	D	Q	J	F	D	U	H	L	V	P	E	S	U	Y

D	J	L	H	T	D	C	J	Q	P	R	E	J	L	D

Determine The Quality Of The Variables (Continued)

- Figure (Memory) – loads on Spatial and Memory
  - Requires remembering a set of figures

Put a check mark (√) in the space after each figure that was on the study sheet. Do not put a check after any figure that you have not studied.



Determine The Quality Of The Variables (Continued)

- Numberf (Memory) – loads on Spatial and Memory
  - Requires remembering a figure and associating it with a number

Here is a list of numbers. Each has a figure, or picture, with it. Study the list so that you can remember the figure that belongs with each number.

After each number draw the figure that belongs with it

Number	Figure	Number	Figure
52		17	
74		65	
12		37	

Deleting Four Items That Have Cross Loadings

Output Excerpts Holzinger-Swineford EFA Using 15 Variables

Promax Rotated Loadings – 4 Factor Solution

	SPATIAL 1	MEMORY 2	SPEED 3	VERBAL 4
VISUAL	**.590**	.040	.078	.034
CUBES	**.566**	-.089	.007	-.012
PAPER	**.419**	.104	.029	.056
LOZENGES	**.734**	-.012	-.014	.028
GENERAL	.128	-.037	.050	**.739**
PARAGRAP	.031	.108	-.118	**.792**
SENTENCE	-.041	-.044	.043	**.878**
WORDC	.132	.008	.158	**.568**
WORDM	.043	.060	-.109	**.826**
ADDITION	-.161	.087	**.698**	.127
COUNTING	.200	-.012	**.841**	-.147
WORDR	.000	**.613**	-.066	.023
NUMBERR	.133	**.585**	-.044	-.104
OBJECT	-.127	**.646**	.144	.019
FIGUREW	.066	**.350**	-.004	.096

Output Excerpts Holzinger-Swineford EFA Using 15 Variables (Continued)

Note that factor structure is maintained and that speed has only two indicators

Promax Factor Correlations

	1	2	3	4
1	1.000			
2	.386	1.000		
3	.258	.355	1.000	
4	.495	.478	.309	1.000

Output Excerpts Holzinger-Swineford EFA Using 15 Variables (Continued)

Estimated Error Variances

VISUAL	CUBES	PAPER	LOZENGES	GENERAL
.576	.714	.738	.452	.346

PARAGRAP	SENTENCE	WORDC	WORDM	ADDITION
.306	.274	.488	.279	.444

COUNTING	WORDR	NUMBERR	OBJECT	FIGUREW
.257	.635	.657	.541	.809

Tests Of Model Fit

Chi-square	48.636 (51) .5681
RMSEA	.000
RMSR	.0275

97

Deleting A Factor With Only Two Items

98

Output Excerpts Holzinger-Swineford EFA Using 13 Variables

Promax Rotated Loadings – 3 Factor Solution

	SPATIAL 1	MEMORY 2	VERBAL 3
VISUAL	**0.577**	0.061	0.035
CUBES	**0.602**	-0.114	-0.039
PAPER	**0.434**	0.115	0.033
LOZENGES	**0.765**	-0.032	-0.010
GENERAL	0.152	-0.029	**0.728**
PARAGRAP	0.009	0.080	**0.777**
SENTENCE	-0.060	-0.015	**0.891**
WORDC	0.149	0.065	**0.572**
WORDM	0.015	0.037	**0.816**
WORDR	-0.023	**0.611**	0.010
NUMBERR	0.116	**0.573**	-0.114
OBJECT	-0.127	**0.678**	0.043
FIGUREW	0.081	**0.351**	0.076

Output Excerpts Holzinger-Swineford EFA Using 13 Variables (Continued)

Promax Factor Correlations

	1	2	3
1	1.000		
2	0.436	1.000	
3	0.540	0.486	1.000

Tests Of Model Fit

Chi-square	39.027 (42) .6022
RMSEA	0.000
RMSR	0.0301

Practical Issues Related To EFA

Choice Of Variables – results can be influenced by the set of variables used.

- EFA requires a set of variables that has been carefully developed to measure certain domains, not just any set of variables.
- Number of factors can be influenced by the number of variables per factor.
- Similar number of variables per factor – at least four or five variables per factor is recommended.

Sample Size

- Advantages of large sample size
  - Sample correlations have smaller sampling variability—closer to population values
  - Reduces Heywood cases, negative residual variances

Practical Issues Related To EFA (Continued)

  - Several observations per estimated parameter are recommended
- Advantages of small sample size
  - Can avoid heterogeneity
  - Can avoid problems with sensitivity of chi-square

Size Of Factor Loadings – no general rules

Elimination Of Factors/Variables

- Drop variables that poorly measure factors
- Drop factors that are poorly measured

Maximum Number Of Factors That Can Be Extracted

$a \le b$ where a = number of parameters to be estimated (H_0)
b = number of variances/covariances (H_1)

$a = p\,m + m\,(m+1)/2 + p - m^2$
$\Lambda \qquad \Psi \qquad \Theta$

$b = p\,(p + 1)/2$

where p = number of observed variables
m = number of factors

Example: $p = 5$ which gives $b = 15$
$m = 1$: $a = 10$
$m = 2$: $a = 14$
$m = 3$: $a = 17$

Even if $a \le b$, it may not be possible to extract m factors due to Heywood cases.

Sample Size

- Stability of sample correlations
  - $V(r) = (1 - \rho^2)^2/n$
  - Example: $\rho = 0.5$, $s.d. = 0.1$, $n = 56$
- Stability of estimates
  - n larger than the number of parameters
  - Example: 5 dimensions hypothesized, 5 items per dimension, number of EFA parameters = 140, n = 140-1400 in order to have 1-10 observations per parameter
- Monte Carlo studies (Muthén & Muthén, 2002)

Further Readings On EFA

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. Multivariate Behavioral Research, 36, 111-150.

Cudeck, R. & O'Dell, L.L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. Psychological Bulletin, 115, 475-487.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4, 272-299.

Gorsuch, R.L. (1983). Factor analysis. 2nd edition. Hillsdale, N.J.: Lawrence Erlbaum.

Kim, J.O. & Mueller, C.W. (1978). An introduction to factor analysis: what it is and how to do it. Sage University Paper series on Quantitative Applications in the Social Sciences, No 07-013. Beverly Hills, CA: Sage.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC: American Psychological Association.

Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA)

Used to study how well a hypothesized factor model fits a new sample from the same population or a sample from a different population. CFA is characterized by allowing restrictions on factor loadings, variances, covariances, and residual variances.

- See if factor models fits a new sample from the same population – the confirmatory aspect
- See if the factor models fits a sample from a different population – measurement invariance
  - Study the properties of individuals by examining factor variances, and covariances
    - Factor variances show the heterogeneity in a population
    - Factor correlations show the strength of the association between factors

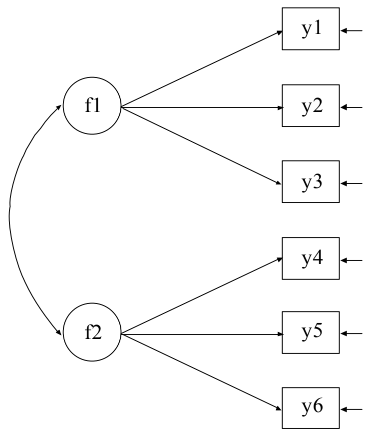
107

Confirmatory Factor Analysis (CFA) (Continued)

- Study the behavior of new measurement items embedded in a previously studied measurement instrument
- Estimate factor scores
- Investigate an EFA more fully

108

A Two-Factor CFA Model



- Squares or rectangles represent observed variables
- Circles or ovals represent factors or latent variables
- Uni-directional arrows represent regressions or residuals
- Bi-directional arrows represent correlations/covariances

109

The CFA Model

The CFA model is the same as the EFA model with the exception that restrictions can be placed on factor loadings, variances, covariances, and residual variances resulting in a more parsimonious model. In addition residual covariances can be part of the model.

Measurement Parameters – describe measurement characteristics of observed variables

- Intercepts
- Factor loadings
- Residual variances

110

The CFA Model (Continued)

Structural Parameters – describe characteristics of the population from which the sample is drawn

- Factor means
- Factor variances
- Factor covariances

Metric Of Factors – needed to determine the scale of the latent variables

- Fix one factor loading to one
- Fix the factor variance to one

111

CFA Model Identification

Necessary Condition For Identification

$a \leq b$ where a = number of parameters to be estimated in H_0
b = number of variances/covariances in H_1

Sufficient Condition For Identification

Each parameter can be solved for in terms of the variances and covariances

112

CFA Model Identification (Continued)

Practical Way To Check

- Program will complain if a parameter is most likely not identified.
- If a fixed or constrained parameter has a modification index of zero, it will not be identified if it is free.

Models Known To Be Identified

- One factor model with three indicators
- A model with two correlated factors each with two indicators

113

CFA Modeling Estimation And Testing

Estimator

In CFA, a covariance matrix is analyzed.

- ML – minimizes the differences between matrix summaries (determinant and trace) of observed and estimated variances/covariances
- Robust ML – same estimates as ML, standard errors and chi-square robust to non-normality of outcomes and non-independence of observations

Chi-square test of model fit

Tests that the model does not fit significantly worse than a model where the variables correlate freely – p-values greater than or equal to .05 indicate good fit

H_0: Factor model

H_1: Free variance-covariance model

If $p < .05$, H_0 is rejected

Note: We want large p

114

CFA Modeling Estimation And Testing (Continued)

Model fit indices (cutoff recommendations for good fit based on Yu, 2002 / Hu & Bentler, 1999; see also Marsh et al, 2004)

- CFI – chi-square comparisons of the target model to the baseline model – greater than or equal to .96/.95
- TLI – chi-square comparisons of the target model to the baseline model – greater than or equal to .95/.95
- RMSEA – function of chi-square, test of close fit – less than or equal to .05 (not good at n=100)/.06
- SRMR – average correlation residuals – less than or equal to .07 (not good with binary outcomes)/.08
- WRMR – average weighted residuals – less than or equal to 1.00 (also good with non-normal and categorical outcomes – not good with growth models with many timepoints or multiple group models)

Degrees Of Freedom For Chi-Square Testing Against An Unrestricted Model

The p value of the χ^2 test gives the probability of obtaining a χ^2 value this large or larger if the H_0 model is correct (we want high p values).

Degrees of Freedom:

(Number of parameters in H_1) – (number parameters in H_0)

Number of H_1 parameters with an unrestricted $\boldsymbol{\Sigma}$: $p\,(p + 1)/2$

Number of H_1 parameters with unrestricted $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:
$p + p\,(p + 1)/2$

A degrees of freedom example – EFA

- $p\,(p + 1)/2 - (p\,m + m\,(m + 1)/2 + p) - m^2$
  Example: if $p = 5$ and $m = 2$, then $df = 1$

Chi-Square Difference Testing Of Nested Models

- When a model H_a imposes restrictions on parameters of model H_b, H_a is said to be nested within H_b
- To test if the nested model H_a fits significantly worse than H_b, a chi-square test can be obtained as the difference in the chi-square values for the two models (testing against an unrestricted model) using as degrees of freedom the difference in number of parameters for the two models
- The chi-square difference is the same as 2 times the difference in log likelihood values for the two models
- The chi-square theory does not hold if H_a has restricted any of the H_b parameters to be on the border of their admissible parameter space (e.g. variance = 0)

117

CFA Model Modification

Model modification indices are estimated for all parameters that are fixed or constrained to be equal.

- Modification Indices – expected drop in chi-square if the parameter is estimated
- Expected Parameter Change Indices – expected value of the parameter if it is estimated
- Standardized Expected Parameter Change Indices – standardized expected value of the parameter if it is estimated

Model Modifications

- Residual covariances
- Factor cross loadings

118

Factor Scores

Factor Score

- Estimate of the factor value for each individual based on the model and the individual's observed scores
- Regression method

Factor Determinacy

- Measure of how well the factor scores are estimated
- Correlation between the estimated score and the true score
- Ranges from 0 to 1 with 1 being best

Uses Of Factor Scores

- Rank people on a dimension
- Create percentiles
- Proxies for latent variables
  - Independent variables in a model – not as dependent variables

119

Technical Aspects Of Maximum-Likelihood Estimation And Testing

120

ML Estimation

The ML estimator chooses parameter values (estimates) so that the likelihood of the sample is maximized. Normal theory ML assumes multivariate normality for $\boldsymbol{y}_i$ and n i.i.d. observations,

$$logL = -c - n / 2 \log |\boldsymbol{\Sigma}| - 1/2\, \boldsymbol{A}, \qquad (1)$$

where $c = n / 2 \log (2\pi)$ and

$$A = \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})' \, \Sigma^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu}) \qquad (2)$$

$$= trace\, [\boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})\, (\boldsymbol{y}_i - \boldsymbol{\mu})'] \qquad (3)$$

$$= n\ trace\, [\boldsymbol{\Sigma}^{-1} (\boldsymbol{S} + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\, (\bar{\boldsymbol{y}} - \boldsymbol{\mu})'] . \qquad (4)$$

121

ML Estimation (Continued)

This leads to the ML fitting function to be minimized with respect to the parameters

$$F_{ML}(\boldsymbol{\pi}) = 1/2\, [ln\, |\boldsymbol{\Sigma}| + trace\, (\boldsymbol{\Sigma}^{-1}\boldsymbol{T}) - ln\, |\boldsymbol{S}| - p], \qquad (5)$$

where

$$\boldsymbol{T} = \boldsymbol{S} + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\, (\bar{\boldsymbol{y}} - \boldsymbol{\mu})'. \qquad (6)$$

When there is no mean structure, $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$, and

$$F_{ML}(\boldsymbol{\pi}) = 1/2\, [ln\, |\boldsymbol{\Sigma}| + trace\, (\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) - ln\, |\boldsymbol{S}| - p]. \qquad (7)$$

122

Model Testing

The standard H_1 model considers an unrestricted mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Under this model $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$ and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{S}$, which gives the maximum-likelihood value

$$logL_{H_1} = -c - n / 2 \; log \; |\boldsymbol{S}| - n / 2 \; p, \qquad (8)$$

Note that

$$F_{ML} \; (\boldsymbol{\pi}) = -logL/n + logL_{H_1} /n, \qquad (9)$$

Letting $\hat{\pi}$ denote the ML estimate under H_0, the value of the likelihood-ratio χ^2-test of model fit for H_0 against H_1 is therefore obtained as $2 \; n \; F_{ML} \; (\hat{\boldsymbol{\pi}})$

123

Model Fit With Non-Normal Continuous Outcomes

- Non-normality robust chi-square testing
  - A robust goodness-of-fit test (cf. Satorra & Bentler, 1988, 1994; Satorra, 1992) is obtained as the mean-adjusted chi square defined as

$$T_m = 2 \; n \; F \; (\hat{\pi}) \; / \; c, \qquad (1)$$

where c is a scaling correction factor,

$$c = tr[\mathbf{U\Gamma}] \; / \; d, \qquad (2)$$

with

$$\mathbf{U} = (\mathbf{W}^{-1} - \mathbf{W}^{-1} \; \boldsymbol{\Delta} \; (\boldsymbol{\Delta}' \mathbf{W}^{-1} \boldsymbol{\Delta})^{-1} \boldsymbol{\Delta}' \mathbf{W}^{-1}) \qquad (3)$$

and where d is the degrees of freedom of the model.

124

Model Fit With Non-Normal Continuous Outcomes (Continued)

- Chi-square difference testing with robust (mean-adjusted) chi-square T_{md} (Satorra, 2000, Satorra & Bentler, 1999)

$$T_{md} = (T_0 - T_1)/c_d, \quad (4)$$

$$= (T_{m0}\, c_0 - T_{m1}\, c_1)/c_d, \quad (5)$$

$$c_d = (d_0\, c_0 - d_1\, c_1)/(d_0 - d_1), \quad (6)$$

where the 0/1 subscript refers to the more/less restrictive model, *c* refers to a scaling correction factor, and *d* refers to degrees of freedom.

125

Common Model Fit Indices

- Root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993; Steiger & Lind, 1980). With continuous outcomes, RMSEA is defined as

$$RMSEA = \sqrt{max[(2\, F_{ML}\, (\hat{\pi})/d - 1/n),0]}\, \sqrt{G} \quad (7)$$

where d is the number of degrees of freedom of the model and G is the number of groups. With categorical outcomes, Mplus replaces d in (7) by $tr[\mathbf{U\Gamma}]$.

- TLI and CFI

$$TLI = (\chi^2_B / d_B - \chi^2_{H_0} / d_{H_0}) / (\chi^2_B / d_B - 1), \quad (8)$$

$$CFI = 1 - max\,(\chi^2_{H_0} - d_{H_0}, 0) / max\,(\chi^2_{H_0} - d_{H_0}, \chi^2_B - d_B, 0), \quad (9)$$

126

Common Model Fit Indices (Continued)

where d_B and d_{H_0} denote the degrees of freedom of the baseline and H_0 models, respectively. The baseline model has uncorrelated outcomes with unrestricted variances and unrestricted means and / or thresholds.

- SRMR (standardized root mean square residual)

$$SRMR = \sqrt{\sum_j \sum_{k \leq j} r_{jk}^2 / e} . \qquad (10)$$

Here, $e = p\,(p + 1)/2$, where p is the number of outcomes and r_{jk} is a residual in a correlation metric.

A New Model Fit Index

WRMR (weighted root mean square residual) is defined as

$$WRMR = \sqrt{\sum_r^e \frac{(s_r - \sigma_r)^2}{v_r} \Big/ e} , \qquad (20)$$

where s_r is an element of the sample statistics vector, $\hat{\sigma}_r$ is the estimated model counterpart, v_r is an estimate of the asymptotic variance of s_r, and the e is the number of sample statistics. WRMR is suitable for models where sample statistics have widely varying variances, when sample statistics are on different scales such as in models with mean structures, with non-normal continuous outcomes, and with categorical outcomes including models with threshold structures.