

# **Psychology 236**

## **Factor Analysis**

### **Class Notes** **Fall, 2004**

**Instructor: Robert MacCallum**

These notes have been compiled collaboratively by  
Michael Browne and Robert MacCallum

These notes are intended for use by students in Psychology 236  
and should not be further copied or distributed  
without permission from the instructor

# Index

Module		Page
1	Introduction	1
2	History and Models	9
3	Matrix Algebra	19
4	The Common Factor Model	32
5	Fitting the Factor Analysis Model Theory and Methodology	46
6	Fitting the Factor Analysis Model Additional Theory and Methodology	67
7	Fitting the Factor Analysis Model: Computer Program Usage and Illustrations	80
8	Estimating the Number of Common factors: Theory and Methodology	105
9	Estimating the Number of Common Factors: Illustrations	116
10	Analytic Rotation	119
11	Principal Components Analysis, Factor Scores and Sample Size	143
12	Confirmatory Factor Analysis: Model Specification	157
13	Confirmatory Factor Analysis: Assessing Goodness of Fit	174
14	Confirmatory Factor Analysis: Computer Program and Illustrations	185
	REFERENCES	213

## Greek Alphabet

A	$\alpha$	Alpha	N	$\nu$	Nu
B	$\beta$	Beta	$\Xi$	$\xi$	Xi
$\Gamma$	$\gamma$	Gamma	O	o	Omicron <sup>†</sup>
$\Delta$	$\delta$	Delta	$\Pi$	$\pi$	Pi
E	$\epsilon$ $\varepsilon$	Epsilon	P	$\rho$	Rho
Z	$\zeta$	Zeta	$\Sigma$	$\sigma$	Sigma
H	$\eta$	Eta	T	$\tau$	Tau
$\Theta$	$\theta$ $\vartheta$	Theta	$\Upsilon$	$\upsilon$	Upsilon
I	$\iota$	Iota <sup>†</sup>	$\Phi$	$\phi$ $\varphi$	Phi
K	$\kappa$	Kappa	X	$\chi$	Chi
$\Lambda$	$\lambda$	Lambda	$\Psi$	$\psi$	Psi
M	$\mu$	Mu	$\Omega$	$\omega$	Omega

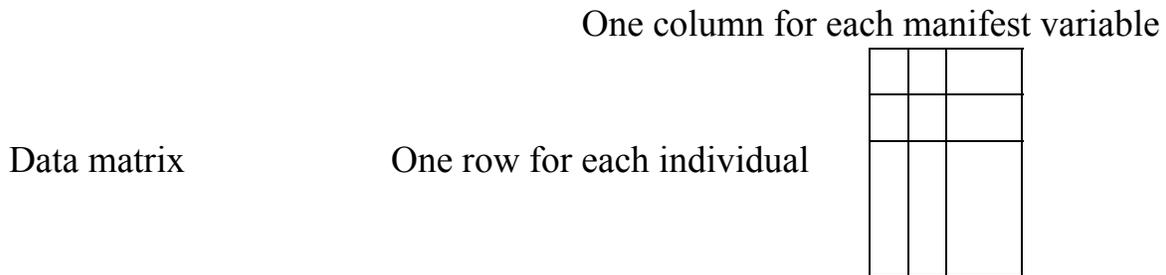
<sup>†</sup>Seldom used

$\mathcal{E}$  Expected value

Module 1 provides a conceptual introduction to factor analysis, focusing on terminology, objectives, and basic principles. Mathematical development of the model and methods begins in Module 2.

Consider first the type of data to which factor analysis is typically applied:

Multivariate Data — Scores for a sample of individuals on each of a number of manifest (measured) variables (e.g. psychological tests)



Entries in the data matrix represent scores for individuals on manifest variables.

A basic premise of factor analysis is that data of this type are not completely random, but rather have systematic aspects that can be identified and studied.

Factor Analysis: One way to study the structure underlying data → Simpler understanding of data.

Factor analysis was originally developed to study the structure of human abilities. Factor analysis models and methods provided a formal basis for various theories about intelligence and the structure of mental abilities. We shall use examples from this domain in this course. Keep in mind, however, that factor analysis is used extensively in other domains, including the study of personality, attitudes, behaviors, etc., across the social and behavioral sciences.

Two aspects of FA

- (1) Theory: Statistical models specifying a structure underlying the data
- (2) Methodology: Computational procedures that enable us to analyse these data and reveal the specified structures.

**Definitions:**

**Domain:** Set of phenomena of interest in a particular research project. e.g. mental abilities, personality traits.

**Manifest variable:** Variable that can be directly measured (measured variable) selected from the domain under investigation.

**Battery of manifest variables:** Set of manifest variables used in a particular study.

$p$  variables

e.g. Domain of Mental Abilities

- Battery:  $p = 4$
1. Paragraph Comprehension
  2. Vocabulary
  3. Arithmetic
  4. Mathematical Problem Solving

Note that empirical studies using factor analysis will typically involve a much larger battery of manifest variables.

**Latent variable:** Variable that cannot be directly observed or measured — Hypothetical construct. A latent variable is a factor in factor analysis. Thus, a factor is a variable and individuals have scores on those factors (although, as we shall see later, those scores can not be determined precisely).

**Population:** Entire set of individuals of interest.

**Sample:** Group of persons selected from the population.  $N$  persons.

**Data matrix:**

$p$  columns (manifest variables)

$\mathbf{X} = N$  rows (individuals)

$x_{11}$	$x_{12}$		$x_{1p}$
		$x_{ij}$	
$x_{N1}$	$x_{N2}$		$x_{Np}$

$x_{ij}$  = Score of person  $i$  on variable  $j$

What can we observe in these data?

1. Observe variation of each variable over individuals. Measured by variance or by s.d.
2. Observe covariation in a pair of variables over individuals. Measured by correlation coefficient.

**Correlation Matrix**

← manifest variables →

$$\mathbf{R} : \begin{array}{cccc}
 1 & r_{12} & r_{13} & r_{1p} \\
 r_{21} & 1 & r_{23} & r_{2p} \\
 r_{31} & r_{32} & 1 & r_{3p} \\
 & & & r_{jk} \\
 r_{p1} & r_{p2} & r_{p3} & 1
 \end{array}$$

$\updownarrow$   
 manifest variables

Diagonal elements are 1, representing correlation of each MV with itself. Off-diagonal elements are simple correlations between MVs; element  $r_{jk}$  is correlation between MVs  $j$  and  $k$ .

Within this matrix we would find varying degrees of correlation among MVs. To understand the pattern of relationships among the MVs, we could try to describe interrelationships between manifest variables in terms of the entries in this correlation matrix. This would be fine with 4 or less variables. If  $p$  is large the number of correlations ( $\frac{p(p-1)}{2}$ ) is too large to understand fully. e.g. If  $p = 20$  then the number of distinct correlations is 190.

For a large number of variables, the pattern of correlations may appear chaotic and difficult to describe or understand. A general principle of factor analysis is that these correlations are structured and can be explained in a simple way. **The objective of factor analysis** is to uncover and understand the structure that produces the covariances (correlations) in our data.

The cornerstone of this objective is the notion of **factors**.

**Factors:**

Latent (unobservable) variables

Hypothetical constructs

A basic principle of factor analysis is that there exists a small number of

factors within a given domain. These factors influence the MVs and thereby produce the variances and covariances (correlations) between manifest variables.

Variation in a MV is due to variation in the factors that influence that MV. For example, variation on a test of arithmetic skills (MV) is attributable in part to variation in underlying mathematical ability (factor). While the main emphasis of factor analysis is on correlations the model does incidentally affect variances.

Correlation between 2 MVs is due to the dependence of those 2 MVs on one or more of the same factors. For example, correlation between an arithmetic skills test and an algebra test (2 MVs) is due in part to the influence of mathematical ability (factor) on both MVs.

The objective is to identify the number and nature of the factors that produce the observed covariation and variation in the MVs.

### **How is this Accomplished ?**

Factor analysis theory: Formulation of a model

### **Principle Underlying Factor Analysis Theory**

Interrelationships between all possible manifest variables in the domain under consideration may be explained by a small (limited) number of factors. The number of factors is considered to be much smaller than the number of MVs. (If this were not the case, there would be little to be gained by doing factor analysis.)

e.g. It is assumed that a *limited* number of fundamental mental abilities will explain interrelationships between *all* ability tests.

It is assumed that a *limited* number of personality traits will explain interrelationships between *all* personality tests.

That is, each MV does not represent a distinct ability or trait.

It was mentioned earlier that factors influence MVs. One objective in factor analysis is to obtain measures of these influences. We measure the influence of factors on manifest variables by means of **factor loadings**. The numerical value of a factor loading indicates the strength of the influence of the factor on the manifest variable. (0 indicates no influence). Factor loadings are equivalent to regression coefficients, representing the influence of a factor (independent variable) on a MV (dependent variable).

The pattern of loadings helps us understand what the factors are. A factor is defined by the subset of manifest variables that are substantially influenced by it.

We shall learn how to obtain estimates of factor loadings, given observed correlations among MVs.

### **Assumptions of Factor Analysis Theory**

Correlations between manifest variables arise because of their dependency on one or more of the same factors.

e.g. Reading comprehension and Vocabulary are highly correlated because both are influenced by a common underlying factor of verbal ability.

### **Aim of a Factor Analysis.**

Given our data we want to determine the *number* and *nature* of the underlying factors and their *pattern of influence* (factor loadings) on the manifest variables.

We want to get a *simple* explanation of relationships in the data using a *small* number of factors.

An assumption of factor analysis theory is that correlations between manifest variables arise because of their dependency on one or more of the same factors.

### **Example:**

Suppose we obtain measures on a sample of individuals for the following 4 tests: paragraph comprehension, vocabulary, arithmetic skills, and mathematical problem solving. The problem solving test involves reading verbally stated problems, formulating them into mathematical terms, and solving them. We observe the following correlation matrix.

	P.C.	Voc.	Arith.	M.P.S.
Paragraph Comp.	1.00			
Vocabulary	.49	1.00		
Arithmetic	.14	.07	1.00	
Math. Prob. Solv.	.48	.42	.48	1.00

The off-diagonal entries in this matrix represent correlations for pairs of tests. For instance, the correlation between the Paragraph Comprehension and Vocabulary tests is .49. Considering the full matrix, we observe varying degrees of correlation among the 4 tests. Given these correlations, we wish to identify the underlying factors. We employ factor analysis methods (to be studied) to extract a factor loading matrix. The rows of this matrix represent the MVs, the columns represent the factors (LVs) and the elements represent the factor loadings. Suppose the following loading matrix is obtained in this problem.

	Factor 1	Factor 2
Paragraph Comp.	.70	.10
Vocabulary	.70	.00
Arithmetic	.10	.70
Math. Prob. Solv.	.60	.60

Entries in this matrix are weights representing the linear influence of each factor on each test.

We wish to interpret the meaning of these 2 factors. Interpretation of a given factor is achieved by inspecting the column of loadings representing that factor and seeking to identify a trait or characteristic whose pattern of influences on the MVs corresponds to the pattern of loadings. For Factor 1, we see high loadings for P.C., Vocabulary, and M.P.S., and a low loading for Arithmetic. What underlying ability corresponds to this pattern? That is, what ability influences the P.C., vocabulary, and M.P.S. tests but is not required by the arithmetic test? We could call this construct "verbal ability." In the same fashion, we consider the 2nd factor. We see high loadings for Arithmetic and M.P.S, and low loadings for the other 2 tests. We recognize that the last 2 tests require mathematical ability and the first 2 do not, so we label the 2nd factor as "mathematical ability."

The variation and covariation of the 4 MVs is explained by these 2 factors (LVs). For example, individuals vary in their scores on the Paragraph Comprehension test because they vary in their levels of underlying verbal ability. The Paragraph Comprehension and Vocabulary tests are correlated because both depend to some degree on the verbal ability factor. Individuals with high verbal ability will tend to do well on both tests and individuals with

low verbal ability will tend to do poorly on both tests, thus giving rise to a correlation between the two tests.

We shall study methods that allow us to seek an interpretable matrix of loadings, given correlations among the MVs.

Mathematical factor analysis models incorporate the above ideas. Although other factor analysis models have been suggested in the past one model is now dominant and has been for at least 50 years: **Common factor model**. In Module 2 we shall begin to study the formal principles and mathematical representation of this model.

When doing factor analysis we must always keep in mind that we are using a model; i.e., factor analysis is based on a model that represents a hypothesized structure of observed phenomena.

Common sense attitude: All mathematical models are wrong to some extent and fit the real world imperfectly. We shall not expect the factor analysis model to explain the data perfectly.

### **Points to consider**

A factor analysis will not necessarily make sense of any available data set. Some planning is required. It is unwise to consider just any correlation matrix as appropriate material for factor analysis. Rather, MVs are selected and a study is designed for the purpose of conducting factor analysis to explore or confirm the number and nature of the factors.

Interpretation of solutions must involve a subjective component. Some solutions will not make sense.

A model that makes sense and does not fit the data reasonably well is useless! The fit of the model to data should be evaluated.

A model that fits the data and does not make sense is also useless!

In factor analysis studies, a range of situations exist regarding the existence of prior hypotheses about the number and nature of the factors.

**Exploratory FA:** The researcher has no prior idea of how many factors there are and where there are zero loadings

**Confirmatory FA:** The number of factors and the positions of zero loadings are prespecified.

} The underlying theory is the same.

The underlying theory is the same.

In general, studies within a domain progress from EFA to CFA.

In this course, we shall first study the theoretical and conceptual foundations of factor analysis, along with the most widely used mathematical model. We shall then study methods for exploratory factor analysis, followed by methods for confirmatory factor analysis.

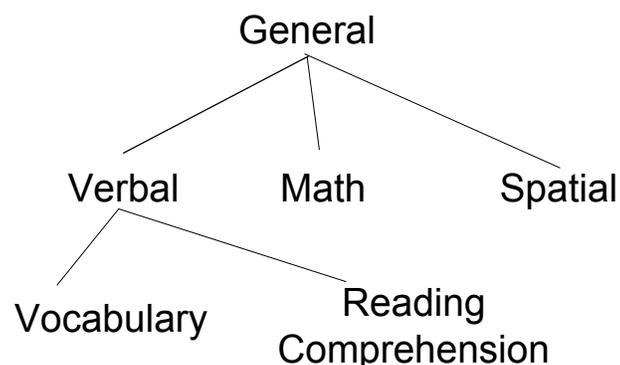
Factor Analysis began with the study of mental abilities, with most of the early developments taking place in Great Britain.

The first factor analysis model appeared in a paper by Charles Spearman in 1904. This model was expressed in mathematical form and expressed the performance on any test as a function of two factors: A **general** ability factor (Spearman's **g**) is regarded as common to all ability tests and a **specific** ability factor is relevant only to the specific test under consideration. This is called the two factor theory of intelligence, with performance on any given test depending on only two factors, **g** and the specific factor for that test.

Spearman's original **two factor theory of intelligence**: Ability tests are correlated because they all depend on the general factor to some degree. It was found later that this model sometimes yielded an unsatisfactory fit to data.

**Thomson's sampling theory of mental abilities** (cf. Thomson, 1956): Thomson disagreed with the notion of a general factor. Thomson believed intelligence consists of a large collection of narrowly defined abilities. Each test makes use of a sample of these abilities. Two tests correlate because the two samples of abilities overlap to some extent.

**Burt and Vernon: Hierarchical model of abilities** (cf. Vernon, 1961). General ability **G** is retained. The human mind is organized in a hierarchy of abilities. General ability is at the top of the hierarchy with more specific abilities lower in the hierarchy.



These developments in factor analysis took place in England in the 1930's and 1940's.

In the 1940's major new developments in factor analysis took place in the US at the University of Chicago.

**L.L. Thurstone: Multiple Factor Analysis** (Thurstone, 1947) (now called Common Factor Analysis)

Thurstone's CF model has been the dominant approach to factor analysis since the 1940s. Thurstone disagreed with the notions of **g** and a hierarchy of abilities. In Thurstone's view, manifest variables depend on two different types of underlying factor:

1. **Common factors** (Primary Mental Abilities) influence more than one manifest variable in the set we are observing. Common factors are called "common" because they are common to more than one MV. Common factors account for the correlations among the manifest variables. MVs are correlated because they depend on one or more of the same common factors. The common factors account for a portion of the variance of each MV. The number of common factors is designated  $m$ , where  $m \ll p$ .
2. **Unique factors** influence only one manifest variable. i.e. Each manifest variable has one unique factor representing that part of the MV not explained by CFs. Unique factors do not explain correlations among MVs. Unique factors account for the variance in each MV that is not accounted for by common factors.

The  $p$  manifest variables depend on  $m$  ( $m \ll p$ ) common factors and  $p$  unique factors.

Thus, there are a total of  $p+m$  factors;  $m$  common factors and  $p$  unique factors.

Each unique factor has two components:

1. Specific factor
2. Error of measurement.

The specific component represents systematic factors affecting only a given MV.

The error component represents random error of measurement (unreliability).

Based on these components, we can break down the variance of a given MV.

$$\text{Observed variance} = \text{Common variance} + \text{Unique variance}$$

$$\text{Unique variance} = \text{Specific variance} + \text{Error variance}$$

$$\text{Observed variance} = \text{Common variance} + \text{Specific variance} + \text{Error variance}$$

Communality = proportion of observed variance due to common factors

$$= \frac{\text{Common variance}}{\text{Observed variance}} = 1 - \frac{\text{Unique variance}}{\text{Observed variance}}$$

Reliability = proportion of observed variance that is systematic (not random error)

$$= \frac{\text{Common variance} + \text{Specific variance}}{\text{Observed variance}} = 1 - \frac{\text{Error variance}}{\text{Observed variance}}$$

This breakdown of variance is defined for each MV.

### Mathematical Expression of Common Factor Model

$$\mathbf{X} = N \begin{matrix} p \\ \boxed{x_{ij}} \end{matrix}$$

$$x_{ij} = \mu_j + \lambda_{j1}z_{i1} + \lambda_{j2}z_{i2} + \cdots + \lambda_{jm}z_{im} + 1u_{ij}$$

Mean + Common Factor Portion + Unique factor portion

Manifest variable:  $x_{ij}$  is the score for person  $i$  on manifest variable  $j$ .

$\mu_j$  is the mean of manifest variable  $j$ .

### Common Factor Score:

$z_{ik}$  is the common factor score for person  $i$  on factor  $k$ .  
Latent variable. It cannot be measured or assessed exactly.  
It is the same for a person across tests. Factor scores are unknown and indeterminate -- we shall never be able to exactly determine their values.

### Common Factor Loading:

$\lambda_{jk}$  is the factor loading of test  $j$  on factor  $k$ . It is the regression weight of test  $j$  on factor  $k$ .  
Same across persons.  
High  $\lambda_{jk} \Rightarrow$  high influence of factor  $k$  on test  $j$ .

### Unique Factor Score:

$u_{ij}$  is the factor score for person  $i$  on unique factor  $j$   
By definition unique factor  $j$  influences manifest variable  $j$  only.  
Large  $u_{ij}$  implies a large unique factor contribution to the score of person  $i$  on manifest variable  $j$ . We would like to include tests,  $x_j$ , for which the  $u_{ij}$  are not too large (unique variance is not large; mean is always zero)

$$u_{ij} = s_{ij} + e_{ij}$$

$s_{ij}$  is the factor score for person  $i$  on specific factor  $j$

$e_{ij}$  is the error term for person  $i$  on test  $j$ .

### Unique factor loading = 1.0.

1.0 is the unique factor loading of test  $j$  on unique factor  $j$ . (The unique factor loading of test  $j$  on unique factor  $l \neq j$  is 0.)

It is important to avoid confusing scores with variances of those scores. For instance,  $u_{ij}$  is the score on unique factor  $j$  for individual  $i$ . The variance of those scores across individuals is the unique variance for MV  $j$ .

Note that the model can be re-written by subtracting the mean from both sides of the equation:

$$x_{ij} - \mu_j = \lambda_{j1}z_{i1} + \lambda_{j2}z_{i2} + \cdots + \lambda_{jm}z_{im} + 1u_{ij}$$

In this form, we see that the model specifies the deviation from the mean as being explained by common and unique factors.

**Critical Assumption:** In the factor analysis model the  $u_{ij}$  and  $u_{il}$  are assumed to be uncorrelated over all persons  $i$  in the population (but no requirement of smallness is imposed on their variances). This implies that all partial correlations between manifest variables eliminating the effect of common factors are zero. The unique portions of the manifest variables are uncorrelated with each other. In other words, unique factors for different MVs are uncorrelated. This assumption refers to the population.

What factors are common and what factors are specific depends on the other manifest variables in the battery. That is, if we change the battery by introducing new MVs or deleting MVs, we can potentially change specific factors to common factors (by adding MVs) or common factors to specific factors (by deleting MVs). For example, in the 4-test battery in Module 1, if we remove the Mathematical Problem Solving test from the battery, then mathematical ability is no longer a common factor, but becomes a specific factor influencing only the Arithmetic test.

Note that this model says that the score on a test does not just occur on its own but rather is “caused” by the person's abilities (factors).

The factor analysis model is always “wrong” to some degree in practice. Consider some ways in which the model is likely to be wrong.

- (1) Linear Model: The model is linear, specifying MVs as linear combinations of factors. The real world is probably not exactly linear.
- (2) In the factor analysis model the number of common factors is assumed to be small (substantially smaller than the number of tests). In the real world the number of “influences” (common factors) on a test score is probably large. We hope that all but a few are negligible, and we seek to identify the major common factors.

Our objective is to identify the major common factors, recognizing that they will not perfectly account for the variation/covariation of the MVs.

### **Relationship with multiple regression.**

The form of the common factor equation looks just like a multiple regression equation. Manifest variables  $x_{ij}$  are like dependent variables, factor variables  $z_{il}$  are like independent variables and factor loadings  $\lambda_{jl}$  are like regression weights. The factor analysis model is like a set of multiple linear regression equations in which the independent variables are unobservable. In multiple linear regression we know the values of the independent variables and can use them to predict values of the dependent variable. Our objective in multiple regression is to determine the optimal regression coefficients. In factor analysis the values of the independent variables are unknown and cannot be used for predictive purposes. We can not directly determine the weights (factor loadings) as we do in regression analysis. Another important difference is that in multiple regression there is no assumption that the  $u_{ij}$  and  $u_{il}$  are uncorrelated (no assumption that partial correlations between dependent variables are zero).

Summary:

Regression:  $DV = f(IVs) + residual$

Factor analysis:  $MV = f(Common\ factors) + unique\ factor$

#### Regression

Dependent variables

Independent variables (observed)

Regression coefficients

Residuals (may be correlated for different dependent variables)

#### Factor Analysis

Measured variables

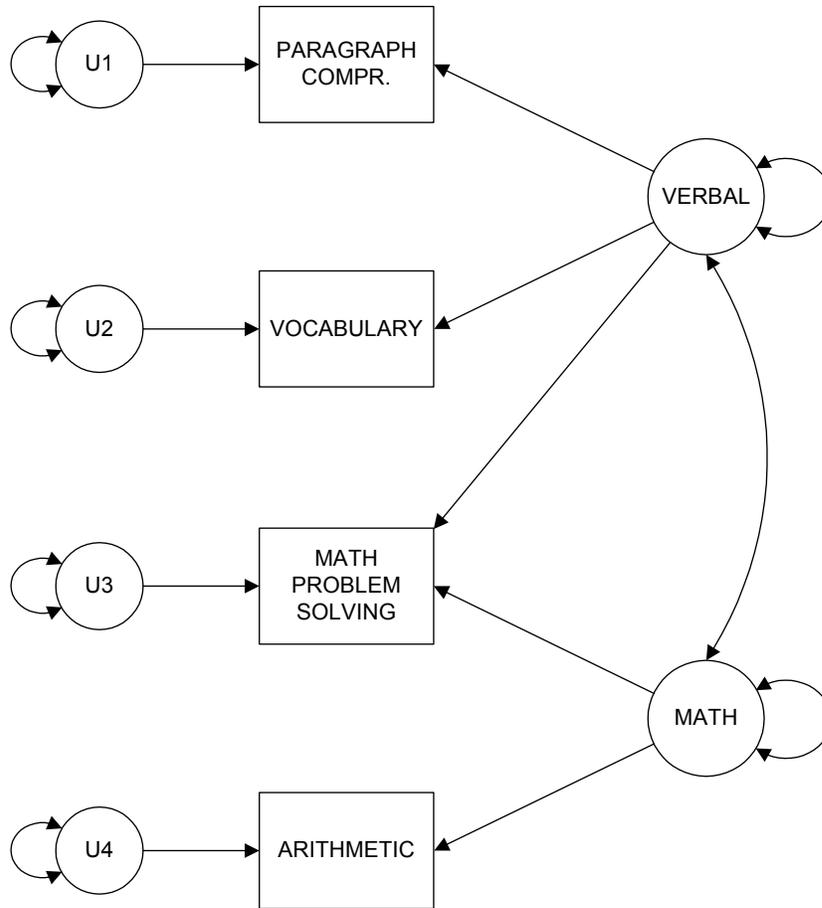
Common factors (unobserved)

Common factor loadings

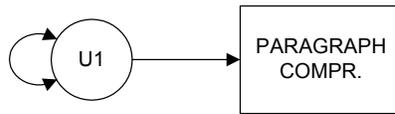
Unique factors (uncorrelated for different MVs)

## Graphical Representation of the Restricted Factor Analysis Model

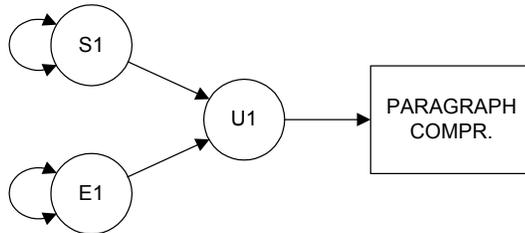
It is common to represent the factor analysis model graphically with a "path diagram." In such diagrams, squares or rectangles represent MVs and circles or ellipses represent LVs. Arrows represent linear influences. Following is an example of a path diagram for a 2-factor model for the example of 4 MVs presented earlier.



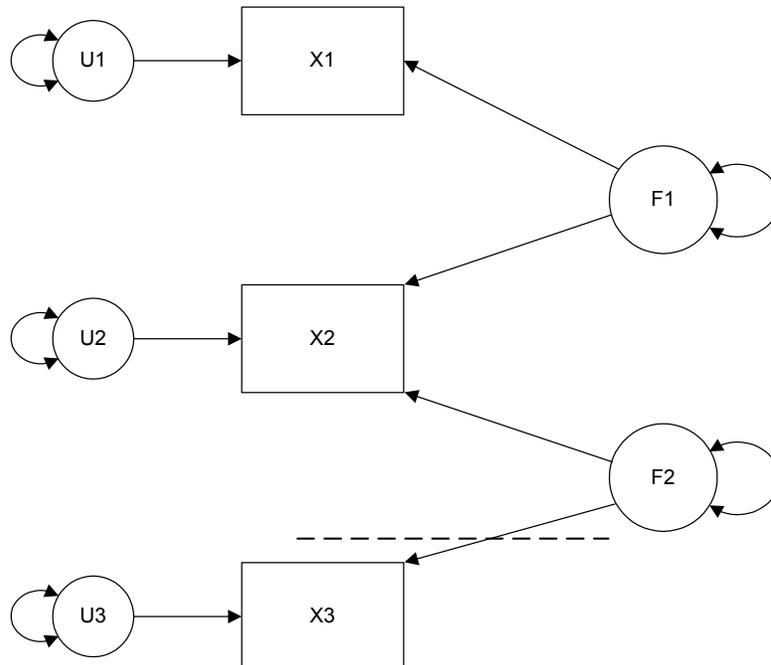
where



stands for



We can also see the effect of altering the test battery. In the example below, if X3 is included, there are 2 common factors. But if X3 is deleted, then factor F2 is no longer a common factor but rather becomes a specific factor.



## Principal Components

Principal Components (PC) is a widely used but poorly understood technique. Many users believe, mistakenly, that PC is the same as factor analysis, or provides a suitable approximation. The method of principal components differs from factor analysis in that no testable model is assumed in principal components. The basic equation is very similar to that for factor analysis

$$x_{ij} = \mu_j + \lambda_{j1}z_{i1} + \lambda_{j2}z_{i2} + \cdots + \lambda_{jm}z_{im} + e_{ij}$$

but there are substantial differences in detail:

- (1) The  $e_{ij}$  are residuals and there is no assumption that  $e_{ij}$  and  $e_{il}$  are uncorrelated over persons  $i$ . Thus, the  $e_{ij}$  term does not correspond to or have the same meaning as the unique factor term in the common factor model.
- (2) The component scores  $z_{ik}$  are linear combinations of the manifest variables  $x_{ij}$  and are not latent variables. They may be computed and are therefore observable. No latent variables are involved in Principal Components. Rather, components represent a mixture of common, specific, and error influences. Factor scores, on the other hand, are indeterminate; they can not be computed exactly.
- (3) Principal components defines components scores  $z_{ik}$  so as to retain as much information in the manifest variables  $x_{ij}$  as possible. That is, the PC approach reconstructs the MV scores as accurately as possible. The residuals are made as small as possible (i.e. Residual means are zero and the sum of residual variances is minimised.) Thus, PC is potentially useful for data reduction; i.e., the reduction of  $p$  MVs to  $m$  components, where the components carry a maximum amount of the information in the original MVs.
- (4) Principal component analysis does not explain manifest variable intercorrelations as effectively as factor analysis. Principal component analysis is useful when one wants to replace a large set of variables by a much smaller set (component scores) retaining as much information as possible (i.e. Data Reduction) Component scores should however not be regarded as measures of latent variables (factors). Factor analysis is useful when the structure of intercorrelations is to be examined. That is, factor analysis reconstructs the MV intercorrelations more accurately than PC.

Principal Components Analysis and Factor Analysis often give similar results but this is not always true. Unsophisticated users often carry out a Principal

Component Analysis and say that a factor Analysis has been carried out. It is incorrect to do this. Some canned factor analysis programs carry out a principal components analysis as a default.

The following table summarizes differences between PC and FA:

<b><u>Principal Components</u></b>	<b><u>Factor Analysis</u></b>
{ Better fit to raw data. { Better fit to MV Variances	Better fit to MV intercorrelations.
No LVs; Component scores are determinate, but components are not LVs.	Factor scores are indeterminate, but factors are interpretable as LVs.
Not a testable model.	Testable model.
Residuals are correlated.	Unique factors uncorrelated.
Useful for data reduction.	Useful for understanding structure of MV intercorrelations.

PC is not a suitable substitute or approximation for FA. It is a different technique, useful for different purposes and providing results that are not necessarily similar to FA results. There is a long-standing controversy in the FA literature. Supporters of PC argue that FA is not a useful model because the factor scores are indeterminate. Supporters of FA counter that we are not generally interested in computing the factor scores; our interest is in the loadings, which can be computed from the MV intercorrelations. Indeterminacy is a non-issue.

Matrix algebra provides a mathematical framework for manipulating arrays of numbers or algebraic symbols. It is especially useful in the representation and development of methods for modeling and analyzing multivariate data.

In Module 2 we presented the common factor model as representing the structure of a score  $x_{ij}$  for individual  $i$  on MV  $j$ . This model applies in theory to every  $x_{ij}$  that is an element of data matrix  $\mathbf{X}$ . Matrix algebra will allow us to express the common factor model for all individuals and MVs simultaneously, rather than for a single score.

**Definitions:**

**Scalar** : a single number, e.g. sample size.

**Matrix** : a rectangular table of quantities known as elements (e.g. numbers, algebraic symbols).

e.g. 
$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} = \begin{bmatrix} 1.0 & 3.0 \\ 0.2 & 4.6 \\ 8.5 & -2.3 \end{bmatrix}$$

$a_{ij}$  = element in row  $i$  and column  $j$

$a_{11} = 1.0, a_{32} = -2.3$

An upper case letter (e.g.  $\mathbf{A}$ ) names the matrix and stands for the whole array of elements.

**Data Matrix** : Each element is a score for an individual on a manifest variable.

$$\mathbf{X} = \begin{matrix} \overbrace{\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}}^{p \text{ manifest variables}} \\ \left. \vphantom{\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \end{bmatrix}} \right\} N \text{ Subjects} \end{matrix}$$

$x_{ij}$  = score for the  $i$ -th individual on the  $j$ -th manifest variable

**Order** : size of a matrix. A matrix with  $N$  rows and  $p$  columns is of order  $N \times p$ . A *scalar may be regarded as a  $1 \times 1$  matrix.*

**Square Matrix** : number of rows = number of columns

**Vector** : Matrix with a single column (**column vector**) or a single row (**row vector**)

e.g. 
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$
 is a  $m \times 1$  column vector.

$$\mathbf{v}' = [v_1 \quad v_2 \quad \cdots \quad v_m]$$
 is a  $1 \times m$  row vector

**Transpose** of a matrix  $\mathbf{A}$  : the transposed matrix  $\mathbf{A}'$  is a new matrix formed by writing the rows of  $\mathbf{A}$  as the columns of  $\mathbf{A}'$ .

e.g. If 
$$\mathbf{A} = \begin{bmatrix} 1.0 & 3.0 \\ 0.2 & 4.6 \\ 8.5 & -2.3 \end{bmatrix}$$
 then its transpose is

$$\mathbf{A}' = \begin{bmatrix} 1.0 & 0.2 & 8.5 \\ 3.0 & 4.6 & -2.3 \end{bmatrix}$$

If 
$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$
 then its transpose is 
$$\mathbf{A}' = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

The transpose of a  $p \times m$  matrix  $\mathbf{A}$  is a  $m \times p$  matrix and is represented by  $\mathbf{A}'$  or by  $\mathbf{A}^T$ .

**Equality of matrices** : Two matrices are equal if they are of the same order and all corresponding elements are equal.

e.g. If  $A = \begin{bmatrix} 1 & 3 \\ 8 & 2 \\ 5 & 7 \end{bmatrix}$  and  $B = \begin{bmatrix} 1 & 3 \\ 8 & 2 \\ 5 & 7 \end{bmatrix}$  then  $A = B$

### Arithmetic operations with matrices :

**Addition and Subtraction:** Add or subtract the corresponding elements, e.g.

$$C = A + B \Rightarrow c_{ij} = a_{ij} + b_{ij} \text{ for all } i \text{ and } j.$$

Corresponding elements of  $A$  and  $B$  are summed to form  $C$ .

$A$  and  $B$  have to be of the same order for the sum to exist.  $C$  is of the same order as  $A$  and  $B$ .

e.g. 
$$\begin{bmatrix} 6 & 8 & 8 \\ 10 & 9 & 11 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 5 \end{bmatrix} + \begin{bmatrix} 4 & 5 & 7 \\ 6 & 8 & 6 \end{bmatrix}$$

Similarly

$$C = A - B \Rightarrow c_{ij} = a_{ij} - b_{ij} \text{ for all } i \text{ and } j.$$

Laws regarding the addition and subtraction of matrices:

1. Commutative Law:  $A+B = B+A$ .
2. Associative Law:  $(A+B)+C = A+(B+C)$
3. Cancellation Law: If  $A+B = A+C$ , then  $B = C$
4. Transpose of a sum of 2 matrices:  $(A+B)' = A'+B'$

The transpose of a sum of matrices equals the sum of the transposes.

Note that Laws 1, 2 & 3 are the same as those that apply to scalars.

### Multiplication of Two Matrices:

Two matrices  $A$  and  $B$  can be multiplied together (under certain conditions specified later) to form a product matrix  $C$ . The  $ij$ -th element of the product,

$C$ , is the sum of cross products of the elements in row  $i$  of  $A$  with the elements in column  $j$  of  $B$ .

$$C = AB \Leftrightarrow c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^p a_{ik}b_{kj}$$

$$i \begin{bmatrix} \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & \dots & a_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{matrix} j \\ \begin{bmatrix} \dots & b_{1j} & \dots \\ \dots & b_{2j} & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots & b_{pj} & \dots \end{bmatrix} \end{matrix} = \begin{bmatrix} \dots & \dots & \dots \\ \dots & \dots & \dots \\ \dots & c_{ij} & \dots \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

$$A \quad B \quad = \quad C$$

e.g.  $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 5 \end{bmatrix} = \begin{bmatrix} 8 & 10 & 28 \\ 4 & 5 & 14 \end{bmatrix}$

e.g.  $\begin{bmatrix} 2 & 4 \\ * & * \end{bmatrix} \begin{bmatrix} * & 3 & * \\ * & 1 & * \end{bmatrix} = \begin{bmatrix} * & 10 & * \\ * & * & * \end{bmatrix}$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} = (2 \times 3) + (4 \times 1) = 10$$

The two matrices must be **conformable for multiplication**. i.e. Number of columns in the first matrix must = Number of rows in the second matrix,

e.g.  $n \times p \quad A \quad p \times m \quad B = n \times m \quad C$

The product has the same number of rows as the first matrix and the same number of columns as the second matrix.

Note that the sequence matters.  $BA$  does not exist if  $m \neq n$ . Square matrices can always be multiplied in both directions.

If  $C = AB$  we say that “ $B$  is *premultiplied* by  $A$ ” or that “ $A$  is *postmultiplied* by  $B$ ”.

Laws regarding the multiplication of matrices:

1. Associative:  $A(BC) = (AB)C$
2. Commutative Law does NOT hold; In general,  $AB$  does not =  $BA$
3. Cancellation does NOT hold;  $AB = AC$  does NOT imply that  $B = C$ . True **only if**  $|A| \neq 0$ . For scalars

$$ab = ac \text{ implies that } b = c \text{ only if } a \neq 0$$

4. Distributive:  $A(B+C) = AB+AC$
5. Transpose of a Product:  $(AB)' = B'A'$

Note: The transpose of a product of matrices equals the product of the transposes in reverse order.

### Multiplication of a Matrix by a Scalar.

$$C = aB \Leftrightarrow c_{ij} = a \times b_{ij}$$

Each element of the matrix  $B$  is multiplied by the scalar  $a$  to form the matrix  $C$ .

e.g. 
$$5 \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 10 & 20 \\ 5 & 15 \end{bmatrix}$$

### Kinds of Matrices.

**Diagonal Matrix** : A square matrix with all off diagonal elements equal to zero.

e.g. 
$$D_a = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

Premultiplication by a diagonal matrix scales rows.

$$D_a B = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{22}b_{21} & a_{22}b_{22} \\ a_{33}b_{31} & a_{33}b_{32} \end{bmatrix}$$

Postmultiplication by a diagonal matrix scales columns.

$$\mathbf{C}\mathbf{D}_a = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \end{bmatrix} \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} = \begin{bmatrix} c_{11}a_{11} & c_{12}a_{22} & c_{13}a_{33} \\ c_{21}a_{11} & c_{22}a_{22} & c_{23}a_{33} \end{bmatrix}$$

**Identity Matrix** : A diagonal matrix with all diagonal elements = 1.

e.g. 
$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Multiplication of a matrix by an identity matrix does not change the matrix (like multiplying any scalar by 1).

If  $\mathbf{A}$  is a  $p \times m$  matrix then  $\mathbf{A}\mathbf{I} = \mathbf{A}$        $\mathbf{I}\mathbf{A} = \mathbf{A}$ .

**Symmetric Matrix**: a *square* matrix,  $\mathbf{S}$ , is symmetric if it is equal to its own transpose;

i.e.  $\mathbf{S} = \mathbf{S}'$ .

e.g. 
$$\mathbf{S} = \begin{bmatrix} 9 & 8 & 5 \\ 8 & 6 & 4 \\ 5 & 4 & 3 \end{bmatrix}$$
 is symmetric.

Elements on opposite sides of the diagonal are mirror images of each other.  
Must be square (necessary but not sufficient condition)

**Correlation matrix**. A correlation matrix,  $\mathbf{R}$ , is a (nonnegative definite, which will be defined later) symmetric matrix with unit diagonals and off diagonal elements,  $r_{ij}$ , that satisfy  $-1 \leq r_{ij} = r_{ji} \leq 1$ ,  $i \neq j$ ,

e.g. 
$$\mathbf{R} = \begin{bmatrix} 1.0 & 0.5 & 0.2 \\ 0.5 & 1.0 & 0.4 \\ 0.2 & 0.4 & 1.0 \end{bmatrix}$$

Note that not every symmetric matrix with unit diagonal elements and off diagonal elements satisfying  $-1 \leq r_{ij} = r_{ji} \leq 1$  is a correlation matrix.

e.g. 
$$\mathbf{R} = \begin{bmatrix} 1.0 & 0.9 & 0.8 \\ 0.9 & 1.0 & 0.1 \\ 0.8 & 0.1 & 1.0 \end{bmatrix}$$
 is not a correlation matrix.

(Eigenvalues = 2.25, 0.90, -0.15. Eigenvalues will be discussed in this module.)

### Orthogonal Matrix .

A *square* matrix  $\mathbf{T}$  is orthogonal if  $\mathbf{T}\mathbf{T}' = \mathbf{I}$  or, equivalently,  $\mathbf{T}'\mathbf{T} = \mathbf{I}$ .

e.g. 
$$\mathbf{T} = \begin{bmatrix} 0.5 & -0.86603 \\ 0.86603 & 0.5 \end{bmatrix} = \begin{bmatrix} \cos 60^\circ & -\sin 60^\circ \\ \sin 60^\circ & \cos 60^\circ \end{bmatrix}$$

is orthogonal.

### Functions of Matrices

#### Determinant

The determinant of a *square* matrix  $\mathbf{A}$  is a scalar function of the elements of  $\mathbf{A}$ . The determinant of  $\mathbf{A}$  is represented by  $|\mathbf{A}|$ . If (and only if)  $|\mathbf{A}| = 0$  then some columns (rows) of  $\mathbf{A}$  may be expressed as linear combinations of other columns (rows). i.e. columns are linearly dependent and rows are linearly dependent.

e.g. Let 
$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 4 \\ 4 & 1 & 5 \\ 2 & 3 & 5 \end{bmatrix}$$
 Note that the last column is the sum of the first two columns.

Therefore  $|\mathbf{A}| = 0$ . (Note also that  $10 \times \text{row 1} - 7 \times \text{row 2} = \text{row 3}$ )

A matrix with a determinant of zero is called *singular*, meaning that there is a linear dependency among its rows (or columns).

**Inverse:** If  $\mathbf{A}$  is a *square* matrix and  $|\mathbf{A}| \neq 0$  then  $\mathbf{A}$  has a *unique* inverse  $\mathbf{A}^{-1}$  such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

e.g. If  $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$

then  $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21} = (2 \times 5) - (3 \times 4) = -2 \neq 0$

Note: this approach to computing a determinant is valid only for a  $2 \times 2$  matrix. Computation of a determinant for a larger matrix is more complex.

Then  $\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} = \begin{bmatrix} -\frac{5}{2} & \frac{3}{2} \\ 2 & -1 \end{bmatrix}$

- Comments:
- (1) Only square matrices have inverses.
  - (2) Some square matrices do not have inverses. (if  $|\mathbf{A}| = 0$ )
  - (3) The inverse of a scalar  $a$  is its reciprocal  $a^{-1} = \frac{1}{a}$

If  $|\mathbf{A}| = 0$  then  $\mathbf{A}$  is said to be singular. If  $|\mathbf{A}| \neq 0$  then  $\mathbf{A}$  is said to be nonsingular. In practice the use of composite variables (e.g.  $x_3 = x_1 + x_2$ ) leads to singular (covariance) matrices.

**Solution of equations:** Consider the matrix equation

$$\mathbf{A}\mathbf{X} = \mathbf{B}$$

where  $\mathbf{X}$  is unknown,  $\mathbf{A}$  and  $\mathbf{B}$  are known, and  $\mathbf{A}$  is square and nonsingular.

We wish to solve for  $\mathbf{X}$ .

Premultiplication of both sides by  $\mathbf{A}^{-1}$  shows that

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

$$\mathbf{I}\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$$

**Eigenvalues and eigenvectors.** Suppose that  $\mathbf{S}$  is a symmetric matrix of order  $p$ . If  $\mathbf{u}$  is a column vector of order  $p$ , and  $\ell$  is a scalar such that

$$\mathbf{S}\mathbf{u} = \ell\mathbf{u}$$

then  $\ell$  is said to be an *eigenvalue* (*characteristic root*) of  $\mathbf{S}$  and  $\mathbf{u}$  is said to be the corresponding *eigenvector* (*characteristic vector*) of  $\mathbf{S}$ . If  $\mathbf{u}'\mathbf{u} = 1$  (the sum of squares of elements of  $\mathbf{u}$  is equal to 1) then  $\mathbf{u}$  is said to be a *standardized* eigenvector of  $\mathbf{S}$ .

Comments:

(1) For a given symmetric matrix  $\mathbf{S}$ , there will generally be more than one solution for the eigenvector and eigenvalue that satisfies the equation above. A  $p \times p$  symmetric matrix  $\mathbf{S}$  has  $p$  eigenvalues

$$\ell_1 \geq \ell_2 \geq \dots \geq \ell_p$$

ordered in sequence from largest to smallest, and associated standardized eigenvectors

$$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$$

The eigenvalues can be arranged in descending order as diagonal entries in a diagonal matrix  $\mathbf{D}_l$ , and the corresponding eigenvectors arranged as columns in matrix  $\mathbf{U}$ .

Matrix  $\mathbf{U}$  is orthogonal, meaning  $\mathbf{U}'\mathbf{U} = \mathbf{I}$ .

Then the eigenstructure of  $\mathbf{S}$  can be given in matrix form as

$$\mathbf{S}\mathbf{U} = \mathbf{U}\mathbf{D}_l$$

In addition, it also holds that

$$\mathbf{S} = \mathbf{U}\mathbf{D}_l\mathbf{U}'$$

This equation is an important one. Equations of this form will turn up when we consider the matrix form of the common factor model.

Note that if both sides of this equation are postmultiplied by  $U$ , we have

$$\begin{aligned}SU &= U D_l U' U \\ &= U D_l I \\ &= U D_l\end{aligned}$$

which is the eigen-equation for  $S$ .

- (2) If all  $p$  eigenvalues of a symmetric matrix  $S$  are positive then  $S$  is said to be positive definite. If all  $p$  eigenvalues of a symmetric matrix  $S$  are nonnegative ( $\geq 0$ ) then  $S$  is said to be nonnegative definite.
- (3) The determinant,  $|S|$ , of a symmetric matrix  $S$  is equal to the product of its eigenvalues.
- (4) If any eigenvalue of  $S$  is zero then  $S$  is singular.
- (5) The trace (sum of diagonal elements) of a symmetric matrix  $S$  is equal to the sum of its eigenvalues.

### **Linear Combinations of Random Variables**

Matrix equations are useful for representing linear combinations of random variables.

Let  $\mathbf{x}$  be a column vector of order  $p$  containing scores for a random individual on variables  $x_1, x_2, \dots, x_p$ .

Let  $\mathbf{z}$  be a column vector of order  $m$  containing scores for a random individual on variables  $z_1, z_2, \dots, z_m$ .

Suppose the variables in  $\mathbf{x}$  are to be represented as linear functions of the variables in  $\mathbf{z}$ . Let  $\mathbf{A}$  be a matrix of order  $p \times m$  containing coefficients  $a_{jk}$  representing the linear effect of  $z_k$  on  $x_j$ .

Let  $\boldsymbol{\mu}$  be a column vector of order  $p$  containing fixed constants  $\mu_1, \mu_2, \dots, \mu_p$

The following matrix equation then represents the variables in  $\mathbf{x}$  as a linear function of the variables in  $\mathbf{z}$ , plus the corresponding constants in  $\boldsymbol{\mu}$ .

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}$$

Schematically:

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_p \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1m} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{p1} & a_{p2} & \cdot & \cdot & a_{pm} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{bmatrix}$$

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}$$

Following the rules for matrix multiplication, we can write an equation for any variable in  $\mathbf{x}$ , such as:

$$x_1 = \mu_1 + a_{11}z_1 + a_{12}z_2 + \dots + a_{1m}z_m$$

$$x_p = \mu_p + a_{p1}z_1 + a_{p2}z_2 + \dots + a_{pm}z_m$$

Thus, the matrix equation above actually represents a set of linear equations, one for each variable in  $\mathbf{x}$ . Each such equation represents a variable in  $\mathbf{x}$  as a linear combination of the variables in  $\mathbf{z}$ , plus a constant in  $\boldsymbol{\mu}$ .

Such equations are commonly used to represent linear statistical models such as regression and factor analysis.

### Representing a Covariance Matrix Using Matrix Algebra

Suppose  $\mathbf{x}$  is a random vector of order  $p$  containing scores on  $p$  variables for a random individual from a population, and  $\boldsymbol{\mu}$  is a random vector of order  $p$

containing the population means of the same  $p$  variables. Then vector  $(\mathbf{x} - \boldsymbol{\mu})$  represents the random vector  $\mathbf{x}$  with population means subtracted.

Consider the following expression:

$$E\{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\}$$

This expectation yields a matrix of order  $p \times p$ , which can be seen to correspond to a population covariance matrix for the  $p$  manifest variables. Call this matrix  $\boldsymbol{\Sigma}_{xx}$ .

To see this, consider the following. Taking the expected value of the cross-product of the vector with  $(\mathbf{x} - \boldsymbol{\mu})$  with itself:

Expected value of

$$\begin{bmatrix} (x_1 - \mu_1) \\ (x_2 - \mu_2) \\ \dots \\ \dots \\ (x_p - \mu_p) \end{bmatrix} [(x_1 - \mu_1) \quad (x_2 - \mu_2) \quad \dots \quad \dots \quad (x_p - \mu_p)]$$

yields a square symmetric matrix  $\boldsymbol{\Sigma}_{xx}$  of the form

$$\begin{bmatrix} \sigma_{11} & & & & \\ \sigma_{21} & \sigma_{22} & & & \\ \dots & \dots & \ddots & & \dots \\ & \sigma_{ij} & & \ddots & \\ & & & & \sigma_{pp} \end{bmatrix}$$

where

$$\sigma_{11} = E(x_1 - \mu_1)^2 = var(x_1)$$

$$\sigma_{ii} = E(x_i - \mu_i)^2 = var(x_i)$$

$$\sigma_{21} = E\{(x_2 - \mu_2)(x_1 - \mu_1)\} = cov(x_2, x_1)$$

$$\sigma_{ij} = E\{(x_i - \mu_i)(x_j - \mu_j)\} = cov(x_i, x_j)$$

Thus, diagonal elements of  $\boldsymbol{\Sigma}_{xx}$  are variances of manifest variables, and off-diagonal elements are covariances for pairs of manifest variables.

In our study of the common factor model we will make use of this representation of a population covariance matrix numerous times.

Note that if the variables in  $x$  have means of zero, then

$$\Sigma = E(xx')$$

is a population covariance matrix.

Also, if variables have been standardized in the population, then

$$P = E(zz')$$

is a population correlation matrix.

**The Data Model in Factor Analysis**

The data model in factor analysis expresses the MVs as a function of the common and unique factors. We saw in Module 2 that this model is given by:

$$\begin{aligned} x_{ij} &= \mu_j + \lambda_{j1}z_{i1} + \lambda_{j2}z_{i2} + \cdots + \lambda_{jm}z_{im} + 1u_{ij} \\ &= \mu_j + \sum_{k=1}^m \lambda_{jk}z_{ik} + u_{ij} \end{aligned} \quad (1)$$

where:

$x_{ij}$  is the score for person  $i$  on manifest variable  $j$ .

$\mu_j$  is the mean of manifest variable  $j$ .

$z_{ik}$  is the common factor score for person  $i$  on factor  $k$ .

$\lambda_{jk}$  is the factor loading of test  $j$  on factor  $k$ .

$u_{ij} = s_{ij} + e_{ij}$  is the factor score for person  $i$  on unique factor  $j$

$s_{ij}$  is the factor score for person  $i$  on specific factor  $j$

$e_{ij}$  is the error term for person  $i$  on test  $j$ .

We are going to consider the model as operating in a population of individuals that is too large for individuals to be numbered individually so that it will be convenient to consider the data model for a typical individual. This is obtained by omitting the subscript  $i$  in (1) so that we have

$$\begin{aligned} x_j &= \mu_j + \lambda_{j1}z_1 + \lambda_{j2}z_2 + \cdots + \lambda_{jm}z_m + 1u_j \\ &= \mu_j + \sum_{k=1}^m \lambda_{jk}z_k + u_j \quad j = 1, \dots, p \end{aligned} \quad (2)$$

where:

$x_j$  is the score for a typical person on manifest variable  $j$ .

$\mu_j$  is the mean of manifest variable  $j$ .

$z_k$  is the common factor score for a typical person on factor  $k$ .

$\lambda_{jk}$  is the factor loading of test  $j$  on factor  $k$ .

$u_j = s_j + e_j$  is the factor score for a typical person on unique factor  $j$

$s_j$  is the factor score for a typical person on specific factor  $j$

$e_j$  is the error term for a typical person on test  $j$ .

In (2) we have a system of  $p$  linear equations; one for each manifest variable  $x_j$ ,  $j = 1, \dots, p$ . We shall express the system of equations as a single matrix equation using matrix notation:

**Data Model :** 
$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{z} + \mathbf{u} \tag{3}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

is a  $p \times 1$  vector representing scores for a typical individual on each of the  $p$  manifest variables,

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

is a  $p \times 1$  vector representing population means for each of the  $p$  manifest variables.

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix} \quad \text{Rectangular matrix with more rows than columns}$$

is a  $p \times m$  ( $p > m$ ) factor matrix with factor loadings as elements.

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$

is a  $m \times 1$  unobservable vector of common factor scores

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix}$$

is a  $p \times 1$  unobservable vector of unique factor scores.

From matrix equation (3) with  $p = 5$  manifest variables and  $m = 3$  common factors, let us extract the equation for the second manifest variable:

$$\begin{bmatrix} x_2 \end{bmatrix} = \begin{bmatrix} \mu_2 \end{bmatrix} + \begin{bmatrix} \lambda_{21} & \lambda_{22} & \lambda_{23} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} u_2 \end{bmatrix}$$

$$x_2 = \mu_2 + \lambda_{21}z_1 + \lambda_{22}z_2 + \lambda_{23}z_3 + u_2$$

The model in (3) represents a typical observation in the population. It will be referred to as the **Data Model**. That is, this model is intended to explain the structure of the raw data, as represented by scores on manifest variables. It is not feasible to use this model in practice because there are too many unknown quantities. Note that only the MV scores will be known. Neither the scores nor the loadings for the factors are known, and it is not possible to estimate all of these values simultaneously. This model must be converted into a different form. We do this by deriving a new form of the model in which we express the structure of the variances and covariances of the MVs, rather than the individual scores on the MVs.

This data model is supplemented by assumptions concerning the joint distribution of the elements of  $\mathbf{z}$  and  $\mathbf{u}$  and generates (implies) a model for the population covariance matrix. This model for the covariance matrix will be known as a **Covariance Structure**. This covariance structure model will be intended to account for the variances and covariances (or correlations) of the manifest variables, and not the raw data themselves.

Prior to carrying out the derivation of the covariance structure model, we first note some important distributional assumptions and notation.

### **Distributional Assumptions and Notation.**

We shall make the following assumptions concerning the common factors  $\mathbf{z}$  and unique factors  $\mathbf{u}$ .

1. The common factors  $\mathbf{z}$  and unique factors  $\mathbf{u}$  are *independently distributed*. Consequently, common factors are uncorrelated with unique factors. This implies that  $\Sigma_{zu} = \mathbf{0} = \Sigma'_{uz}$ .
2. The  $p$  unique factors forming elements of  $\mathbf{u}$  are mutually independent. Consequently, unique factors for different manifest variables are uncorrelated with each other. This implies that the covariance matrix  $\Sigma_{uu}$  is diagonal. We shall write this covariance matrix as  $\Sigma_{uu} = \mathbf{D}_\psi$ .
3. For identification purposes we shall suppose that the common factors  $\mathbf{z}$  and unique factors  $\mathbf{u}$  are standardised to have zero mean vectors. That is, all factors will be defined as having means of zero :
 
$$\boldsymbol{\mu}_z = \mathbf{0} \quad \boldsymbol{\mu}_u = \mathbf{0}.$$
4. For identification purposes we shall also suppose that the common factors  $\mathbf{z}$  are standardised to have unit variances (i.e., variances of 1.0).

The following notation will be employed:

The manifest variable covariance matrix:  $\Sigma = \Sigma_{xx}$

The common factor covariance matrix:  $\Phi = \Sigma_{zz}$

The unique factor covariance matrix (diagonal in form, with diagonal elements representing unique variances of manifest variables):  $\mathbf{D}_\psi = \Sigma_{uu}$

Because of assumption 4, the diagonal elements of  $\Phi$  are required to be equal to one,  $\varphi_{jj} = 1, j = 1, \dots, m$ . Consequently  $\Phi$  is a factor correlation matrix.

## Derivation of Mean Structure and Covariance Structure from Data Model

The mean structure and covariance structure are derived from the data model. To carry out this derivation, we begin by restating the data model:

### **Data Model:**

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{z} + \mathbf{u}$$

First consider derivation of a mean structure. That is, we want an equation that represents the mean vector  $\boldsymbol{\mu}$  of the manifest variables. To obtain this, we can simply take the expectation of both sides of the above equation:

$$E\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}E\mathbf{z} + E\mathbf{u}$$

Given the assumptions stated above, this simplifies to:

$$\boldsymbol{\mu}_x = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{0} + \mathbf{0}$$

$$\boldsymbol{\mu}_x = \boldsymbol{\mu}$$

This "mean structure" implies that the means of the manifest variables are not functions of a smaller ( $< p$ ) set of parameters, and are unrestricted by the model.

Now we consider the derivation of the covariance structure. The objective here is to obtain an equation for  $\boldsymbol{\Sigma}_{xx}$ , the covariance matrix of the manifest variables.

Beginning with the data model, subtracting the mean vector  $\boldsymbol{\mu}$  from both sides yields:

$$\mathbf{x} - \boldsymbol{\mu} = \boldsymbol{\Lambda}\mathbf{z} + \mathbf{u}$$

This equation expresses deviation of individual scores from population means as a linear function of factor scores and factor loadings.

Next we postmultiply each side of this equation by its respective transpose:

$$(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = (\boldsymbol{\Lambda}\mathbf{z} + \mathbf{u})(\boldsymbol{\Lambda}\mathbf{z} + \mathbf{u})'$$

We modify the last term on the right side by using first the rule that the transpose of the sum of matrices is the sum of the transposes. This yields:

$$(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = (\boldsymbol{\Lambda}\mathbf{z} + \mathbf{u})((\boldsymbol{\Lambda}\mathbf{z})' + \mathbf{u}')$$

We next apply the rule that the transpose of a product of matrices is equal to the product of the transposes in the opposite order:

$$(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = (\boldsymbol{\Lambda}\mathbf{z} + \mathbf{u})(\mathbf{z}'\boldsymbol{\Lambda}' + \mathbf{u}')$$

We next expand the right side of the equation:

$$(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = \boldsymbol{\Lambda}\mathbf{z}\mathbf{z}'\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathbf{z}\mathbf{u}' + \mathbf{u}\mathbf{z}'\boldsymbol{\Lambda}' + \mathbf{u}\mathbf{u}'$$

Note that this equation includes random variables in  $\mathbf{x}$ ,  $\mathbf{z}$ , and  $\mathbf{u}$ . The next step in the derivation is to take expected values of both sides of the equation.

$$E(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' = \boldsymbol{\Lambda}E(\mathbf{z}\mathbf{z}')\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}E(\mathbf{z}\mathbf{u}') + E(\mathbf{u}\mathbf{z}')\boldsymbol{\Lambda}' + E(\mathbf{u}\mathbf{u}')$$

This expression simplifies considerably when we note that each expectation is a covariance matrix of one kind or another:

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{zu} + \boldsymbol{\Sigma}_{uz}\boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_{uu}$$

We can simplify further by noting that covariances of common with unique factors are zero, and also by using the notation defined above for  $\boldsymbol{\Sigma}_{zz}$  and  $\boldsymbol{\Sigma}_{uu}$ :

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}\mathbf{0}' + \mathbf{0}\boldsymbol{\Lambda}' + \mathbf{D}_{\psi}$$

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{D}_{\psi} \tag{9}$$

Equation (9) is the **factor analysis covariance structure**, an extremely important equation.

This equation represents the factor structure of the population covariance matrix of the manifest variables. It says that the variances and covariances in  $\boldsymbol{\Sigma}$  are a function of the common factor loadings (in  $\boldsymbol{\Lambda}$ ), the common factor variances and covariances (in  $\boldsymbol{\Phi}$ ), and the unique variances (in  $\mathbf{D}_{\psi}$ ). Thus, we have derived a model to explain the variances and covariances of the manifest

variables. It is especially important to note that this covariance structure model does not contain any factor scores at all, neither common nor unique. Thus, manifest variable variances and covariances do not depend on factor scores.

Note also that the above model represents common factors as potentially being intercorrelated (given in  $\Phi$ ). In unrestricted factor analysis it is common practice to assume (initially at least) that the common factors are uncorrelated so that  $\Phi = I$  and the Factor Analysis Covariance Structure becomes:

$$\Sigma = \Lambda\Lambda' + D_\psi \quad (10)$$

Using the model in either (9) or (10) we can recognize and define the breakdown of the observed variance in each MV into common and unique portions. For each MV  $x_j$ , the observed variance is given in the corresponding diagonal element of  $\Sigma$ ,  $\sigma_{jj}$ . The unique variance of  $x_j$  is given by the corresponding diagonal element of  $D_\psi$ , which is designated  $\psi_{jj}$ . And the common variance of  $x_j$  is given by the the corresponding diagonal element of the matrix  $\Lambda\Phi\Lambda'$ . Thus

$$\sigma_{jj} = (\Lambda\Phi\Lambda')_{jj} + \psi_{jj}$$

For uncorrelated factors,  $\Phi$  would be an identity matrix.

Recall that the communality of MV  $x_j$ , denoted  $h_{jj}$ , is the ratio of common variance to observed variance. Thus

$$\begin{aligned} h_{jj} &= \frac{[\Lambda\Phi\Lambda']_{jj}}{\sigma_{jj}} \\ &= 1 - \frac{\psi_{jj}}{\sigma_{jj}} \end{aligned} \quad (11)$$

### **The Factor Analysis Correlation Structure.**

Factor loadings are regression weights, representing the influences of factors (independent variables) on manifest variables (dependent variables). It was seen that the factor covariance matrix  $\Phi$  is standardised to have unit variances for identification purposes so that the independent variables (factors) are standardised. It is convenient in most situations to standardise the dependent variables (manifest variables) as well so that the factor loadings are standardised regression weights. This is achieved by converting the manifest

variable covariance matrix into a correlation matrix. It will be seen that, in general, the factor analysis structure also applies to the manifest variable correlation matrix when it applies to the manifest variable covariance matrix.

Consider first the conversion of the covariance matrix  $\Sigma$  into a correlation matrix  $P$ . Let  $D_\sigma$  be a diagonal matrix formed from the diagonal elements of  $\Sigma$ ,

$$D_\sigma = \text{Diag}[\Sigma].$$

The diagonal elements of  $D_\sigma$  therefore are variances of the manifest variables. We can define a diagonal matrix  $D_\sigma^{-1/2}$ , which contains reciprocals of standard deviations of manifest variables on the diagonal. In matrix notation the manifest variable population covariance matrix  $\Sigma$  may be transformed into a correlation matrix  $P$  as follows:

$$P = D_\sigma^{-1/2} \Sigma D_\sigma^{-1/2} \quad (12)$$

For example, if  $p = 3$ , we have

$$\begin{aligned} P &= \begin{bmatrix} 1.0 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1.0 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1.0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\sigma_{33}}} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\sigma_{33}}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{13}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \frac{\sigma_{23}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{33}}} \\ \frac{\sigma_{31}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{11}}} & \frac{\sigma_{32}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{22}}} & \frac{\sigma_{33}}{\sqrt{\sigma_{33}}\sqrt{\sigma_{33}}} \end{bmatrix} \end{aligned}$$

Note that in this matrix, the diagonal entries will be 1.0 and the off-diagonals will be correlation coefficients, where a correlation is equivalent to a covariance divided by the standard deviations of the variables involved.

Substitution of (9) into (12) shows that the factor analysis structure for the correlation matrix is

$$\begin{aligned}
P &= D_\sigma^{-1/2} (\Lambda \Phi \Lambda' + D_\psi) D_\sigma^{-1/2} \\
&= \Lambda^* \Phi \Lambda^{*'} + D_\psi^*
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
\Lambda^* &= D_\sigma^{-1/2} \Lambda \\
D_\psi^* &= D_\sigma^{-1/2} D_\psi D_\sigma^{-1/2} = D_\sigma^{-1} D_\psi.
\end{aligned}$$

i.e.

$$\begin{aligned}
\begin{bmatrix} \lambda_{11}^* & \lambda_{12}^* & \cdots & \lambda_{1m}^* \\ \lambda_{21}^* & \lambda_{22}^* & \cdots & \lambda_{2m}^* \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1}^* & \lambda_{p2}^* & \cdots & \lambda_{pm}^* \end{bmatrix} &= \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\lambda_{11}}{\sqrt{\sigma_{11}}} & \frac{\lambda_{12}}{\sqrt{\sigma_{11}}} & \cdots & \frac{\lambda_{1m}}{\sqrt{\sigma_{11}}} \\ \frac{\lambda_{21}}{\sqrt{\sigma_{22}}} & \frac{\lambda_{22}}{\sqrt{\sigma_{22}}} & \cdots & \frac{\lambda_{2m}}{\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda_{p1}}{\sqrt{\sigma_{pp}}} & \frac{\lambda_{pp}}{\sqrt{\sigma_{pp}}} & \cdots & \frac{\lambda_{pm}}{\sqrt{\sigma_{pp}}} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\begin{bmatrix} \psi_{11}^* & 0 & \cdots & 0 \\ 0 & \psi_{22}^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{pp}^* \end{bmatrix} &= D_\sigma^{-1/2} D_\psi D_\sigma^{-1/2} = \begin{bmatrix} \frac{\psi_{11}}{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \frac{\psi_{22}}{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\psi_{pp}}{\sigma_{pp}} \end{bmatrix}
\end{aligned}$$

The factor loadings for the correlation structure (in  $\Lambda^*$ ) are equivalent to the factor loadings for the covariance structure divided each divided by the standard deviation of the corresponding manifest variable. And the unique variances for the correlation structure (in  $D_\psi$ ) are equivalent to the unique variances for the covariance structure, each divided by the variance of the corresponding manifest variable. Thus, the unique variances for the correlation structure actually correspond to proportions of variance of each manifest variable attributable to unique factors.

Note that, just as with the covariance structure model, we usually initially define common factors as uncorrelated (i.e.,  $\Phi = I$ ) in the correlation structure form of the model. Thus, the orthogonal factor model in the form of a correlation structure would be given by:

$$\begin{aligned} P &= D_\sigma^{-1/2}(\Lambda\Lambda' + D_\psi)D_\sigma^{-1/2} \\ &= \Lambda^*\Lambda^{*'} + D_\psi^* \end{aligned}$$

Given the loading matrix and unique variances from a covariance structure, it is straightforward to obtain the corresponding matrices for a correlation structure.

Communalities may then be expressed as

$$\begin{aligned} h_{jj} &= \rho^2(x_j, z) = [\Lambda^*\Phi\Lambda^{*'}]_{jj} \\ &= 1 - \psi_{jj}^* \end{aligned}$$

If  $\Phi = I$  then

$$\begin{aligned} h_{jj} &= \rho^2(x_j, z) = [\Lambda^*\Lambda^{*'}]_{jj}^* \\ &= \sum_{k=1}^m \lambda_{jk}^{*2} \end{aligned} \tag{14}$$

It is usually assumed that  $\Phi = I$  and the usual definition of a communality is given by (14).

Let us consider the correlation structure for orthogonal factors more closely so as to clearly understand the association between the elements of  $\mathbf{P}$  and those of  $\mathbf{A}^*$  and  $\mathbf{D}_\psi$ . Following is a schematic representation of the correlation structure  $\mathbf{P} = \mathbf{A}^* \mathbf{A}^{*'} + \mathbf{D}_\psi^*$  :

$$\begin{bmatrix} 1 & & & & \\ \rho_{21} & 1 & & & \\ \cdot & \cdot & 1 & & \\ \cdot & \rho_{ij} & \cdot & 1 & \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} = \begin{bmatrix} \lambda_{11}^* & \cdot & \lambda_{1m}^* \\ \lambda_{21}^* & \cdot & \lambda_{2m}^* \\ \cdot & \cdot & \cdot \\ \lambda_{i1}^* & \cdot & \lambda_{im}^* \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \lambda_{11}^* & \lambda_{21}^* & \lambda_{j1}^* & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_{1m}^* & \lambda_{2m}^* & \lambda_{jm}^* & \cdot & \cdot \end{bmatrix} + \begin{bmatrix} \psi_{11}^* & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \psi_{pp}^* \end{bmatrix}$$

Considering diagonal entries of  $\mathbf{P}$ :

$$\begin{aligned} \rho_{11} = 1.0 &= \lambda_{11}^{*2} + \dots + \lambda_{1m}^{*2} + \psi_{11}^* \\ &= h_{11}^* + \psi_{11}^* \\ \rho_{jj} = 1.0 &= \lambda_{j1}^{*2} + \dots + \lambda_{jm}^{*2} + \psi_{jj}^* \\ &= h_{jj}^* + \psi_{jj}^* \end{aligned}$$

According to these expressions, a diagonal element of  $\mathbf{P}$ , which is equal to 1.0 and can be thought of as the variance of a standardized manifest variable, is equal to the sum of the squared factor loadings for that manifest variable, plus the standardized unique variance. In turn, the sum of the squared factor loadings for a given manifest variable is equal to the communality of the manifest variable.

Considering next off-diagonal elements of  $\mathbf{P}$ :

$$\begin{aligned} \rho_{21} &= \lambda_{21}^* \lambda_{11}^* + \dots + \lambda_{2m}^* \lambda_{1m}^* \\ \rho_{jk} &= \lambda_{j1}^* \lambda_{k1}^* + \dots + \lambda_{jm}^* \lambda_{km}^* \end{aligned}$$

Thus, a correlation between two manifest variables is equal to the sum of products for those two variables on the  $m$  common factors. If two manifest variables have high loadings on one or more of the same factors, the sum of products of loadings will tend to be high. If two manifest variables do not load highly on any of the same factors, the sum of products of loadings will be low. In this sense, the common factors account for the correlations among the manifest variables.

We next show how the correlation structure just presented is exemplified by the illustration presented in Module 1. Recall that that example involved a correlation matrix for four mental tests (Paragraph Comprehension, Vocabulary, Arithmetic, and Mathematical Problem Solving). That correlation matrix corresponds to  $\mathbf{P}$  in the correlation structure and was given as follows in Module 1:

$$\begin{array}{l} \text{PC} \\ \text{VOC} \\ \text{AR} \\ \text{MPS} \end{array} \begin{bmatrix} 1.00 & & & \\ .49 & 1.00 & & \\ .14 & .07 & 1.00 & \\ .48 & .42 & .48 & 1.00 \end{bmatrix}$$

The factor loading matrix for this example in Module 1 corresponds to  $\mathbf{\Lambda}^*$ . That matrix and corresponding communalities and unique variances are as follows:

Factor Loading Matrix

$\mathbf{\Lambda}^*$	<u>Communalities</u>	<u>Uniquenesses</u>
PC $\begin{bmatrix} .70 & .10 \end{bmatrix}$	$.7^2 + .1^2 = .50$	$1 - .50 = .50$
VOC $\begin{bmatrix} .70 & .00 \end{bmatrix}$	$.7^2 + .0^2 = .49$	$1 - .49 = .51$
AR $\begin{bmatrix} .10 & .70 \end{bmatrix}$	$.1^2 + .7^2 = .50$	$1 - .50 = .50$
MPS $\begin{bmatrix} .60 & .60 \end{bmatrix}$	$.6^2 + .6^2 = .72$	$1 - .72 = .28$

Communalities are computed from Equation (14).

The unique variances are diagonal elements of matrix  $D_\psi$ . That is,

$$D_\psi = \begin{bmatrix} .50 & 0 & 0 & 0 \\ 0 & .51 & 0 & 0 \\ 0 & 0 & .50 & 0 \\ 0 & 0 & 0 & .28 \end{bmatrix}$$

Correlation structure model:  $\mathbf{P} = \mathbf{\Lambda}^* \mathbf{\Lambda}'^* + \mathbf{D}_\psi^*$

Note that for an off-diagonal element of  $\mathbf{P}$ , the corresponding entry in  $\mathbf{D}_\psi^*$  is zero, implying that the unique variances do not contribute to explaining correlations between MVs.

The model implies that each element of  $\mathbf{P}$  can be reconstructed as a function of the elements of  $\mathbf{A}^*$  and  $\mathbf{D}_\psi$ . Consider for example element  $\rho_{12}$  :

$$\begin{aligned}\rho_{12} &= \lambda_{11}\lambda_{21} + \lambda_{12}\lambda_{22} \\ &= (.70)(.70) + (.10)(.00) \\ &= .49\end{aligned}$$

This correlation is of moderate size because MVs 1 and 2 depend fairly strongly on the same common factor.

This correlation reconstructed by the model matches the actual correlation in  $\mathbf{P}$ . The same holds for all elements of  $\mathbf{P}$ . Thus, for this example, the model exactly accounts for the correlations among the MVs.

This completes our introduction to the common factor model. Note that we have considered several forms of the model. First we expressed the model as a data model. Then, after making some assumptions and imposing some identification conditions, we derived a covariance structure model, specifying common factors as either correlated or uncorrelated. We then showed how the covariance structure model can be transformed into a correlation structure, with common factors again either correlated or uncorrelated. Keep in mind that these models are defined with reference to a population.

## Summary of Representations of Common Factor Model

We have now defined the common factor model in several different forms. Following is a brief summary.

### Data Model

Expresses MVs as linear combinations of common and unique factors.

Single element form:

$$x_j = \mu_j + \lambda_{j1}z_1 + \lambda_{j2}z_2 + \dots + \lambda_{jm}z_m + u_j$$

Matrix form:  $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{z} + \mathbf{u}$

### Covariance Structure

Derived from data model. Expresses covariances among MVs as function of model parameters.

Correlated common factors:

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{D}_\psi$$

Uncorrelated common factors:

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \mathbf{D}_\psi$$

### Correlation Structure

A rescaling of covariance structure.

Expresses correlations among MVs as function of model parameters.

Correlated common factors:

$$\mathbf{P} = \boldsymbol{\Lambda}^*\boldsymbol{\Phi}\boldsymbol{\Lambda}^{*'} + \mathbf{D}_\psi^*$$

Uncorrelated common factors:

$$\mathbf{P} = \boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*'} + \mathbf{D}^*$$

# PSY 820      Fitting the Factor Analysis Model      Module 5

## Theory and Methodology

In previous modules we have studied the common factor model. We now begin the study of methods for fitting the model to data. In Modules 5 and 6 we shall study the theoretical basis for fitting the model to the data. In Module 7 we shall study computer programs and examples.

We must distinguish between model and method. The model represents the hypothesized structure of the data. The common factor model is a mathematical expression that represents data (covariances or correlations among MVs) as a function of parameters whose values are not known (factor loadings, unique variances). Various methods are available for fitting the model to data, providing estimates of model parameters and information about the degree to which the model fits the data. In studying methods for fitting the common factor model to data, we begin by considering the theoretical case where the population correlation matrix, designated  $\mathbf{P}$ , is known, and the model holds exactly in the population. These conditions will never hold in practice, but this simple case provides a basis for us to begin to understand the principles of fitting the model to data. Later we shall drop these theoretical notions and consider the empirical case where  $\mathbf{P}$  is not known and the model does not hold exactly.

### 1. Properties of the Population Correlation Matrix.

In a hypothetical ideal situation where the factor analysis model holds exactly, a population correlation matrix  $\mathbf{P}$  has the structure:

$$\mathbf{P} = \mathbf{A}\mathbf{A}' + \mathbf{D}_\psi \quad (1)$$

Alternatively and equivalently:  $\mathbf{P} - \mathbf{D}_\psi = \mathbf{A}\mathbf{A}'$

Note: Strictly we should have  $\mathbf{A}^*$  and  $\mathbf{D}_\psi^*$  but we shall omit the \*'s here for simplicity. In the whole of this lecture please assume that  $\mathbf{A}$  stands for  $\mathbf{A}^*$  and  $\mathbf{D}_\psi$  for  $\mathbf{D}_\psi^*$ .

#### 1.1 Rotational Indeterminacy of the Factor matrix.

If a single  $p \times m, m \geq 2$ , factor matrix  $\mathbf{A}_1$  can be found so that  $\mathbf{P} = \mathbf{A}_1 \mathbf{A}'_1 + \mathbf{D}_\psi$ , and if  $m \geq 2$ , then there are infinitely many other  $p \times m$  factor matrices,  $\mathbf{A}_2, \mathbf{A}_3, \dots$  such that

$$\mathbf{P} = \mathbf{A}_2 \mathbf{A}'_2 + \mathbf{D}_\psi = \mathbf{A}_3 \mathbf{A}'_3 + \mathbf{D}_\psi = \dots$$

If  $\mathbf{P} = \mathbf{A}_1 \mathbf{A}'_1 + \mathbf{D}_\psi = \mathbf{A}_2 \mathbf{A}'_2 + \mathbf{D}_\psi$  then  $\mathbf{A}_2 = \mathbf{A}_1 \mathbf{T}$  where  $\mathbf{T}$  is a  $m \times m$  orthogonal matrix (i.e.  $\mathbf{T} \mathbf{T}' = \mathbf{I}$ ).

To see that  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are equally good solutions, consider:

$$\begin{aligned} \mathbf{A}_2 \mathbf{A}'_2 &= \mathbf{A}_1 \mathbf{T} (\mathbf{A}_1 \mathbf{T})' \\ &= \mathbf{A}_1 \mathbf{T} \mathbf{T}' \mathbf{A}'_1 \\ &= \mathbf{A}_1 \mathbf{I} \mathbf{A}'_1 \\ &= \mathbf{A}_1 \mathbf{A}'_1 \end{aligned}$$

Another way to say this is as follows: Suppose we find one solution  $\mathbf{A}_1$  that fits the model so that  $\mathbf{P} = \mathbf{A}_1 \mathbf{A}'_1 + \mathbf{D}_\psi$ . Suppose we then choose any matrix  $\mathbf{T}$  such that  $\mathbf{T} \mathbf{T}' = \mathbf{I}$  and define  $\mathbf{A}_2 = \mathbf{A}_1 \mathbf{T}$ . Then  $\mathbf{A}_2$  is an equally good solution to  $\mathbf{A}_1$  in the sense that  $\mathbf{A}_2 \mathbf{A}'_2 = \mathbf{A}_1 \mathbf{A}'_1$ . Thus, if one solution exists, we can find alternative equally good solutions by this method.

Since there are an infinite number of orthogonal matrices  $\mathbf{T}$  of order  $m$ , there will be an infinite number of matrices  $\mathbf{A}$  that fit the common factor model for  $\mathbf{P}$ . This is called rotational indeterminacy. As we shall learn in our study of rotation in later Modules, the procedure described here amounts to rotation of a given solution to produce alternative equally good solutions.

The problem caused by rotational indeterminacy at this point is that we must resolve it in order to find a single solution for  $\mathbf{A}$ . In effect, we are confronted with a model where we seek matrix  $\mathbf{A}$  but where an infinite number of solutions exist. We need to be able to find one of those solutions. Thus we need some criterion for choosing a unique solution. One way to achieve this is by the use of the eigenvectors and eigenvalues of  $\mathbf{P}$ .

## 1.2 Relationships between Eigenvalues, Eigenvectors and $\mathbf{A}$

In the module on matrix algebra we defined eigenvalues and eigenvectors of a symmetric matrix,  $\mathbf{S}$ . The eigenstructure of  $\mathbf{S}$  is given by

$$\mathbf{S} = \mathbf{U} \mathbf{D}_t \mathbf{U}'$$

where columns of  $U$  are eigenvectors and diagonal elements of  $D_l$  are eigenvalues.

In the present context, let the matrix  $(P - D_\psi)$  be considered the symmetric matrix of interest. Note that  $(P - D_\psi)$  is the population correlation matrix, with unique variances subtracted from the diagonal, leaving communalities on the diagonal. According to the common factor model,  $(P - D_\psi) = \Lambda\Lambda'$ , so we wish to find a unique matrix  $\Lambda$  that satisfies this model.

If it is known that  $P$  and  $D_\psi$  satisfy the common factor model in equation (1) for  $m$  factors, a factor matrix  $\Lambda$  may be constructed from the  $m$  largest eigenvalues,  $\ell_1, \ell_2, \dots, \ell_m$  and corresponding standardized eigenvectors,  $u_1, u_2, \dots, u_m$  of  $(P - D_\psi)$ . Note that the  $p - m$  smallest eigenvalues of  $(P - D_\psi)$  are all  $= 0$ .

Following material presented in the module on matrix algebra, we can construct the diagonal matrix of eigenvalues,  $D_l$ , and the matrix of corresponding eigenvectors,  $U$ . Then,

$$(P - D_\psi) = U D_l U'$$

(Note that we are not covering computational methods for obtaining eigenvalues and eigenvectors.)

We then define a diagonal matrix  $D_l^{1/2}$  containing square roots of eigenvalues.

We can then define a factor loading matrix as

$$\Lambda = U D_l^{1/2}$$

That is, each column of  $\Lambda$  is equal to an eigenvector multiplied by the square root of the corresponding eigenvalue. Let  $\lambda_j$  be the  $p \times 1$  vector formed from the  $j$ -th column of  $\Lambda$ . Then the columns of  $\Lambda$  are defined by,

$$\lambda_j = u_j \sqrt{\ell_j} \quad j = 1, \dots, m \quad (2)$$

Then

$$\begin{aligned} \Lambda\Lambda' &= U D_l^{1/2} (U D_l^{1/2})' \\ &= U D_l^{1/2} D_l^{1/2} U' \\ &= U D_l U' \\ &= (P - D_\psi) \end{aligned}$$

This particular  $\mathbf{A}$  will also satisfy the following additional equation

$$\mathbf{A}'\mathbf{A} = \mathbf{D}_\ell \quad (3)$$

so that  $\mathbf{A}'\mathbf{A}$  is a diagonal matrix, or, equivalently, all sums of cross-products of elements in different columns of  $\mathbf{A}$  are zero (i.e. columns of  $\mathbf{A}$  are orthogonal). There are  $\frac{1}{2}m(m-1)$  equations in (3) since the  $m \times m$  symmetric matrix,  $\mathbf{A}'\mathbf{A}$ , has  $\frac{1}{2}m(m-1)$  nonduplicated off-diagonal elements. These equations are known as *identification conditions* and serve to select a particular  $\mathbf{A}$  from the infinitely many possibilities satisfying

$$\mathbf{P} - \mathbf{D}_\psi = \mathbf{A}\mathbf{A}' \quad (4)$$

Keep in mind that using this procedure for obtaining  $\mathbf{A}$  requires that we know  $\mathbf{D}_\psi$ . The resulting  $\mathbf{A}$  provides a unique solution for the model.

Following is an example of applying this procedure to the illustration presented in Module 1. The correlation matrix for 4 mental tests was given as:

$$\mathbf{P} = \begin{bmatrix} 1.00 & & & \\ .49 & 1.00 & & \\ .14 & .07 & 1.00 & \\ .48 & .42 & .48 & 1.00 \end{bmatrix}$$

Assume unique variances are known and given by (from Module 4):

$$\mathbf{D}_\psi = \begin{bmatrix} .50 & 0 & 0 & 0 \\ 0 & .51 & 0 & 0 \\ 0 & 0 & .50 & 0 \\ 0 & 0 & 0 & .28 \end{bmatrix}$$

Population Correlation Matrix with Communalities in Diagonal:

$$\mathbf{P} - \mathbf{D}_\psi = \begin{bmatrix} .50 & & & \\ .49 & .49 & & \\ .14 & .07 & .50 & \\ .48 & .42 & .48 & .72 \end{bmatrix}$$

We can then obtain the eigenvalues and eigenvectors of  $(\mathbf{P} - \mathbf{D}_\psi)$  :

Nonzero Eigenvalues of  $(\mathbf{P} - \mathbf{D}_\psi)$  :

$$\mathbf{D}_l = \begin{bmatrix} 1.662 & 0 \\ 0 & .548 \end{bmatrix} \text{ (Note: two eigenvalues were zero)}$$

Corresponding Eigenvectors of  $(\mathbf{P} - \mathbf{D}_\psi)$ :

$$\mathbf{U} = \begin{bmatrix} .502 & -.386 \\ .461 & -.500 \\ .353 & .731 \\ .641 & .259 \end{bmatrix}$$

The factor loading matrix can then be obtained by  $\mathbf{A} = \mathbf{U}\mathbf{D}_l^{1/2}$  :

$$\mathbf{A} = \begin{bmatrix} .647 & -.285 \\ .594 & -.370 \\ .455 & .541 \\ .826 & .192 \end{bmatrix}$$

This solution has the property that  $\mathbf{A}\mathbf{A}' = \mathbf{P} - \mathbf{D}_\psi$  . The 2-factor model exactly fits the population correlation matrix.

This loading matrix is a transformation of the loading matrix presented in Module 1. There is a  $2 \times 2$  orthogonal transformation matrix  $\mathbf{T}$  that transforms one into the other by  $\mathbf{A}_2 = \mathbf{A}_1\mathbf{T}$ . Specifically:

$$\begin{bmatrix} .647 & -.285 \\ .594 & -.370 \\ .455 & .541 \\ .826 & .192 \end{bmatrix} = \begin{bmatrix} .70 & .10 \\ .70 & .00 \\ .10 & .70 \\ .60 & .60 \end{bmatrix} \begin{bmatrix} .848 & -.529 \\ .529 & .848 \end{bmatrix}$$

Matrices  $\mathbf{A}_1$  and  $\mathbf{A}_2$  both provide an exact solution to the model  $\mathbf{P} - \mathbf{D}_\psi = \mathbf{A}\mathbf{A}'$ . The eigenvector eigenvalue approach allowed us to find the unique solution represented above.

### 1.3 Unique Variances and Communalities.

It was seen in Section 1.2 that a factor matrix  $\mathbf{A}$  can be computed using eigenvalues and eigenvectors of  $(\mathbf{P} - \mathbf{D}_\psi)$  if the diagonal matrix  $\mathbf{D}_\psi$  of scaled (see Module 4) unique variances is known. Equivalently  $\mathbf{A}$  can be computed if the communalities

$$h_{ii} = 1 - \psi_{ii}$$

are known since

$$\begin{aligned} \mathbf{P} - \mathbf{D}_\psi &= \begin{bmatrix} 1 - \psi_{11} & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & 1 - \psi_{22} & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 - \psi_{pp} \end{bmatrix} \\ &= \begin{bmatrix} h_{11} & \rho_{12} & \dots & \rho_{1p} \\ \rho_{21} & h_{22} & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & h_{pp} \end{bmatrix} \end{aligned} \quad (5)$$

i.e.  $(\mathbf{P} - \mathbf{D}_\psi)$  has communalities as diagonal elements.

The problem therefore is to obtain the unique variances or equivalently communalities. In the earlier days of factor analysis this was regarded as a real problem, referred to as the communality problem. Many approximations were suggested and the one which has been most widely used (and is still used in most factor analysis packages as an initial approximation) is the squared multiple correlation (SMC).

We saw in Module 4 that the first communality is the squared multiple correlation,  $\rho^2(x_j, z_1, z_2, \dots, z_m)$  of the  $j$ -th manifest variable,  $x_j$ , on the  $m$  unobservable factors,  $z_1, z_2, \dots, z_m$ . There is an interesting relationship between that value and the squared multiple correlation obtained by predicting the  $j$ -th manifest variable from the  $p - 1$  remaining manifest variables. Louis Guttman (1940) showed that *if the factor analysis structure (1) applies to the manifest variable population correlation matrix,  $\mathbf{P}$* , then the squared multiple correlation of the  $j$ -th manifest variable,  $x_j$ , on the  $p - 1$  remaining manifest variables,  $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p$  is a lower bound for the  $j$ -th communality:

$$\rho^2(x_j \cdot x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p) \leq \rho^2(x_j \cdot z_1, z_2, \dots, z_m) = h_{jj} \quad (6)$$

This bound approaches the communality as the number of manifest variables increases provided that the number of factors does not change. If  $\mathbf{P}$  is known, these SMCs may be computed from

$$\rho^2(x_j \cdot x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p) = 1 - \frac{1}{\rho^{jj}} \quad (7)$$

where  $\rho^{jj}$  is the  $j$ -th diagonal element of  $\mathbf{P}^{-1}$ . The corresponding approximation to the scaled unique variance is

$$\psi_{jj} = 1 - h_{jj} \approx \frac{1}{\rho^{jj}} \quad (8)$$

Note that (6) applies *only* in the population and *only if* the population correlation matrix  $\mathbf{P}$  satisfies (1) exactly. Following is an example of obtaining these SMCs using the population correlation matrix from our 4-test example. The matrix  $\mathbf{P}$  is given by

$$P = \begin{bmatrix} 1.00 & & & \\ .49 & 1.00 & & \\ .14 & .07 & 1.00 & \\ .48 & .42 & .48 & 1.00 \end{bmatrix}$$

The inverse of this matrix is (note: computational methods not covered):

$$P^{-1} = \begin{bmatrix} 1.502 & & & \\ - .512 & 1.423 & & \\ .089 & .183 & 1.341 & \\ - .549 & - .440 & - .763 & 1.814 \end{bmatrix}$$

Squared Multiple Correlations for each MV can then be computed as follows. For each MV, the SMC is shown to be  $\leq$  the known communality.

$$\begin{aligned} \rho_{x_1 \cdot x_2, x_3, x_4}^2 &= 1 - 1/\rho^{11} = 1 - (1/1.502) = .334 \\ &\leq h_{11} = .50 \end{aligned}$$

$$\begin{aligned}\rho_{x_2.x_1,x_3,x_4}^2 &= 1 - 1/\rho^{22} = 1 - (1/1.423) = .297 \\ &\leq h_{22} = .49\end{aligned}$$

$$\begin{aligned}\rho_{x_3.x_1,x_2,x_4}^2 &= 1 - 1/\rho^{33} = 1 - (1/1.341) = .254 \\ &\leq h_{33} = .50\end{aligned}$$

$$\begin{aligned}\rho_{x_4.x_1,x_2,x_3}^2 &= 1 - 1/\rho^{44} = 1 - (1/1.814) = .449 \\ &\leq h_{44} = .72\end{aligned}$$

These values could then be used as estimates of the communalities and substituted into the diagonal of  $\mathbf{P}$ . The resulting matrix could then be factored using the eigenvector-eigenvalue approach described earlier in this module.

Of course, it is not possible to obtain these true population SMCs unless  $\mathbf{P}$  is known. In practice, when  $\mathbf{P}$  is unknown, we can apply this same procedure to a sample correlation matrix,  $\mathbf{R}$ , to obtain sample SMCs. It is now common practice to use sample SMCs as approximations for the communalities.

There is another way of computing communalities that is exact but has not been widely known and is not yet in any of the commercial programs or any of the text books. It was first suggested by Albert (1944) as a population result and more recently by Ihara and Kano (1986) as a means of estimating communalities. Certain practical difficulties are solved and a clear account of the method is given in Cudeck (1991).

We shall call this method of obtaining the communality the partitioning method. This method requires that  $p \geq 2m + 1$ , where  $m$  is the number of factors. Suppose that  $\mathbf{P}$  satisfies (1); i.e., that the model holds exactly in the population. Consider the computation of the  $j$ -th communality. The method requires that three subsets of manifest variables be selected from the  $p$  available manifest variables. The first subset consists of only the  $j$ -th variable and the second and third each consist of  $m$  (different) selected variables. Let the column vectors of correlations of the first set ( $j$ -variable) with each of the other two sets be denoted by  $\rho_{21}$  and  $\rho_{31}$  respectively and the  $m \times m$  matrix of correlations between the two sets by  $\mathbf{P}_{23}$  respectively. Then if (1) is true, it can be shown that

$$h_{jj} = \boldsymbol{\rho}'_{31} \mathbf{P}_{23}^{-1} \boldsymbol{\rho}_{21} \quad (9)$$

so that the corresponding scaled unique variance is

$$\psi_{jj} = 1 - h_{jj} = 1 - \boldsymbol{\rho}'_{31} \mathbf{P}_{23}^{-1} \boldsymbol{\rho}_{21} \quad (10)$$

Note that sets 2 and 3 must be chosen in such a manner that  $|\mathbf{P}_{23}| \neq 0$  but that there is not a unique manner in which sets 2 and 3 may be chosen. When  $\mathbf{P}$  satisfies (1) the same result will be obtained from (9) using different partitions of variables. This is no longer true when  $\mathbf{P}$  is replaced by a sample correlation matrix  $\mathbf{R}$  that does not satisfy (1).

This approach for calculation of communalities could be used in conjunction with the eigenvector-eigenvalue approach for calculation of a factor loading matrix. Given  $\mathbf{P}$  satisfying the model, one could obtain exact communalities by the partitioning method. These could then be placed in the diagonal of  $\mathbf{P}$ , yielding  $(\mathbf{P} - \mathbf{D}_\psi)$ . Then matrix  $\mathbf{A}$  could be obtained from the eigenvectors and eigenvalues of  $(\mathbf{P} - \mathbf{D}_\psi)$  as discussed in section 1.2.

Following is an example of computing communalities via the partitioning method, again using the population correlation matrix from our 4-test example. To make the example work, however, we must introduce a 5th MV because the partitioning method requires that  $p \geq 2m + 1$ . Since  $m = 2$ , we need at least 5 MVs. Following is the original population correlation matrix with a 5th MV added:

$$P = \begin{bmatrix} 1.00 & \vdots & .49 & .14 & \vdots & .48 & .22 \\ \dots & \vdots & \dots & \dots & \vdots & \dots & \dots \\ .49 & \vdots & 1.00 & .07 & \vdots & .42 & .14 \\ .14 & \vdots & .07 & 1.00 & \vdots & .48 & .58 \\ \dots & \vdots & \dots & \dots & \vdots & \dots & \dots \\ .48 & \vdots & .42 & .48 & \vdots & 1.00 & .60 \\ .22 & \vdots & .14 & .58 & \vdots & .60 & 1.00 \end{bmatrix} = \begin{bmatrix} 1 & \vdots & & & \vdots & \boldsymbol{\rho}_{13} \\ \dots & \vdots & \dots & \dots & \vdots & \dots & \dots \\ \boldsymbol{\rho}_{21} & \vdots & 1 & & \vdots & \mathbf{P}_{23} \\ & \vdots & & 1 & \vdots & & \\ \dots & \vdots & \dots & \dots & \vdots & \dots & \dots \\ & \vdots & & & \vdots & 1 & \\ & \vdots & & & \vdots & & 1 \end{bmatrix}$$

To compute the communality of the first MV, we must define 3 sets of MVs:

Set 1: MV #1  
 Set 2: MV #2 and #3  
 Set 3: MV #4 and #5

Note that sets 2 and 3 could be defined differently.

The corresponding correlation matrices then are:

$$\rho_{21} = \begin{bmatrix} .49 \\ .14 \end{bmatrix} \quad \rho_{31} = \begin{bmatrix} .48 \\ .22 \end{bmatrix} \quad P_{23} = \begin{bmatrix} .42 & .14 \\ .48 & .58 \end{bmatrix}$$

$$P_{23}^{-1} = \begin{bmatrix} 3.277 & -.791 \\ -2.712 & 2.373 \end{bmatrix}$$

The communality of MV #1 can be computed as follows:

$$h_{11} = \rho_{13} P_{23}^{-1} \rho_{21}$$

$$= [.48 \quad .22] \begin{bmatrix} 3.277 & -.791 \\ -2.712 & 2.373 \end{bmatrix} \begin{bmatrix} .49 \\ .14 \end{bmatrix} = .498$$

Within rounding error, this result matches the known true communality of .50 for MV#1. This procedure can be applied to obtain the communality of each MV in turn. The resulting values could be inserted into the diagonal of  $\mathbf{P}$  to produce  $(\mathbf{P} - \mathbf{D}_\psi)$ , and that factors could be obtained using the eigenvalue-eigenvector method.

#### 1.4 Principal Components.

The computations for principal components are almost the same as those for factor analysis, even though the rationales for the two methods are completely different. This similarity in computational methods gives some researchers the mistaken belief that principal components is a form of factor analysis. The difference between the computational methods involves communalities. Factor analysis methods incorporate estimation of communalities. In principal components, communalities are not part of the model and are not estimated. Thus, components analysis can be conducted by obtaining eigenvalues and eigenvectors of  $\mathbf{P}$  rather than of  $(\mathbf{P} - \mathbf{D}_\psi)$ . Note that  $\mathbf{P}$  is often referred to as an "unreduced" correlation matrix, whereas

$(\mathbf{P} - \mathbf{D}_\psi)$  is referred to as a "reduced" correlation matrix. If the  $\ell_j$  and  $u_j$  are the largest eigenvalues of  $\mathbf{P}$ , rather than  $(\mathbf{P} - \mathbf{D}_\psi)$ , component loadings are obtained. In the computations of principal components analysis, the diagonal elements of  $\mathbf{P}$  remain equal to one and are not replaced by communalities. For example, recall in the example above that the reduced correlation matrix,  $(\mathbf{P} - \mathbf{D}_\psi)$ , had two nonzero eigenvalues, 1.662 and .548. The unreduced correlation matrix,  $\mathbf{P}$ , on the other hand, has four nonzero eigenvalues: 2.081, 1.037, .377, and .505. The corresponding eigenvectors are as follows:

$$\begin{bmatrix} .531 & -.345 & .309 & .710 \\ .492 & -.474 & .224 & -.695 \\ .362 & .777 & .503 & -.112 \\ .588 & .230 & -.776 & .010 \end{bmatrix}$$

Obviously, these eigenvalues and eigenvectors of  $\mathbf{P}$  are different from the eigenvalues and eigenvectors of  $(\mathbf{P} - \mathbf{D}_\psi)$  shown earlier.

The component loading matrix obtained by multiplying the eigenvectors by square roots of eigenvalues, using only the first two components, is as follows:

$$\begin{bmatrix} .766 & -.351 \\ .709 & -.482 \\ .522 & .791 \\ .848 & .234 \end{bmatrix}$$

Note differences between this matrix of component loadings and the matrix of factor loadings presented earlier -- repeated here for convenience:

$$\begin{bmatrix} .647 & -.285 \\ .594 & -.370 \\ .455 & .541 \\ .826 & .192 \end{bmatrix}$$

Despite differences such as those in this example, many investigators seem to consider the difference between analysis of  $\mathbf{P}$  and analysis of  $(\mathbf{P} - \mathbf{D}_\psi)$  to be a rather minor issue. Ignoring this issue leads to the confusion of principal components analysis with factor analysis. It should be remembered, however,

that although the computations for the two methods are similar, the ways in which the results are interpreted are different. (See Module 2).

## 2. Fitting the Structural Model to the Sample Correlation Matrix.

Our study of factor analysis methods to this point has been limited to the ideal case where the population correlation matrix  $\mathbf{P}$  is available. Returning to methods for fitting the common factor correlational structure, consider the real world where we only have a sample correlation matrix  $\mathbf{R}$  and are not even sure whether it has been drawn from a population with a correlation matrix that satisfies (1). That is, the model does not hold exactly in the population, or in the sample.

In general it will not be possible to find a  $\hat{\mathbf{A}} : p \times m$  and  $\hat{\mathbf{D}}_\psi : p \times p$  that exactly reproduce our sample correlation matrix. Instead, we shall need to obtain a solution such that we obtain a good approximation to  $\mathbf{R}$  according to the factor analysis correlation structure:

$$\mathbf{R} \approx \hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{\mathbf{D}}_\psi \quad (11)$$

We want  $m \ll p$  for a parsimonious representation but there is usually no exact solution,  $\hat{\mathbf{A}}$ ,  $\hat{\mathbf{D}}_\psi$ , for (11) if  $m < p - 1$ . The best we can realistically hope for is to find a solution which provides a close approximation. The degree of approximation is reflected in the residual matrix, defined as  $\mathbf{R} - (\hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{\mathbf{D}}_\psi)$ . The elements of this matrix indicate the degree to which a particular solution does not fit each corresponding element of  $\mathbf{R}$ .

What happens if we apply the computations specified in Section 1 of this module for an ideal population correlation matrix  $\mathbf{P}$  to  $\mathbf{R}$ ? Assume initially that some  $\hat{\mathbf{D}}_\psi$  is given. That is, we obtain some initial estimates of communalities and substitute them into the diagonal of  $\mathbf{R}$ , to obtain  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ . Let  $\hat{\ell}_1 \geq \hat{\ell}_2 \geq \dots \geq \hat{\ell}_p$  and  $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_p$  be the eigenvalues and corresponding standardized eigenvectors of  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ . In general, there will not exist a  $\hat{\mathbf{D}}_\psi$  such that the  $(p - m)$  smallest eigenvalues of  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$  are zero.

Let  $\hat{\mathbf{A}}$  be defined by,

$$\hat{\lambda}_j = \hat{\mathbf{u}}_j \sqrt{\hat{\ell}_j} \quad j = 1, \dots, m \quad (12)$$

or, equivalently, by

$$\hat{\mathbf{A}} = \hat{\mathbf{U}}_m \mathbf{D}_{\hat{\ell}_m}^{1/2} \quad (13)$$

Then  $\hat{\mathbf{A}}$  minimizes the residual sum of squares

$$RSS = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \left[ (\mathbf{R} - \hat{\mathbf{D}}_\psi) - \hat{\mathbf{A}} \hat{\mathbf{A}}' \right]_{ij}^2 \quad (14)$$

the sum of squares of elements of the residual matrix  $(\mathbf{R} - \hat{\mathbf{D}}_\psi) - \hat{\mathbf{A}} \hat{\mathbf{A}}'$ , given  $\hat{\mathbf{D}}_\psi$ . That is, this solution for  $\mathbf{A}$  yields a minimum sum of squared residuals, conditional on a given set of prior communality estimates.

An estimate  $\hat{\mathbf{D}}_\psi$  may be obtained either from (8) (SMCs for communalities) or (10) (partitioning method) replacing  $\mathbf{P}$  by  $\mathbf{R}$ . Then  $\hat{\mathbf{A}}$  is obtained from (13). This method is known as the **Principal Factor Method using Prior Communality Estimates**.

In practice, this method is applied by first obtaining prior communality estimates, then inserting those estimates into the diagonal of  $\mathbf{R}$  to yield  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ . Then  $\hat{\mathbf{A}}$  is obtained from the eigenvalues and eigenvectors of  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ , providing a least squares fit to  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ . Keep in mind that this is a least squares solution for  $\hat{\mathbf{A}}$ , conditional on the prior communality estimates.

Consider an example of this procedure. In previous examples we made use of the population correlation matrix from the 4-test example in Module 1. Suppose we instead have a sample correlation matrix,  $\mathbf{R}$ , that is not the same as the population matrix,  $\mathbf{P}$ :

$$\mathbf{R} = \begin{bmatrix} 1.00 & & & \\ .51 & 1.00 & & \\ .11 & .17 & 1.00 & \\ .41 & .52 & .39 & 1.00 \end{bmatrix}$$

We wish to extract factors, but we do not know the true communalities, and we also do not know whether the model holds exactly in the population. One method for estimating communalities is to obtain sample SMCs. These can be computed using diagonal elements of  $\mathbf{R}^{-1}$ :

$$\hat{h}_{11} = \hat{\rho}_{x_1.x_2,x_3,x_4}^2 = .290$$

$$\hat{h}_{22} = \hat{\rho}_{x_2.x_1,x_3,x_4}^2 = .377$$

$$\hat{h}_{33} = \hat{\rho}_{x_3.x_1,x_2,x_4}^2 = .156$$

$$\hat{h}_{44} = \hat{\rho}_{x_4.x_1,x_2,x_3}^2 = .390$$

(Note: An alternative procedure would have been to use the partitioning method to obtain communality estimates. Those estimates would have been different than the sample SMCs.)

Sample correlation matrix with SMC communality estimates in diagonal:

$$\mathbf{R} - \hat{\mathbf{D}}_{\psi} = \begin{bmatrix} .290 & & & \\ .510 & .377 & & \\ .110 & .170 & .156 & \\ .410 & .520 & .390 & .390 \end{bmatrix}$$

Eigenvalues of  $(\mathbf{R} - \hat{\mathbf{D}}_{\psi})$  :

$$\hat{\mathbf{D}}_l = \begin{bmatrix} 1.438 & 0 \\ 0 & .184 \end{bmatrix}$$

(Note: remaining 2 eigenvalues were negative.)

Eigenvectors of  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$  corresponding to 2 positive eigenvalues :

$$\mathbf{U} = \begin{bmatrix} .492 & - .462 \\ .571 & - .331 \\ .296 & .754 \\ .586 & .330 \end{bmatrix}$$

Sample factor loading matrix, retaining 2 factors:

$$\widehat{\mathbf{A}} = \widehat{\mathbf{U}} \widehat{\mathbf{D}}_l^{1/2} = \begin{bmatrix} .590 & - .198 \\ .685 & - .142 \\ .355 & .323 \\ .703 & .141 \end{bmatrix}$$

Note that the sums of squares for each column of this matrix correspond to the eigenvalues.

Reconstructed  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi) = \widehat{\mathbf{A}}\widehat{\mathbf{A}}'$  :

$$\begin{bmatrix} .387 & & & \\ .432 & .489 & & \\ .146 & .198 & .231 & \\ .387 & .461 & .295 & .514 \end{bmatrix}$$

This matrix  $\widehat{\mathbf{A}}\widehat{\mathbf{A}}'$  contains:

(a) on the diagonal: communalities for this 2-factor solution. These values are the sums of squares of the loadings in each row of  $\widehat{\mathbf{A}}$ .

(b) off the diagonal: reconstructed correlations among MVs as approximated from this 2-factor solution. These values are sums of products for pairs of rows of  $\widehat{\mathbf{A}}$ . Note that the reconstructed correlations do not match the sample correlations in  $\mathbf{R}$ , meaning that the 2-factor model does not fit the sample data exactly. Lack of fit is reflected in the residual matrix.

Residual Matrix  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi) - \widehat{\mathbf{A}}\widehat{\mathbf{A}}'$  :

$$\begin{bmatrix} - .097 & & & & \\ .078 & - .113 & & & \\ - .036 & - .028 & - .075 & & \\ .023 & .059 & .095 & - .125 & \end{bmatrix}$$

The residual matrix indicates how closely the obtained 2-factor solution reconstructs the original correlation matrix with communality estimates in the diagonal.

The obtained solution has produced a **residual matrix with minimum sum of squares**, counting both off-diagonal and diagonal elements. That is, given the initial estimates of communalities contained in  $\widehat{\mathbf{D}}_\psi$ , the solution for  $\widehat{\mathbf{A}}$  provides a least squares fit to  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ . So this is a least squares solution conditional on the values in  $\widehat{\mathbf{D}}_\psi$ .

An alternative approach to fitting the model could be defined in terms of estimating both  $\mathbf{D}_\psi$  and  $\mathbf{A}$  so as to minimize RSS. We may minimize RSS with respect to both  $\widehat{\mathbf{D}}_\psi$  and  $\widehat{\mathbf{A}}$  using an iterative procedure. Given  $\widehat{\mathbf{A}}$ , estimates of the communalities are obtained by

$$\widehat{h}_{jj} = \sum_{l=1}^m \widehat{\lambda}_{jl}^2, \quad j = 1, \dots, p \quad (15)$$

with corresponding scaled unique variances given by

$$\widehat{\psi}_{jj} = 1 - \sum_{l=1}^m \widehat{\lambda}_{jl}^2, \quad j = 1, \dots, p \quad (16)$$

These communality (scaled unique variance) estimates conditionally minimize the RSS given  $\widehat{\mathbf{A}}$ .

The following procedure may therefore be used to minimize

$$\begin{aligned}
 RSS &= \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p [\mathbf{R} - \hat{\mathbf{D}}_\psi - \hat{\mathbf{A}}\hat{\mathbf{A}}']_{ij}^2 \\
 &= F_{\text{OLS}}(\mathbf{R}, \hat{\mathbf{P}}) = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p [\mathbf{R} - \hat{\mathbf{P}}]_{ij}^2 \quad (17) \\
 &\quad \text{where } \hat{\mathbf{P}} = \hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{\mathbf{D}}_\psi
 \end{aligned}$$

with respect to both  $\hat{\mathbf{D}}_\psi$  and  $\hat{\mathbf{A}}$ :

- (1) Obtain initial communality estimates. (e.g. using SMCs or the partitioning method). Insert these into the diagonal of  $\mathbf{R}$  to obtain  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ .
- (2) Obtain  $\hat{\mathbf{A}}$  from (14). That is, obtain factor loading estimates by the principal factors method, from the eigenvalues and eigenvectors of  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ .
- (3) Obtain new communality estimates from (15), by computing sums of squares of the loadings in each row of  $\hat{\mathbf{A}}$ . These are considered as new estimates and are inserted into the diagonal of  $\mathbf{R}$  to obtain a new  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ .
- (4) Go back to (2).

Continue this process until the communalities obtained on successive iterations do not differ appreciably (i.e. convergence has occurred).

This method is known as the Iterative Principal Factor Method or as Thomson's method (as it was first suggested by Sir Godfrey Thomson). The resulting  $\hat{\mathbf{D}}_\psi$  and  $\hat{\mathbf{A}}$  are known as Ordinary Least Squares (OLS) estimates as they minimize RSS. Other more efficient computing procedures for obtaining the same OLS estimates are available.

The function  $F_{\text{OLS}}(\mathbf{R}, \hat{\mathbf{P}})$  in (17) that is minimized is a **discrepancy function** because it (i) is always greater than or equal to zero and (ii) is equal to zero if and only if  $\mathbf{R} = \hat{\mathbf{P}}$ . The value of this function indicates degree of lack of fit of the model.

It can happen that some OLS communality estimates will be greater than one (scaled unique variance estimates less than zero). These estimates

are not admissible and a **Heywood case** is said to have occurred when this happens. Methods for constraining  $h$  to be  $\leq 1$  (unique variances  $\geq 0$ ) are available.

## Summary

We have considered the problem of fitting the common factor model for several cases.

Case 1:  $\mathbf{P}$  and  $\mathbf{D}_\psi$  known (This is a theoretical case and will never hold in practice.)

Obtain eigensolution of  $(\mathbf{P} - \mathbf{D}_\psi)$ , and compute factor loadings according to equation (2).

Case 2:  $\mathbf{P}$  known,  $\mathbf{D}_\psi$  not known (another theoretical case)

Obtain communalities using SMCs or by the partitioning method. The SMCs are lower bounds under conditions specified earlier, and the partitioning method provides exact communalities if the model holds exactly.

Insert the communalities into the diagonal of  $\mathbf{P}$  to yield  $(\mathbf{P} - \mathbf{D}_\psi)$ .

Obtain eigensolution of  $(\mathbf{P} - \mathbf{D}_\psi)$ , and compute factor loadings according to equation (2).

Case 3:  $\mathbf{P}$  not known; Given sample correlation matrix  $\mathbf{R}$ . (Usual case in practice.)

Method 1: Principal factors with prior communality estimates.

Estimate communalities using SMCs or partitioning method, applied to  $\mathbf{R}$ .

Insert communality estimates into diagonal of  $\mathbf{R}$ , to yield  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ .

Obtain eigensolution from  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ , and compute factor loading matrix  $\widehat{\mathbf{A}}$ .

This solution minimizes RSS, given  $\widehat{\mathbf{D}}_\psi$ .

Method 2: Iterative principal factors (Ordinary Least Squares).

Estimate communalities using SMCs or partitioning method, applied to  $\mathbf{R}$ .

Insert communality estimates into diagonal of  $\mathbf{R}$ , to yield  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ .

Obtain eigensolution from  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ , and compute factor loading matrix  $\widehat{\mathbf{A}}$ .

Compute new communality estimates by summing squares of entries in each row of  $\widehat{\mathbf{A}}$ .

Insert new communality estimates into diagonals of  $\mathbf{R}$ , to obtain new  $(\mathbf{R} - \widehat{\mathbf{D}}_\psi)$ .

Compute new loading matrix, new communalities, and repeat process until convergence.

This solution yields  $\widehat{\mathbf{A}}$  and  $\widehat{\mathbf{D}}_\psi$  that minimize RSS.

Many of the older software packages offer the two types of principal factoring methods described in Case 3. Squared multiple correlations are used for communalities in Case 3, Method 1. CEFA has a related version of Method 1 in which the partitioning method is used to estimate communalities and conditional maximum likelihood is used instead of principal factors. CEFA also can yield the ordinary least squares solution of Method 2 but uses a different algorithm.

## **PSY 820 Fitting the Factor Analysis Model Module 6 Additional Theory and Methodology**

In Module 5 we considered ordinary least squares estimates of parameters in the factor analysis model, paying particular attention to the iterated principal factor procedure for computing these estimates. An important aspect of these methods is that they require no distributional assumptions regarding the manifest variables. In this module we shall consider maximum likelihood (ML) estimates under the assumption of normality for the data. This is equivalent to the assumption of a Wishart distribution for the sample covariance matrix so that these estimates are often referred to as MWL (Maximum Wishart Likelihood) estimates (see Module 7). ML estimation is widely used in statistics for estimating model parameters. At the end of the lecture the OLS and Maximum Likelihood Approaches will be compared.

### **General Principles of MWL Estimation:**

Let us first consider general principles of ML estimation. Then we shall apply these principles to the problem of estimating parameters of the common factor model.

1. Assume a Random Sample from some defined population.
2. Assume that the distribution of the manifest variables has a specific form (typically multivariate normality).

We write the likelihood function for the distribution:

$$L = \text{Likelihood function} = \text{function}(\text{data}, \text{parameters})$$

The likelihood function indicates the likelihood of the observed data, given values of the model parameters. The method of ML estimation is based on the following concept: Given our data, we obtain the values of the parameters that maximize the value of  $L$ . This gives us the maximum likelihood estimates of the parameters, i.e., the parameter values that maximise the likelihood of our data under our assumptions about the population.

## Maximum Likelihood Estimation in Factor Analysis

ML estimation in factor analysis is a relatively old idea, but did not become practical until the development of sophisticated computational methods and modern computers. ML estimation in factor analysis was considered by D.N. Lawley in 1940 but the iterative algorithm then suggested was not very effective. Furthermore computers were not yet available and computations were not feasible. Another more effective iterative algorithm for maximum likelihood estimation was suggested by Howe and by Bargmann in unpublished Ph.D. theses in the early 1950s. The algorithm now generally used was published independently in three papers around 1970 — Jöreskog, Jennrich & Robinson, Clarke. (The most commonly cited reference is Jöreskog, 1967.) It makes use of a trick to reduce the number of parameters in the iteration and obtains conditional maximizers using eigenvalues and eigenvectors. It is effective.

Assumption. Manifest variables follow a multivariate normal distribution in the population under consideration. This implies that each individual variable follows a normal distribution, and all regressions between pairs of variables are linear.

Based on this assumption we can write the multivariate normal likelihood function, which would have the following form:

$$L = f(\mathbf{S}, \mathbf{\Sigma})$$

where  $\mathbf{S}$  is the sample covariance matrix and  $\mathbf{\Sigma}$  is the population covariance matrix.

Although we will not use the following full expression in ensuing developments, it is informative to consider the formal representation of this likelihood function. The joint likelihood for a sample of observations under the assumption of multivariate normality is given by:

$$L = \prod_{i=1}^N \frac{|\mathbf{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})' \mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right]$$

The value of the likelihood function is a measure of joint likelihood of an observed sample of data (represented by the individual data vectors  $\mathbf{x}_i$ ) given the parameters in the population covariance matrix  $\mathbf{\Sigma}$  and the population mean vector  $\boldsymbol{\mu}$ . Because the population means are not relevant in the usual common factor model, they do not need to be considered further in the present context.

From this expression, we can derive another equation that defines  $L$  as a function of the sample covariance matrix  $\mathbf{S}$  and the population covariance matrix  $\mathbf{\Sigma}$ . That is,  $L = f(\mathbf{S}, \mathbf{\Sigma})$ .

It is then assumed that the common factor model for a specified number of factors,  $m$ , holds in the population. That is:  $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda} + \mathbf{D}_\psi$ , meaning that any lack of fit of the model in the sample is due to sampling error only. Thus, the likelihood function could be re-expressed as

$$L = f(\mathbf{S}, \mathbf{\Lambda}, \mathbf{D}_\psi)$$

That is, the value of the likelihood function is dependent on the sample covariance matrix (which is known) and the factor loadings and unique variances (which are not known). The problem is to find best estimates of these parameters; i.e., estimates that maximize the value of the likelihood function,  $L$ .

Maximum likelihood estimates,  $\hat{\mathbf{\Lambda}}$ ,  $\hat{\mathbf{D}}_\psi$  are chosen so that the likelihood function  $L(\mathbf{S}, \hat{\mathbf{\Sigma}})$  with  $\mathbf{S}$  representing the sample covariance matrix and with  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}} + \hat{\mathbf{D}}_\psi$  is maximised.

The computational procedure for ML estimation works in general as follows. Given  $\mathbf{S}$ , we can choose estimates of  $\mathbf{\Lambda}$  and  $\mathbf{D}_\psi$ . We call these estimates  $\hat{\mathbf{\Lambda}}$  and  $\hat{\mathbf{D}}_\psi$ . From these estimates we can compute a reconstructed covariance matrix  $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}} + \hat{\mathbf{D}}_\psi$ . We can then compute a value of the likelihood function,  $L = f(\mathbf{S}, \hat{\mathbf{\Sigma}})$ . Given  $\mathbf{S}$ , the task is to find  $\hat{\mathbf{\Lambda}}$  and  $\hat{\mathbf{D}}_\psi$  so that the resulting  $\hat{\mathbf{\Sigma}}$  yields the largest possible value of  $L = f(\mathbf{S}, \hat{\mathbf{\Sigma}})$ .

Rather than maximizing the likelihood function, it has been found more convenient to work with an alternative function that is inversely related to the likelihood function. This alternative function is given by

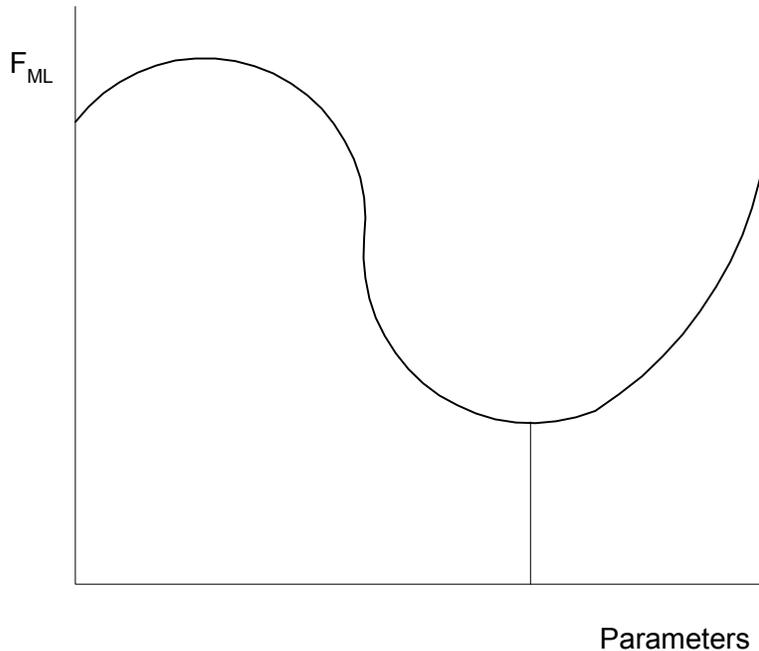
$$\begin{aligned} F_{\text{ML}}(\mathbf{S}, \mathbf{\Sigma}) &= -2 \times \log_e(L) \\ &= -2 \times \log_e f(\mathbf{S}, \mathbf{\Sigma}) \\ &= \log_e |\mathbf{\Sigma}| - \log_e |\mathbf{S}| + \text{tr}[(\mathbf{S} - \mathbf{\Sigma})\mathbf{\Sigma}^{-1}] \end{aligned} \quad (1)$$

where  $\text{tr}[\mathbf{A}]$  is the trace (sum of diagonal elements) of a square matrix  $\mathbf{A}$ . The function  $F_{\text{ML}}$  is inversely related to  $L$ , meaning that as  $L$  increases,  $F_{\text{ML}}$  decreases. Thus, maximizing  $L$  is equivalent to minimizing  $F_{\text{ML}}$ .

The ML estimation problem is then reformulated as a problem of finding  $\hat{\Lambda}$  and  $\hat{D}_\psi$  such that the resulting reconstructed covariance matrix  $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}' + \hat{D}_\psi$  yields a minimum value of  $F_{ML}$  when substituted into (1).

The function  $F_{ML}(\mathcal{S}, \hat{\Sigma})$  in (1) is always  $\geq 0$  (when  $\mathcal{S}$  and  $\hat{\Sigma}$  are positive definite) and is  $= 0$  if and only if  $\mathcal{S} = \hat{\Sigma}$ . It therefore is a **discrepancy function**. An alternative discrepancy function, the ordinary least squares discrepancy function  $F_{OLS}(\mathcal{S}, \hat{\Sigma})$ , was considered in Module 5.

The ML discrepancy function can be considered as having a geometric representation, where one axis represents the value of the function and the other axes represent the parameters to be estimated.



As the values of the parameters change, the value of the discrepancy function changes. We seek parameter values that yield the minimum value of the function.

Following is a summary of the computational method for ML factor analysis. We first obtain an initial estimate of the unique variances in  $\widehat{\mathbf{D}}_\psi$  by one of the methods considered in the previous chapter (e.g., SMCs or partitioning method). Given any such  $\widehat{\mathbf{D}}_\psi$  we can then obtain a conditional ML solution for  $\mathbf{A}$ , conditional on the given  $\widehat{\mathbf{D}}_\psi$ . The conditional minimizer  $\widehat{\mathbf{A}}$  of  $F_{ML}(\mathbf{S}, \widehat{\mathbf{\Sigma}})$  may be obtained from the  $m$  largest eigenvalues and corresponding eigenvectors of  $\widehat{\mathbf{D}}_\psi^{-1/2}(\mathbf{S} - \widehat{\mathbf{D}}_\psi)\widehat{\mathbf{D}}_\psi^{-1/2}$ . Modern iterative computing procedures use this fact to increase efficiency. At this point, a reconstructed covariance matrix  $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{A}}\widehat{\mathbf{A}}' + \widehat{\mathbf{D}}_\psi$  can be calculated along with a corresponding value of the discrepancy function  $F_{ML}$ . This completes the first iteration. A new  $\widehat{\mathbf{D}}_\psi$  is then obtained by calculating the communalities using the current  $\widehat{\mathbf{A}}$ , and a new  $\widehat{\mathbf{A}}$  is calculated, and then a new value of  $F_{ML}$ , which will generally be smaller than the previous value. This completes the second iteration. Iteration is continued until convergence occurs. Convergence is defined as a change in the value of the discrepancy function that is lower than some small pre-set criterion. Usually convergence is quite rapid but Heywood cases can occur just as in Iterated Principal Factor Analysis.

After convergence, the procedure yields ML estimates of factor loadings and unique variances. These values, when substituted into the definition of  $F_{ML}$ , yield a minimum value of that function. Given the sample covariance matrix, no other parameter estimates would yield a smaller value of this function.

After convergence the likelihood ratio test statistic is given by  $(N - 1)F_{ML}(\mathbf{S}, \widehat{\mathbf{\Sigma}})$ . This statistic will be used for testing hypotheses about model fit.

### An Alternative Method: Conditional ML (PACE)

An alternative to the use of the iterative maximum likelihood procedure is to use the partitioning estimate of  $\widehat{\mathbf{D}}_\psi$  and to obtain the corresponding factor matrix from eigenvalues and eigenvectors of  $\widehat{\mathbf{D}}_\psi^{-1/2}(\mathbf{S} - \widehat{\mathbf{D}}_\psi)\widehat{\mathbf{D}}_\psi^{-1/2}$  as suggested by Cudeck (1991). This is a single iteration procedure. This method is implemented in the PACE option of CEFA (see Module 7) and has the advantages that it is noniterative and also that Heywood cases seldom occur. This approach can be understood as a *conditional* ML approach. The

factor loadings are ML estimates, conditional on the partitioning estimate of the unique variances. The solution as a whole, however, when considering the estimates of factor loadings and unique variances, is not a true ML solution because the estimates of both sets of parameters have not been obtained so as to minimize the ML discrepancy function.

Once a solution is obtained by this method, the partitioned estimates of the unique variances ( $\widehat{\mathbf{D}}_\psi$ ) and conditional ML estimates of factor loadings ( $\widehat{\mathbf{A}}$ ) can be used to compute a value of the ML discrepancy function, as in (1). Note, however, that the resulting value of  $F_{ML}$  will be minimized only with respect to  $\widehat{\mathbf{A}}$ , and not with respect to  $\widehat{\mathbf{D}}_\psi$ . Thus, this value will be larger than it would be for a true ML solution for all model parameters. Browne has developed a correction for this phenomenon. That is, the value of the discrepancy function obtained by the PACE procedure can be corrected to provide an estimate of a true ML value of the discrepancy function, thus making it feasible to conduct approximate significance tests and calculate appropriate fit measures, as discussed below.

### Analysis of Covariance vs. Correlation Matrices

These developments have been presented in the context of analysis of a covariance matrix,  $\mathbf{S}$ . With many factor analysis models (including the *unrestricted factor analysis model*), maximum likelihood estimates of the scaled factor matrix  $\mathbf{A}^*$  and scaled uniqueness covariance matrix  $\widehat{\mathbf{D}}_\psi^*$  in the factor analysis correlation structure may be obtained by replacing  $\mathbf{S}$  by  $\mathbf{R}$  and  $\mathbf{\Sigma}$  by  $\mathbf{P}$  in (1) and minimizing  $F_{ML}(\mathbf{R}, \widehat{\mathbf{P}})$  — that is by inputting  $\mathbf{R}$  instead of  $\mathbf{S}$  when using a maximum likelihood factor analysis computer program. (This does not always work, however, and there are some *restricted* factor analysis models where the the minimisation of  $F_{ML}(\mathbf{R}, \widehat{\mathbf{P}})$  will not yield maximum likelihood estimates and will not yield a likelihood ratio test statistic.)

## **Summary**

We have now considered four different methods for fitting the common factor model to sample data:

Principal factors with prior communality estimates (noniterative).

Iterative principal factors (ordinary least squares).

Maximum likelihood.

Conditional maximum likelihood using partitioning method for estimation of unique variances.

Keep in mind that these are four different *methods* for fitting the same *model* to a given set of data. Other methods exist for fitting the model, beyond those described in these notes. The principal factors methods and ML method are commonly available in statistical software packages.

## **Comparison between OLS solution and ML solution**

The most commonly used methods in practice are OLS and ML. It is important to understand that OLS and ML are two different methods for fitting the same model to data. Both are methods for estimating the parameters of the common factor model. Given sample data, the two methods will typically provide slightly different estimates of factor loadings and unique variances. Each is best by its own definition of optimality. Both are widely used in the applied factor analysis literature.

OLS makes no assumptions about the distribution of data to obtain estimates. If normality assumptions are made one can obtain a test of fit and confidence intervals on parameters. However, the test of fit is essentially an approximation to that given by ML and the confidence intervals will be wider than those for ML.

ML does make assumptions about the distribution of the data but then provides tests of fit, confidence intervals, etc.

OLS is more robust and does not break down as often as ML if  $\mathbf{R}$  is not positive definite

Thomson's method (iterated principal factors) is an algorithm for obtaining OLS estimates that is easy to understand, works fairly well and was known long before adequate computational methods for maximum likelihood factor analysis were available.

ML provides a capability for inferential statistics; e.g., tests of model fit and confidence intervals for some measures of fit, such as RMSEA.

Normal distribution implies that ML will give correct tests, confidence intervals etc.

### **BUT**

Non-normal distribution does not imply that ML will not give correct tests, confidence intervals etc. The critical assumption in unrestricted factor analysis is not normality but independence of  $\mathbf{z}$  and  $\mathbf{u}$  and independence of elements of  $\mathbf{u}$ .

### **Methods for Evaluating Fit**

Following is a presentation of some methods for evaluating fit of factor analysis solutions obtained by the ML or PACE methods.

#### **1. Test of perfect fit:** Direct hypothesis test of our model

$H_0: \Sigma = \Lambda\Lambda' + D_\psi$  i.e., FA Model holds exactly in the population with  $m$  factors

$H_1: \Sigma$  has no particular structure

Note that the role of  $H_0$  here is reversed from its usual role in research settings. We hope to support  $H_0$  rather than to reject it. Rejection of  $H_0$  implies rejection of the model. Failure to reject  $H_0$  implies that the model is plausible.

Under  $H_0$  and as  $N$  becomes large, the likelihood ratio test statistic  $(N - 1)F_{ML} \sim \chi^2$  with degrees of freedom given by

$$d = \frac{1}{2}p(p + 1) - \left\{ p + pm - \frac{1}{2}m(m - 1) \right\}$$

$$= \frac{(p - m)^2 - (p + m)}{2}$$

The expression for  $d$  follows the general principle for defining degrees of freedom as the number of data values minus the number of parameters being estimated. In the present context, the number of data values is the number of distinct entries in the covariance matrix, which is  $\frac{1}{2}p(p + 1)$ . (This is the number of entries in the diagonal and below; remember, this is a symmetric matrix, so we don't count all the entries in the full square matrix.) The effective number of parameters being estimated is  $\left\{ p + pm - \frac{1}{2}m(m - 1) \right\}$ .

This is the number of factor loadings ( $pm$ ) plus the number of unique variances ( $p$ ), minus a term that corrects for rotational indeterminacy of the loading matrix.

If  $(N - 1)F_{ML}$  is significant, we reject  $H_0$  — i.e. We reject the factor analysis model with  $m$  factors.

If  $(N - 1)F_{ML}$  is not significant we cannot reject  $H_0$  — i.e. We have failed to show that the factor analysis model with  $m$  factors does not fit exactly in the population. The hypothesis of exact fit is not rejected.

**Problem:** Although this may seem like a nice clean way to resolve the question of fit of the model, it doesn't work well. The primary problem is that we don't believe  $H_0$  anyway—we know it's not true. Therefore there is not much point in testing it. Note that the likelihood ratio test statistic is defined as  $(N - 1)F_{ML}$ , which follows a  $\chi^2$  distribution if the model is correct. Note that, even if  $F_{ML}$  is small, meaning that the discrepancy or lack of fit is small, a sufficiently large  $N$  will result in a significant  $\chi^2$ , resulting in turn in rejection of the model. Thus, if  $N$  is sufficiently large, virtually any parsimonious model will be rejected. The test essentially tells us whether the sample size is sufficiently large for the statistical test to tell us what we already know.

The best we can hope for is that the model fits really well but we cannot depend on this test to give us an idea of whether the model with  $m$  factors yields a good fit.

Instructor's opinion: The use of tests of perfect fit tests in factor analysis has done more harm than good.

We need alternative approaches for the assessment of model fit.

## 2. RMSEA = Root Mean Square Error of Approximation (Steiger & Lind, 1980; Browne & Cudeck, 1992)

$\hat{F}$ : Sample value of  $F_{ML}$  is known ( $\hat{F} \geq 0$ )

$F_o$ : Population value of  $F_{ML}$  is unknown ( $F_o \geq 0$ ). This is the value of  $F_{ML}$  that would be obtained if the model were fitted to the population covariance or correlation matrix.

$d$  = degrees of freedom

RMSEA :  $\varepsilon = \sqrt{\frac{F_o}{d}}$  Measure of discrepancy per degree of freedom

This is a population measure. We would like to be able to estimate  $\varepsilon$  in a sample. It appears that this could be achieved by substituting  $\hat{F}_{ML}$ , the sample value of the discrepancy function, for  $F_o$  in the equation for  $\varepsilon$ . However, this is not appropriate because  $\hat{F}$  is not an unbiased estimate of  $F_o$ .

$\hat{F}_o = \text{Max}\{\hat{F} - \frac{d}{N-1}, 0\}$  is a less biased estimate of  $F_o$ .

That is, we use the value of  $(\hat{F} - \frac{d}{N-1})$  as our estimate of  $F_o$ . However, if this value is negative, we use zero as our estimate of  $F_o$  since  $F_o$  can not be negative.

We substitute this estimate of  $F_o$  into the definition of  $\varepsilon$  to obtain a point estimate of RMSEA:  $\hat{\varepsilon} = \sqrt{\frac{\hat{F}_o}{d}}$

**Guidelines for interpretation of RMSEA:** (Browne & Cudeck, 1992)

$$\varepsilon \left\{ \begin{array}{l} < .05 \text{ indicates close fit} \\ .05 - .08 \text{ indicates reasonable fit.} \\ .08 - .10 \text{ indicates mediocre fit.} \\ > .10 \text{ indicates unacceptable fit.} \end{array} \right.$$

These are guidelines: they should not be used as cutoffs (for example, there isn't much difference between .049 and .052)

Note the role of degrees of freedom in the behavior of RMSEA. As one introduces more parameters into the model, for instance by increasing the number of factors, one loses degrees of freedom. The RMSEA index takes into account the differences in degrees of freedom for different models. For instance, if models for  $m = 2$  and  $m = 3$  yielded the same fit (same value of  $\widehat{F}_0$ ), the 2-factor model would yield the smaller RMSEA since it would have larger  $d$ . Thus, the 2-factor model would be preferred. Generally, for the same level of discrepancy as reflected in the discrepancy function value, a model with fewer parameters (more degrees of freedom) is preferred over a model with more parameters (fewer degrees of freedom).

**Confidence Intervals for RMSEA:** A very useful property of RMSEA is that it is possible to calculate confidence intervals for the true value of  $\varepsilon$ . Thus, in addition to the point estimate, we can obtain an indication of the precision of that estimate. The confidence interval for RMSEA is provided by the CEFA software.

e.g.  $\widehat{\varepsilon}_a = 0.06$   
 CI of (0.00 ; 0.14)  $\rightarrow$  We do not know whether the model fits well or badly  
 CI of (0.05 ; 0.07)  $\rightarrow$  We know that the model fits reasonably well

We strongly urge the use of such confidence intervals in practice. They provide all the information contained in a test of fit, plus additional information. If one really wishes to use some sort of hypothesis test of fit, we

strongly discourage the test of exact fit and instead recommend some alternatives that are based on more empirically sensible hypotheses.

**Test of Close Fit:**  $H_0: \varepsilon_a \leq 0.05$        $H_1: \varepsilon_a > 0.05$

This a more reasonable null hypothesis than for the test of perfect Fit (i.e.,  $H_0: \varepsilon_a = 0$ )

If *significant*: Reject  $H_0$  of close fit  $\Rightarrow$  We are reasonably sure that *Fit is not close*.

If not significant: Cannot Reject  $H_0$ , Model may fit closely but *we are not sure*. Look at confidence interval.

The test of close fit is provided by the CEFA software.

**Test of Not Close Fit:** (discussed in paper by MacCallum, Browne, & Sugawara, *Psychological Methods*, 1996) This test involves reversing the role of the null hypothesis from its conventional role in likelihood ratio tests. The null hypothesis is constructed as indicating absence of close fit:

$H_0: \varepsilon_a \geq 0.05$        $H_1: \varepsilon_a < 0.05$

If *significant*: Reject  $H_0$  of not close fit  $\Rightarrow$  We are reasonably sure that *Fit is close*.

If not significant: Cannot Reject  $H_0$ , Model may not fit closely but we are not sure.

Note: The confidence interval for RMSEA indicates the outcome of the test of exact fit, close fit, not-close fit, or any other hypothesis you would want to construct about RMSEA.

The outcome of the test of exact fit is indicated by the lower bound of the confidence interval. If that bound is 0, then exact fit is not rejected. If the lower bound is  $> 0$ , then exact fit is rejected.

If the lower limit of the 90% confidence interval is 0 then the hypothesis of perfect fit is not rejected at the 5% level.

If the lower limit of the 90% confidence interval is  $> 0$  then the hypothesis of perfect fit is rejected at the 5% level.

The outcome of the test of close fit is indicated by the lower bound of the confidence interval. If that bound is  $< 0.05$ , then close fit is not rejected. If that bound is  $> 0.05$ , then close fit is rejected.

If the lower limit of the 90% confidence interval is  $< 0.05$  then the hypothesis of close fit is not rejected at the 5% level.

If the lower limit of the 90% confidence interval is  $> 0.05$  then the hypothesis of close fit is rejected at the 5% level.

The outcome of the test of not-close fit is indicated by the upper bound of the confidence interval. If that bound is  $< 0.05$ , then not-close fit is rejected. If that bound is  $> 0.05$ , then not-close fit is not rejected.

If the upper limit of the 90% confidence interval is  $> 0.05$  then the hypothesis of not-close fit is not rejected at the 5% level.

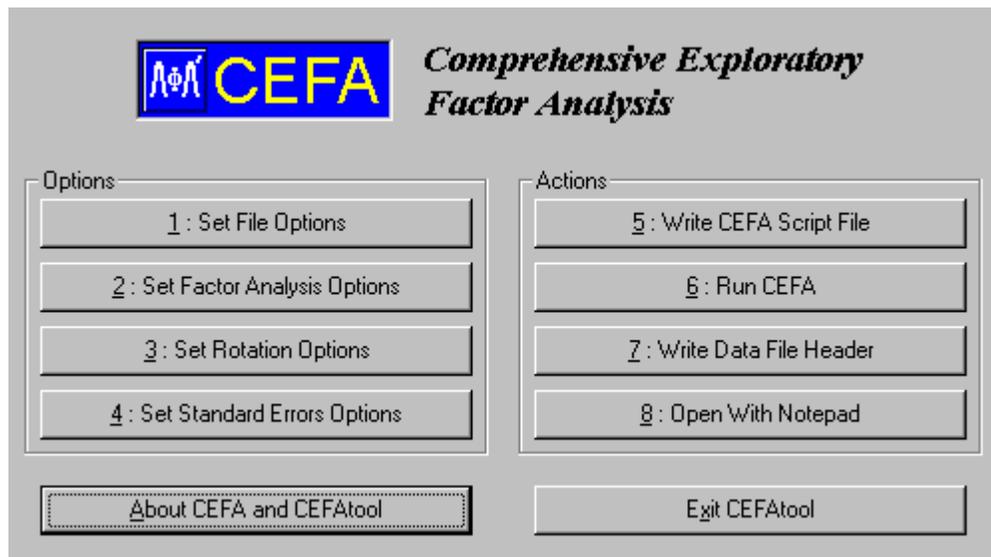
If the upper limit of the 90% confidence interval is  $< 0.05$  then the hypothesis of not-close fit is rejected at the 5% level.

*The confidence interval contains all information in the tests of fit plus additional information. This instructor would like to replace all tests of fit by confidence intervals.*

## PSY 820 Fitting the Factor Analysis Model Module 7 Computer Program Usage and Illustrations

A new exploratory factor analysis program CEFA (Comprehensive Exploratory Factor Analysis) will be employed in this course. This program operates under Windows 95/98 and can fit the exploratory factor analysis model using several discrepancy functions as well as a non iterative PACE method, rotate the factor matrix according to several simplicity criteria and obtain standard errors for rotated factor loadings.

CEFA is a FORTRAN 90 program for carrying out factor analysis, rotation and standard error computations, inputting data and outputting results. This may be run under the DOS prompt or Windows. It is more convenient, however, to run CEFA directly in the Windows environment using the Graphical User Interface, CEFAtool. To bring up this interface click on the CEFAtool icon . You will then see the main window



**Main Window**

Typically the Main Window will remain open until further CEFA runs are no longer required. It may be closed by clicking on the Exit CEFATool button.

CEFA requires two files, a data file, \*.dat, that contains the data to be analyzed, and a script or command file, \*.cmd, that gives instructions to CEFA on how to carry out the analysis. CEFA also produces an output file, \*.out, containing the results of the analysis.

The data file may be prepared using any text editor, for example Notepad. If you use a word processor, please ensure that the file is a text file and is not written in a special format.

Here we shall give an example of how CEFA may be used to carry out a factor analysis of unpublished data due to Holzinger reported in Harman (1960). This example is based on the files example.dat and example.cmd that come with CEFA. The names of the files have been changed here to Holzinger.\* since minor alterations have been made.

File: Holzinger.dat

```

696 9          (Number of cases, variables)
1             (Datatype)
0 0          (Random seed)
1             (Variable names)
WrdMean SntComp OddWrds
MxdArit Remndrs MissNum
Gloves Boots Hatchts
1             (Factor names)
Verb Arith Spat
1             (Target/Order matrix)

9 0 0        (A Target matrix is used for Target rotation. An
9 0 0        Order matrix is used to specify the ordering of
9 0 0        columns of a rotated matrix. 9 refers to an
0 9 0        unspecified element. Any element ≠ 9 represents a
0 9 0        specified value to be matched as closely as
0 9 0        possible.)
0 0 9
0 0 9
0 0 9

```

```

1
.75 1
.78 .72 1
.44 .52 .47 1
.45 .53 .48 .82 1
.51 .58 .54 .82 .74 1
.21 .23 .28 .33 .37 .35 1
.30 .32 .37 .33 .36 .38 .45 1
.31 .30 .37 .31 .36 .38 .52 .67 1

```

*(Sample correlation matrix)*

**Header information:**

Datatype: 1 for correlations/covariances, 2 for raw data, 3 for a factor loading matrix. If type 1 is specified a correlation or covariance matrix should be given at the end of the file as shown here. If type 2 is specified the data matrix should be given at the end of the file. If type 3 is specified a factor matrix (unrotated or orthogonally rotated) should be given at the end of the file.

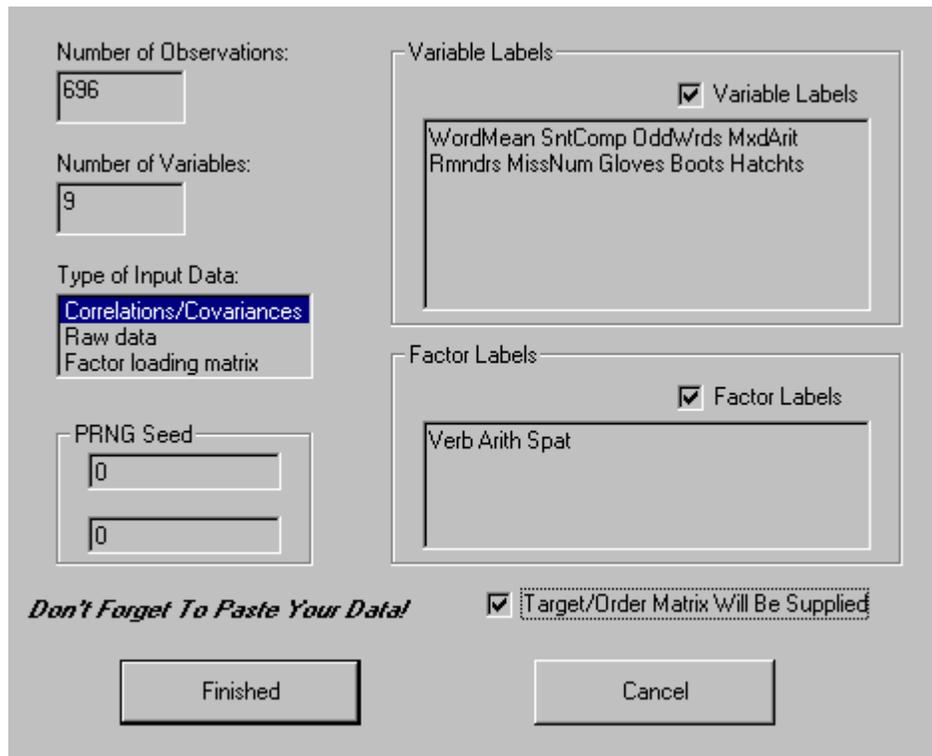
Random seed: Used for random starts in rotation. 0 0 means choose one based on the system clock; otherwise type in values given in a previously obtained output file.

Variable names: 1 if provided, 0 if not. If = 1, type names on following lines in free-format separated by blanks.

Factor names: 1 if provided, 0 if not. If = 1, type names on following lines in free-format separated by blanks.

Target/Order matrix: 1 if provided, 0 if not. If = 1, type the matrix on the following lines

Alternatively, the lines preceding the Target/Order matrix may be prepared using option 7 of the CEFAtool Main Window.

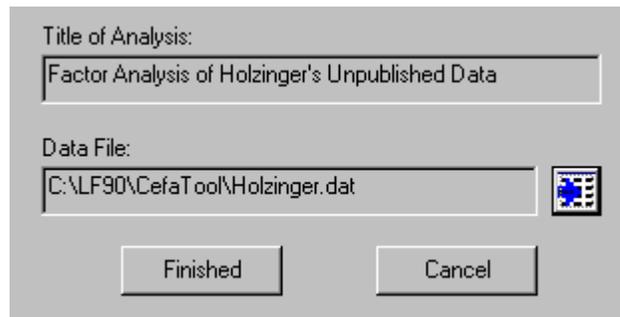


### Window 7: Write Initial Lines of the Data File

To prepare the initial lines of the Data file, click on 7: Write Data File Header from the Main Window. The window shown above appears.

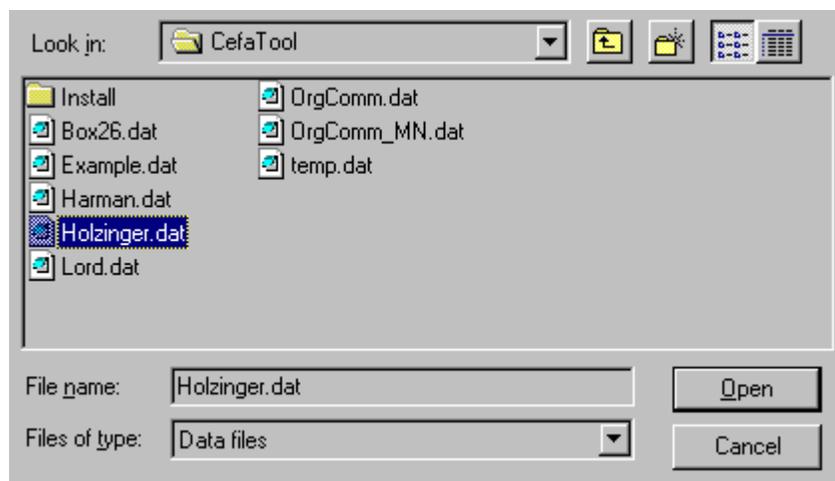
Fill in the required information and click on Finished. A file name is requested and the first lines of the data file are written. This file should then be opened with Notepad and the Target/Order matrix (if appropriate) and data should be typed in.

The data file may be used for several runs. A new script file is required for each run. Although the script file may be prepared using a text editor (e.g. Notepad) alone it is more convenient to prepare it using options 1-5 of the Main Window. First click on 1: Set File Options in the Main window. This will bring up the following window. In the figure shown, an appropriate title has been typed and the name and path of the Data file has been entered.



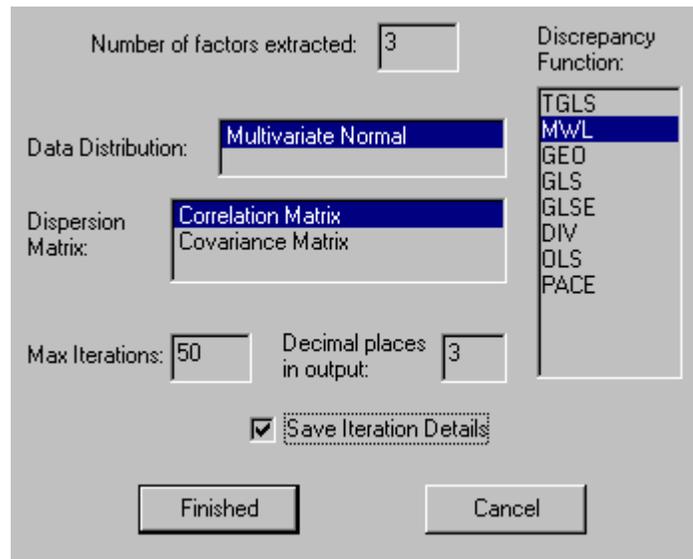
### Window 1: File Options

This is done by clicking on the icon on the right of Window 1, choosing the appropriate file from the list that appears



### Window 1a

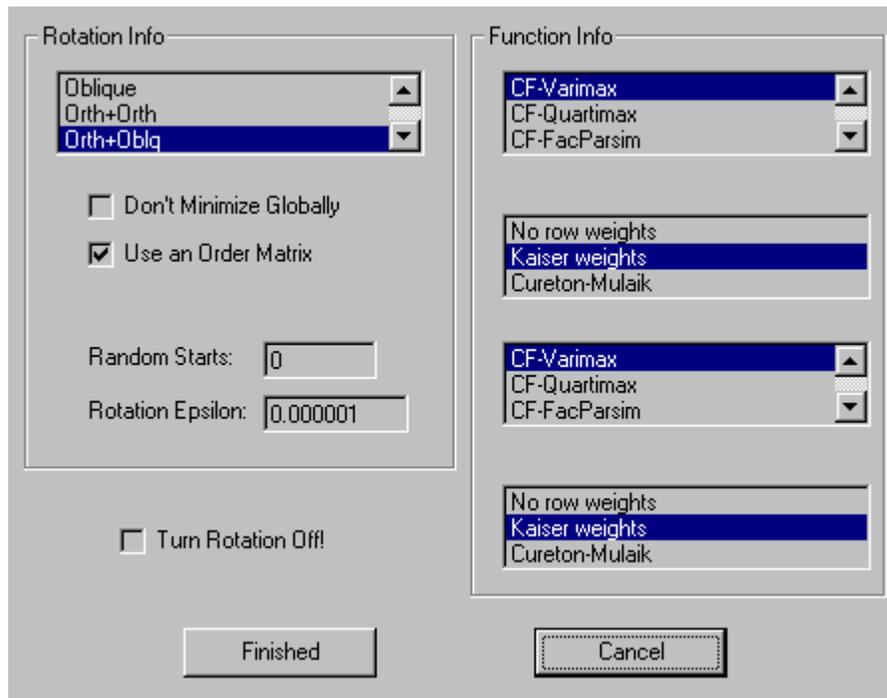
(see Window 1a) and clicking on the Open button on the right. Then close Window 1 by clicking on the Finished Button. Now click on 2: Set Factor Analysis Options in the Main Window and you will see Window 2.



## Window 2: Factor Analysis Options

In this Window the Number of factors extracted has been set to 3 and the defaults of Correlation Matrix for Dispersion Matrix, 50 for Maximum Number of Iterations , 3 for Decimal places in output and MWL for Discrepancy Function have been kept. The only option available in the current version of CEFA for Data Distribution is Multivariate Normal. In the present course the MWL, OLS and PACE options for estimation of factor loadings will be of primary interest and the Covariance Matrix option for Dispersion Matrix will not be used. Click on Save Iteration Details to obtain iteration details for both the Factor Analysis and Rotation procedures. Close Window 2 by clicking on the Finished button.

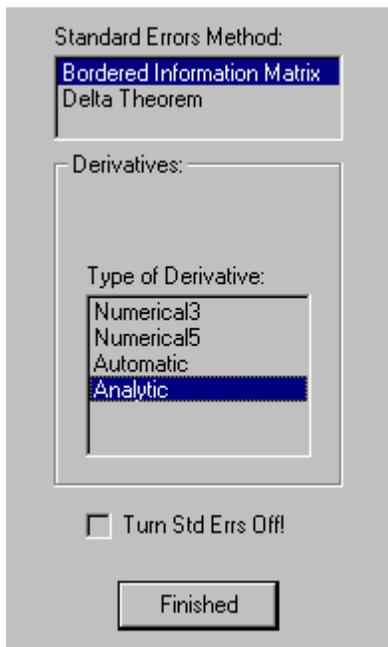
Now click on 3: Set Rotation Options in the Main Window to Obtain Window 3. In the window shown,



### Window 3: Rotation Options

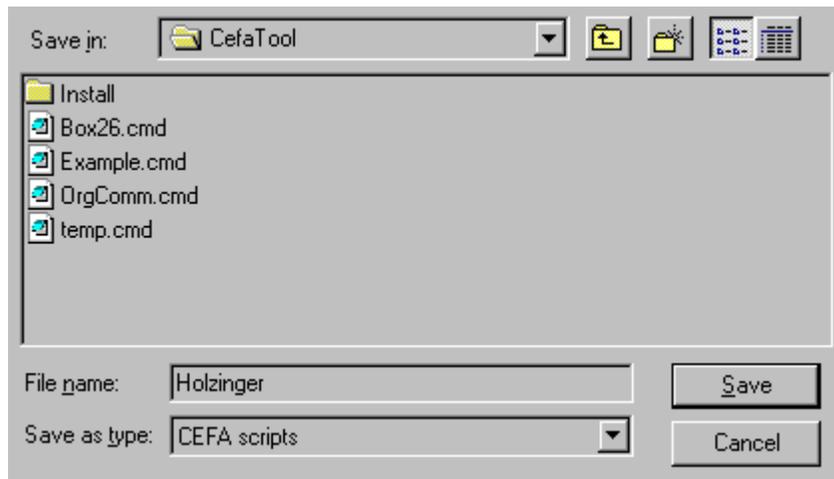
Orth+Oblq (orthogonal rotation followed by oblique rotation) has been chosen, the order matrix given in the Holzinger.dat file will be used, and the Crawford-Ferguson equivalent of Varimax rotation (CF-Varimax) has been chosen for both orthogonal and oblique rotation. More about this will be said in Module 10 on rotation. Several other rotation functions are available. One that will be dealt with in Module 10 is Target Rotation (Target). This is not visible in the picture but could be found on the screen by moving the highlighted area with the down arrow of the keyboard past CF-FacParsim. The default value, 0.0000001, of the rotation convergence criterion (Rotation Epsilon) yields a very accurate solution. With large problems, it could be made a little larger if time is a consideration. In those situations where one wishes to investigate possible local minima of the rotation criterion, Random Starts should be assigned a positive value (e.g. 10). The two initial random numbers required are then obtained from the Data File. When a previous run with random starts is to be duplicated, the previously obtained initial random numbers should be given in the Data File. If no rotation is required, click on Turn Rotation Off!

After clicking on Finished to close Window 3, go back to the Main window, click on 4: Set Standard Errors Options and obtain Window 4. The options chosen, Bordered Information Matrix and Analytic Derivatives, are recommended for general use in this course. If the requirements for correct standard errors are not met (for example, if the PACE or OLS options have been used in Window 2 or if a principal component matrix has been input with Datatype 3 in the Data file), ensure that a ✓ occurs next to Turn Rotation Off! by clicking if necessary. Note that if two consecutive rotations



#### **Window 4: Standard Errors Options**

have been chosen in Window 3, standard errors are only provided for the second option. After closing Window 4 with Finished, go to the Main Window and click on 5: Write CEFA Script File. This gives Window 5.



### Window 5: Write CEFA Script File

The file name for the script file must be specified either by clicking on a name in the list or by typing it.

The following script file is produced by CEFAtool when Windows 1-5 have been completed as shown earlier. The instructions produced appear in a typewriter font. Annotations have been inserted in italics.

---

File: Holzinger.cmd

Note that any line in the script not beginning with a keyword, or a blank line, is treated as a comment.

TITLE       Factor Analysis of Holzinger's Unpublished Data

*Up to 80 characters may be used for a descriptive title.*

DATAFILE   C:\LF90\CefaTool\Holzinger.dat

*The name of the data file need not include the full path, if CEFA.exe is invoked from the same directory.*

NUMFACT    3

*The number of factors extracted. This should match the number of columns of any supplied factor or order/target matrices in the data file!*

DISCFUN    2

*The discrepancy function can be: 1 (TGLS), 2 (MWL), 3 (GEO), 4 (GLS), 5 (GLSE), 6 (DIV), 7 (OLS), or 8 (PACE).*

DISPMAT 1  
*The dispersion matrix used in the exploratory factor analysis can be 1 (correlations) or 2 (covariances).*

DATADIST 1  
*At present, the data distribution can only be 1 (multivariate normal)*

MAXITER 50  
*The maximum number of iterations for the Newton algorithm in the exploratory factor analysis step of CEFA. (50 is the default)*

ITERINFO 1  
*This can be 1 (save iteration details) or 0. ITERINFO applies simultaneously to iterations for the factor extraction procedure and for the rotation procedure.*

FNAME CF-VARIMAX  
*The rotation criterion name for the first rotation can be: CF-VARIMAX, CF-QUARTIMAX, CF-FACPARSIM, CF-EQUAMAX, CF-PARSIMAX, MINENT1, INFOMAX, GEOMIN, CRAWFER, or TARGET. The first five are special cases of CRAWFER.*

ROWWT1 2  
*Row weights for the first rotation may be 1 (no row weights), 2 (Kaiser weights), or 3 (Cureton-Mulaik weights).*

FNAME2 CF-VARIMAX  
*If TYPROT is 3 or 4, a second rotation criterion is required.*

ROWWT2 2  
*Row weights for the second rotation may be 1 (no row weights), 2 (Kaiser weights), or 3 (Cureton-Mulaik weights).*

TYPROT 4  
*The type of rotation can be 1 (orthogonal), 2 (oblique), 3 (orthogonal followed by another orthogonal), or 4 (orthogonal followed by oblique).*

ROTEPS 0.000001  
 I

RNDSTARTS 0  
 I

USEORDMAT 1  
*This can be 1 (sort columns of the rotated factor loading matrix using the specified order matrix), or 0 (sort columns in descending order of sums of squares). The order matrix is provided in the data file.*

NOMIN 0  
*This can be 1 (find the closest local minimum instead of optimizing the rotation criterion), or 0.*

SERMETH 1  
*This can be 1 (use a bordered information matrix), or 2 (use the Delta theorem). The*

latter option is available for the Swain family of discrepancy functions, but not OLS or PACE.

TYPDER 4

The derivatives used for computation of standard errors can be 1 (3-point numerical), 2 (5-point numerical), 3 (automatic), or 4 (analytic).

DEREPS 0.001

The perturbation size for numerical derivatives. Not used since analytic derivatives are required.

NUMDP 3

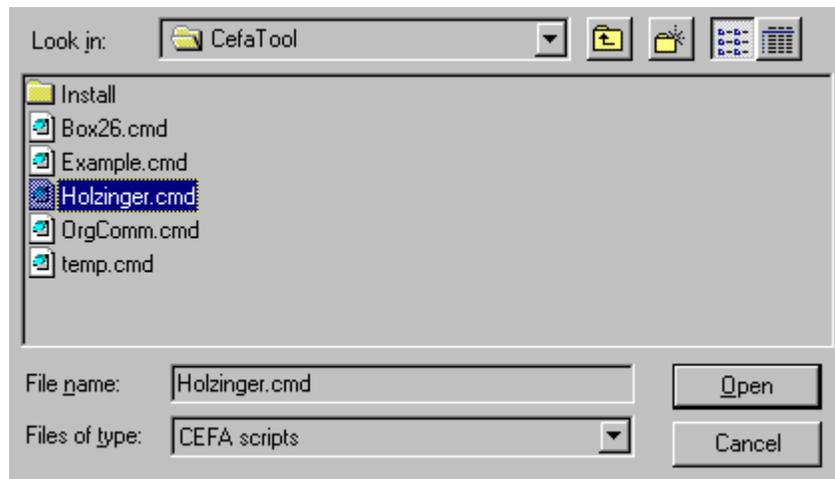
Number of decimal places *REQUIRED* in the output.

STOP

The last keyword in the script file must be *STOP*.

---

Once you have prepared the data and script files you may run CEFA. Click on 6: Run CEFA in the Main Window. First the following window appears.

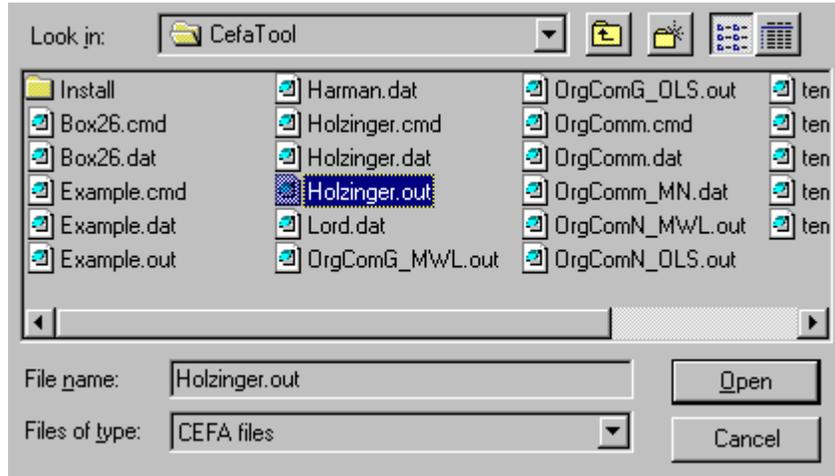


### Window 6a: Specify Script File to be Used in Run

Either type in the Script File name (the .cmd may be omitted) or click on an existing file that will be overwritten. Click on Open and the following window appears.



When the last line appears press the **Enter Key**. You may then open the output file using 8. Open with Notepad in the Main Window. The following window appears.



### Window 8: Open a File with Notepad

Click on the appropriate file name and then click on **Open**. Notepad opens the required file.

The CEFA run specified earlier produces the following output file. (The font has been changed to a smaller font for lines that would not fit on the page using the standard font.)

File: Holzinger.out

```

O-----O
| CEFA: Comprehensive Exploratory Factor Analysis |
|
|   Release Version 1.01   |
|   August 1998           |
|
|   Programming:         |
| Krishna Tateneni, Michael W. Browne, |
| Gerhard Mels, and Robert Cudeck.   |
|
|   Mathematical Specification:   |
| Michael W. Browne, Krishna Tateneni, |

```

| Robert Cudeck, and Gerhard Mels. |  
o-----o

Date: 09/10/1998  
Time: 22:31:33

\*\*\*\*\*  
\* Factor Analysis of Holzinger's Unpublished Data \*  
\*\*\*\*\*

o=====o  
| Details of Analysis |  
o=====o

Data file : C:\LF90\CefaTool\Holzinger.dat

Number of observations : 696  
Number of variables : 9  
Number of factors : 3

Data follow multivariate normal distribution

Discrepancy function : MWL  
Dispersion matrix : Correlations  
Max EFA iterations : 50  
Save iteration information

Rotation type : Orthogonal+Oblique  
Sort columns using : Order matrix

Rotation Criterion : CF-VARIMAX  
Row weights : Kaiser

Rotation Criterion 2 : CF-VARIMAX  
Row weights : Kaiser

Rotation convergence : 0.100E-05

Std Errors using : Bordered information matrix  
Derivatives : Analytic

```

o=====o
| Sample Correlation Matrix |
o=====o

```

	WrdMean	SntComp	OddWrds	MxdArit	Remndrs	MissNum	Gloves	Boots	Hatchts
WrdMean	1.000								
SntComp	0.750	1.000							
OddWrds	0.780	0.720	1.000						
MxdArit	0.440	0.520	0.470	1.000					
Remndrs	0.450	0.530	0.480	0.820	1.000				
MissNum	0.510	0.580	0.540	0.820	0.740	1.000			
Gloves	0.210	0.230	0.280	0.330	0.370	0.350	1.000		
Boots	0.300	0.320	0.370	0.330	0.360	0.380	0.450	1.000	
Hatchts	0.310	0.300	0.370	0.310	0.360	0.380	0.520	0.670	1.000

Eigenvalues of Sample Correlation Matrix:

```

0.48E+01 0.14E+01 0.11E+01 0.56E+00 0.32E+00 0.27E+00 0.26E+00
0.21E+00 0.14E+00

```

```

o=====o
| Target/Order Matrix |
o=====o

```

	Verb	Arith	Spat
WrdMean	9.000	0.000	0.000
SntComp	9.000	0.000	0.000
OddWrds	9.000	0.000	0.000
MxdArit	0.000	9.000	0.000
Remndrs	0.000	9.000	0.000
MissNum	0.000	9.000	0.000
Gloves	0.000	0.000	9.000
Boots	0.000	0.000	9.000
Hatchts	0.000	0.000	9.000

\*\*\*\*\*

\* Exploratory Factor Analysis Details \*

\*\*\*\*\*

=====0  
 | Non iterative Unique Variances, Communalities, and SMCs |  
 =====0

Variable	Unique Variance	Communality	SMC
WrdMean	0.177	0.823	0.684
SntComp	0.294	0.706	0.648
OddWrds	0.248	0.752	0.672
MxdArit	0.063	0.937	0.775
Remndrs	0.268	0.732	0.704
MissNum	0.246	0.754	0.724
Gloves	0.631	0.369	0.325
Boots	0.427	0.573	0.484
Hatchts	0.207	0.793	0.525

=====0  
 | Iteration Details for 3 Factors |  
 =====0

-----  
 MWL Iteration Details  
 -----

Iter	Disc Fun	Max Grad	Step	NSH	NZP	POB
0	0.132764741017	0.8371908				
1	0.089219229954	1.4320760	8.6E-1	0	0	1
2	0.027357386318	0.3266576	1.0E+0	0	0	0
3	0.016690327543	0.0530229	1.0E+0	0	0	0
4	0.016359700002	0.0022382	1.0E+0	0	0	0
5	0.016359082741	0.0000046	1.0E+0	0	0	0
6	0.016359082738	0.0000000	1.0E+0	0	0	0

=====0  
 | MWL Unrotated Factor Loadings |

o=====o

	Fac1	Fac2	Fac3
WrdMean	0.678	0.533	-0.273
SntComp	0.717	0.377	-0.231
OddWrds	0.699	0.486	-0.164
MxdArit	0.900	-0.340	-0.033
Remndrs	0.836	-0.200	0.028
MissNum	0.862	-0.128	0.002
Gloves	0.421	0.089	0.431
Boots	0.478	0.246	0.540
Hatchts	0.483	0.302	0.673

o=====o

| MWL Unique Variances and Communalities |

o=====o

Variable	Unique Variance	Communality
WrdMean	0.181	0.819
SntComp	0.290	0.710
OddWrds	0.248	0.752
MxdArit	0.073	0.927
Remndrs	0.260	0.740
MissNum	0.241	0.759
Gloves	0.629	0.371
Boots	0.420	0.580
Hatchts	0.222	0.778

o=====o

| MWL Scaled Standard Deviations |

o=====o

1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

o=====o

| MWL Matrix of Residuals |

o=====o

	WrdMean	SntComp	OddWrds	MxdArit	Remndrs	MissNum	Gloves	Boots	Hatchts
WrdMean	0.000								
SntComp	-0.001	0.000							
OddWrds	0.002	-0.003	0.000						
MxdArit	0.002	-0.005	0.000	0.000					
Remndrs	-0.003	0.012	-0.003	0.000	0.000				
MissNum	-0.005	0.011	0.000	0.001	-0.006	0.000			
Gloves	-0.006	-0.006	0.012	-0.005	0.024	-0.002	0.000		
Boots	-0.008	0.009	0.005	0.001	-0.005	-0.001	-0.006	0.000	
Hatchts	0.005	-0.005	-0.005	0.000	-0.002	0.002	-0.001	0.001	0.000

Value of the maximum absolute residual = 0.0237

o=====o  
| Measures of Fit |  
o=====o

Sample discrepancy function value : 0.01635908

Population discrepancy function value, Fo  
Bias adjusted point estimate : 0.000  
90 percent confidence interval : ( 0.000; 0.016)

Root mean square error of approximation  
Steiger-Lind RMSEA = SQRT(Fo/DF)  
Point estimate : 0.000  
90 percent confidence interval : ( 0.000; 0.037)

Expected cross-validation index  
Point estimate (modified AIC) : 0.111  
90 percent confidence interval : ( 0.112; 0.129)  
ECVI (modified AIC) for the saturated model : 0.129

Chi-square test statistic : 11.370

Exceedance Probabilities  
Perfect fit (Ho: RMSEA = 0.0) : 0.498  
Close fit (Ho: RMSEA <= 0.05) : 0.995

Multiplier for obtaining test statistic : 695.0  
Degrees of freedom : 12  
Effective number of parameters : 33

\*\*\*\*\*

\* Rotation Details \*

\*\*\*\*\*

o=====o  
| Row Weights (Not Necessarily Used!) |  
o=====o

Kaiser weights, Cureton-Mulaik weights, and  
their products (final Cureton-Mulaik weights)

WrdMean	1.105	0.712	0.787
SntComp	1.187	0.479	0.568
OddWrds	1.153	0.531	0.612
MxdArit	1.039	0.649	0.674
Remndrs	1.163	0.427	0.496
MissNum	1.148	0.302	0.346
Gloves	1.641	0.833	1.366
Boots	1.313	0.882	1.158
Hatchts	1.134	0.965	1.094

o=====o  
 | Iteration Information |  
 o=====o

Cycle	FunVal	MinDelta	MaxDelta	BMin	BMax
1	4.411363549826	0.100818615066	0.299376266561	6	6
2	4.311529083293	0.008900917219	0.062770187651	6	6
3	4.311105523592	0.000724662766	0.003666643119	5	6
-----					
Switched to Minimum Bracketing Procedure...					
-----					
4	4.311102073949	0.000065628980	0.000344054302	5	5
5	4.311102048857	0.000005567484	0.000029024751	5	5
6	4.311102048670	0.000000479260	0.000002521433	5	5
7	4.311102048669	0.000000041104	0.000000215414	3	5

o=====o  
 | CF-VARIMAX Rotated Factor Matrix |  
 o=====o

	Verb	Arith	Spat
WrdMean	0.865	0.217	0.152
SntComp	0.754	0.341	0.156
OddWrds	0.794	0.253	0.243
MxdArit	0.245	0.913	0.182
Remndrs	0.285	0.770	0.256
MissNum	0.362	0.747	0.263
Gloves	0.094	0.230	0.556
Boots	0.188	0.166	0.719
Hatchts	0.171	0.122	0.857

CF-VARIMAX Criterion: 4.31110204866866

o=====o  
 | Factor Correlations |  
 o=====o

	Verb	Arith	Spat
Verb	1.000		
Arith	0.000	1.000	
Spat	0.000	0.000	1.000

Cycle	FunVal	MinDelta	MaxDelta	BMin	BMax
1	2.018089064117	0.115999590199	0.320204158735	5	6
2	1.922289220701	0.005520547315	0.075933030096	5	6
3	1.922002501854	0.000500000000	0.005161351897	5	6
-----					
Switched to Minimum Bracketing Procedure...					
-----					
4	1.921982807520	0.000121157339	0.000992152826	5	6
5	1.921981464742	0.000045066359	0.000320612909	5	6
6	1.921981416578	0.000011328529	0.000055951565	5	5
7	1.921981414949	0.000002250726	0.000009766887	5	5
8	1.921981414886	0.000000438514	0.000001829679	5	5
9	1.921981414883	0.000000087346	0.000000367692	4	5

o=====o  
| CF-VARIMAX Rotated Factor Matrix |  
o=====o

	Verb	Arith	Spat
WrdMean	0.897	0.008	0.012
SntComp	0.739	0.178	0.008
OddWrds	0.796	0.052	0.112
MxdArit	0.012	0.959	-0.006
Remndrs	0.086	0.772	0.095
MissNum	0.178	0.726	0.097
Gloves	-0.022	0.150	0.545
Boots	0.080	0.026	0.720
Hatchts	0.057	-0.044	0.879

CF-VARIMAX Criterion: 1.92198141488323

o=====o  
| Factor Correlations |  
o=====o

	Verb	Arith	Spat
Verb	1.000		
Arith	0.485	1.000	
Spat	0.329	0.387	1.000

\*\*\*\*\*  
\* Standard Errors Details \*

\*\*\*\*\*

Maximum constraint violation: 0.995E-06

Elapsed Seconds: 0

o=====o  
| Standard Errors after Rotation |  
o=====o

	Verb	Arith	Spat
WrdMean	0.018	0.018	0.017
SntComp	0.022	0.024	0.021
OddWrds	0.020	0.020	0.020
MxdArit	0.017	0.016	0.014
Remndrs	0.021	0.020	0.020
MissNum	0.022	0.021	0.020
Gloves	0.029	0.032	0.034
Boots	0.027	0.025	0.031
Hatchts	0.026	0.024	0.033

	Verb	Arith	Spat
Verb	0.000		
Arith	0.025	0.000	
Spat	0.032	0.028	0.000

o=====o  
| One-At-A-Time 90% Confidence Intervals |  
o=====o

	Verb	Arith	Spat
WrdMean	( 0.867; 0.927)	(-0.022; 0.038)	(-0.016; 0.040)
SntComp	( 0.704; 0.775)	( 0.139; 0.217)	(-0.027; 0.042)
OddWrds	( 0.763; 0.829)	( 0.018; 0.086)	( 0.079; 0.146)
MxdArit	(-0.015; 0.040)	( 0.932; 0.986)	(-0.029; 0.018)
Remndrs	( 0.052; 0.121)	( 0.740; 0.805)	( 0.063; 0.127)
MissNum	( 0.141; 0.214)	( 0.692; 0.760)	( 0.064; 0.130)
Gloves	(-0.070; 0.025)	( 0.096; 0.203)	( 0.489; 0.600)
Boots	( 0.035; 0.126)	(-0.016; 0.068)	( 0.669; 0.772)
Hatchts	( 0.015; 0.099)	(-0.083;-0.004)	( 0.824; 0.933)

Arith ( 0.441; 0.525)  
Spat ( 0.276; 0.381) ( 0.340; 0.432)

o=====o  
| CEFA Completed |  
o=====o

**PSY 820     Estimating the Number of Common Factors**  
**Theory and Methodology** **Module 8**

The objective of exploratory factor analysis is to determine the number and nature of the major common factors. Therefore, a key step in an exploratory factor analysis is the decision as to the appropriate number of factors to retain. The subsequent step in such analyses involves rotation, and results of rotation are dependent on the number of factors retained. There is a great deal of research literature on the number-of-factors problem. Unfortunately, much of this literature is directed toward development of mechanical rules that will reliably indicate the "correct" number of factors to retain. The instructor considers such efforts to be misguided for several reasons. First, as will be discussed below, there is generally no correct or true number of factors. Second, mechanical rules are not fail-safe; no such rule will work well all of the time. Rather, the decision about the number of factors must be based first on a recognition that the model does not hold exactly for any parsimonious number of factors, and must consider several kinds of information. One must use informed judgment to assess this information and make a decision. In some cases there may be no extremely clear answer as to the best number of factors, and a final decision must be based on interpretability of rotated solutions based on different numbers of factors.

The applied factor analysis literature is for the most part characterized by an unfortunate mechanical approach to the number of factors question. This problem is exacerbated by widely used software that makes automated decisions about the number of factors. Many investigators use simple rules of thumb mechanically, without careful judgment or understanding. Such an approach is generally ineffective and can result in misleading and distorted solutions and interpretations. Rules of thumb do provide useful information, but various criteria must be considered in combination with each other and with the investigator's judgment as well as the interpretability of the resulting factors. There is a tradeoff to be considered: we are seeking a number of factors such that the resulting solution fits the data reasonably well but still provides a degree of parsimony. Ideally what we will find is a number of factors such that good fit of the model to the data is obtained using that number, while using fewer factors results in substantially poorer fit and using more factors does not result in substantially better fit. And we would also find those retained factors to be substantively meaningful and interpretable.

We will not identify a parsimonious number of factors that explains our data perfectly. As mentioned several times, we must recognize that the common factor model does not hold exactly in either the population or the sample. Lack of fit of the factor analysis model to a sample correlation matrix is due to both Sampling Error and Error of Approximation

**Sampling Error:** Even if the Model fitted exactly in the Population

$$\mathbf{P} = \mathbf{A}\mathbf{A}' + \mathbf{D}_\psi \quad (1)$$

it wouldn't hold exactly in the sample because

$$\mathbf{R} - \mathbf{P} = \mathbf{E}$$

where the off diagonal elements of  $\mathbf{E}$  are random sampling errors (not independent).

**Error of Approximation.** Even if the Population correlation matrix  $\mathbf{P}$  were available, it is not realistic to suppose that it satisfies the factor analysis model (with few factors) exactly. All that can be hoped is that it can be reasonably approximated by the factor analysis model: that is, it is possible to find a  $\mathbf{A} : p \times m$  and a  $\mathbf{D}_\psi$  such that

$$\mathbf{P} \approx \mathbf{A}\mathbf{A}' + \mathbf{D}_\psi \quad (2)$$

Causes of error of approximation may include such phenomena as nonlinear influences of factors on manifest variables, and the influences of large numbers of minor common factors that can not be represented in a parsimonious model.

**“True” number of Factors?** If  $m$  is taken to be equal to  $p - 1$  then the factor analysis model will fit exactly to any population correlation matrix and every sample correlation matrix. That is, it is not a falsifiable model. The factor analysis model is only of interest if  $m \ll p$  since the whole aim is parsimony. It is reasonable only to suppose that the model holds as an approximation (2) with  $m \ll p$ .

Therefore, *there is no unambiguously correct number of factors*. There is *no true number* of factors. Consequently, any rule that attempts to provide the number of factors cannot ever be completely satisfactory.

We must clearly define our objective as identifying the number of *major* common factors. These factors help explain the correlations among the manifest variables, but will probably not do so perfectly. If there is some optimal number of major factors, then it would be the case that a model with fewer factors would do considerably worse at explaining the correlations among the manifest variables, and a model with more factors would not do much better. This point is relevant when evaluating some kinds of information to aid in making this decision.

The information available for aiding in the decision about the number of factors depends on the methods of factor extraction employed. In OLS methods, users typically focus on eigenvalues. In ML and PACE, users focus on tests and measures of model fit. Each of these kinds of information will be discussed in turn.

### **Some rules of thumb for choosing the number of factors that are based on eigenvalues.**

In much of the applied factor analysis literature, the decision about the number of factors is based on some simple "rule of thumb." These rules may have some theoretical foundation, but in general they can not be relied upon to provide a reliable indicator of the best number of factors.

(1) **Number of eigenvalues of the correlation matrix that are greater than one.** Guttman (1954) has shown that in the idealized situation where the model is exact (1) this is a lower bound to the number of factors. That is, when the common factor model with  $m$  factors holds exactly in the population, then the number of eigenvalues of the unreduced population correlation matrix  $\mathbf{P}$  that are  $\geq 1.0$  will be a lower bound for  $m$ . This number is called Guttman's "weakest lower bound." This bound approaches  $m$  if  $p$  increases and  $m$  remains constant. Kaiser (1960) has suggested applying this rule in practice when a sample correlation matrix is used.

Although there is a theoretical foundation for this rule in an ideal case in the population, it has come to be used routinely in factor analysis of sample

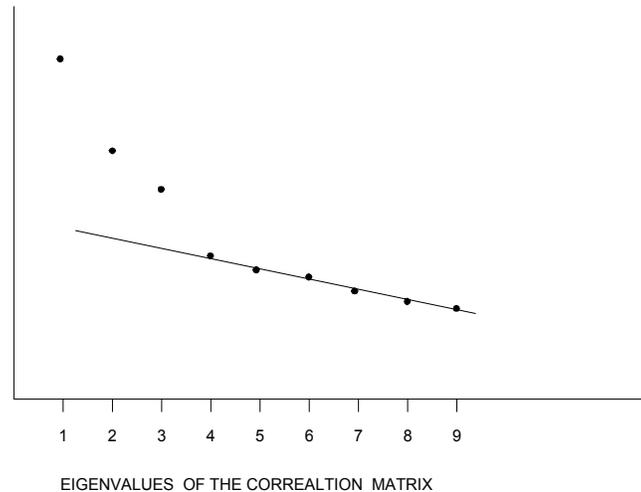
correlation matrices. It is probably the most widely used rule for determining the number of factors. This wide usage is unjustified in that it has been shown that the rule is highly fallible (Cattell & Vogelmann, 1977; Hakstian, Rogers, & Cattell, 1982; Linn, 1968; Tucker, Koopman, & Linn, 1969; Zwick & Velicer, 1982, 1986). It is biased upwards in small samples (overestimate of lower bound in idealized situation). In various conditions it may result in overfactoring or underfactoring. It can suggest an initial approximation but should never be believed without further evidence.

A common misuse of the eigenvalues-greater-than-one rule exists in the applied factor analysis literature. Often, users will conduct a (noniterative) principal factors analysis using prior communality estimates, and will then consider the eigenvalues of the reduced sample matrix ( $\mathbf{R} - \hat{\mathbf{D}}_\psi$ ). They will count how many of these eigenvalues are  $>1.0$  and will retain that number of factors. This approach is doubly misguided. First, it is a purely mechanical approach to a problem that requires reason and judgment. Second, it has no theoretical foundation at all, since the theoretical foundation of the eigenvalues-greater-than-one rule is with reference to the eigenvalues of  $\mathbf{R}$ , with unities in the diagonal, and not the eigenvalues of  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ , which has communality estimates in the diagonal.

The eigenvalue-greater-than-one rule is widely used in practice, both correctly as applied to  $\mathbf{R}$  and incorrectly as applied to  $(\mathbf{R} - \hat{\mathbf{D}}_\psi)$ . Regardless of whether it is used correctly or not, there is much evidence that it does not work consistently well and probably should be abandoned (see references given above). Its wide usage in the literature suggests that many published applications of factor analysis probably are characterized by a poor decision about the number of factors.

Guttman (1954) also provided a proof regarding a different lower bound for the number of factors. He showed that the number of positive eigenvalues of the population correlation with diagonal elements replaced by SMCs (that is, matrix  $(\mathbf{R} - \mathbf{D}_\psi)$ ), is a better lower bound when the common factor model in (1) holds exactly in the population with  $m$  factors. This is Guttman's "strongest lower bound," and will always produce a value greater than or equal to the weakest lower bound. This rule is not applied in practice to sample correlation matrices as the number of factors chosen is usually obviously too large (worse overestimate of better lower bound in idealized situation).

(2) **Scree “test”:** (Cattell, 1966) The scree test is a procedure for inspecting the series of eigenvalues in an effort to identify the last large discontinuity in their sequence. A plot is constructed using the sequence of eigenvalues of correlation matrix ( $\mathbf{R}$ ), or correlation matrix with communality estimates in diagonal ( $\mathbf{R} - \widehat{\mathbf{D}}_\psi$ ). The horizontal axis is the eigenvalue number (1, 2, 3, etc.), and the vertical axis is the eigenvalue magnitude or value. For example:



We then choose number of factors that precedes the last large drop, in this case 3.

In making use of such eigenvalue plots in practice, one must consider which eigenvalues to plot. For instance, we could plot the eigenvalues of  $\mathbf{R}$ , or the eigenvalues of ( $\mathbf{R} - \widehat{\mathbf{D}}_\psi$ ). It is probably more sensible to plot the latter, especially in the case of a noniterative principal factors analysis with prior communality estimates, since those eigenvalues correspond to the extracted principal factors. However, users must keep in mind that any such results are dependent on our choice of communality estimates.

This is a subjective procedure and does not have a strong theoretical justification but can be used to give an initial approximation. There is only an informal rationale for this approach. Recall that eigenvalues represent the variance accounted for by a particular factor. The rationale is that, if there exists a specific number  $m$  of major common factors, then there will be  $m$  relatively large eigenvalues, and when the  $m$ th factor is extracted the eigenvalues will show a sharp drop and remaining eigenvalues will be relatively small. Those remaining factors represent "noise," or effects of

random or small systematic influences. Thus, we look for the last sharp drop in the eigenvalue plot. The procedure has been found to work well in practice when a clear number of strong common factors are present (Tucker et al., 1969; Hakstian et al., 1982).

**Goodness of fit tests.** (Maximum Likelihood as well as other members of the Swain family & PACE)

**1. Test of perfect fit:** Direct hypothesis test of the factor analysis model with  $m$  factors.

$H_0: \Sigma = \Lambda\Lambda' + D_\psi$  i.e., FA Model holds exactly in the population with  $m$  factors

$H_1: \Sigma$  has no particular structure

Under  $H_0$ , the likelihood ratio test statistic  $(N - 1)F_{ML} \sim \chi^2$  with degrees of freedom given by

$$d = \frac{1}{2}p(p + 1) - \left\{ p + pm - \frac{1}{2}m(m - 1) \right\}$$

$$= \frac{(p - m)^2 - (p + m)}{2}$$

If  $N^*F_{ML}$  is significant, we reject  $H_0$  — i.e. We reject the factor analysis model with  $m$  factors.

Although the standard multiplier of  $F_{ML}$  is  $N - 1$ , some alternative multipliers have been proposed so as to produce a test statistic that more closely approximates a chi-square distribution.

One important alternative is given by  $N^* = N - \frac{2p+11}{6} - \frac{2m}{3}$

This is called the Bartlett correction. It is provided in some commercial software (SAS) and is justified for *unrestricted* factor analysis when the model holds exactly in the population.

If  $N^*F_{ML}$  is not significant we cannot reject  $H_0$  — i.e. We have failed to show that the factor analysis model with  $m$  factors does not fit exactly in the population.

This test is often used in empirical research to estimate the number of factors. The procedure involves a sequence of tests, beginning with  $m = 0$ . At this first step, we are testing the hypothesis that the model holds exactly with 0 common factors. Equivalently, we are testing the hypothesis that the manifest variables are uncorrelated in the population. (If there are no factors, then there are no correlations among manifest variables.) We conduct the test as shown above. If the resulting test statistic is not significant, we do not reject the hypothesis of exact fit for  $m = 0$ , and we stop. However, if we do reject this hypothesis, which will usually be the case, we are rejecting the hypothesis that the model holds with  $m = 0$ , which implies that  $m > 0$ . Therefore, we set  $m = 1$  and repeat the process. If we do not reject  $H_0$  for  $m = 1$ , then  $m = 1$  is plausible and we stop. If we reject  $H_0$  for  $m = 1$ , then  $m = 1$  is insufficient and we increase  $m$  to 2 and repeat the test. We continue the sequence of tests until we obtain a test statistic that is not significant, and we retain that number of factors.

Although this might seem like an elegant approach with a strong theoretical/statistical basis, there are major problems. In general, as mentioned earlier, we discourage the use of mechanical rules for making the number-of-factors decision, and this is a purely mechanical rule. More specifically, there are difficulties with the test being used.

**Problem:** We don't believe  $H_0$  anyway—we know it is not true. Therefore there is not much point in testing it. The test essentially tells us whether the sample size is sufficiently large for the statistical test to tell us what we already know.

The best we can hope for is that the model fits really well but we cannot depend on this test to give us an idea of whether the model with  $m$  factors yields a good fit.

Difficulties with a sequence of tests of perfect fit for  $m = 1, 2, 3, \dots$

- (1) Theoretical properties not known. Tests in the sequence are not independent.
- (2) More factors chosen as  $N$  increases (power of test increases and small error of approximation is detected). That is, for any particular number of factors, the test statistic will be larger as sample size increases, thus implying

greater likelihood of rejecting  $H_0$  for that number of factors. Therefore, the number of factors determined by this approach will increase as  $N$  increases. This may lead to over factoring with large  $N$ , and under factoring with small  $N$ .

Instructor's opinion: The use of tests of perfect fit in factor analysis does more harm than good.

We need to consider other kinds of information to help us make the decision about the number of factors. We can make use of information about goodness of fit as described in Module 6.

**2. RMSEA = Root Mean Square Error of Approximation**  
(Steiger & Lind, 1980; Browne & Cudeck, 1992)

$\hat{F}$ : Sample value of  $F_{ML}$  is known ( $\hat{F} \geq 0$ )

$F_o$ : Population value of  $F_{ML}$  is unknown ( $F_o \geq 0$ )

$d$  = degrees of freedom

RMSEA :  $\varepsilon_a = \sqrt{\frac{F_o}{d}}$  Measure of *discrepancy per degree of freedom*

$\hat{F}$  is a biased estimate of  $F_o$ .  $\hat{F}_o = \text{Max}\left\{\hat{F} - \frac{d}{N-1}, 0\right\}$  is a less biased estimate of  $F_o$ .

Point estimate of RMSEA:  $\hat{\varepsilon}_a = \sqrt{\frac{\hat{F}_o}{d}}$

- Guidelines:**
- $\varepsilon_a < .05$  indicates close fit.
  - $.05 - .08$  indicates reasonable fit.
  - $.08 - .10$  indicates mediocre fit.
  - $> .10$  indicates unacceptable fit.

These are guidelines: they should not be used as cutoffs (for example, there isn't much difference between .049 and .052)

In Module 6 information was presented regarding confidence intervals and hypothesis tests (close-fit, not-close-fit) using RMSEA. That information is applicable in the present context.

One can use RMSEA to aid in the decision about the number of factors. A sequence of factor analysis solutions would be obtained, beginning with 1 factor and proceeding through some number of factors that is as large as would reasonably be of interest. Values of RMSEA (along with confidence intervals and hypothesis tests) can then be inspected in search of an appropriate number of factors. Ideally we would find a number such that good fit was obtained using that number, while using fewer factors would result in substantially poorer fit and using more factors would not result in substantially better fit.

### **Tucker-Lewis (1973) Coefficient**

One widely used index for aiding in determining the number of factors in ML factor analysis was proposed by Tucker and Lewis (1973). This index is based on comparing the fit of a model with  $m$  factors to two reference point models. The first is a worst case model, a model with 0 (zero) factors. The zero factor model is more generally referred to as a "null model." This model implies that the MVs are uncorrelated in the population. The second reference point model is a best case model, or ideal model, that would fit exactly in the population.

The definition of the Tucker-Lewis index is based on properties of the  $\chi^2$  test statistic. When a given model is correct,  $\mathcal{E}(\chi^2) = d$ . Thus, we would expect the ratio  $\chi^2/d \approx 1$  for a correct model. When a model is incorrect, this ratio becomes  $\gg 1$ . Suppose we are evaluating the fit of a model with  $m$  factors. Call this model  $m$ . Define

$$Q_m = \chi_m^2/d_m$$

For the null model or zero factor model, define

$$Q_0 = \chi_0^2/d_0$$

Since the null model would probably fit the data very poorly, we expect  $Q_0$  to be very large. For an ideal model, we would expect  $Q \approx 1$ . The Tucker-Lewis index for the  $m$ -factor model is then defined as

$$TL_m = \frac{Q_0 - Q_m}{Q_0 - 1}$$

The rationale for this expression is as follows. The numerator represents the improvement in fit of the  $m$ -factor model over the zero-factor model. The denominator represents the improvement in fit of an ideal model over the zero-factor model. Thus, the ratio indicates where the  $m$ -factor model lies on a continuum where the reference points are the zero-factor model and an ideal model. If this index is used in practice, it is desirable that its value be well above .90 to indicate good fit, preferably in the range of .94 or better (Hu & Bentler, 1999). One important limitation of this index is that no confidence interval is available.

In practice, one could use the T-L index to aid in determining the number of factors. The value of the index could be computed for successive numbers of factors, seeking the level of  $m$  where the T-L index indicated adequate fit of the model.

### **How to choose the number of factors to be employed.**

- (1) Consider prior hypotheses about the number of factors.
- (2) Possibly consider rules of thumb, such as eigenvalue plot, to aid in obtaining a rough approximation of the appropriate number of factors.
- (3) Consider results of tests of model fit and measures of model fit, including RMSEA and the T-L index.
- (4) Based on these findings, establish an optimal number of factors, or a very small range.

Extract this number as well as one or two fewer and one or two more. Try to interpret factor loadings after rotation. Did you choose the right number or do you want to change your mind.

- (5) If you increase  $m$  can you still interpret the results. Avoid increasing until some factors have nonzero loadings for single variables (unique factors).

A final decision is based on a combination of interpretability, model fit, and parsimony.

In general, underfactoring is a more serious error than overfactoring (Wood, Tataryn, & Gorsuch, 1996). Underfactoring will result in rotated solutions where factors have collapsed (been combined), or the structure of loadings has become complex and difficult to interpret. Overfactoring will often result in rotated solutions where one or more factors has no high loadings, or only a single high loading. Such factors obviously do not represent major common factors, since they do not exert strong influences on more than one manifest variable.

**Worst possible approach.** Use some mechanical rule by itself (e.g. number of eigenvalues greater than one, sequence of tests until  $H_0$  is not rejected at 5% level, increase  $m$  until  $\hat{\varepsilon}_a < 0.05$ ). Instead, users must make a careful, reasoned decision about the number of factors. One must consider various sources of information (information about fit, interpretability of rotated solutions, etc.).

If your final choice takes many things into consideration but still agrees with some rule there is no problem.

## PSY 820 Estimating the Number of Common Factors: Illustrations

## Module 9

### Holzinger's Unpublished Data (See Module 7)

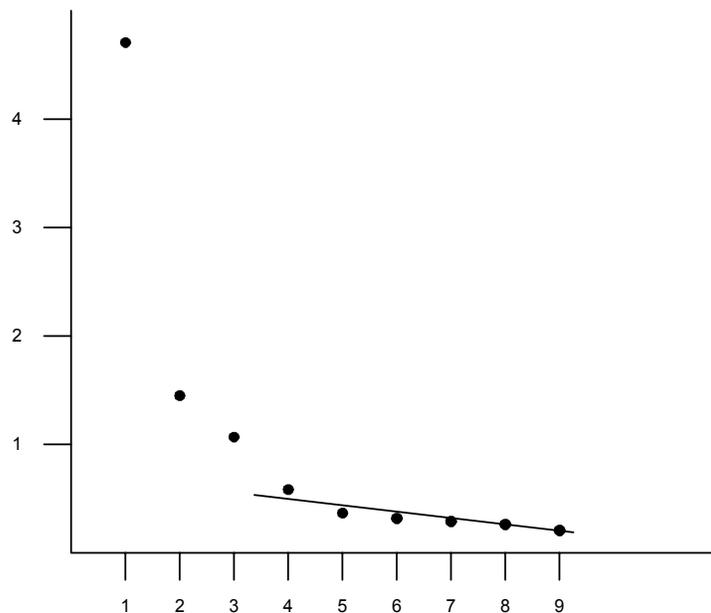
Here we consider how to evaluate several sources of information to aid in a decision about the number of factors to retain in our analysis of the Holzinger data. First consider rules of thumb regarding eigenvalues of  $R$ .

#### Eigenvalues of $R$

1	2	3	4	5	6	7	8	9
4.771	1.405	1.055	0.563	0.325	0.273	0.255	0.209	0.145

There are 3 eigenvalues  $>1$ , implying we should retain 3 factors.

Now consider a scree plot of these eigenvalues.



The last large drop in the series of eigenvalues lies between the third and fourth eigenvalues, implying we should retain 3 factors.

Note that for this same example, the eigenvalues for the reduced sample correlation matrix, with SMCs in the diagonal, are as follows:

4.42, 0.92, 0.75, -.02, -.06, -.08, -.11, -.11, -.18

Thus, there are only 3 positive eigenvalues, and a scree plot of these eigenvalues would also indicate 3 factors.

Let us now consider information provided by ML factor analysis of the same data.

ML and PACE solutions were obtained for 2, 3, and 4 factors for the Holzinger data, and associated tests and measures of fit were calculated. Following is a table containing this information:

**Measures of Fit**

	$m$	$\hat{F}$	Test St.	d.f.	# par	$p>$	RMSEA	Conf. Interval	T-L Index
ML	0	5.907	4105.69	36	9	0.000			
ML	2	0.8954	622.311	19	26	0.000	0.214	(0.199; 0.228)	0.71
	3	0.0164	11.370	12	33	0.498	0.000	(0.000; 0.037)	1.00
	4	0.0076	5.310	6	39	0.505	0.000	(0.000; 0.046)	1.00
PACE	2	1.2468	866.493	19	26	0.000	0.253	(0.239; 0.268)	
	3	0.0160	11.124	12	33	0.518	0.000	(0.000; 0.036)	
	4	0.0096	6.692	6	39	0.350	0.013	(0.000; 0.052)	

The first column indicates the number of factors specified in each solution. (Note that the first row of entries provides results for a zero-factor model. These values are used in the calculation of the Tucker-Lewis index.)

The next column indicates the value of the sample discrepancy function. Note that this value decreases as the number of factors is increased, indicating better absolute fit.

The next column provides values of the  $\chi^2$  test statistic. The associated degrees of freedom are provided in the next column, followed by the number of parameters for the specified model. Note that as number of factors increases, the value of the test statistic decreases, number of parameters increases, and degrees of freedom decreases.

The next column provides  $p$ -values for the test of exact fit for each number of factors. Results indicate that the hypothesis of exact fit in the population is rejected for 2 factors, not rejected for 3 factors, meaning a 3-factor model is plausible.

The next two columns provide point and interval estimates of RMSEA. Results indicate that fit of the 2-factor model is very poor, but that fit of the 3-factor model is excellent. For 3-factors, the estimated RMSEA is 0, and the

upper bound of the CI is .037, indicating that we can have high confidence that a 3-factor model fits very well in the population.

The last column shows values of the Tucker-Lewis index. These results clearly indicate that a two-factor model fits inadequately, but that a three-factor model does very well.

Thus, all of our results indicate that the appropriate number of factors for this set of data is 3.

A major aspect of exploratory factor analysis methodology is rotation. Rotation is a procedure that is applied to the factor loading matrix to enhance its interpretability, so that we can more accurately understand the nature of the major common factors. The process and methods of rotation, and issues that must be considered, are often poorly understood. As a result, the rotation process is often dealt with inadequately in practice, resulting in poorly structured solutions that may be misleading or difficult to interpret. To begin our study of rotation, we first review the meaning and interpretation of factor loadings.

Recall the factor analysis data model:

$$x_j = \mu_j + \lambda_{j1}z_1 + \lambda_{j2}z_2 + \cdots + \lambda_{jm}z_m + u_j \quad (1)$$

where:

$x_j$  is the score on manifest variable  $j$ .

$\mu_j$  is the mean of manifest variable  $j$ .

$z_k$  is the common factor score on factor  $k$ .

$\lambda_{jk}$  is the factor loading of test  $j$  on factor  $k$ .

$u_j$  is the score on unique factor  $j$

We assume that both the observed variables (dependent variables) and common factors (independent variables) have been standardized to have unit variance. i.e.  $\text{Cov}(\mathbf{x}, \mathbf{x}') = \mathbf{P}$  and  $\text{Cov}(\mathbf{z}, \mathbf{z}') = \mathbf{\Phi}$  have unit diagonal elements. Thus, the matrices  $\mathbf{P}$  and  $\mathbf{\Phi}$  are correlation matrices.

It can be seen from (1) that an increase of one unit in  $z_k$  would result in  $x_j$  increasing by  $\lambda_{jk}$  units. Thus the regression weight  $\lambda_{jk}$  is a measure of the influence of the independent variable (factor)  $z_k$  on the dependent variable (manifest variable)  $x_j$ . If  $\lambda_{jk}$  is zero then  $z_k$  has no influence on  $x_j$ . If  $\lambda_{jk}$  is positive then an increase in  $z_k$  is accompanied by an increase in  $x_j$ . If  $\lambda_{jk}$  is negative then an increase in  $z_k$  is accompanied by a decrease in  $x_j$ . Thus factor loadings are directional (factor  $\rightarrow$  manifest variable) measures of the effect of factors on manifest variables.

Because the loadings are equivalent to regression weights (standardized in the present context, because we have defined factors and manifest variables

as standardized), we can say certain things about their magnitude. Recall that in multiple regression, if independent variables are uncorrelated, then standardized regression coefficients can be interpreted simultaneously as regression weights and as simple correlations between independent variables and dependent variables. In the present context, this means that if common factors (independent variables) are orthogonal ( $\Phi = I$ ), then loadings (regression weights) can either be interpreted as weights representing the linear effect of a factor on a manifest variable, or as simple correlations between factors and manifest variables. Thus, when factors are orthogonal, loadings are bounded by +1 and  $-1$ .

What if factors are not orthogonal (off-diagonals of  $\Phi$  not zero)? In multiple regression, if independent variables are correlated, then standardized regression weights are not equivalent to simple correlations, and are not bounded by +1 and  $-1$ . Thus, in factor analysis, if factors are not orthogonal, then loadings are interpreted as weights representing linear influences of factors on manifest variables, but are not equivalent to simple correlations of factors with manifest variables and are not bounded by +1 and  $-1$ . This point often confuses researchers in empirical studies when they are surprised to see a few loadings  $>1$  after doing oblique rotation.

### **Interpretation of Factors:**

The interpretation of the factor analysis model (1) is greatly simplified if most of the loadings,  $\lambda_{jk}$ , corresponding to manifest variable,  $x_j$ , are zero. This means that the manifest variable is influenced by a small number of factors only. For, example,  $x_j$  is influenced by the first factor,  $z_1$ , alone if

$$\lambda_{j2} = \lambda_{j3} = \dots = \lambda_{jm} = 0$$

so that (1) becomes

$$x_j = \mu_j + \lambda_{j1}z_1 + u_j.$$

We use the factor loadings to understand what the factors are, and to name them. What is the first factor? We look at the first column of factor loadings in the factor matrix  $\mathbf{A}$ . High loadings show which variables are strongly influenced by the first factor. Near zero loadings show which variables are not influenced appreciably by the first factor. A characteristic that is shared by the variables that have high loadings will suggest what effect the factor has and enable the investigator to name this factor.

This requires knowledge of the manifest variables. A similar procedure is used for each of the other factors.

Consider the following example from Module 1.

	Verbal Ability	Mathematical Ability
Paragraph Comp.	.70	.10
Vocabulary	.70	.00
Arithmetic	.10	.70
Math. Prob. Solv.	.60	.60

In interpreting factor loading matrices, the following facts should be remembered. They apply both to unrotated and rotated factor matrices. It is permissible to:

- (1) Change the sign of all elements in any column. This change simply reverses the meaning of the factor; e.g., a factor interpreted as "verbal ability" when loadings are positive would be interpreted as "lack of verbal ability" when signs of all loadings are changed. No changes need be made to  $\hat{\Phi}$  if the factors are uncorrelated:  $\hat{\Phi} = \mathbf{I}$ .

If factors are correlated,  $\hat{\Phi} \neq \mathbf{I}$ , then a change of sign of all elements in a column of the factor matrix,  $\hat{\Lambda}$ , must be accompanied by sign changes in the corresponding row and column of the factor correlation matrix,  $\hat{\Phi}$ . For example suppose that

$$\hat{\Lambda} = \begin{bmatrix} -.1 & -.8 & .1 \\ .1 & -.7 & -.1 \\ .1 & -.6 & .1 \\ .7 & .1 & -.1 \\ .6 & .1 & -.1 \\ .7 & -.1 & .1 \\ .1 & .1 & .6 \\ -.1 & -.1 & .5 \\ .1 & .1 & .9 \end{bmatrix} \quad \hat{\Phi} = \begin{bmatrix} 1.0 & -.4 & .3 \\ -.4 & 1.0 & -.2 \\ .3 & -.2 & 1.0 \end{bmatrix}$$

and the signs of all elements in column 2, of  $\hat{\Lambda}$  are changed.

Then

$$\hat{\Lambda} \rightarrow \begin{bmatrix} -.1 & .8 & .1 \\ .1 & .7 & -.1 \\ .1 & .6 & .1 \\ .7 & -.1 & -.1 \\ .6 & -.1 & -.1 \\ .7 & .1 & .1 \\ .1 & -.1 & .6 \\ -.1 & .1 & .5 \\ .1 & -.1 & .9 \end{bmatrix} \quad \hat{\Phi} \rightarrow \begin{bmatrix} 1.0 & -.4 & .3 \\ .4 & -1.0 & .2 \\ .3 & -.2 & 1.0 \end{bmatrix} \rightarrow \begin{bmatrix} 1.0 & .4 & .3 \\ .4 & 1.0 & .2 \\ .3 & .2 & 1.0 \end{bmatrix}$$

Note that the sign of the diagonal element of  $\hat{\Phi}$  is changed twice, once with the row and once with the column. It therefore regains its original positive sign.

(2) Reorder the columns of  $\hat{\Lambda}$  in any convenient order. No changes need be made to  $\hat{\Phi}$  when  $\hat{\Phi} = I$ .

If factors are correlated,  $\hat{\Phi} \neq I$ , then an interchange of columns of  $\hat{\Lambda}$  must be accompanied by interchanges of the corresponding rows and columns of the factor correlation matrix,  $\hat{\Phi}$ . For example suppose that the first two columns of  $\hat{\Lambda}$  are interchanged:

$$\hat{\Lambda} \rightarrow \begin{bmatrix} .8 & -.1 & .1 \\ .7 & .1 & -.1 \\ .6 & .1 & .1 \\ -.1 & .7 & -.1 \\ -.1 & .6 & -.1 \\ .1 & .7 & .1 \\ -.1 & .1 & .6 \\ .1 & -.1 & .5 \\ -.1 & .1 & .9 \end{bmatrix} \quad \hat{\Phi} \rightarrow \begin{bmatrix} .4 & 1.0 & .2 \\ 1.0 & .4 & .3 \\ .3 & .2 & 1.0 \end{bmatrix} \rightarrow \begin{bmatrix} 1.0 & .4 & .2 \\ .4 & 1.0 & .3 \\ .2 & .3 & 1.0 \end{bmatrix}$$

Note that the diagonal elements of  $\hat{\Phi}$  are once again equal to one after the two interchanges have been carried out.

Neither the column sign change (reflection) nor the column reordering will alter the fit of the model. That is, the product  $\hat{\Lambda}\hat{\Phi}\hat{\Lambda}'$  will not be affected by

these changes so that  $\widehat{\Sigma} = \widehat{\Lambda}\widehat{\Phi}\widehat{\Lambda}' + \widehat{D}_\psi$  does not change. There is nothing irrevocable about the ordering and signs within columns of a rotated solution output by a computer program. (The only exception this instructor can think of is in target rotation if specified elements in the target are non-zero. This seldom happens.)

### Uncorrelated Factors: Orthogonal Rotation

In general, it is inappropriate to attempt to interpret the unrotated factor loading matrix. Factor extraction methods (such as ML and principal factors) are not designed to produce an interpretable solution directly. The original (unrotated) factor matrix yielded by a typical factor analysis procedure (e.g. maximum likelihood, principal factors) has very few loadings that are close to zero. This means that most manifest variables have substantial regression weights on all factors. Therefore, the unrotated solution must be transformed, or rotated, to enhance its interpretability. For example, consider the maximum likelihood factor matrix with  $m = 3$  for the Holzinger data considered in Module 7.

#### Unrotated ML Solution (Holzinger's unpublished data) : $\widehat{\Lambda}_U$

	I	II	III
1 Word_Mean	0.678	0.533	-0.273
2 Sent_Comp	0.717	0.377	-0.231
3 Odd_Words	0.699	0.486	-0.164
4 Mixed_Arit	0.900	-0.340	-0.033
5 Remainders	0.836	-0.200	0.028
6 Miss_Numb	0.862	-0.128	0.002
7 Gloves	0.421	0.089	0.431
8 Boots	0.478	0.246	0.540
9 Hatchets	0.483	0.302	0.673

This matrix is obtained under the assumption that factors are uncorrelated, that is

$$\hat{\Phi} = I.$$

Note that all the loadings on the first factor are moderate to high, and that the second factor contains a mixture of moderate positive and negative loadings. These factors are not interpretable.

In Module 5 we discussed rotational indeterminacy of factor loading matrices. That is, if  $\hat{P} = \hat{A}_U \hat{A}'_U + \hat{D}_\psi$  then there are other factor matrices,  $\hat{A}_R$ , such that

$$\hat{P} = \hat{A}_U \hat{A}'_U + \hat{D}_\psi = \hat{A}_R \hat{A}'_R + \hat{D}_\psi. \quad (2)$$

Thus  $\hat{A}_U$  and  $\hat{A}_R$  are equivalent in terms of fitting the model to the data, and

$$\hat{A}_R = \hat{A}_U \mathbf{T} \quad (3)$$

where  $\mathbf{T}$  is an orthogonal matrix satisfying

$$\mathbf{T}\mathbf{T}' = I \quad (4)$$

Given an unrotated loading matrix  $\hat{A}_U$ , we are therefore entitled to find an orthogonal matrix  $\mathbf{T}$  so that the elements of  $\hat{A}_R$  are easier to interpret. The process shown in (3) is called **orthogonal rotation** and the matrix  $\mathbf{T}$  is known as a **transformation matrix**.

Because of (2), this rotation process does not alter the fit of the model. All measures of fit and tests of fit considered in Module 8 are *not* at all affected by the rotation process. Communalities, also, are not affected

$$h_{jj} = \left[ \hat{A}_R \hat{A}'_R \right]_{jj} = \left[ \hat{A}_U \hat{A}'_U \right]_{jj}. \quad (5)$$

Obviously, there is an infinite number of potential transformation matrices  $\mathbf{T}$  to choose from. Each different  $\mathbf{T}$  would result in a different  $\hat{A}_R$ . We want to find a  $\mathbf{T}$  such that the resulting  $\hat{A}_R$  is interpretable, so that we can understand the nature of the major common factors.

How can a more easily interpreted solution be defined? If the magnitude of some loadings is large and other loadings are close to zero, and if patterns of loadings can be detected then interpretation is easier. The pattern of zero loadings is important. For example if all elements in the first column of the factor matrix are large and all others are close to zero, interpretation is difficult. Thurstone (1947) has defined requirements of a factor matrix with **simple structure** which is more easily interpretable.

1. Each row of  $\hat{\mathbf{A}}_R$  should contain at least one zero.
2. Each column of  $\hat{\mathbf{A}}_R$  should contain at least  $m$  zeros.
3. Every pair of columns of  $\hat{\mathbf{A}}_R$  should have several rows with a zero in one column but not the other.
4. If  $m \geq 4$ , every pair of columns of  $\hat{\mathbf{A}}_R$  should have several rows with zeros in both columns.
5. Every pair of columns of  $\hat{\mathbf{A}}_R$  should have few rows with nonzero loadings in both columns.

(In practical applications of these rules “small” loadings are interpreted as zeros )

In general, these criteria imply that each factor should be represented by high loadings for a distinct subset of manifest variables, and low loadings for remaining manifest variables. Also, these subsets defining different factors should not overlap a great deal. In addition, each manifest variable should be influenced by only a subset of the common factors. It is informative to note that the principles of simple structure do not state or imply that each manifest variable should have only one high factor loading; i.e., that each manifest variable should be strongly influenced by only one factor. This is a common misunderstanding regarding the notion of simple structure. Empirical researchers often think, mistakenly, that this condition is the definition of simple structure. To the contrary, factor analytic theory and the principles of simple structure clearly allow for the possibility that a given manifest variable is influenced by more than one major common factor. Such occurrences are common. A solution that has the characteristic of each manifest variable having only one high loading, and the rest very small, is called an *independent cluster* solution, meaning that each factor is defined by an independent cluster of manifest variables. Such a solution corresponds to a particularly simple kind (i.e., a special case) of simple structure, but the principles of simple structure do not require such a condition to hold.

In the 1940s-1950s, researchers carried out factor rotation by "graphical" or "hand" methods (Thurstone, 1947). The factor solution was represented geometrically. Plots were made that were essentially scatter plots of factor loadings. There would be a dimension for each factor and a point for each manifest variable, with the coordinates of the points defined by the loadings. Then the factor dimensions would be rotated in this space so that they passed through clusters of points, yielding rotated dimensions that would exhibit improved simple structure.

In effect, the researcher would be seeking orientations for the rotated factors that would satisfy Thurstone's principles of simple structure. The elements of the transformation matrix,  $\mathbf{T}$ , were then determined based on angles of rotation, and the loadings on the rotated factors were computed. This process was subjective and time consuming, becoming especially difficult as  $m$  increased beyond 2 or 3. Subjectivity was problematic because different investigators conducting graphical rotation on the same problem might produce somewhat different final solutions. This occurred because the procedure required subjective judgments and also because the principle of simple structure involved multiple criteria stated in a verbal form.

In the 1950s, researchers began developing automated ways to carry out the rotation process. The guiding principle in this movement was to establish an objective, mathematical definition of simple structure. In general terms, this is achieved by defining a function of the elements of  $\hat{\mathbf{A}}_R$  that attempts to measure simple structure. We shall refer to such a function as a *simplicity function* (or simplicity criterion). The transformation matrix,  $\mathbf{T}$ , is then constructed so as to maximize this simplicity function. This general approach to rotation is called *analytical rotation*. Such rotations are completely objective in that they do not require the user to make any subjective judgments about ideal locations for rotated factors, and the methods can be computerized so that they can be carried out very rapidly.

Sometimes it is convenient to minimize a function that measures complexity of the pattern of loadings in  $\hat{\mathbf{A}}$ . Such a function is known as a *complexity function*. Maximization of a simplicity function,  $f(\mathbf{A})$ , is equivalent to minimizing the complexity function,  $k - f(\mathbf{A})$ , where  $k$  is any convenient constant. Consequently, any pattern of rotated loadings obtained by maximizing a simplicity function may also be obtained by minimizing a corresponding complexity function.

Since the signs of factor loadings depend on arbitrary choices of scoring direction for factor variables, it is usual to base simplicity criteria on squared factor loadings,  $\hat{\lambda}_{jk}^2$ . Consequently the criterion is unaffected by the signs of factor loadings

The first simplicity criterion suggested was the **Quartimax** criterion which is the sum of fourth powers of factor loadings.

$$Q = \sum_{j=1}^p \sum_{k=1}^m \left( \hat{\lambda}_{jk}^2 \right)^2 \quad (6)$$

where  $\hat{\lambda}_{jk}$  is an element of  $\hat{\mathbf{A}}_R$ . Criteria that are equivalent to the Quartimax criterion were published independently in four papers that appeared around 1954 (Carroll, 1953; Neuhaus and Wrigley, 1954; Saunders, 1953; Ferguson, 1954). An equivalent form of this criterion is the overall variance of all squared factor loadings

$$\frac{1}{pm} \sum_{j=1}^p \sum_{k=1}^m \left( \hat{\lambda}_{jk}^2 - \bar{\hat{\lambda}}^2 \right)^2 \quad \text{where } \bar{\hat{\lambda}}^2 = \frac{1}{pm} \sum_{j=1}^p \sum_{k=1}^m \hat{\lambda}_{jk}^2.$$

As the rotated factor loadings become more spread out from each other (e.g., each loading either very high or very low), the variance will increase.

Quartimax rotation methods then solve the following problem: Given an unrotated factor loading matrix,  $\hat{\mathbf{A}}_U$ , find a transformation matrix  $\mathbf{T}$  such that the value of  $Q$  is maximized in  $\hat{\mathbf{A}}_R = \hat{\mathbf{A}}_U \mathbf{T}$ .

It was soon found that maximizing this criterion yielded a rotated factor matrix,  $\hat{\mathbf{A}}_R$ , that tends to have a column of large factor loadings with small factor loadings in all other columns (general factor).

Kaiser in 1958 then suggested the **Varimax** criterion which is the sum of the  $m$  within column variances of squared factor loadings

$$V = \sum_{k=1}^m \frac{1}{p} \sum_{j=1}^p \left( \hat{\lambda}_{jk}^2 - \bar{\hat{\lambda}}_{\cdot k}^2 \right)^2 \quad \text{where } \bar{\hat{\lambda}}_{\cdot k}^2 = \frac{1}{p} \sum_{j=1}^p \hat{\lambda}_{jk}^2. \quad (7)$$

This criterion may be written in other forms but remains equivalent to the sum of the variances of the squared factor loadings on each factor. As simple

structure of each factor improves, the squared loadings on those factors become more variable (some loadings very high, the rest very low). If we sum the variances of the squared loadings across the  $m$  factors, we have a measure of simplicity. This criterion is sometimes known as the raw Varimax criterion because it is applied to the “raw” factor loadings (before processing). The varimax rotation method then solves the following problem: Given an unrotated factor loading matrix,  $\hat{\mathbf{A}}_U$ , find a transformation matrix  $\mathbf{T}$  such that the value of  $V$  is maximized by  $\hat{\mathbf{A}}_R = \hat{\mathbf{A}}_U \mathbf{T}$ .

Kaiser found that raw varimax rotation worked fairly well but that loadings in rows with small communalities did not tend to exhibit a simple structure. He therefore standardized rows of the factor matrix before rotation by dividing factor loadings by the square roots of communalities before rotation. The rotation was then carried out on the rescaled loadings. After rotation, the rotated factor loadings were then brought back to the original scale by multiplying by the same square roots of communalities. This is often known as Normal Varimax. Here it will be referred to as “Varimax with row standardization”. This process of rotation with row standardization is common practice and is carried out in most computer programs.

Varimax rotation often works well but is by no means infallible. It has virtually monopolized orthogonal rotation. The Varimax rotation (with row standardization) of the Holzinger factor matrix is given in the next table.

**Varimax Rotation (Holzinger's unpublished data) :  $\hat{A}_R$**

	I	II	III
1 Word_Mean	0.865	0.217	0.152
2 Sent_Comp	0.754	0.341	0.156
3 Odd_Words	0.794	0.253	0.243
4 Mixed_Arit	0.245	0.913	0.182
5 Remainders	0.285	0.770	0.256
6 Miss_Numb	0.362	0.747	0.263
7 Gloves	0.094	0.230	0.556
8 Boots	0.188	0.166	0.719
9 Hatchets	0.171	0.122	0.857

Consider the interpretation of these factors. The first factor exhibits high loadings on the first 3 tests and relatively low loadings on the remaining 6 tests. The first 3 tests involve verbal skills, and the remaining 6 tests do not. We could consider the first factor a "verbal skills" factor. Factor 2 shows high loadings for tests 4, 5, and 6, and relatively low loadings for the remaining tests. This factor could be called a "numerical facility" factor, based on the nature of tests 4, 5, and 6. Factor 3 shows high loadings for the last 3 tests and relatively low loadings for the other 6 tests. The last 3 tests are very similar to each other, all involving identification of spatial relations among various types of figures. We could call Factor 3 a "spatial relations" factor. Clearly, this rotated solution is much more interpretable than the unrotated solution presented earlier. However, note that the simple structure in this solution is not especially good in one important sense: the low loadings in this solution are not especially low. There are many "low" loadings here in the range of .2 to .3. We shall address this issue further subsequently.

It is advisable to carry out rotations for various values of  $m$  and to decide on the most interpretable solution. Frequently under factoring gives very poor results with several factors collapsing into one.

The following symptoms are diagnostic of over factoring:

- (1) A column of the factor matrix has a single element that is not near zero (Unique factor)
- (2) All elements in a column are near zero.

Varimax rotations (with row standardization) for  $m = 2$  and  $m = 3$  are given below.

**Varimax Rotation (Holzinger's unpublished data) : 2 Factors & 4 Factors**

	I	II	I	II	III	IV
1 Word_Mean	0.729	0.299	0.872	0.208	0.151	-0.046
2 Sent_Comp	0.637	0.416	0.757	0.339	0.157	0.085
3 Odd_Words	0.969	0.246	0.792	0.249	0.243	-0.008
4 Mixed_Arit	0.258	0.895	0.241	0.945	0.177	-0.130
5 Remainders	0.286	0.823	0.292	0.761	0.262	0.124
6 Miss_Numb	0.353	0.806	0.370	0.730	0.268	0.056
7 Gloves	0.200	0.351	0.094	0.231	0.557	0.077
8 Boots	0.298	0.331	0.189	0.164	0.715	-0.019
9 Hatchets	0.300	0.321	0.172	0.117	0.862	-0.044

Considering the 2-factor solution, we see that the "spatial relations" factor has been lost. Tests 7, 8, and 9 have small to moderate loadings on both factors. Considering the 4-factor solution, we see that Factor 4 has no high loadings, indicating that this factor is not a major common factor and should not be retained. These solutions provide further evidence that the 3-factor solution is best.

Factor loadings are given here to three decimal places to minimize retyping from the computer output. For a journal article it is preferable to report loadings to only two places.

## Comments on orthogonal rotation

Orthogonal rotation methods rotate the factor loading matrix  $\hat{\Lambda}$  under the restriction that  $\mathbf{T}\mathbf{T}' = \mathbf{I}$ . This restriction implies that  $\hat{\Phi} = \mathbf{I}$ , meaning that the rotated factors are uncorrelated.

Question: In practice, is there a reasonable justification for imposing this restriction?

Keep in mind that factors are latent variables. Is it reasonable to assume or restrict these latent variables to be uncorrelated? In exploratory factor analysis, when we are seeking to identify the number and nature of the major common factors and we are doing rotation to aid us in interpreting the meaning of those factors, how would we know *a priori* that those factors are uncorrelated?

Answer: Rarely, if ever, would this restriction be justified. In fact, in domains in which factor analysis is widely used (e.g., mental abilities, personality, attitude, etc.) we would anticipate that factors would be correlated to some degree.

Nevertheless, reviews of applied factor analysis literature show that orthogonal rotation is routinely used in practice, perhaps as often as 80% of applications (Ford, MacCallum, & Tait, 1986; Fabrigar et al., 1999). Why?

Possible explanations:

It's what everybody else does, so I'll do it.

Keeping things simple.

Computer software options (some do varimax as default).

Ignorance about implications, alternatives.

"Believing" or "Wanting" factors to be uncorrelated.

Early success and preeminence of varimax.

None of these explanations provides sufficient justification for using blind orthogonal rotation in practice. If the restriction to orthogonality is difficult to justify, then what should we do instead?

Drop the restriction, allowing rotated factors to be correlated.

This implies oblique rotation.

Result: Better simple structure; easier interpretation; more accurate identification of factors and relationships among factors; more realistic approach to search for factors.

Oblique rotation does not require that rotated factors be correlated. It allows them to be correlated. If the best simple structure corresponds to orthogonal factors, those orthogonal factors can be found using oblique rotation.

### Correlated Factors: Oblique Rotation

Although the unrotated factor matrix corresponds to orthogonal factors it is possible to allow factors to become correlated during rotation. This often results in a clearer simple structure.

The oblique rotation is given by

$$\hat{\Lambda}_R = \hat{\Lambda}_U \mathbf{T} \quad (8)$$

$$\hat{\Phi} = \mathbf{T}^{-1} \mathbf{T}'^{-1}$$

where the diagonal elements of  $\mathbf{T}^{-1} \mathbf{T}'^{-1}$  must all be equal to one since  $\hat{\Phi}$  must be a correlation matrix:

$$\text{diag}[\mathbf{T}^{-1} \mathbf{T}'^{-1}] = \mathbf{I} \quad (9)$$

Matrix  $\hat{\Lambda}_R$  is the rotated factor loading matrix, and  $\hat{\Phi}$  is the factor intercorrelation matrix.

Note that we do not impose the restriction that  $\mathbf{T} \mathbf{T}' = \mathbf{I}$ , so the rotated factors are not required to be uncorrelated.

The oblique rotation problem is to find a transformation matrix  $\mathbf{T}$  such that equation (9) holds and which yields a rotated factor matrix,  $\hat{\Lambda}_R$ , and associated factor correlation matrix,  $\hat{\Phi}$ , so as to optimize some simplicity criterion.

The reproduced correlation matrix,  $\hat{\mathbf{P}}$ , and the fit of the model to the data are not affected by this transformation. That is:

$$\hat{\mathbf{P}} = \hat{\Lambda}_U \hat{\Lambda}_U' + \hat{\mathbf{D}}_\psi = \hat{\Lambda}_R \hat{\Phi} \hat{\Lambda}_R' + \hat{\mathbf{D}}_\psi. \quad (10)$$

This can be seen as follows:

$$\hat{\Lambda}_R \hat{\Phi} \hat{\Lambda}_R' = (\hat{\Lambda}_U \mathbf{T})(\mathbf{T}^{-1} \mathbf{T}'^{-1})(\hat{\Lambda}_U \mathbf{T})'$$

$$\begin{aligned}
&= \widehat{\Lambda}_U \mathbf{T} \mathbf{T}^{-1} \mathbf{T}'^{-1} \mathbf{T}' \widehat{\Lambda}'_U \\
&= \widehat{\Lambda}_U \mathbf{I} \mathbf{I} \widehat{\Lambda}'_U \\
&= \widehat{\Lambda}_U \widehat{\Lambda}'_U
\end{aligned}$$

The communalities are not influenced by oblique rotation. In a solution where factors are intercorrelated, communalities are given by:

$$h_{jj} = \left[ \widehat{\Lambda}_R \widehat{\Phi} \widehat{\Lambda}'_R \right]_{jj} = \left[ \widehat{\Lambda}_U \widehat{\Lambda}'_U \right]_{jj} \quad (11)$$

instead of by (5). Thus, when factors are not orthogonal, we can not compute communalities simply by summing squares of loadings in each row of the loading matrix. Rather, we must take into account the fact that the factors are intercorrelated.

Early methods of oblique rotation (published before 1966) were unsatisfactory. Most operated on a matrix known as a reference structure, primarily because ways of imposing the fairly complicated constraints in (9) had not yet been discovered. Two methods of oblique rotation that did not operate directly on the reference structure, and were less unsatisfactory than those that did, were the Hendrickson-White (1964) PROMAX procedure and the Harris-Kaiser (Harris & Kaiser, 1964) orthoblique procedure. These two procedures are still available in some computer packages but are not always satisfactory and are not recommended by this instructor. Promax and Harris-Kaiser rotations are available in SAS.

Jennrich and Sampson (1966) provided the first method of oblique rotation that is satisfactory in most applications. This is the procedure employed for the Jennrich-Sampson Direct Quartimin rotation. It can be generalized to other simplicity criteria. The advantage of the Jennrich-Sampson approach is that the constraints in (9) are imposed *directly* and no attempt is made to make use of the reference structure.

In commercial packages that use the Jennrich-Sampson approach, the family of Oblimin complexity criteria is usually permitted. The oblimin family of rotations employs a simplicity criterion that includes a constant,  $\gamma$ , whose value can be set by the user. Altering the value of  $\gamma$  alters the simplicity criterion. Different values of  $\gamma$  specify different members of this family of rotations.

This family contains the Quartimin criterion as a special case ( $\gamma = 0$ ). The Direct Quartimin rotation is usually the default option. In a report the heading “Direct Oblimin Rotation” would be uninformative — in most cases a Direct Quartimin rotation has been carried out but, if not, the value of the parameter  $\gamma$  should be given.

Not all members of the Oblimin family give satisfactory results. If the parameter  $\gamma$  is too large, no minimum of the function may exist. The Quartimin criterion is the recommended member of this family. Row standardization may be carried out in the same manner as for orthogonal rotation.

There is another more satisfactory family of rotation criteria that also includes the Quartimin criterion as a special case. This is the Crawford-Ferguson (Crawford & Ferguson, 1970) family. A minimum exists for each member of this family. This approach is available in CEFA.

### Crawford-Ferguson family of rotation criteria

$$f_{\kappa} = (1 - \kappa) \sum_{j=1}^p \sum_{\substack{k=1 \\ k \neq \ell}}^m \sum_{\ell=1}^m \lambda_{jk}^2 \lambda_{j\ell}^2 + \kappa \sum_{k=1}^m \sum_{\substack{j=1 \\ j \neq h}}^p \sum_{h=1}^p \lambda_{jk}^2 \lambda_{hk}^2 \quad 0 \leq \kappa \leq 1 \quad (12)$$

$$= (1 - \kappa) \times \text{Row Parsimony (Quartimin)} + \kappa \times \text{Column Parsimony}$$

This is a complexity function, meaning that a smaller value of this function implies a simpler pattern of factor loadings. The first component of  $f_{\kappa}$  is a measure of row parsimony and would attain the lower bound of zero if each row (manifest variable) of the factor matrix had a single non-zero loading. The second is a measure of column parsimony and would attain the lower bound of zero if each column (factor) had a single nonzero loading.

This family has a single parameter  $\kappa$  ( $0 \leq \kappa \leq 1$ ). The user can set the value of  $\kappa$ , with different values of  $\kappa$  defining different members of the family. The user can also specify, for any value of  $\kappa$ , whether rotation is to be orthogonal or oblique. When  $\kappa = 0$  and we conduct an oblique rotation, the complexity criterion is the Quartimin criterion. When  $\kappa = 0$  and we conduct orthogonal rotation, we are doing quartimax rotation. That is, minimizing the Quartimin complexity criterion in orthogonal rotation with  $\kappa = 0$  is equivalent to maximizing the Quartimax criterion defined earlier.

Other correspondences between members of the Crawford-Ferguson family and known orthogonal rotation criteria are shown below.

$\kappa = 0$	Quartimax
$\kappa = \frac{1}{p}$	Varimax (Kaiser)
$\kappa = \frac{m}{2p}$	Equamax (Saunders)
$\kappa = \frac{m-1}{p+m-2}$	Parsimax (Crawford & Ferguson)
$\kappa = 1$	Factor Parsimony (Crawford & Ferguson)

Note that if  $\kappa = 1/p$  and we carry out oblique rotation, we are in effect conducting an oblique analogue of the Varimax criterion in (7). Maximization of  $V$  in (7) in oblique rotation gives unsatisfactory results but minimization of  $f_{\kappa=1/p}$  gives satisfactory results in both orthogonal and oblique rotation.

It is interesting, also, that  $f_{\kappa=0}$  does not give very satisfactory results in orthogonal rotation (Quartimax) but does give good results in oblique rotation (Direct Quartimin). The most widely used member of the Oblimin family was direct quartimin ( $\kappa = 0$  in the Crawford Ferguson family). Other promising rotation criteria are now available in CEFA (see Module 7) and it is worth trying them out.

CEFA is able to carry out orthogonal or oblique rotations using any member of the Crawford-Ferguson family as a rotation criterion. In CEFA output the names of rotation criteria listed above are preceded by CF to indicate the corresponding member of the Crawford-Ferguson family. Thus the Crawford-Ferguson complexity function with  $\kappa = 1/p$  is referred to as CF-Varimax.

The direct Quartimin rotation (oblique CF-Quartimax in CEFA), with row standardization, of the Holzinger factor matrix ( $m = 3$ ) is given in the following table.

### **Direct Quartimin Rotation (Holzinger's unpublished data)**

	I	II	III
1 Word_Mean	0.944	- 0.050	- 0.026
2 Sent_Comp	0.767	0.139	- 0.031

3 Odd_Words	0.831	- 0.002	0.081
4 Mixed_Arit	- 0.046	<u>1.011</u>	- 0.048
5 Remainders	0.040	0.804	0.062
6 Miss_Numb	0.138	0.749	0.062
7 Gloves	- 0.055	0.134	0.560
8 Boots	0.053	- 0.012	0.743
9 Hatchets	0.026	- 0.091	0.912

**Factor Correlation Matrix**

	I	II	III
I	1.000	0.587	0.431
II	0.587	1.000	0.483
III	0.431	0.483	1.000

It can be seen that the simple structure is substantially more satisfactory than that obtained in orthogonal rotation. The first factor is clearly a Verbal factor, the second an Arithmetic factor and the third a Spatial relations factor. Note especially that the low factor loadings are much smaller in this solution than they were in the varimax solution, yielding much cleaner simple structure and much clearer definition of the factors. Note also that there is one factor loading >1.0, which is permissible in an oblique solution. There is additional information to be interpreted, given in the elements of the factor correlation matrix  $\hat{\Phi}$ . The three factors are more or less equally correlated. The cleaner simple structure and the substantial positive intercorrelations among the factors reveal that the varimax solution was misleading. It would be inappropriate to interpret the major factors underlying these 9 tests as being uncorrelated with each other.

In oblique rotation it is also permissible to reflect (change all signs in a column) and reorder factors. However the same reflections and reorderings applied to the factor matrix,  $\hat{A}$ , must also simultaneously be applied to both the rows and the columns of the factor correlation matrix,  $\hat{\Phi}$ .

It is also advisable in oblique rotation to carry out rotations for different values of  $m$ . The Direct Quartimin rotations of the Holzinger data are shown in the following table for  $m = 2$  and  $m = 4$ .

**Direct Quartimin Rotation (Holzinger's unpublished data) : 2 Factors & 4 Factors**

	I	II	I	II	III	IV
1 Word_Mean	0.750	0.057	0.948	-0.025	-0.027	-0.064
2 Sent_Comp	0.580	0.242	0.788	0.071	-0.033	0.101
3 Odd_Words	<u>1.071</u>	-0.114	0.829	0.002	0.081	-0.017
4 Mixed_Arit	-0.118	1.004	-0.052	1.089	-0.031	-0.097
5 Remainders	-0.048	0.902	0.087	0.662	0.081	0.184
6 Miss_Numb	0.042	0.852	0.179	0.648	0.084	0.101
7 Gloves	0.075	0.351	-0.046	0.076	0.562	0.093
8 Boots	0.205	0.284	0.047	0.004	0.742	-0.029
9 Hatchets	0.213	0.270	0.012	-0.059	0.923	-0.064

**Factor Correlation Matrices**

	I	II	I	II	III	IV
I	1.000	0.653	1.000	0.571	0.440	0.246
II	0.653	1.000	0.571	1.000	0.462	0.512
III			0.440	0.462	1.000	0.274
IV			0.246	0.512	0.274	1.000

Two factors seem insufficient since the Arithmetic and Spatial reasoning tests both load on Factor I, and the spatial relations factor has been lost. Four factors seem too many since all loadings on Factor IV are near zero.

The example above, and the issues discussed earlier differentiating orthogonal and oblique rotation, should make it clear that oblique rotation is preferable in empirical applications of exploratory factor analysis.

Nevertheless, as discussed earlier, far more orthogonal rotations than oblique rotations are reported in the literature. For some time this may have been due in part to the fact that no generally satisfactory method of oblique rotation was available.

This is no longer the case, and it would be advisable to report an orthogonal rotation only if a direct oblique rotation gave essentially zero factor correlations.

### Target Rotation

The exploratory methods of factor analysis described up to this point can be modified to be used in a confirmatory manner by means of target rotation. In target rotation the user constructs a target matrix, which is a matrix specifying the general pattern of loadings expected. Specification of the target is based on prior hypotheses about the nature of the underlying common factors and their pattern of influences on the manifest variables. The target matrix is usually a matrix with two kinds of elements: (1) zeros, indicating positions where the user expects or seeks very small loadings; (2) unspecified or free entries, indicating positions where the user does not expect or seek very small loadings. In constructing the target, the user must also obviously specify the number of factors (number of columns in the target matrix). A target that is suitable for the Holzinger data is presented in the following table. Question marks (?) stand for unspecified elements and zeros (0) for specified elements. The target defines a Verbal Factor, an Arithmetic Factor and a Spatial Relations factor.

#### Target Matrix (Holzinger's unpublished data)

	Verbal	Arith	Spat_Rel
Word_Mn	?	0	0
Sent_Cmp	?	0	0
Odd_Wrds	?	0	0
Mixed_Ar	0	?	0
Remndrs	0	?	0
Mis_Nmb	0	?	0
Gloves	0	0	?
Boots	0	0	?
Hatchets	0	0	?

The objective is to rotate our unrotated solution from the Holzinger data so that it matches this target as closely as possible. More specifically, we wish the rotated solution to exhibit low loadings in the locations where the zeros are specified in the target matrix. Loadings corresponding to a ? may be large or small.

Target rotation is carried out by finding a transformation matrix  $T$  such that in  $\widehat{A}_U T$  the sum of squares of the loadings corresponding to the zeros in the target is minimized. The root mean square of these elements then provides a summary measure of the degree to which the rotation has yielded small loadings in the desired locations. Both oblique and orthogonal rotations to partially specified targets may be carried out using CEFA (see Module 7). CEFA requires that an unspecified element be indicated by a 9 instead of by a ?.

Following are the orthogonal and oblique target rotations to the target specified above.

**Target Rotation (Holzinger's unpublished data)**

	<b>Orthogonal</b>			<b>Oblique</b>		
	Verbal	Arith	Spat_Rel	Verbal	Arith	Spat_Rel
Word_Mn	0.864	0.241	0.118	0.957	-0.065	-0.033
Sent_Cmp	0.751	0.363	0.118	0.774	0.130	-0.038
Odd_Wrds	0.794	0.281	0.208	0.839	-0.013	0.074
Mixed_Ar	0.232	0.926	0.124	-0.067	<u>1.028</u>	-0.057
Remndrs	0.276	0.789	0.204	0.021	0.819	0.054
Mis_Nmb	0.354	0.768	0.211	0.123	0.762	0.054
Gloves	0.102	0.264	0.540	-0.072	0.147	0.560
Boots	0.201	0.211	0.704	0.038	0.000	0.744
Hatchets	0.189	0.175	0.843	0.009	-0.076	0.913

**Factor Correlation Matrix**

Verbal	Arith	Spat_Rel
1.000	0.613	0.452
0.613	1.000	0.488
0.452	0.488	1.000

Root Mean Squared Deviation    0.227

0.072

As can be expected, the fit of the rotated matrix to the specified elements of the target in terms of root mean squared deviation is better for the oblique target rotation.

### **Additional Facilities in CEFA**

There are some additional rotation criteria, that give good results and are worth trying out, but do not belong to the Crawford Ferguson family. These include McKeon's Infomax criterion, Yates' Geomin criterion and McCammon's minimum entropy criterion (orthogonal rotation only), which are available in CEFA. Additional information about these criteria may be found in Browne (2001).

When several rotation criteria are applied with the same data, the mathematical equivalence of different orderings and reflections of the same factor matrix can be confusing. Many different orderings of columns and different column sign allocations yield the same optimum rotation function value. Consequently it is difficult to match solutions produced by different rotation criteria. In order to ensure that the ordering of columns and within column loading signs match as closely as possible between different solutions an order matrix may be used. This a matrix with some specified loadings, not necessarily zero, and possibly some unspecified loadings (?). It looks just like a target matrix but is used for a different purpose. After rotation, the columns of the factor matrix are reordered and reflected so as to match the order matrix as closely as possible, but no further rotation (see (3), (8) ) is carried out. In target rotation, a rotation is carried out so as to match the target matrix as closely as possible and no subsequent reordering or reflection is necessary. A target/order matrix may be included in the CEFA data file. If any rotation, except target rotation, is carried out, this matrix is used as an order matrix. If a target rotation is carried out, the matrix is used as a target matrix.

With some data sets, particularly if no independent cluster solution exists, the complexity criterion may have several different, and non-equivalent, local minima. Completely different solutions may therefore exist, and the one with the lowest complexity function value may not yield the simplest pattern of rotated loadings. It is therefore advisable to use several different starting points to verify whether or not the final solution remains the same. CEFA has

a facility for carrying out random initial rotations to produce different starting points.

Whenever parameters are estimated it is helpful to have information about the standard error of the estimates and to be able to impose confidence intervals on the population parameters. CEFA is able to produce standard errors for rotated factor loadings and factor correlations, and provides associated confidence intervals.

### **"Factor Analysis" Methodology in Published Applications**

Now that we have completed study of methodology for exploratory factor analysis, let us consider the methods that are most commonly used in the applied literature. A great many published applications of "factor analysis" employ the following sequence of methods: Principal components analysis; retain those components with eigenvalues  $>1.0$ ; varimax rotation. And many applications that do not employ this entire sequence of procedures do employ at least one of them.

From our study of methods, it should be clear that this choice of procedures represents poor technique. Components analysis is not a latent variable model; it is a procedure that ignores the presence of unique factors and extracts components that do not correspond to latent variables and do not optimally account for the intercorrelations among the manifest variables. Retaining components with eigenvalues  $>1.0$  is a purely mechanical rule that does not reliably result in an appropriate decision. Varimax rotation imposes the seldom justifiable restriction that latent variables be uncorrelated. Because of these problems, this set of techniques is generally not useful for determining the number and nature of the major common factors underlying a battery of manifest variables. Use of any of these techniques, let alone the whole set, can easily result in failure to clearly identify those factors. Solutions obtained by such methods are likely to be uninterpretable or misleading. Researchers wishing to conduct exploratory factor analysis should employ an appropriate method for fitting the common factor model to their data, should consider information about fit and interpretability in making a reasoned decision about the number of factors, and should generally employ oblique rotation to allow for the likely possibility that the common factors are intercorrelated.

In 1967, Armstrong published an article in *The American Statistician* entitled "Derivation of Theory by Means of Factor Analysis, or Tom Swift and His Electric Factor Analysis Machine." The intent of the article was to show that factor analysis is not useful for identifying underlying latent variables. Armstrong generated artificial data that had a known underlying factor structure. He then analyzed the data using components analysis, kept those components with eigenvalues  $>1.0$ , and did varimax rotation. He believed he was doing "factor analysis" since he was using methods commonly used in the applied literature.

He showed that the results did not correspond at all to the known structure, and was highly critical of "factor analysis" methodology for this failure. However, if Armstrong's data are analyzed instead using the common factor model, if a careful decision is made about the number of factors retained based on information about model fit, and if oblique rotation is employed, then the results correspond very well to the known structure in the data. Thus, Armstrong unwittingly showed how important is choice of technique in factor analysis; he showed how poor choice of technique yields poor solutions. This point is examined closely by Preacher & MacCallum (2003). Unfortunately, such techniques remain fairly commonly used in the applied literature, and solutions in such papers should be considered highly suspect.

Fabrigar et al. (1999) provide an excellent review of issues involved in the application of exploratory factor analysis in psychological research. They review methodological issues carefully and thoroughly, provide guidance for investigators regarding choices and decisions in factor analysis, present results of a review of empirical applications of factor analysis, and demonstrate importance of methodological decisions via a reanalysis of some published data. Researchers planning to use factor analysis in practice should give careful consideration to the perspective and advice offered in this article.

# PSY 820 Principal Components Analysis, Factor Scores and Sample Size

## Module 11

### Comparison of Factor Analysis and Principal Components Analysis

For simplicity of exposition we assume that the elements of the vector,  $\mathbf{x}$ , of manifest variables are standard scores. i.e.  $\Sigma = P, \mu = \mathbf{0}$ .

#### Factor Analysis

Data model:  $\mathbf{x} = \Lambda\mathbf{z} + \mathbf{u}$        $\Lambda : p \times m, \mathbf{x} : p \times 1, \mathbf{z} : m \times 1,$   
 $\mathbf{u} : p \times 1$

Factor scores forming  $\mathbf{z}$  are not observable.

Assumptions:      (i)  $\text{Cov}(\mathbf{u}, \mathbf{u}') = D_\psi$  is diagonal  
                         (ii)  $\text{Cov}(\mathbf{z}, \mathbf{u}') = \mathbf{0}$

Consequence:  $P = \Lambda\Lambda' + D_\psi$

This is the correlation structure, derived from the data model. This is a testable model.

When this model is fit to sample data, factor analysis methods choose  $\hat{\Lambda}$  so as to make the off-diagonal elements of  $(R - \hat{\Lambda}\hat{\Lambda}')$  as small as possible in some sense. Thus the whole purpose of factor analysis is to explain correlation coefficients between manifest variables as well as possible.

#### Principal Components

Equation:  $\mathbf{x} = \tilde{\Lambda}\tilde{\mathbf{z}} + \tilde{\mathbf{u}}$        $\tilde{\Lambda} : p \times m, \mathbf{x} : p \times 1, \tilde{\mathbf{z}} : m \times 1, \tilde{\mathbf{u}} : p \times 1$

Component scores forming  $\tilde{\mathbf{z}}$  are linear combinations of the manifest variables and are therefore observable

$$\tilde{\mathbf{z}} = B'\mathbf{x} \quad B' : m \times p$$

Matrix  $B'$  is a matrix of coefficients that is multiplied by MV scores in  $\mathbf{x}$  to produce the component scores. Computation of  $B'$  will be defined shortly.

The component scores,  $\mathbf{z}$ , are defined so as to maximise the average of squared multiple correlations of manifest variables on  $m$  components,

$$\text{ASMC} = \frac{1}{p} \sum_{j=1}^p r^2(x_j, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m) \quad (1)$$

That is, we can conceive of the components as independent variables and the MVs as dependent variables in the components equation  $\mathbf{x} = \mathbf{\Lambda}\mathbf{z} + \mathbf{u}$ . In turn, we can define a SMC for each MV reflecting the proportion of variance in each MV accounted for by the components. ASMC is then the average of these SMCs across the  $p$  MVs. Components are chosen so as to maximize this quantity.

Thus the  $m$  component scores are chosen so as to predict the  $p$  manifest variable scores as precisely as possible. The aim of principal components is data reduction with minimal loss of information concerning the original variables.

It turns out that we can obtain a solution for  $\mathbf{B}'$  as follows:

$$\mathbf{B}' = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'$$

and that

$$\text{Cov}(\mathbf{z}, \mathbf{u}') = \mathbf{0}$$

Consequence:  $\mathbf{P} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$

Off diagonal elements of  $\mathbf{\Psi}$  are not assumed to be zero, or even small, but that the sum of diagonal elements of  $\mathbf{\Psi}$  is required to be small.

By substituting the definition of  $\mathbf{B}'$  into the equation for component scores, we obtain the following equation for computing component scores:

$$\mathbf{z} = (\mathbf{\Lambda}'\mathbf{\Lambda})^{-1}\mathbf{\Lambda}'\mathbf{x}$$

Thus, component scores can be computed exactly as linear combinations of the MVs.

The presentation based on (1) is correct (e.g. Meredith) but not standard. Most text books present the original derivation of principal components that involves

the maximisation of variances of the  $\tilde{z}$ 's. The original derivation does not allow rotation of  $\tilde{\mathbf{A}}$ . The one based on (1) does.

Orthogonal rotation:  $\tilde{\Phi} = \mathbf{I}$

Oblique rotation:  $\text{diag}[\tilde{\Phi}] = \mathbf{1}$

In reporting results from principal components solutions, it is common practice to report "variance accounted for" by each component. For the unrotated solution represented by  $\tilde{\mathbf{A}}$ , variance accounted for corresponds to the eigenvalue for each component. One way to understand this is in relation to the computation of communalities from  $\tilde{\mathbf{A}}$ . Summing squares in each row of  $\tilde{\mathbf{A}}$  yields the squared multiple correlations,  $r^2(x_j, \tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_m)$ , or communality equivalents. Note that summing the communalities across the  $p$  manifest variables would yield a total of variance accounted for by the common components. An alternative way to arrive at this same total would be to sum the squares in each column of  $\tilde{\mathbf{A}}$ , and then sum these values. The total would be the same, and the sum of squares for each column of  $\tilde{\mathbf{A}}$  could be interpreted as variance explained by each principal component. These column sums of squares are equal to the eigenvalues of the corresponding components. Rotation of the component loading matrix alters variance accounted for by the individual components, so eigenvalues can be interpreted as variance accounted for only for unrotated components. The average variance accounted for by the  $m$  retained components, however, does not change with rotation.

Residual covariance matrices for Holzinger's Unpublished Data (see Module 7) are shown below:

#### Holzinger's Unpublished Data

Residual Covariance Matrix,  $\hat{\Sigma}_{uu} = (\mathbf{R} - \hat{\mathbf{A}}\hat{\mathbf{A}}')$ : M.L. Factor Analysis,  $m = 3$ .

	1	2	3	4	5	6	7	8	9
Word_Mean	0.181								
Sent_Comp	-0.001	0.290							
Odd_Words	0.002	-0.003	0.248						
Mixed_Arit	0.002	-0.005	0.000	0.073					
Remainders	-0.003	0.012	-0.003	0.000	0.260				
Miss_Numb	-0.005	0.011	0.000	0.001	-0.006	0.241			
SGloves	-0.006	-0.006	0.012	-0.005	0.024	-0.002	0.629		
Boots	-0.008	0.009	0.005	0.001	-0.005	-0.001	-0.006	0.420	
Hatchets	0.005	-0.005	-0.005	0.000	-0.002	0.002	-0.001	0.001	0.222

#### Holzinger's Unpublished Data

Residual Covariance Matrix,  $S_{\ddot{u}\ddot{u}} = (\mathbf{R} - \ddot{\mathbf{A}}\ddot{\mathbf{A}}')$ : Principal Components,  $m = 3$

	1	2	3	4	5	6	7	8	9
Word_Mean	0.135								
Sent_Comp	-0.075	0.192							
Odd_Words	-0.064	-0.089	0.165						
Mixed_Arit	0.006	-0.017	0.006	0.093					
Remainders	0.007	-0.007	0.003	-0.053	0.154				
Miss_Numb	-0.008	-0.020	-0.007	-0.039	-0.096	0.165			
SGloves	0.042	0.033	0.034	-0.036	-0.029	-0.040	0.388		
Boots	-0.026	-0.001	-0.029	0.025	0.007	0.012	-0.190	0.267	
Hatchets	-0.005	-0.010	-0.023	0.013	0.012	0.017	-0.146	-0.091	0.210

It can be seen that the off-diagonal elements of  $\hat{\Sigma}_{uu}$  are smaller than those of  $S_{\ddot{u}\ddot{u}}$ . Thus factor analysis yields a better explanation of the correlations than principal components. On the other hand the diagonal elements of  $S_{\ddot{u}\ddot{u}}$  are smaller than those of  $\hat{\Sigma}_{uu}$ . This is because the average squared multiple correlation of manifest variables on the observable component scores is higher than the (estimated) average squared multiple correlation of manifest variables with the unobservable factors. In this

example the ASMC of manifest variables on 3 components is

$$\begin{aligned}
 \text{ASMC} &= \frac{1}{9} \sum_{i=1}^9 r^2(x_i, \check{z}_1, \check{z}_2, \check{z}_3) \\
 &= \frac{1}{9} \left\{ (1 - .135) + (1 - .192) + (1 - .165) + (1 - .093) \right. \\
 &\quad \left. + (1 - .154) + (1 - .165) + (1 - .388) + (1 - .267) + (1 - .210) \right\} \\
 &= 0.80 \\
 &= \frac{\ell_1 + \ell_2 + \ell_3}{9} = \frac{4.8 + 1.4 + 1.1}{9}
 \end{aligned}$$

Although principal component scores are used for data reduction, it is not always clear that this is a good thing. For example, multiple regression equations are sometimes carried out on few principal components of many available independent variables in an attempt to avoid shrinkage of the multiple correlation coefficient. Although this is often beneficial, examples can be constructed where this procedure will yield a very poor multiple correlation coefficient.

The critical point is that components analysis is something conceptually and mathematically distinct from factor analysis. Components are not latent variables (common factors), but rather represent a mix of common, specific, and error influences. Components loadings tend to be overestimates of corresponding factor loadings (Widaman, 1993). Components analysis is not recommended when the common factor model is appropriate and when the objective of the researcher is to identify factors that account for correlations among the manifest variables (Gorsuch, 1990; Snook & Gorsuch, 1989; Tucker et al., 1969).

## Factor Scores

Recall the common factor model:  $\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{u}$

In this model, vector  $\mathbf{z}$  contains scores on common factors. In some applications of factor analysis, researchers desire to obtain these scores and conduct further analysis on them. For example, researchers might wish to use factor scores as independent or dependent variables in regression models, or as dependent variables in ANOVA.

However, these scores can not be determined exactly. They are indeterminate. Mathematically, the reason for this is as follows: If  $\mathbf{A}$  is known there are  $p$  equations and  $p + m$  unknowns ( $\mathbf{z}$ ,  $\mathbf{u}$ ) in

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \mathbf{u}.$$

More conceptually, the common factor model defines the  $p$  MVs as linear functions of  $m$  common factors and  $p$  unique factors. Thus, the model has  $p$  MVs and  $(p + m)$  latent variables. It is not possible to determine the scores on the  $(p + m)$  LVs from only  $p$  MVs. The issue of indeterminacy of factor scores has a long and controversial history in the factor analysis literature. Some critics have argued that this is a fatal flaw in the common factor model and that researchers should therefore not use this model. Proponents of factor analysis often counter that this problem involves only the factor scores and does not affect the covariance or correlation structure models. Thus, there is no problem doing factor analysis using those structural models. The indeterminacy problem arises only at the point where one wishes to obtain factor scores.

Most text books deal with the “estimation” of factor scores and most computer packages have programs for calculating these. There are, however, serious difficulties associated with doing this. For simplicity assume that  $\mathbf{A}$  and  $\mathbf{D}_\psi$  are known. (The situation is no better, but more difficult to investigate, when  $\mathbf{A}$  and  $\mathbf{D}_\psi$  are replaced by estimates  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{D}}_\psi$ ).

There are two methods for estimating  $\mathbf{z}$  in general use:

## Regression method

$$\hat{z}_R = \Lambda'(\Lambda\Lambda' + D_\psi)^{-1}x$$

The rationale of the regression method is as follows: The factors are considered to be dependent variables and the MVs are considered to be independent variables. (This reverses their role in the data model.) We then use a multiple regression framework to obtain the most accurate "predictions" of the factor scores from the MV scores. Application of this approach yields the equation above for predicting factor scores.

## Least squares method (Bartlett's method)

$$\hat{z}_B = (\Lambda'\Lambda)^{-1}\Lambda'x$$

This method provides factor score estimates that, when substituted into the factor analysis data model, yield the most accurate reconstruction of the scores on the MVs, in a least squares sense.

It should be borne in mind that

$$\hat{z}_R \neq z$$

$$\hat{z}_B \neq z$$

$$\hat{z}_R \neq \hat{z}_B$$

and that

$$\text{Cov}(\hat{z}_R, \hat{z}'_R) \neq \Phi$$

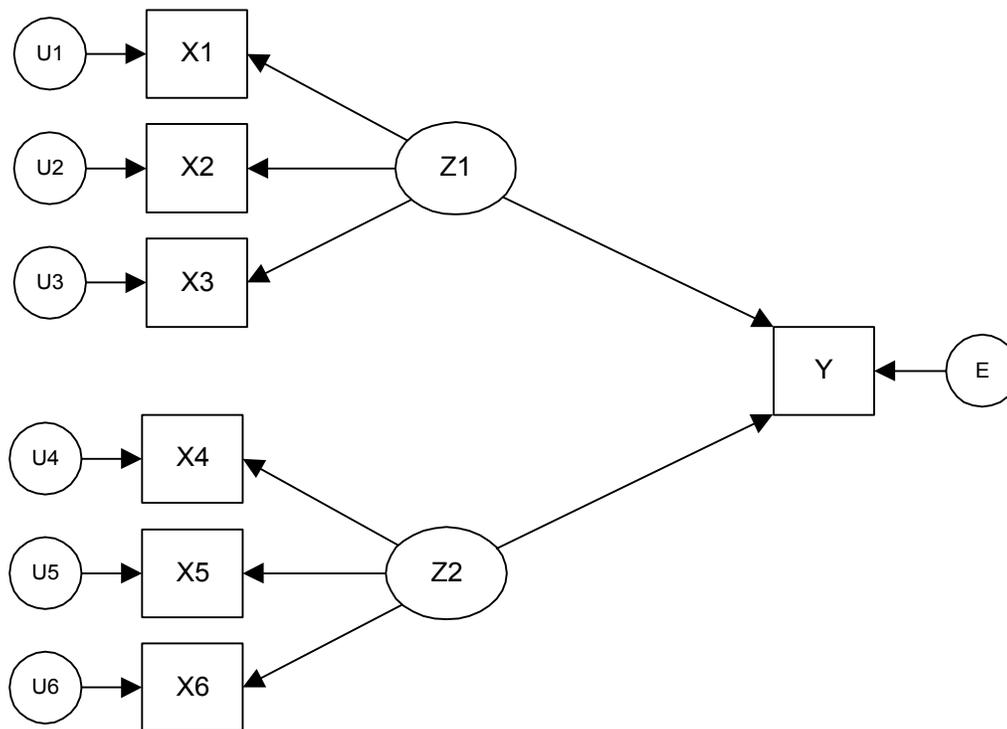
$$\text{Cov}(\hat{z}_B, \hat{z}'_B) \neq \Phi$$

Thus the estimated factor scores do not have the same covariance matrices as the factor scores (e.g.  $\Phi = I$ ) so that the estimated factor scores cannot be regarded as adequate replacements for the factors.

So, what should a researcher do if he or she wishes to investigate questions about scores on factors? Fortunately, methodological advances in recent years have made it possible in most situations to address such questions without having to obtain factor scores. That is, we can investigate questions about relationships of factors to other variables without obtaining factor scores themselves. This can be done using structural equation modeling.

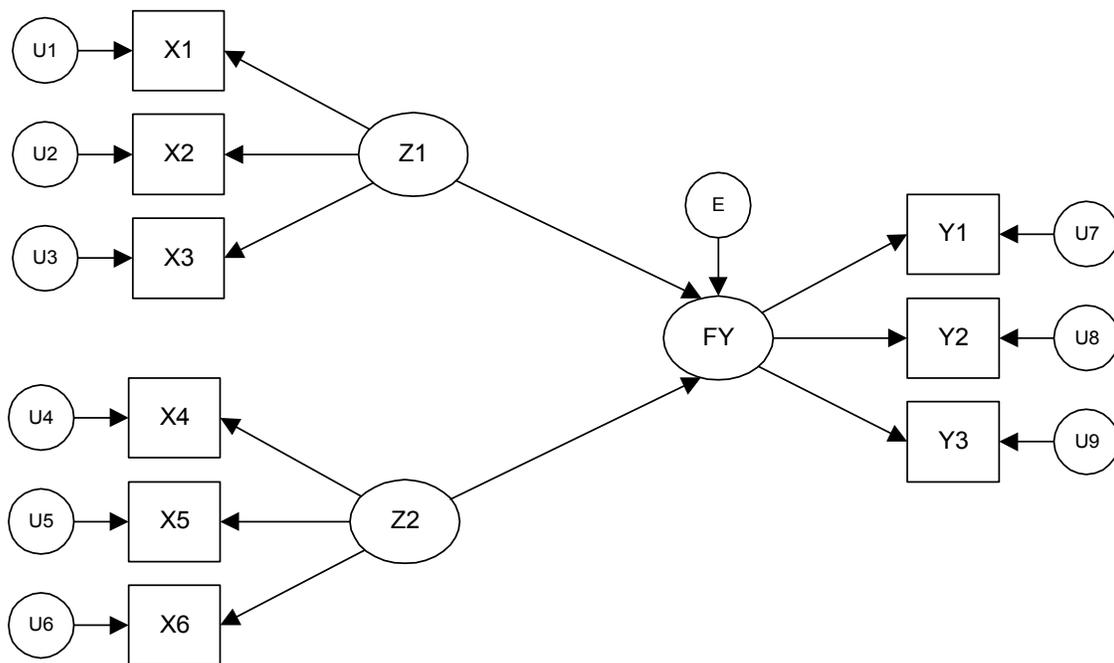
For example, consider a case where a researcher identifies two common factors accounting for relationships among 6 MVs, and wishes to use these two factors as predictors of some other MV. This is a "factor analysis regression" model and can be specified as a single structural equation model as shown in the following path diagram. This model can be fitted to data as a single model, and effects of the factors on the Y variable can be estimated, without ever needing to determine scores on the two factors.

**Path Diagram for Factor Analysis Regression**



For another example, suppose an investigator wished to study linear regression relationships among factors; e.g., two factors accounting for the relationships among 6 MVs predicting another factor that accounts for the relationships among 3 other MVs. This model also can be specified as a structural equation model, shown in the path diagram below, and can be fitted to data as a single model. Effects of the "independent" factors on the "dependent" factor can be estimated without determining factor scores.

**Path Diagram for Regression of One Factor on Others**



### General Advice about Factor Scores

In general, researchers using factor analysis should avoid situations involving estimation of factor scores. Questions about relationships among factors or of factors to other MVs can usually be addressed using structural equation modeling. It is risky to obtain factor score estimates and conduct further analysis on them, and even more unjustified to interpret factor score estimates for specific individuals. These score estimates do not correspond to true factor scores.

A thorough discussion of the issue of indeterminacy of factor scores is provided by Steiger (1979) and also in a special issue of *Multivariate Behavioral Research*, 1996, #4.

## Composite Scores

Researchers often use a simple "substitute" for factor scores. To illustrate, consider the 4-test example used in Module 1 where we identified two factors, a verbal factor and a math factor. A researcher wishing to obtain scores for individuals on the Mathematical Ability factor would often do the following:

- (1) Determine which MVs had high loadings on that factor. In this case, the math factor had two high loadings, one on an Arithmetic test, and one on a Mathematical Problem Solving test.
- (2) Obtain standard scores for the individuals on those MVs.
- (3) For each individual, sum the standard scores on those MVs.

The result is a composite score representing the sum of the variables that had high loadings on the Mathematical Ability factor. This procedure is often used in practice. Researchers carry out a factor analysis. For each factor they determine which variables have high loadings. Then, for each individual, they sum standard scores on those variables to obtain composite scores representing the factors.

Although these composite scores may have some empirical meaning, it is essential to understand that they are not factor scores, nor are they direct estimates of factor scores as defined in the data model. These composite scores are simply combinations of standardized MVs. They should not be called factor scores. They should be called composites or scales or subscales, for example. Computation of such composites, in effect, jumps out of the factor analysis model. Further analysis of such scores is not part of a factor analysis.

If investigators wish to study questions about relationships of factors to other LVs or MVs, we strongly recommend use of structural equation modeling rather than composite scores.

## Sample Size in Factor Analysis

If we are going to conduct a factor analysis, how large of a sample do we need in our study? What is usually meant by this question is how large must  $N$  be to assure accurate recovery of the major factors that exist in the population? This is an old question in factor analysis, and one that has not been dealt with adequately in the literature. Reviews of literature on sample size in factor analysis are provided by MacCallum, Widaman, Zhang, & Hong (1999) and by Velicer and Fava (1998).

There are a variety of "rules of thumb" regarding sample size. Many references provide recommendations in terms of minimum  $N$ , or else minimum ratio of  $N$  to  $p$ . For example:

Gorsuch (1983) and Kline (1979):  $N \geq 100$ ;  $N/p \geq 5$

Guilford (1954):  $N \geq 200$ ;

Cattell (1978):  $N \geq 250$ ;  $N/p \geq 3$ , prefer  $N/p \geq 6$

Comrey & Lee (1992): Suggested a scale for levels of  $N$  in factor analysis:

100 = poor; 200 = fair; 300 = good; 500 = very good;  
>1,000 = excellent.

Everitt (1975):  $N/p \geq 10$

Obviously, these recommendations are inconsistent. MacCallum et al. (1999) argue that such recommendations are not useful because they are based on a misconception. That misconception is that the minimum level of  $N$  (or of  $N/p$ ) necessary to achieve adequate recovery of population factors is invariant across studies. MacCallum et al. show that this view is incorrect and that the necessary  $N$  is highly dependent on several specific aspects of a given study. Under some conditions, relatively small samples may be entirely adequate, while under other conditions very large samples may be inadequate.

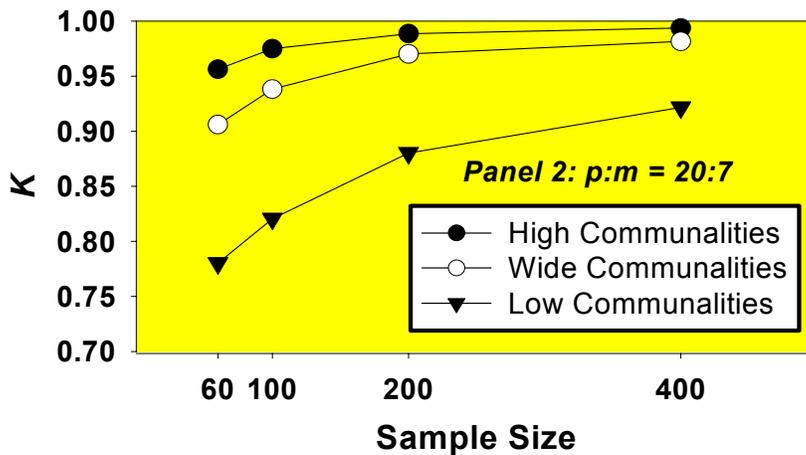
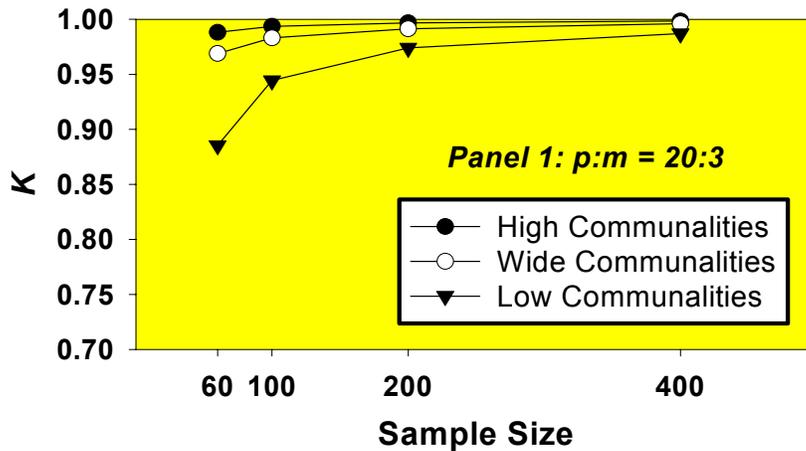
To investigate this issue, MacCallum et al. make use of a mathematical theoretical framework originally proposed by MacCallum and Tucker (1991). Using this approach, they show that the following effects should hold:

When communalities are high (e.g.,  $>.7$  or so) and factors are well determined (adequately represented by at least 4-5 indicators each), then accurate recovery

of population factors can be achieved with relatively small samples (e.g.,  $N = 60$ ).

However, when communalities are low (e.g.,  $<.4$  or so) and factors are not well determined, then accurate recovery of population factors may require extremely large samples (e.g.,  $N > > 400$ ).

MacCallum et al. conducted a sampling study using artificial data to determine whether these effects would actually hold in the analysis of data. The effects were validated. A simplified version of some of their results is presented in the following figure. In each figure the vertical axis is a measure of recovery of population factors, with values of in the mid-.90s or higher indicating accurate recovery. The horizontal axis represents sample size. The first panel represents results for designs with 20 variables and 3 factors, and the second panel represents results for designs with 20 variables and 7 factors. Within each panel there are three conditions of communality level: high (communalities  $.7 - .9$ ), wide (communalities in the range of  $.2$  to  $.8$ ), and low (communalities  $.2 - .4$ ).



These results clearly show the very strong effect of communality level on the influence of sample size, and a lesser effect of degree of overdetermination of factors. Under conditions of high communality, good to excellent recovery of factors was consistently obtained even when  $N$  was not large. As communalities became lower, much larger levels of  $N$  were needed, and this effect was amplified when factors were not so well determined. When factors were not well determined and communalities were low, rather poor recovery of factors was obtained even with very large  $N$ .

This study, including both the mathematical framework and the simulation study, considered the general case where the common factor model holds exactly in the population. Such a situation will not hold in empirical research.

A follow-up study by MacCallum, Widaman, Preacher, & Hong (2001) extended this work to the case where the model does not hold exactly in the population. The extended theoretical framework resulted in the same predicted effects of communality and overdetermination, and those effects were verified in two sampling studies, one using artificial data and one using empirical data. Results of these studies make it very clear that rules of thumb regarding sample size in factor analysis are invalid, misleading, and not useful. They should be abandoned.

# **PSY 820      Confirmatory Factor Analysis      Module 12**

## **Model Specification**

### **Introduction**

The overall difference between exploratory and confirmatory factor analysis involves the incorporation of prior hypotheses into the model specification and estimation process. In EFA, the investigator is typically exploring to determine the number and nature of the major common factors. In CFA, the investigator has a specific prior hypothesis about the number and nature of the factors, and that hypothesis is incorporated explicitly into model specification and estimation.

Both EFA and CFA employ the same general model: the common factor model. However, CFA requires some additional assumptions concerning the positions of (usually) zero loadings reflecting prior hypotheses. The methodology of confirmatory factor analysis also is quite different. It is no longer possible to obtain minimum discrepancy (maximum likelihood, ordinary least squares) estimates of factor loadings once the corresponding estimates of unique variances are available. It is then necessary to include all estimates simultaneously in the iterative process so that confirmatory factor analysis tends to be slower than exploratory factor analysis even though the number of parameters estimated typically is smaller.

### **Software**

While the original confirmatory factor analysis programs (Jöreskog) dealt only with factor analysis, more general structural equation models that include confirmatory factor analysis as special cases have since been developed (LISREL, COSAN, Bentler-Weeks, RAM). This course will employ a computer program, RAMONA, that can fit more general structural equation models but will make use of it only for confirmatory factor analysis.

## **Exploratory (Unrestricted) Factor Analysis**

In exploratory factor analysis, there typically are no firm prior assertions concerning the number of factors or the positions of zero loadings. While the investigator should have some idea about what is to be expected, no firm statement of these expectations is necessary and the results of the analysis can suggest unexpected modifications to original theories.

Sometimes exploratory factor analysis is used when firm hypotheses about the positions of zero loadings are available, for example when target rotation is used. There are, however, no rigorous tests for these hypotheses concerning the positions of zero loadings. In blind rotation, one never knows whether the failure of factor loadings to be near zero is due to incorrect hypotheses or to a rotation criterion that is not adequate in the particular situation considered. Even in target rotation there is no test of fit to the partially specified target and decisions have to be based entirely on judgement.

On the other hand, target rotation can suggest specific modifications to hypotheses concerning positions of zero loadings. Large residuals corresponding to specified zero loadings can suggest that they should no longer be specified to be zero. Also near-zero values of unspecified loadings can suggest that they should be specified to be zero. A new partially specified target can be constructed and a second target rotation carried out. It must be recognized, though, that such an approach is data-driven and that such a modified target does not represent a prior hypothesis.

## **Confirmatory (Restricted) Factor Analysis**

In confirmatory factor analysis, firm prior hypotheses concerning the number of factors and the positions of zero loadings are required. One advantage is that overall measures of fit of the model are available. On the other hand modifications of the model are suggested less clearly than in exploratory factor analysis (although model modification indices are available in some programs).

## Constraints, Restrictions and Identification Conditions

In both exploratory and confirmatory factor analysis, the specification of the model involves certain types of constraints.

Constraints are equations that refer to the parameters of a model and which are required to be satisfied.

e.g.  $\varphi_{11} = 1$  or  $\lambda_{21} = \lambda_{31}$  or  $\lambda_{12} = 0$

There are two kinds of constraints: restrictions and identification conditions.

Identification conditions are imposed only to select one particular solution from a whole class of possible solutions. Identification conditions do not affect the implied correlation matrix ( $\hat{P}$ ), do not affect the fit of the model, and are not testable.

Restrictions usually represent aspects of a prior hypothesis and are imposed so as to represent that hypothesis. They do affect the implied correlation matrix, do affect the fit of the model, and are testable.

For any given number of factors:

In exploratory factor analysis, we impose only identification conditions. That is, we impose just enough constraints to define a unique solution, and no more. There are many different ways to do this. Details are given below. Because we impose only identification conditions, this sort of analysis is often referred to as "unrestricted factor analysis."

For instance, the restriction that  $\Phi = I$  is an identification condition. Setting the diagonal elements of  $\Phi$  equal to 1 defines the latent variables as having unit variance, thus giving them a scale. Setting the off-diagonals of  $\Phi$  equal to zero defines the factors (before rotation) as being orthogonal. Without such identification conditions, we can't find a solution to the model.

In confirmatory factor analysis, we typically impose both identification conditions and restrictions. Identification conditions serve to select one solution from among a large class of solutions, and restrictions represent

aspects of our prior hypothesis. The addition of these restrictions means that a confirmatory factor analysis solution will generally fit data more poorly than will an exploratory factor analysis solution for the same number of factors. Because we impose restrictions beyond simple identification conditions, this type of analysis is often referred to as "restricted factor analysis."

### **Exploratory (Unrestricted) Factor Analysis -- Identification conditions**

There are essentially two distinct types of identification conditions in EFA. One type has the purpose of setting a scale for each factor. Factors are latent variables without an inherent scale. In order to estimate factor loadings and scores, the factors must be given a scale. The simplest and most common way to do this is to set the diagonal elements of  $\Phi$  equal to 1.0. This defines the factors as each having a variance of 1.0; since their means are zero, this step defines the factors as being standardized variables.

A second set of identification conditions is necessary in order to address the rotational indeterminacy problem. That is, there are an infinite number of equally good solutions, all rotations of each other. We must impose identification conditions so as to uniquely define one of the solutions. Once we have one solution, we are free to apply rotation techniques to that solution.

There are a wide variety of ways to impose these identification conditions. One general approach is to set equal to zero  $m(m - 1)$  parameters in  $\Phi$  and  $\Lambda$ . However, the selection of the parameters that are set to zero is not arbitrary. Here are two systematic ways to define this set of identification conditions.

If  $\Phi = \mathbf{I}$ , one column of  $\Lambda$  has no zeros; one column of  $\Lambda$  should have *at least* one zero, one should have at least two zeros, . . . , and one should have at least  $m - 1$  zeros

$$\begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \\ \vdots & \vdots & \vdots \end{bmatrix} \text{ or } \begin{bmatrix} \lambda_{11} & 0 & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & 0 \\ \lambda_{31} & \lambda_{32} & 0 \\ \lambda_{41} & \lambda_{42} & \lambda_{43} \\ \vdots & \vdots & \vdots \end{bmatrix}$$

If  $\text{Off}(\Phi) \neq \mathbf{0}$ , each column of  $\Lambda$  should have *at least*  $m - 1$  zeros that can be arranged as follows

e.g. 
$$\begin{bmatrix} \lambda_{11} & 0 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & 0 & \lambda_{33} \\ \lambda_{41} & \lambda_{42} & \lambda_{43} \\ \vdots & \vdots & \vdots \end{bmatrix} \text{ or } \begin{bmatrix} 0 & \lambda_{12} & \lambda_{13} \\ 0 & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & 0 & \lambda_{33} \\ \lambda_{41} & 0 & \lambda_{43} \\ \lambda_{51} & \lambda_{52} & 0 \\ \lambda_{61} & \lambda_{62} & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

Column and row interchanges (reorderings of factors and/or variables) are permissible.

If only these minimal constraints shown are imposed, these constraints act as identification conditions and not as restrictions. One could conduct an unrestricted factor analysis of this sort using a program that normally is used for restricted (confirmatory) factor analysis. If one imposes only the constraints just described, then the solution is unrestricted. Different unrestricted solutions are just rotations of each other.

Finally, it should be noted that in unrestricted factor analysis by maximum likelihood, as implemented in programs like CEFA, identification is achieved via the following identification conditions:

$$\Phi = I \Rightarrow \text{diag}[\Phi] = \mathbf{1} \text{ and } \text{Off}(\Phi) = \mathbf{0}$$

$$\text{OffDiag}(\Lambda' D_{\psi}^{-1} \Lambda) = \mathbf{0}$$

The resulting ML solution would be a rotation of other ML solutions obtained by imposing the identification conditions described above.

Also, note that solutions obtained by imposing different identification conditions would all fit the data equally well.

### Confirmatory Factor Analysis Model

Both CFA and EFA employ the common factor model. But in CFA, additional restrictions are imposed and some identification conditions are discarded as they are no longer necessary.

As before the data model is

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{z} + \mathbf{u}$$

and generates the covariance structure

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \mathbf{D}_\psi$$

The matrices  $\boldsymbol{\Lambda}$ ,  $\boldsymbol{\Phi}$ , and  $\mathbf{D}_\psi$  are "parameter matrices." A prior hypothesis about the number and nature of the common factors is incorporated into the analysis via specification of elements of these three matrices. In confirmatory factor analysis restrictions are imposed on some of the parameters of the model so as to represent aspects of our prior hypothesis. Generally these restrictions are of the form  $\lambda_{ij} = 0$  and assign zero values to some factor loadings. Other restrictions are possible, however:

e.g.  $\psi_{11} = \psi_{22} = \psi_{33}$       or       $\lambda_{21} = \lambda_{31}$

Some of the identification conditions used in unrestricted factor analysis are no longer required in confirmatory factor analysis where additional restrictions have been imposed. In confirmatory factor analysis the identification conditions  $\text{diag}[\boldsymbol{\Phi}] = \mathbf{1}$  are still generally imposed, to establish a scale for each factor, but  $\text{Off}(\boldsymbol{\Phi}) = \mathbf{0}$  will not be employed if factors are to be allowed to be correlated.

In confirmatory factor analysis, we typically impose restrictions on the elements of  $\boldsymbol{\Lambda}$ ,  $\boldsymbol{\Phi}$ , and  $\mathbf{D}_\psi$ , and the number of these restrictions will exceed the minimum number for an unrestricted solution, given above as  $m(m - 1)$ .

If we impose fewer restrictions, it will not be possible to obtain a solution due to identification problems. For example requiring only one factor loading to be zero would not be sufficient in confirmatory factor analysis. (RAMONA would warn you of redundant parameters in this case). If all restrictions are of the form  $\lambda_{ij} = 0$  the following (necessary) conditions should be satisfied:

Keep in mind that the fit of the restricted model will generally be worse than that of the unrestricted exploratory factor analysis model. And a restricted solution cannot be obtained by rotation of an unrestricted solution.

## Model Specification

Model specification refers to the procedure for specifying one's prior hypothesis formally in terms of the parameter matrices of the common factor model. With regard to computer software, specification of the model is done algebraically in some programs and graphically in others. If a graphical model specification procedure is employed the elements of the matrices of the model are still stored in the computer and employed in the iterative computational procedure.

We first consider the algebraic approach to model specification. This approach requires us to specify information about the elements in the three parameter matrices:

$\Lambda$  Factor matrix (matrix of factor loadings)

$\Phi$  Factor correlation matrix (may be constrained to be equal to  $\mathbf{I}$ ).

$D_\psi$  Diagonal covariance matrix of unique factors.

It is possible to allow unique factors to be correlated but this will not be considered in this course.

Each element of each of these three matrices is specified to have one of two forms:

**Free parameters:** Values are unknown and must be estimated.

**Fixed parameters:** Numerical values are specified a priori by the investigator.

Examples of fixed parameters:

$$\lambda_{12} = 0$$

$$\varphi_{11} = 1$$

It is also possible to require some parameters to be equal.

e.g.  $\psi_{11} = \psi_{22} = \dots = \psi_{pp}$

### Example of Restricted Factor Analysis

Consider a situation where there are six manifest variables,  $X_1, \dots, X_6$ , and two factors,  $F_1, F_2$ .

The factor loadings of  $X_1, X_2, X_3$  on  $F_1$  are substantial and those of  $X_4, X_5, X_6$  on  $F_1$  are (essentially) zero.

The factor loadings of  $X_1, X_2, X_3$  on  $F_2$  are (essentially) zero and those of  $X_4, X_5, X_6$  on  $F_2$  are substantial.

$F_1$  and  $F_2$  are correlated.

### Matrix specification

Matrix  $\Lambda$  will have 6 rows and 2 columns and will have the following pattern of fixed and free parameters:

$$\Lambda = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}$$

Matrix  $\Phi$  will be  $2 \times 2$  and have unities in the diagonal (fixed parameters), and a free correlation parameter off the diagonal:

$$\Phi = \begin{bmatrix} 1 & \varphi_{12} \\ & 1 \end{bmatrix}$$

Matrix  $D_\psi$  will be  $6 \times 6$ , with zeros off the diagonal (fixed parameters) and unique variances on the diagonal (free parameters):

$$D_\psi = \begin{bmatrix} \psi_{11} & 0 & 0 & 0 & 0 & 0 \\ & \psi_{22} & 0 & 0 & 0 & 0 \\ & & \psi_{33} & 0 & 0 & 0 \\ & & & \psi_{44} & 0 & 0 \\ & & & & \psi_{55} & 0 \\ & & & & & \psi_{66} \end{bmatrix}$$

Note that the diagonal elements of  $\Phi$  are constrained to be = 1 for identification purposes. Note also that factor loadings on unique factors are *implied* to be = 1, also for identification purposes :

$$\mathbf{x} = \boldsymbol{\mu} + \Lambda \mathbf{z} + \mathbf{I} \mathbf{u} \quad \Rightarrow \quad \Sigma = \Lambda \Phi \Lambda' + \mathbf{I} D_\psi \mathbf{I}'$$

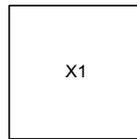
(These implied unique factor loadings of 1 are not specified in the input file.)

## Model Specification by Means of a Path Diagram.

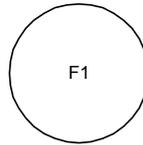
It is popular to specify confirmatory factor analysis models using path diagrams. When we learn to use the RAMONA program we shall see that RAMONA input commands are translated directly from a path diagram.

The following representations are employed in a path diagram. (Read the RAMONA User's Guide, Section 3, for a detailed description.)

Manifest Variable:



Latent Variable:



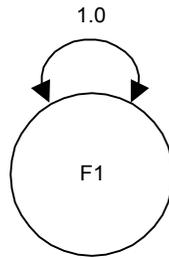
Regression weight, factor loading (directional influence from one variable to another):



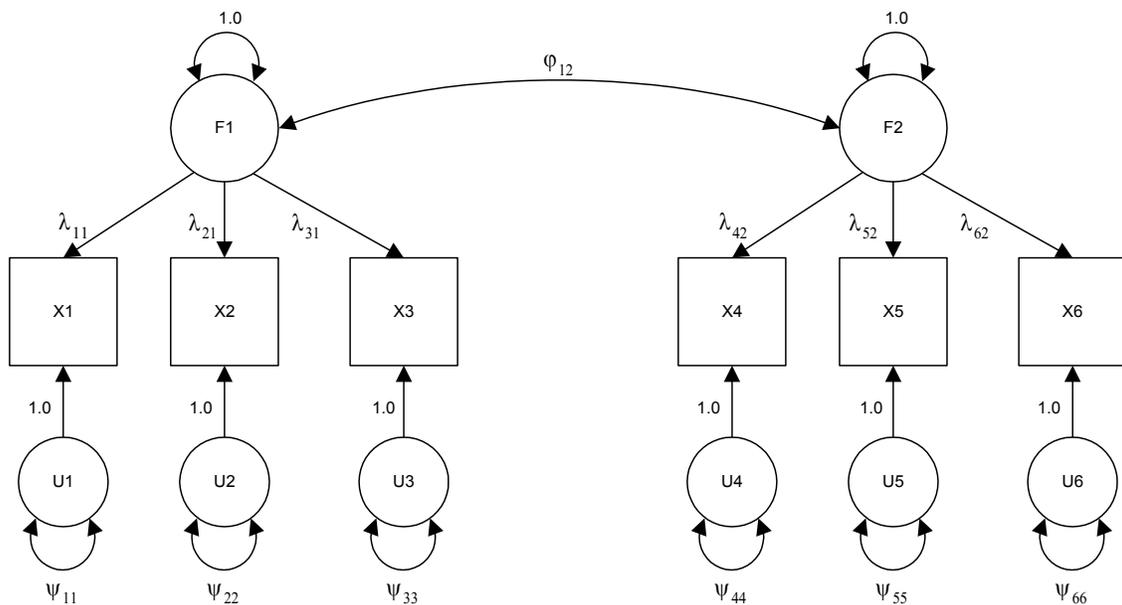
Covariance (Correlation)



## Variance



The path diagram for the example shown earlier in matrix form would look like this:



Note that all arrows corresponding to zero parameters are omitted.

Note that this path diagram shows the relationships that are inherent in the data model representation of the common factor model. The diagram shows each manifest variable as being influenced by common and unique factors. The path diagram goes beyond the data model in the sense that it also shows the variances of the unique factors as well as the variances and covariances of the common factors. Note the mapping between the path diagram and the parameters in the parameter matrices.

### Treatment of unique factors in model specification.

The model above shows each unique factor as having a variance represented as a free parameter, and a loading specified as fixed at 1.0. This is typical model specification for unique factors. In terms of the common factor model in data-model form, the unique factors implicitly have a weight (loading) of 1.0. This is reflected in the data model:

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda} \mathbf{z} + \mathbf{u}$$

Note that the unique factor term has an implied weight of 1.0, rather than a coefficient to be estimated.

However, an alternative representation of unique factors is possible.

For simplicity consider the situation where there is a single common factor,  $z_1$ , and suppose the unique factor weight is not set to 1.0 :

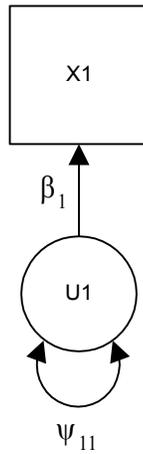
$$\begin{aligned} x_j &= \mu_j + \lambda_j z_1 + \beta_j u_j \\ &= \mu_j + \lambda_j z_1 + \beta_j^* u_j^* \end{aligned}$$

where  $\beta_j^* = k \times \beta_j$  and  $u_j^* = \frac{u_j}{k}$  and  $k$  is any nonzero scalar. That is, the unique factor weight and unique factor score could both be rescaled so that their product remains unchanged.

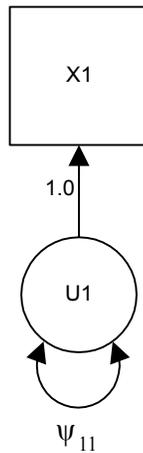
$$\text{If } \text{Var}(u_j) = \psi_{jj} \text{ then } \text{Var}(u_j^*) = \psi_{jj}^* = \frac{\psi_{jj}}{k^2} .$$

$$\text{i.e. } \text{Var}(\beta_j u_j) = \beta_j^2 \psi_{jj} = \beta_j^{*2} \psi_{jj}^* = \text{Var}(\beta_j^* u_j^*)$$

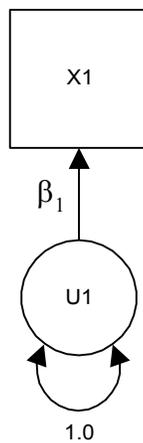
Thus there is an indeterminacy if one tries to estimate both  $\beta_j$  and  $\psi_{jj}$ .



We must therefore employ an identification condition: Either  $\beta_j = 1$  (usual)

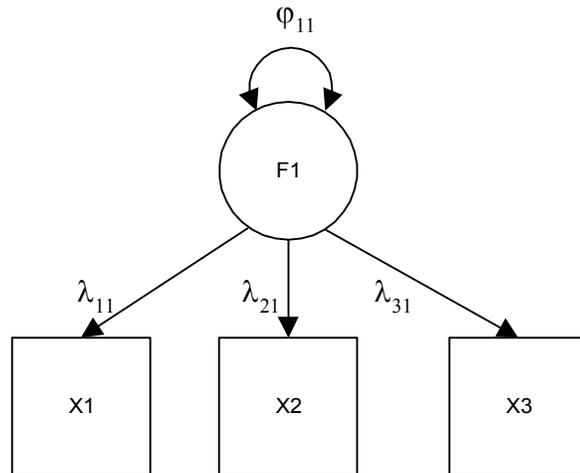


or  $\psi_{jj} = 1$  (equivalent and permissible).

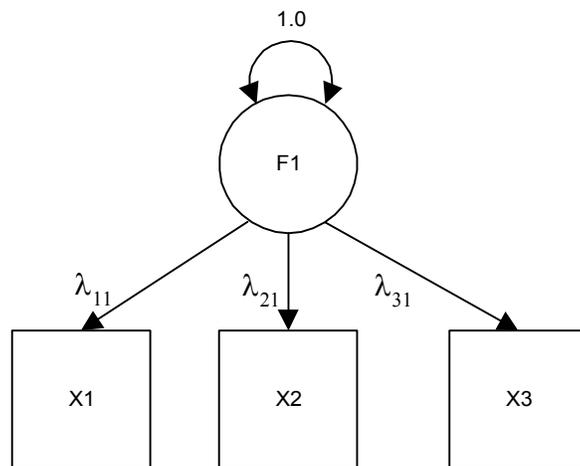


## Treatment of common factors in model specification.

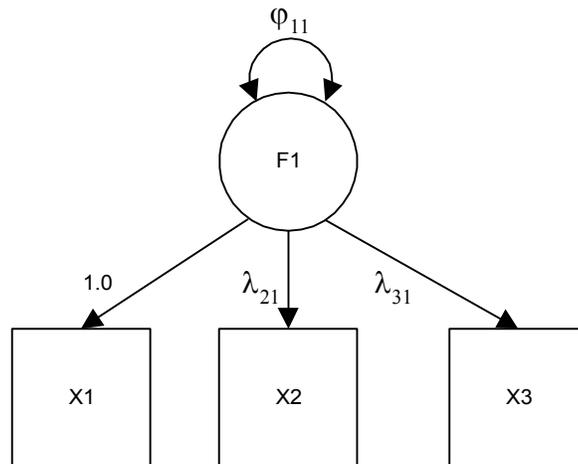
A similar indeterminacy exists for common factors.



This may be resolved either by setting the variance,  $\varphi_{11}$ , equal to 1



or by setting *one* of the factor loadings equal to 1



It is usual to employ one of two possible identification conditions:

1)  $\beta_j = 1$  with  $\psi_{jj}$  to be estimated

OR

2)  $\varphi_{kk} = 1$  with all  $\lambda_{jk}$ ,  $j = 1, \dots, p$ , to be estimated.

Under this second specification, common factor variances are constrained to be 1 for identification purposes, whereas unique factor variances would be free parameters.

In general, the scale of a latent variable (common factor or unique factor) must be set. This is usually done either by setting the variance of the latent variable or by setting the value of one of the emitted arrows.

## Estimation

Once a model is properly specified, it can be fitted to sample data.

Considering the model in covariance structure form:

$$\Sigma = \Lambda\Phi\Lambda' + D_\psi$$

The specified model defines fixed and free parameters in the three parameter matrices. Then, given a sample covariance matrix,  $\mathbf{S}$ , we seek optimal estimates of the parameters so that

$$\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{D}_{\psi}$$

matches  $\mathbf{S}$  as closely as possible. The difference between  $\hat{\Sigma}$  and  $\mathbf{S}$  is measured by a discrepancy function. The same types of discrepancy functions are used here as those in unrestricted factor analysis, most commonly maximum likelihood and ordinary least squares. Also available are normal theory generalized least squares (GLS) and asymptotically distribution free (ADF).

Given  $\mathbf{S}$  and the specified model, we seek estimates of the free parameters in  $\mathbf{\Lambda}$ ,  $\mathbf{\Phi}$ , and  $\mathbf{D}_{\psi}$ , so that the resulting  $\hat{\Sigma}$  yields the smallest possible value of the discrepancy function.

The fixed parameters are not estimated -- they retained the values specified by the investigator in the original model.

Results of this procedure include the parameter estimates, approximate standard errors of those parameter estimates, and measures of overall model fit.

### **Correlation Structures and Covariance Structures.**

Most discrepancy functions (Maximum Likelihood, Normal Theory Generalised Least Squares) in current use are intended for covariance matrices and not correlation matrices. Replacement of the covariance matrix by a correlation matrix without taking special steps will result in incorrect standard errors. (In models with equality constraints on parameters or equality restrictions (not identification conditions) to nonzero values, other difficulties may arise as well).

In RAMONA correlation matrices can be correctly handled (DISP = CORR). A factor analysis correlation structure

$$\mathbf{P} = \mathbf{\Lambda}^*\mathbf{\Phi}\mathbf{\Lambda}^{*'} + \mathbf{D}_{\psi}^*$$

is assumed, equality constraints are imposed to ensure that

$$\text{diag}[\mathbf{P}] = \mathbf{1},$$

and the model is handled as a covariance structure using

$$\Sigma = \mathbf{D}_\sigma \mathbf{P} \mathbf{D}_\sigma$$

The diagonal elements of  $\mathbf{D}_\sigma$  represent manifest variable standard deviations and are regarded as nuisance parameters (parameters that have to be included in the model but are not used for interpretation).

Note that the user of RAMONA need not specify  $\mathbf{D}_\sigma$ . Elements of  $\mathbf{A}^*$ ,  $\Phi$  and  $\psi^*$  alone need be specified.

# PSY 820      Confirmatory Factor Analysis      Module 13

## Assessing Goodness of Fit

### Confirmatory Factor Analysis Review:

#### 3 Steps:

- A. Specify Model
- B. Fit Model to Data (Parameter Estimation)
- C. Evaluate Solution

#### A. Model Specification:

Two Frameworks

1. Matrix Representation :  $\Sigma = \Lambda\Phi\Lambda' + D_\psi$  ( Three Parameter Matrices)  
Elements in  $\Lambda$ ,  $\Phi$ ,  $D_\psi$  are specified as “Fixed” or “Free”
2. Path Diagrams

#### B. Fit the Model to Data

Given  $R$ , obtain estimates of the FREE parameters in  $\Lambda$ ,  $\Phi$ ,  $D_\psi$  so that the resulting  $\hat{P}$  is as close as possible to  $R$ , i.e., minimize  $F$

It is **DANGEROUS** to carry out Exploratory Factor Analyses with a blind rotation to decide on the number of factors and the positions of fixed zero loadings and check the model by carrying out a confirmatory factor analysis using the **SAME** data. This capitalizes on chance and invalidates statistical tests, standard errors etc. Check a model obtained by means of an Exploratory Factor Analyses by carrying out a Confirmatory Factor Analysis using **NEW** data.

You should be able to specify the model before you collect your data if you want to carry out statistical tests (or use any distribution theory). It is in order to **prespecify** and then compare several models.

## C. Evaluate the Solution:

### I. Evaluate estimates of free parameters and obtain related information

- factor loadings in  $\Lambda$
- factor intercorrs in  $\Phi$
- unique variances in  $D_\psi$

#### 1. Examine the values of the estimates carefully (Do not just decide whether or not the model fits the data)

Evaluate the magnitude and sign of the parameter estimates. Are any of the parameter estimates on (or outside) their logical bounds.

e.g. Zero (negative) unique variance estimates,  $\hat{\psi}_{ii} = 0$ .

Factor intercorrelation  $|\hat{\varphi}_{ij}| = 1$  ( $|\hat{\varphi}_{ij}| > 1$ )

The solution is called an Improper or Inadmissible Solution when one or more parameter estimates fall on (or outside) the logical bounds. In situations like this one we should be cautious in proceeding with interpretation, because something may be wrong with the model or the data. (small samples with complicated models, multi-trait multi-method matrices)

RAMONA confines estimates to the admissible region (within or on the boundary) but will tell you how many estimates are on the boundary.

LISREL: allows estimates to go beyond the boundary.

Even if all parameter estimates are admissible, do they make sense? Can you explain the values of all parameters (except nuisance parameters)? Does it make sense to have a negative correlation coefficient between two method factors (peer ratings and supervisor ratings)? Why is a unique variance high?

#### 2. Use of the Standard Error for each Parameter Estimate. (Optional.

Many people do this but MWB thinks this is optional)

( Standard error = estimate of the standard deviation of the parameter estimate over repeated sampling)

How much would the estimate vary across samples? A high value implies an unstable estimate. A low value implies a stable (consistent

over resampling) estimate.

For simplicity of notation let  $\theta_j$  represent a free element of  $\Lambda$ ,  $\Phi$  or  $D_\psi$ .

Parameter:  $\theta_j$  free element of  $\Lambda$ ,  $\Phi$  or  $D_\psi$ .

Estimate:  $\hat{\theta}_j$

Standard Error :  $s(\hat{\theta}_j)$  Can be used to test the significance of each parameter estimate (one at a time, maybe Bonferroni) significance level.

$$H_{o_j} : \theta_j = 0$$

$H_{a_j} : \theta_j \neq 0$  (i.e.  $\theta_j < 0$  or  $\theta_j > 0$ ) Note that this test is **two sided!**

N.B. **NOT**  $H_o : \theta_j = 0, j = 1, \dots, q$

Test statistic:  $t = \frac{\hat{\theta}_j}{s(\hat{\theta}_j)}$  is approximately standard normal under  $H_{o_j}$  if

$N$  is large

Critical Value: 1.64 (10% level)      1.96 (5% level)

If  $|t| > 1.96$  we reject  $H_{o_j}$  at the 5% level      (If  $|t| > 1.64$  we reject  $H_{o_j}$  at the 10% level.

Usually, all  $|t|$ -values will be  $> > > 2$

These are rough significance tests and we do not know anything about the family wise error rate.

### 3. Confidence Intervals for Each Parameter (one at a time).

The confidence interval contains **more** information than the test given in 2 above. If the 90% confidence interval is given the result of the statistical test at the 10% level can be deduced. On the other hand if the result of the statistical test is given (Reject  $H_{o_j}$  or Do not Reject  $H_{o_j}$  at the 10% level), the 90% confidence interval **cannot** be deduced.

90% Confidence interval  $(\hat{\theta}_{jL}; \hat{\theta}_{jU})$

$$\text{Prob}(\hat{\theta}_{jL} < \theta_j < \hat{\theta}_{jU}) = .90$$

90% probability that the interval contains  $\theta_j$ , the true parameter value.

A narrow confidence interval implies a more precise estimate of the parameter. A few wide confidence intervals suggest some unstable parameter estimates. If all confidence intervals are wide, a larger sample is suggested.

Saying that the confidence interval contains 0 ( $\hat{\theta}_{jL} < 0$  and  $\hat{\theta}_{jU} > 0$ )

is equivalent to saying that

$H_{0j}$  is not rejected at the **10%** level (the test statistic  $t_j$  is not significant at the 10% level)

Otherwise if  $\hat{\theta}_{jL} > 0$  or  $\hat{\theta}_{jU} < 0$ ,  $H_{aj}$  is accepted at the **10%** level

i.e. the confidence interval contains all the information of a significance test plus additional information.

**Problem:** Many people go straight to the Goodness of Fit Tests and miss all this information.

## II. Investigation of Overall Goodness of Fit

Bad fit means the model is of no use.

Good fit does not tell you that the model is “true” or “correct” or even “usable”. There may be other models that fit the data just as well, and there are definitely other models that fit better (Models with more free parameters)

Tests of overall fit. Does the model *as a whole* provide a good fit to the data.

See Module 8.

**RMSEA** :  $\varepsilon_a = \sqrt{\frac{F_0}{d}}$  Measure of *discrepancy per degree of freedom*

Guidelines :

$\varepsilon_a < .05$  indicates close fit.

$0.05 < \varepsilon_a < 0.08$  indicates reasonable fit.

$0.08 < \varepsilon_a < 0.10$  indicates mediocre fit.

$\varepsilon_a > 0.10$  indicates unacceptable fit.

Test statistic:  $(N - 1)\hat{F}$

Degrees of freedom :

$$d = \frac{1}{2}p(p + 1) - q \quad (q = \text{Number of free parameters in } \mathbf{A}, \boldsymbol{\Phi}, \mathbf{D}_\psi)$$

### 1. Test of perfect fit:

$H_0 : \varepsilon_a = 0$  (equivalently  $F_0 = 0$ )

$H_a : \varepsilon_a > 0$  (equivalently  $F_0 > 0$ )

sided!

Note that this test is **one**

Even with a good model,  $H_0$  is never exactly true in the population. A good model can be expected to fit  $\boldsymbol{\Sigma}$  well, not perfectly. As in exploratory factor analysis, this test is sensitive to sample size. The power of the test increases as  $N$  increases. If  $N$  is large it is likely that  $H_0$  will be rejected even if the lack of fit,  $\varepsilon_a$ , is very close to zero. In practical applications it is unlikely that  $H_0$  will fail to be rejected if  $N$  is very large.

### 2. Test of close fit:

$H_0 : \varepsilon_a \leq 0.05$

$H_a : \varepsilon_a > 0.05$

Note that this test is **one** sided!

It is reasonable to expect that this  $H_0$  can be true in some populations. It is possible in practice for  $H_0$  to fail to be rejected even if  $N$  is very large.

### 3. Confidence interval on $\varepsilon_a$ .

90% Confidence interval  $(\hat{\varepsilon}_{aL}; \hat{\varepsilon}_{aU})$  where  $\hat{\varepsilon}_{aL}$  cannot be  $< 0$  and

$$\hat{\varepsilon}_{aU} > \hat{\varepsilon}_{aL}$$

$$\text{Prob}(\hat{\varepsilon}_{aL} < \varepsilon_a < \hat{\varepsilon}_{aU}) = .90$$

90% probability that the interval contains  $\varepsilon_a$ , the population value of the RMSEA.

If  $\hat{\varepsilon}_{aL} = 0$  then the hypothesis of perfect fit is not rejected at the **5%** level.

If  $\hat{\varepsilon}_{aL} > 0$  then the hypothesis of perfect fit is rejected at the **5%** level.

If  $\hat{\varepsilon}_{aL} \leq 0.05$  then the hypothesis of close fit is not rejected at the **5%** level.

If  $\hat{\varepsilon}_{aL} > 0.05$  then the hypothesis of close fit is rejected at the **5%** level.

### 4. Extension of the Tucker-Lewis index

In Module 8 we defined the Tucker-Lewis index of fit, often used in exploratory factor analysis as an aide to determining the number of factors to retain. This measure can be generalized to use in confirmatory factor analysis. Bentler and Bonett (1980) generalized this measure to the context of structural equation modeling (the topic of Psychology 231) and called it the "non-normed fit index," or NNFI. However, confirmatory factor analysis is a special case of SEM. Thus, the NNFI can be used, and in fact is quite widely used, as a measure of model fit in confirmatory factor analysis.

Recall that this index is based on the notion of comparing the model under study to two reference models: (1) a bad or worst-case model; (2) an ideal model.

The worst-case model is usually called a null model. It is typically defined as a model specifying that the MVs are uncorrelated in the population. Such a model can be fit to data using RAMONA by specifying that each MV has variance, but specifying no other parameters to be estimated.

The ideal model does not exist in reality but is defined theoretically as the "true model" that holds exactly in the population.

The definition of the NNFI is based on properties of the  $\chi^2$  test statistic. When a given model is correct,  $\mathcal{E}(\chi^2) = d$ . Thus, we would expect the ratio  $\chi^2/d \approx 1$  for a correct model. When a model is incorrect, this ratio becomes  $\gg 1$ . Suppose we are evaluating the fit of Model H in confirmatory factor analysis. Define

$$Q_H = \chi_H^2/d_H$$

For the null model, define

$$Q_N = \chi_N^2/d_N$$

Since the null model would probably fit the data very poorly, we expect  $Q_N$  to be very large.

For an ideal model, we would expect  $Q \approx 1$ .

NNFI is then defined as

$$\text{NNFI} = \frac{Q_N - Q_H}{Q_N - 1}$$

The rationale for this expression is as follows. The numerator represents the improvement in fit of Model H over the null model. The denominator represents the improvement in fit of an ideal model over the null model. Thus, the ratio indicates where Model H lies on a continuum where the reference points are the null model and an ideal model. If this index is used in practice, it is desirable that NNFI be well above .90 to indicate good fit. There is no confidence interval available for this index.

## 5. Look at the Residual Matrix ( $R - \hat{P}$ )

Each element is a residual or difference between a sample correlation and the corresponding correlation reproduced from the model. We look for large residuals and note any patterns (set of high residuals)

### Cautionary Note on Chi-square Based Fit Indices and Residuals

Situations can arise where the RMSEA, and other fit indices that are functions of the  $\chi^2$  test such as those in paragraph 4 above, indicate a poor fit of the model when all elements of the residual matrix are tiny. This can happen when some unique variances are very small and will happen if all unique variances are very small, indicating very accurate measurements. Situations like this seldom occur when psychological measurements are employed because very accurate measurements are seldom possible. They can occur with physiological measurements.

An intuitive explanation of this phenomenon is that the  $\chi^2$ -based fit indices measure detectability of misfit whereas the residuals measure misfit directly. If very accurate measurements are made, one can be sure that there is some misfit of the model even if the amount of misfit is small.

More information about this phenomenon may be found in Browne, MacCallum, Kim, Andersen and Glaser (2002).

## Model Comparison

It is advisable to try out several models for the data and compare them for interpretability and goodness of fit.

### A. Nested Models

Model B is said to be nested within model A if Model B is obtained from Model A by imposing additional restrictions. Free parameters in Model B are a subset of the free parameters of Model A. We shall use the notation  $\text{Model B} \subset \text{Model A}$  to stand for the statement “Model B is nested within Model A”.

Examples:

- (1) Any (truly) restricted factor analysis model is nested within the unrestricted factor analysis model as it involves additional restrictions, usually of the form  $\lambda_{ij} = 0$ .
- (2) If Model A is a restricted factor analysis model with correlated factors and Model B is obtained from Model A by requiring that all factor intercorrelations be zero, then  $\text{Model B} \subset \text{Model A}$ .

If  $\text{Model B} \subset \text{Model A}$  then:

$df(B) > df(A)$	Degrees of freedom
$q(B) < q(A)$	Number of free parameters
$\hat{F}(B) \geq \hat{F}(A)$	Sample discrepancy function value

The converse does **not** apply. For example

$$df(B) \geq df(A) \not\Rightarrow \text{Model B} \subset \text{Model A}$$

## Difference Test for the Equal Fit of Two *Nested* Models

Null and Alternative Hypotheses

$$H_0 : F_0(B) = F_0(A)$$

$$H_a : F_0(B) > F_0(A)$$

Test statistic

$$(N - 1)\widehat{F}(B) - (N - 1)\widehat{F}(A) = (N - 1)\{\widehat{F}(B) - \widehat{F}(A)\}$$

$$= \chi_B^2 - \chi_A^2$$

Asymptotic distribution under  $H_0$  : Chi squared

Degrees of freedom

$$df = df(B) - df(A) = q(A) - q(B)$$

If the test statistic exceeds the critical value then the null hypothesis is rejected (alternative hypothesis is accepted) and we can say the more parsimonious model, B, does not fit as well as the less parsimonious model, A. If the test statistic does not exceed the critical value,  $H_0$  is not rejected and we can say that sample size is not large enough for us to show that model B does not fit as well model A.

This test has the usual difficulties associated with a test of a point hypothesis. If  $N$  is large enough then the null hypothesis is very likely to be rejected. This phenomenon is not usually as noticeable as in the case of the test of perfect fit as the degrees of freedom are typically smaller and the power of the test lower.

Using this testing procedure it is possible to perform tests of the difference between nested models in a variety of interesting cases. For example:

- a) Different confirmatory (restricted) models; e.g., correlated factors vs. uncorrelated factors.
- b) Exploratory vs. confirmatory models; for a given number of factors,  $m$ , any restricted factor analysis model is nested within the unrestricted model.
- c) Exploratory models with different numbers of factors; a model with  $(m-1)$  factors is nested within a model with  $m$  factors.

### **Comment on Comparing Non-Nested Models**

If Model A and Model B are not nested, no formal statistical tests are available. Point estimates of  $\varepsilon_a(A)$  and  $\varepsilon_a(B)$  may be obtained, together with associated confidence intervals, and compared.

### **Problem Approach that is often used but should not be.**

Start with a model and fit it to the data. If it does not fit well, change the model. Model modification indices are available in some programs to help with this process. This data driven model modification process capitalizes on chance. The model modifications made are very unlikely to be supported outside the sample you are using.

A better strategy is to formulate a set of alternative models before analysing the data.

# **PSY 820    Confirmatory Factor Analysis    Module 14**

## **Computer Program and Illustrations**

This course will use the computer program RAMONA. One of its main advantages is that it can correctly analyze a correlation matrix. Alternatives include:

LISREL (Jöreskog & Sörbom, SSI)

EQS (Bentler)

MPlus (Muthén)

CALIS (SAS)

AMOS (Arbuckle, SPSS)

SePath (Steiger, Statistica)

### Obtaining and installing RAMONA

Information on obtaining RAMONA will be provided by the instructor.

Follow instructions for installation on a PC.

Double-click the SYSTAT icon to open SYSTAT.

This will start the SYSTAT package. From this point, there are three steps for conducting covariance structure analysis:

- 1) Create and save data file. (.syd)
- 2) Create and save command file. (.syc)
- 3) Execute program; save and print results.

## Creating a data file

Beginning from the main SYSTAT screen, click on FILE and then choose NEW... DATA. This will open a data worksheet and you can type your data matrix into this worksheet.

At the top of each column in the worksheet is a space for variable names. Type the variable names in first.

Beginning in the first row for data values, type the appropriate values. For a correlation or covariance matrix, it is necessary to type only the diagonal entries and those below the diagonal. Entries above the diagonal can be represented by periods: "."

For example, the correlation matrix for the Holzinger example would have the following form:

	1	2	3	4	5
	WORDMN	SENTCMP	ODDWORDS	MXDARIT	REMNDRS
1	1.000	.	.	.	.
2	0.750	1.000	.	.	.
3	0.780	0.720	1.000	.	.
4	0.440	0.520	0.470	1.000	.
5	0.450	0.530	0.480	0.820	1.000
6	0.510	0.580	0.540	0.820	0.740
7	0.210	0.230	0.280	0.330	0.370
8	0.300	0.320	0.370	0.330	0.360
9	0.310	0.300	0.370	0.310	0.360

To save this data file, click on FILE and choose SAVE AS. Provide a name for the data file, ending in .SYC. Before saving, click on OPTIONS and indicate the type of matrix, such as correlation or covariance. This step is absolutely critical. If you forget to do this, your correlation matrix will be treated by SYSTAT as if it contains raw data and all analyses will be completely incorrect (but usually, it will not run).

Note that if you wish to edit a previously created data file, simply open that file using WINDOW, DATA from the main screen, and choose the file you wish to open. Then save it as the appropriate type.

## Creating a command file

The command file is the file that instructs SYSTAT to run RAMONA and tells the program what data to use and what model to fit. To create a new command file, first click WINDOW and then choose COMMAND EDITOR. A window will open and you can then type your command file into this window. Following is an annotated example of a command file for the Holzinger three correlated factors model. Many of the statements in the command file are based explicitly on the form of the path diagram. To aid in translation of a path diagram into a command file, we include the path diagram for the Holzinger three correlated factors model on a subsequent page.

```
title 'Holzinger example'
```

*This is a title statement providing whatever information the user wants, limited to one line.*

```
ramona
```

*This statement tells SYSTAT to access RAMONA.*

```
fpath 'insert path to data file here '
```

*This statement directs SYSTAT to the data file to be used.*

```
use holznger
```

*This statement tells RAMONA to use the data set in the filename given. This file should contain a covariance or correlation matrix. RAMONA assumes that this file name has a suffix of .SYD .*

```
output holznger.out
```

*This statement specifies the name of an output file, in which RAMONA output will be saved when the program is executed.*

manifest = wordmn sentcmp oddwords mxdarit remndrs misnum gloves boots  
hatchets

*This statement specifies the names of the MVs in the model (Optional -- RAMONA considers all variables names found in the data file to be the manifest variable names).*

latent = VerbFact AritFact SpatFact u1 u2 u3 u4 u5 u6 u7 u8 u9

*This statement specifies the names of the LVs in the model, including error terms (Optional -- RAMONA considers all variable names NOT found in the data file to be the latent variable names).*

*The following set of statements specify the model. The first statement includes the word "model" to initiate the specification. Each statement then represents one or more paths in the path diagram. For each one-directional path, the variable on the left receives the arrow(s) from the variable(s) on the right. Characteristics of each parameter are represented in parentheses. For a statement of the form  $Y \leftarrow X(a,b)$ , the values of  $a$  and  $b$  define the nature of the parameter representing the effect of  $X$  on  $Y$ . The entry " $a$ " can be specified as either  $*$ ,  $0$ , or a nonzero integer. If " $a$ " is specified as  $*$ , then the parameter is free, to be estimated. If " $a$ " is set to  $0$ , then the parameter is fixed. If " $a$ " is set to some nonzero integer, that integer serves as a label for that parameter. When other parameters are assigned the same integer label, the parameters with the same label are constrained to be equal. The entry " $b$ " can be specified as either  $*$  or a numerical value. Specifying " $b$ " as  $*$  directs the program to establish an initial estimate for the parameter. Specifying " $b$ " as a numerical value directs the program to set the parameter to that value. If the parameter is fixed (" $a$ "= $0$ ), then this value becomes the fixed value of the parameter. If the parameter is free, then this value becomes the initial estimate of the parameter. To illustrate, a free parameter to be estimated is designated as  $(*,*)$ . A parameter fixed at  $1.0$  is designated  $(0,1.0)$ . Two different parameter constrained to be equal could be designated  $(1,*)$ . The status  $(*,*)$  is default and can be omitted.*

*The following lines specify the one-directional paths:*

```
model wordmn <- VerbFact u1(0,1),  
      sentcmp <- VerbFact u2(0,1),  
      oddwords <- VerbFact u3(0,1),  
      mxdarit <- AritFact u4(0,1),
```

```

remndrs <- AritFact    u5(0,1),
missnum <- AritFact    u6(0,1),
gloves  <- SpatFact    u7(0,1),
boots   <- SpatFact    u8(0,1),
hatchets <- SpatFact   u9(0,1),

```

*The following lines specify the two-directional paths:*

```

VerbFact <-> SpatFact  AritFact  VerbFact(0,1),
AritFact <-> SpatFact  AritFact(0,1),
SpatFact <-> SpatFact(0,1),
u1 <-> u1,
u2 <-> u2,
u3 <-> u3,
u4 <-> u4,
u5 <-> u5,
u6 <-> u6,
u7 <-> u7,
u8 <-> u8,
u9 <-> u9

```

*This completes the model specification. Note that each of the path specification lines ended with a comma, until the last one which does not end with a comma.*

```
print=medium
```

*This statement tells RAMONA what we want printed. By specifying "long" instead of "medium," we obtain additional information about the iterative history and the estimated covariance matrix of the parameter estimates.*

```
estimate / disp=corr method=mwl ncases=696
```

*disp=corr means we shall analyze a correlation matrix.*

*Alternatives: disp=cov*

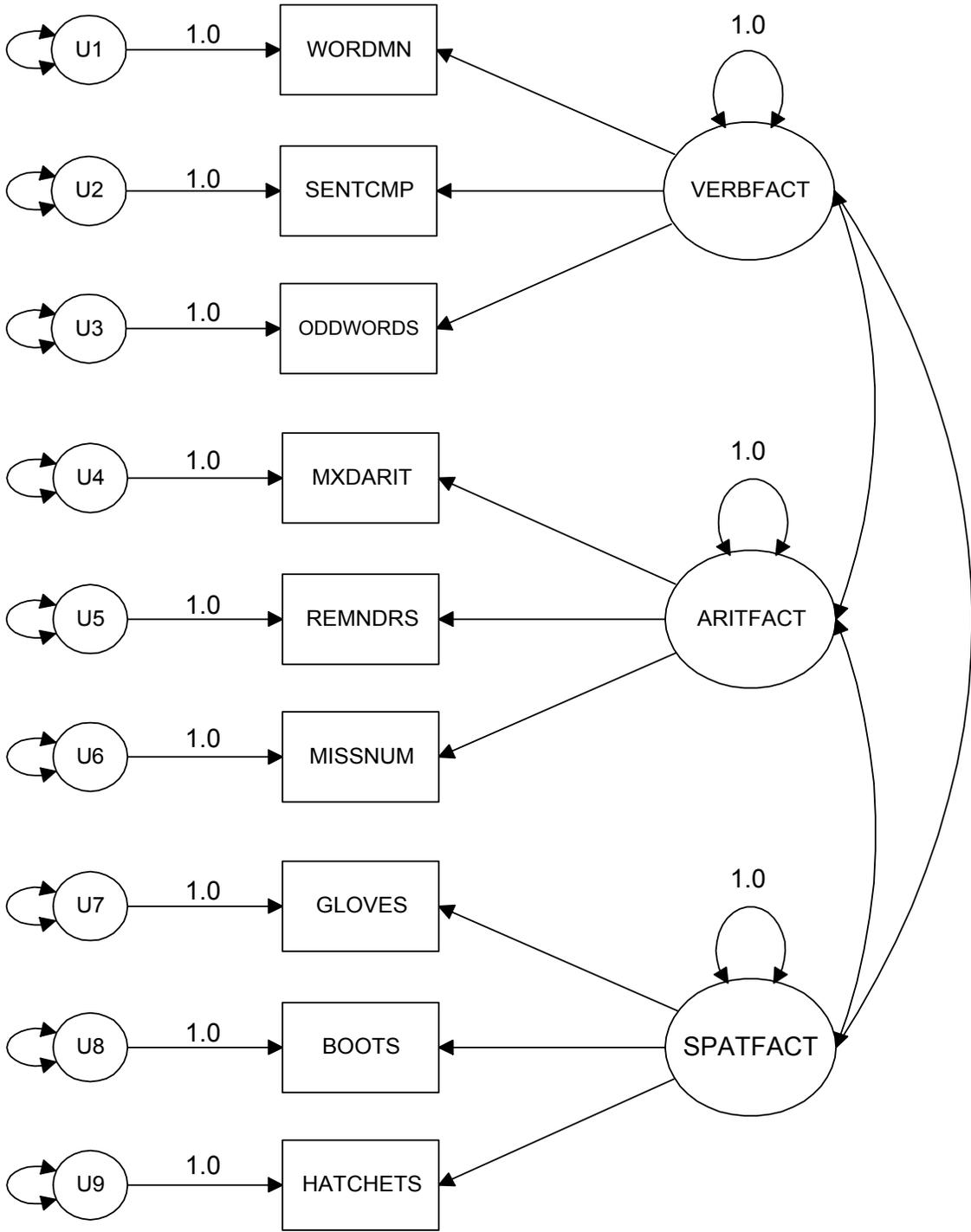
*method=mwl means we shall use maximum likelihood estimation (default).*

*Alternatives: gls,ols,adfg,adfu*

*ncases=696 specifies sample size.*

*Additional specifications available on "estimate" line:*

*maxit*= (max number of iterations; default=100)  
*convg*=(convergence criterion; default=0.0001)  
*confi*= (confidence interval width; default=90)



Path Diagram: Holzinger Data.  
3 Correlated Factors

Following is another version of this input file, without comments and making use of default representation for free parameters in the model specification:

```
ramona
use holznger
output holznger.out
manifest = wordmn sentcmp oddwords mxdarit remndrs missnum gloves boots
         hatchets
latent = VerbFact AritFact SpatFact u1 u2 u3 u4 u5 u6 u7 u8 u9
MODEL   wordmn <- VerbFact  u1(0,1),
        sentcmp <- VerbFact  u2(0,1),
        oddwords <- VerbFact  u3(0,1),
        mxdarit <- AritFact   u4(0,1),
        remndrs <- AritFact   u5(0,1),
        missnum <- AritFact   u6(0,1),
        gloves <- SpatFact   u7(0,1),
        boots <- SpatFact   u8(0,1),
        hatchets <- SpatFact  u9(0,1),
        VerbFact <-> SpatFact  AritFact  VerbFact(0,1),
        AritFact <-> SpatFact  AritFact(0,1),
        SpatFact <-> SpatFact(0,1),
        u1 <-> u1,
        u2 <-> u2,
        u3 <-> u3,
        u4 <-> u4,
        u5 <-> u5,
        u6 <-> u6,
        u7 <-> u7,
        u8 <-> u8,
        u9 <-> u9
print = medium
estimate / disp=corr ncases=696
```

Once the command file is typed, you should save it. From the command editor, click FILE and choose SAVE AS, and provide a file name. The file name should end with the extension .SYC .

Note that if you wish to edit a command file that was created previously, you may simply open it from the main SYSTAT window by choosing WINDOW and COMMAND EDITOR, and then selecting the file you want.

### Executing the command file

From the COMMAND EDITOR window, choose FILE and then choose SUBMIT WINDOW. The program will execute. You'll see results flowing by on the screen on the MAIN window, behind the COMMAND EDITOR window. When execution is complete, results will have been saved in the output filename you specified.

Alternatively, you can go back to the MAIN window, choose FILE and then choose SUBMIT FILE, and specify the name of the command file you wish to run. With this method, you will be able to see results flowing by on the screen without the interference of the COMMAND EDITOR window.

### Viewing and printing your results

Your output file can now be viewed or printed. From the SYSTAT main window, you can click FILE and choose OPEN and OUTPUT, and then choose the file you wish to open. You can then view this file, or print it by clicking FILE and PRINT.

Alternatively, you may exit SYSTAT and then view and print your file from another editor or word processor. After closing SYSTAT, you can go to a word processor or editor and open your output file. In a word processor you may wish to change the font of the output file so that columns of information will line up correctly. Set the font of the file to Courier. Alter the type size as you wish for viewing and printing.

## Overdetermined Restricted F.A. 3 Correlated Factors

**Command File: HOLZNGER.SYC**

TITLE 'Holzinger Data: Overdetermined Restricted F.A. 3 Correlated Factors'

RAMONA

USE HOLZNGER

OUTPUT HOLZNGER.OUT

MANIFEST = WORDMN SENTCMP ODDWORDS MXDARIT REMNDRS MISSNUM  
GLOVES BOOTS HATCHETS

LATENT = VERBFACT ARITFACT SPATFACT U1 U2 U3 U4 U5 U6 U7 U8 U9

```
MODEL WORDMN <- VERBFACT U1(0,1),
      SENTCMP <- VERBFACT U2(0,1),
      ODDWORDS <- VERBFACT U3(0,1),
      MXDARIT <- ARITFACT U4(0,1),
      REMNDRS <- ARITFACT U5(0,1),
      MISSNUM <- ARITFACT U6(0,1),
      GLOVES <- SPATFACT U7(0,1),
      BOOTS <- SPATFACT U8(0,1),
      HATCHETS <- SPATFACT U9(0,1),
      VERBFACT <-> SPATFACT ARITFACT VERBFACT(0,1),
      ARITFACT <-> SPATFACT ARITFACT(0,1),
      SPATFACT <-> SPATFACT(0,1),
      U1 <-> U1,
      U2 <-> U2,
      U3 <-> U3,
      U4 <-> U4,
      U5 <-> U5,
      U6 <-> U6,
      U7 <-> U7,
      U8 <-> U8,
      U9 <-> U9
```

PRINT = MEDIUM

ESTIMATE / DISP=CORR NCASES=696

## Output File: HOLZNGER.OUT

Holzinger Data: Overdetermined Restricted F.A. 3 Correlated Factors

Variables in the SYSTAT Correlation file are:

WORDMN SENTCMP ODDWORDS MXDARIT REMNDRS MISSNUM  
GLOVES BOOTS HATCHETS

There are 9 manifest variables in the model. They are:

WORDMN SENTCMP ODDWORDS MXDARIT REMNDRS MISSNUM GLOVES  
BOOTS HATCHETS

There are 12 latent variables in the model. They are:

VERBFACT U1 U2 U3 ARITFACT U4 U5 U6 SPATFACT U7 U8 U9

RAMONA options in effect are:

Display Corr  
Method MWL  
Start Rough  
Convg 0.000100  
Maximum iterations 100  
Number of cases 696  
Restart No  
Confidence Interval 90.000000

Number of manifest variables = 9

Total number of variables in the system = 30.

Reading correlation matrix...

Details of Iterations

Iter	Method	Discr. Funct.	Max.R.Cos.	Max.Const.	NRP	NBD
0	OLS	2.849300	0.0			
1(0)	OLS	0.973282	0.836463	0.194565	0	0
2(0)	OLS	0.052976	0.362324	0.008532	0	0
3(0)	OLS	0.045465	0.007787	0.000069	0	0
3(0)	MWL	0.305789	0.273349	0.000069	0	0
4(0)	MWL	0.199505	0.144735	0.002787	0	0
5(0)	MWL	0.191620	0.034364	0.000170	0	0
6(0)	MWL	0.190928	0.012441	0.000023	0	0
7(0)	MWL	0.190865	0.003330	0.000002	0	0
8(0)	MWL	0.190858	0.001154	0.000000	0	0
9(0)	MWL	0.190857	0.000346	0.000000	0	0
10(0)	MWL	0.190857	0.000117	0.000000	0	0
11(0)	MWL	0.190857	0.000038	0.000000	0	0
12(0)	MWL	0.190857	0.000013	0.000000	0	0

Iterative procedure complete.

Convergence limit for residual cosines = 0.000100 on 2 consecutive iterations.

Convergence limit for variance constraint violations = 5.00000E-07

Value of the maximum variance constraint violation = 3.21914E-11

Holzinger Data: Overdetermined Restricted F.A. 3 Correlated Factors

Sample Correlation Matrix :

	WORDMN	SENTCMP	ODDWORDS	MXDARIT	REMNDRS
WORDMN	1.000				
SENTCMP	0.750	1.000			
ODDWORDS	0.780	0.720	1.000		
MXDARIT	0.440	0.520	0.470	1.000	
REMNDRS	0.450	0.530	0.480	0.820	1.000
MISSNUM	0.510	0.580	0.540	0.820	0.740
GLOVES	0.210	0.230	0.280	0.330	0.370
BOOTS	0.300	0.320	0.370	0.330	0.360
HATCHETS	0.310	0.300	0.370	0.310	0.360
	MISSNUM	GLOVES	BOOTS	HATCHETS	
MISSNUM	1.000				
GLOVES	0.350	1.000			
BOOTS	0.380	0.450	1.000		
HATCHETS	0.380	0.520	0.670	1.000	

Number of cases = 696.

Reproduced Correlation Matrix :

	WORDMN	SENTCMP	ODDWORDS	MXDARIT	REMNDRS
WORDMN	1.000				
SENTCMP	0.747	1.000			
ODDWORDS	0.770	0.738	1.000		
MXDARIT	0.517	0.496	0.511	1.000	
REMNDRS	0.484	0.464	0.478	0.809	1.000
MISSNUM	0.489	0.469	0.484	0.817	0.765
GLOVES	0.246	0.236	0.243	0.275	0.257
BOOTS	0.319	0.305	0.315	0.356	0.333
HATCHETS	0.343	0.329	0.339	0.383	0.359

	MISSNUM	GLOVES	BOOTS	HATCHETS
MISSNUM	1.000			
GLOVES	0.260	1.000		
BOOTS	0.337	0.477	1.000	
HATCHETS	0.363	0.514	0.664	1.000

Residual Matrix (correlations) :

	WORDMN	SENTCMP	ODDWORDS	MXDARIT	REMNDRS
WORDMN	-0.000				
SENTCMP	0.003	-0.000			
ODDWORDS	0.010	-0.018	-0.000		
MXDARIT	-0.077	0.024	-0.041	-0.000	
REMNDRS	-0.034	0.066	0.002	0.011	-0.000
MISSNUM	0.021	0.111	0.056	0.003	-0.025
GLOVES	-0.036	-0.006	0.037	0.055	0.113
BOOTS	-0.019	0.015	0.055	-0.026	0.027
HATCHETS	-0.033	-0.029	0.031	-0.073	0.001

	MISSNUM	GLOVES	BOOTS	HATCHETS
MISSNUM	-0.000			
GLOVES	0.090	-0.000		
BOOTS	0.043	-0.027	-0.000	
HATCHETS	0.017	0.006	0.006	-0.000

Value of the maximum absolute residual = 0.113.

ML Estimates of Free Parameters in Dependence Relationships

Path	Param #	Point Estimate	90.00% Conf. Int. Lower	90.00% Conf. Int. Upper	Standard Error	T Value
WORDMN <- VERBFACT	1	0.883	0.863	0.902	0.012	73.22
SENTCMP <- VERBFACT	2	0.846	0.823	0.869	0.014	61.51
ODDWORDS <- VERBFACT	3	0.872	0.851	0.893	0.013	69.67
MXDARIT <- ARITFACT	4	0.929	0.915	0.944	0.009	107.41
REMNDRS <- ARITFACT	5	0.870	0.851	0.889	0.011	76.11
MISSNUM <- ARITFACT	6	0.879	0.861	0.897	0.011	80.42
GLOVES <- SPATFACT	7	0.607	0.561	0.654	0.028	21.39
BOOTS <- SPATFACT	8	0.785	0.748	0.822	0.023	34.87
HATCHETS <- SPATFACT	9	0.846	0.811	0.881	0.021	39.89

Scaled Standard Deviations (nuisance parameters)

Variable	Estimate
WORDMN	1.000
SENTCMP	1.000
ODDWORDS	1.000
MXDARIT	1.000
REMNDRS	1.000
MISSNUM	1.000
GLOVES	1.000
BOOTS	1.000
HATCHETS	1.000

Values of Fixed Parameters in Dependence Relationships

Path	Value
WORDMN <- U1	1.000
SENTCMP <- U2	1.000
ODDWORDS <- U3	1.000
MXDARIT <- U4	1.000
REMNDRS <- U5	1.000
MISSNUM <- U6	1.000
GLOVES <- U7	1.000
BOOTS <- U8	1.000
HATCHETS <- U9	1.000

Holzinger Data: Overdetermined Restricted F.A. 3 Correlated Factors  
ML estimates of free parameters in variance/covariance relationships

Path	Param #	Point Estimate	90.00% Conf. Int. Lower	Standard Error	T Value
VERBFACT <->SPATFACT	10	0.460	0.398	0.517	0.036 12.71
VERBFACT <->ARITFACT	11	0.631	0.585	0.672	0.027 23.76
ARITFACT <->SPATFACT	12	0.487	0.428	0.542	0.035 14.09
U1 <->U1	13	0.221	0.189	0.259	0.021 10.39
U2 <->U2	14	0.284	0.249	0.325	0.023 12.23
U3 <->U3	15	0.239	0.206	0.278	0.022 10.97
U4 <->U4	16	0.136	0.112	0.165	0.016 8.46
U5 <->U5	17	0.243	0.213	0.278	0.020 12.23
U6 <->U6	18	0.227	0.197	0.260	0.019 11.78
U7 <->U7	19	0.631	0.577	0.691	0.034 18.31
U8 <->U8	20	0.383	0.329	0.446	0.035 10.84
U9 <->U9	21	0.284	0.231	0.350	0.036 7.93

Values of Fixed Parameters in Variance/Covariance Relationships

Path	Value
VERBFACT <->VERBFACT	1.000
ARITFACT <->ARITFACT	1.000
SPATFACT <->SPATFACT	1.000

Equality Constraints on Variances

Constraint	Lagrange Value	Multiplier	Standard Error
WORDMN <-> WORDMN	1.0000	-0.000	0.000
SENTCMP <-> SENTCMP	1.0000	0.000	0.000
ODDWORDS <-> ODDWORDS	1.0000	0.000	-0.000
MXDARIT <-> MXDARIT	1.0000	0.000	-0.000
REMNDRS <-> REMNDRS	1.0000	-0.000	0.000
MISSNUM <-> MISSNUM	1.0000	-0.000	0.000
GLOVES <-> GLOVES	1.0000	0.000	-0.000
BOOTS <-> BOOTS	1.0000	-0.000	0.000
HATCHETS <-> HATCHETS	1.0000	-0.000	0.000

Holzinger Data: Overdetermined Restricted F.A. 3 Correlated Factors

Maximum Likelihood Discrepancy Function

Measures of fit of the model

-----  
Sample Discrepancy Function Value : 0.191 (1.908572E-01)

Population discrepancy function value, Fo

Bias adjusted point estimate : 0.156  
90.000 percent confidence interval : (0.110,0.214)

Root mean square error of approximation

Steiger-Lind : RMSEA = SQRT(Fo/DF)  
Point estimate : 0.081  
90.000 percent confidence interval : (0.068,0.094)

Expected cross-validation index

Point estimate (modified aic) : 0.251  
90.000 percent confidence interval : (0.205,0.309)  
CVI (modified AIC) for the saturated model : 0.129

Test statistic: : 132.646

Exceedance probabilities:-

Ho: perfect fit (RMSEA = 0.0) : 0.0  
Ho: close fit (RMSEA <= 0.050) : 0.000

Multiplier for obtaining test statistic = 695.000

Degrees of freedom = 24

Effective number of parameters = 21

### **Calculation of the NNFI:**

The fit measures provided by RAMONA do not include the NNFI. It can be calculated by hand easily. We first obtain the discrepancy function value, test statistic, and df for the null model. These were given in a table in Module 9:

Null Model:  $\hat{F}_N = 5.907$ ; Test statistic  $\chi_N^2 = 4105.689$ ;  $df_N = 36$

$$Q_N = \chi_N^2 / df_N = 114.047$$

Tested Model:  $\hat{F}_H = .191$ ; Test statistic  $\chi_H^2 = 132.646$ ;  $df_H = 24$

$$Q_H = \chi_H^2 / df_H = 5.527$$

$$\text{NNFI} = \frac{Q_N - Q_H}{Q_N - 1} = \frac{114.047 - 5.527}{114.047 - 1} = .960$$

This result indicates that this restricted factor model fits the data quite well according to the NNFI.

## Effect of Extracting 2 Factors on Residuals

**Command File: HOLZNGE2.SYC**

OUTPUT HOLZNGE2.OUT

TITLE 'Holzinger Data: Restricted F.A. 2 Correlated Factors'

RAMONA

USE HOLZNGER

```
MODEL WORDMN <- VERBFACT U1(0,1),
      SENTCMP <- VERBFACT U2(0,1),
      ODDWORDS <- VERBFACT U3(0,1),
      MXDARIT <- ARITFACT U4(0,1),
      REMNDRS <- ARITFACT U5(0,1),
      MISSNUM <- ARITFACT U6(0,1),
      GLOVES <- ARITFACT VERBFACT U7(0,1),
      BOOTS <- ARITFACT VERBFACT U8(0,1),
      HATCHETS <- ARITFACT VERBFACT U9(0,1),
      VERBFACT <-> ARITFACT VERBFACT(0,1),
      ARITFACT <-> ARITFACT(0,1),
      U1 <-> U1,
      U2 <-> U2,
      U3 <-> U3,
      U4 <-> U4,
      U5 <-> U5,
      U6 <-> U6,
      U7 <-> U7,
      U8 <-> U8,
      U9 <-> U9
```

PRINT = LONG

ESTIMATE / DISP=CORR NCASES=696

**Selected Output from: HOLZNGE2.OUT**

Holzinger Data: Restricted F.A. 2 Correlated Factors  
Residual Matrix (correlations) :

	WORDMN	SENTCMP	ODDWORDS	MXDARIT	REMNDRS
WORDMN	-0.000				
SENTCMP	0.009	-0.000			
ODDWORDS	0.010	-0.018	-0.000		
MXDARIT	-0.074	0.027	-0.042	-0.000	
REMNDRS	-0.036	0.064	-0.004	0.016	-0.000
MISSNUM	0.019	0.109	0.051	0.008	-0.028
GLOVES	-0.051	-0.020	0.021	-0.041	0.019
BOOTS	-0.050	-0.015	0.022	-0.052	-0.002
HATCHETS	-0.038	-0.034	0.023	-0.066	0.004

	MISSNUM	GLOVES	BOOTS	HATCHETS
MISSNUM	-0.000			
GLOVES	-0.005	-0.000		
BOOTS	0.014	0.274	-0.000	
HATCHETS	0.020	0.346	0.470	-0.000

Value of the maximum absolute residual = 0.470.

Maximum Likelihood Discrepancy Function

Measures of fit of the model

-----  
Sample Discrepancy Function Value : 0.816 (8.155288E-01)

Root mean square error of approximation

Steiger-Lind : RMSEA = SQRT(Fo/DF)

Point estimate : 0.184

90.000 percent confidence interval :(0.171,0.198)

Test statistic: : 566.793

Exceedance probabilities:-

Ho: perfect fit (RMSEA = 0.0) : 0.0

Ho: close fit (RMSEA <= 0.050) : 0.000

Multiplier for obtaining test statistic = 695.000

Degrees of freedom = 23

Effective number of parameters = 22

## Just Determined Constrained Factor Analysis: 3 Correlated Factors

Holzinger Data: Just Determined Constrained F.A. 3 Correlated Factors

### Adequately (But Minimally) Specified Factor Matrix

WORDMN	x	0	*
SENTCMP	x	*	0
ODDWORDS	x	*	*
MXDARIT	0	x	*
REMNDRS	*	x	0
MISSNUM	*	x	*
GLOVES	0	*	x
BOOTS	*	0	x
HATCHETS	*	*	x

---

```

MODEL  WORDMN <- VERBFACT  ARITFACT(0,0) SPATFACT(*,0) U1(0,1),
        SENTCMP <- VERBFACT  ARITFACT(*,0) SPATFACT(0,0) U2(0,1),
        ODDWORDS <- VERBFACT  ARITFACT(*,0) SPATFACT(*,0) U3(0,1),
        MXDARIT <- VERBFACT(0,0) ARITFACT  SPATFACT(*,0) U4(0,1),
        REMNDRS <- VERBFACT(*,0) ARITFACT  SPATFACT(0,0) U5(0,1),
        MISSNUM <- VERBFACT(*,0) ARITFACT  SPATFACT(*,0) U6(0,1),
        GLOVES <- VERBFACT(0,0) ARITFACT(*,0) SPATFACT  U7(0,1),
        BOOTS <- VERBFACT(*,0) ARITFACT(0,0) SPATFACT  U8(0,1),
        HATCHETS <- VERBFACT(*,0) ARITFACT(*,0) SPATFACT  U9(0,1),
        VERBFACT <-> SPATFACT  ARITFACT  VERBFACT(0,1),
        ARITFACT <-> SPATFACT  ARITFACT(0,1),
        SPATFACT <-> SPATFACT(0,1),
        U1 <-> U1,
        U2 <-> U2,
        ⋮ ⋮ ⋮
        U9 <-> U9
    
```

Note: (0,0) denotes a factor loading fixed at 0.

(\* ,0) denotes a starting value of 0 for a free factor loading

Details of Iterations						
Iter	Method	Discr. Funct.	Max.R.Cos.	Max.Const.	NRP	NBD
0	OLS	3.488050		0.083333		
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9(0)	MWL	<b>0.016359</b>	0.000002	0.000000	0	0

Holzinger Data: Just Determined Constrained F.A. 3 Correlated Factors

ML Estimates of Free Parameters in Dependence Relationships

Point 90.00% Conf. Int. Standard T

Path	Param #	Estimate	Lower	Upper	Error	Value
WORDMN <- VERBFACT	1	0.896	0.866	0.926	0.018	48.98
WORDMN <- SPATFACT	2	0.029	-0.037	0.095	0.040	0.73
SENTCMP <- VERBFACT	3	0.736	0.693	0.779	0.026	28.29
SENTCMP <- ARITFACT	4	0.182	0.125	0.239	0.035	5.26
ODDWORDS <- VERBFACT	5	0.799	0.754	0.845	0.028	28.89
ODDWORDS <- ARITFACT	6	0.044	-0.012	0.100	0.034	1.29
ODDWORDS <- SPATFACT	7	0.122	0.058	0.186	0.039	3.16
MXDARIT <- ARITFACT	8	1.019	0.981	1.058	0.024	43.08
MXDARIT <- SPATFACT	9	-0.126	-0.186	-0.066	0.037	-3.44
REMNDRS <- VERBFACT	10	0.081	0.033	0.129	0.029	2.78
REMNDRS <- ARITFACT	11	0.816	0.781	0.852	0.022	37.47
MISSNUM <- VERBFACT	12	0.173	0.127	0.219	0.028	6.16
MISSNUM <- ARITFACT	13	0.766	0.721	0.811	0.027	27.98
MISSNUM <- SPATFACT	14	0.010	-0.045	0.065	0.033	0.29
GLOVES <- ARITFACT	15	0.139	0.061	0.218	0.048	2.91
GLOVES <- SPATFACT	16	0.528	0.463	0.593	0.040	13.33
BOOTS <- VERBFACT	17	0.112	0.014	0.210	0.059	1.88
BOOTS <- SPATFACT	18	0.722	0.668	0.776	0.033	22.15
HATCHETS <- VERBFACT	19	0.097	-0.012	0.206	0.066	1.46
HATCHETS <- ARITFACT	20	-0.079	-0.165	0.006	0.052	-1.52
HATCHETS <- SPATFACT	21	0.889	0.810	0.969	0.048	18.42

ML estimates of free parameters in variance/covariance relationships

Point 90.00% Conf. Int. Standard T

Path	Param #	Estimate	Lower	Upper	Error	Value
VERBFACT <->SPATFACT	22	0.287	0.145	0.416	0.083	3.46
VERBFACT <->ARITFACT	23	0.503	0.437	0.563	0.038	13.14
ARITFACT <->SPATFACT	24	0.498	0.397	0.587	0.058	8.62
U1 <->U1	25	0.181	0.147	0.224	0.023	7.82
U2 <->U2	26	0.290	0.255	0.330	0.023	12.73
U3 <->U3	27	0.248	0.214	0.287	0.022	11.12
U4 <->U4	28	0.073	0.048	0.112	0.019	3.87
U5 <->U5	29	0.260	0.229	0.297	0.021	12.66
U6 <->U6	30	0.241	0.212	0.275	0.019	12.74
U7 <->U7	31	0.629	0.575	0.687	0.034	18.48
U8 <->U8	32	0.420	0.362	0.487	0.038	11.09
U9 <->U9	33	0.222	0.158	0.312	0.046	4.84

Holzinger Data: Just Determined Constrained F.A. 3 Correlated Factors  
Maximum Likelihood Discrepancy Function

Measures of fit of the model

-----  
Sample Discrepancy Function Value : 0.016 (1.635908E-02)

Population discrepancy function value, Fo  
Bias adjusted point estimate : 0.0  
90.000 percent confidence interval : (0.0,0.016)

Root mean square error of approximation  
Steiger-Lind : RMSEA = SQRT(Fo/DF)  
Point estimate : 0.0  
90.000 percent confidence interval : (0.0,0.037)

Expected cross-validation index  
Point estimate (modified aic) : 0.111  
90.000 percent confidence interval : (0.112,0.129)  
CVI (modified AIC) for the saturated model : 0.129

Test statistic: : 11.370  
Exceedance probabilities:-  
Ho: perfect fit (RMSEA = 0.0) : 0.498  
Ho: close fit (RMSEA <= 0.050) : 0.995

Multiplier for obtaining test statistic = 695.000  
Degrees of freedom = 12  
Effective number of parameters = 33

## Underdetermined but Restricted Factor Analysis: 3 Correlated Factors

Holzinger Data: Underdetermined but Restricted F.A. 3 Correlated Factors

### Inadequately Specified Factor Matrix

	WORDMN	x	0	*
	SENTCMP	x	0	*
	ODDWORDS		x	*
*				
	MXDARIT	0	x	*
	REMNDRS	0	x	*
	MISSNUM	*	x	*
	GLOVES	0	*	x
	BOOTS	*	0	x
	HATCHETS	*	*	x

### Adequately Specified F. M.

	x	0	*
	x	*	0
	*	x	*
	0	x	*
	*	x	0
	*	x	*
	0	*	x
	*	0	x
	*	*	x

6 zero loadings in both cases. No specified zero loadings in 3rd column of inadequately specified matrix.

Holzinger Data: Underdetermined but Restricted F.A. 3 Correlated Factors

#### Details of Iterations

Iter	Method	Discr. Funct.	Max.R.Cos.	Max.Const.	<b>NRP</b>	<b>NBD</b>
0	OLS	3.350133		0.083333		
1(0)	OLS	1.667710	0.683411	0.477437	2	0
2(0)	OLS	0.031204	0.437997	0.060703	2	0
3(0)	OLS	0.011899	0.064035	0.004662	2	0
4(0)	OLS	0.011748	0.012795	0.000798	2	0
5(0)	OLS	0.011745	0.000375	0.000008	2	0
5(0)	MWL	0.106750	0.288558	0.000008	2	0
6(0)	MWL	0.068049	0.115931	0.002637	2	0
7(0)	MWL	0.065503	0.026643	0.000998	2	0
8(0)	MWL	0.065315	0.009283	0.000430	2	0
9(0)	MWL	0.065298	0.002061	0.000009	2	0
10(0)	MWL	0.065296	0.000663	0.000000	2	0
11(0)	MWL	0.065296	0.000168	0.000000	2	0
12(0)	MWL	0.065296	0.000054	0.000000	2	0
13(0)	MWL	<b>0.065296</b>	0.000014	0.000000	<b>2</b>	<b>0</b>

Holzinger Data: Underdetermined but Restricted F.A. 3 Correlated Factors  
ML Estimates of Free Parameters in Dependence Relationships

Path	Param #	Point Estimate	90.00% Conf. Lower	Int. Upper	Standard Error	T Value
WORDMN <- VERBFACT	1	0.912	0.876	0.949	0.022	40.73
WORDMN <- SPATFACT	2	-0.064	-0.132	0.003	0.041	-1.56
SENTCMP <- VERBFACT	3	0.857	0.820	0.895	0.023	37.62
SENTCMP <- SPATFACT	4	-0.025	-0.093	0.042	0.041	-0.62
ODDWORDS <- VERBFACT	5	0.869	0.817	0.921	0.032	27.50
ODDWORDS <- ARITFACT	6	-0.046	-0.117	0.024	0.043	-1.08
ODDWORDS <- SPATFACT	7 **	0.071			0.0	
MXDARIT <- ARITFACT	8	1.260	1.186	1.335	0.045	27.77
MXDARIT <- SPATFACT	9	-0.514	-0.592	-0.437	0.047	-10.91
REMNDRS <- ARITFACT	10	1.067	1.009	1.124	0.035	30.32
REMNDRS <- SPATFACT	11	-0.323	-0.397	-0.249	0.045	-7.15
MISSNUM <- VERBFACT	12	0.168	0.119	0.216	0.029	5.70
MISSNUM <- ARITFACT	13	0.946	0.899	0.994	0.029	32.79
MISSNUM <- SPATFACT	14 **	-0.295			0.0	
GLOVES <- ARITFACT	15	0.161	0.060	0.262	0.061	2.61
GLOVES <- SPATFACT	16	0.482	0.389	0.574	0.056	8.55
BOOTS <- VERBFACT	17	0.082	-0.024	0.189	0.065	1.27
BOOTS <- SPATFACT	18	0.723	0.659	0.787	0.039	18.51
HATCHETS <- VERBFACT	19	0.059	-0.060	0.178	0.073	0.82
HATCHETS <- ARITFACT	20	-0.102	-0.216	0.011	0.069	-1.49
HATCHETS <- SPATFACT	21	0.928	0.812	1.044	0.070	13.18

\*\* Denotes a redundant parameter

ML estimates of free parameters in variance/covariance relationships

Path	Param #	Point Estimate	90.00% Conf. Lower	Int. Upper	Standard Error	T Value
VERBFACT <-> SPATFACT	22	0.425	0.284	0.547	0.080	5.28
VERBFACT <-> ARITFACT	23	0.621	0.554	0.680	0.038	16.25
ARITFACT <-> SPATFACT	24	0.722	0.655	0.778	0.037	19.41
U1 <-> U1	25	0.213	0.180	0.253	0.022	9.68
U2 <-> U2	26	0.283	0.247	0.324	0.023	12.12
U3 <-> U3	27	0.239	0.205	0.279	0.022	10.81
U4 <-> U4	28	0.083	0.059	0.118	0.017	4.77
U5 <-> U5	29	0.256	0.224	0.292	0.020	12.48
U6 <-> U6	30	0.238	0.209	0.271	0.019	12.76
U7 <-> U7	31	0.631	0.577	0.689	0.034	18.63
U8 <-> U8	32	0.420	0.362	0.487	0.038	11.06
U9 <-> U9	33	0.222	0.158	0.313	0.046	4.80

Holzinger Data: Underdetermined but Restricted F.A. 3 Correlated Factors  
Maximum Likelihood Discrepancy Function

Measures of fit of the model  
-----

Sample Discrepancy Function Value : 0.065 (6.529623E-02)

Population discrepancy function value, Fo  
Bias adjusted point estimate : 0.045  
90.000 percent confidence interval :(0.021,0.080)

Root mean square error of approximation  
Steiger-Lind : RMSEA = SQRT(Fo/DF)  
Point estimate : 0.057  
90.000 percent confidence interval :(0.039,0.076)

Expected cross-validation index  
Point estimate (modified aic) : 0.155  
90.000 percent confidence interval :(0.130,0.189)  
CVI (modified AIC) for the saturated model : 0.129

Test statistic: : 45.381  
Exceedance probabilities:-  
Ho: perfect fit (RMSEA = 0.0) : 0.000  
Ho: close fit (RMSEA <= 0.050) : 0.248

Multiplier for obtaining test statistic = 695.000  
Degrees of freedom = 14  
Effective number of parameters = 31

## Just Determined Constrained Factor Analysis: 3 Uncorrelated Factors

Holzinger Data: Just Determined Constrained F.A. 3 Uncorrelated Factors  
**Adequately (but minimally) Specified Factor Matrix**

WORDMN	x	0	*
SENTCMP	x	*	0
ODDWORDS	x	*	*
MXDARIT	*	x	0
REMNDRS	*	x	*
MISSNUM	*	x	*
GLOVES	*	*	x
BOOTS	*	*	x
HATCHETS	*	*	x

### Details of Iterations

Iter	Method	Discr. Funct.	Max.R.Cos.	Max.Const.	NRP	NBD
0	OLS	5.958467		0.083333		
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9(0)	MWL	<b>0.016359</b>	0.000003	0.000000	<b>0</b>	<b>0</b>

Maximum Likelihood Discrepancy Function  
 Measures of fit of the model

Sample Discrepancy Function Value : 0.016 (1.635908E-02)

Root mean square error of approximation  
 Steiger-Lind : RMSEA = SQRT(Fo/DF)  
 Point estimate : 0.0  
 90.000 percent confidence interval : (0.0,0.037)

Test statistic: : 11.370  
 Exceedance probabilities:-  
 Ho: perfect fit (RMSEA = 0.0) : 0.498  
 Ho: close fit (RMSEA <= 0.050) : 0.995

Multiplier for obtaining test statistic = 695.000  
 Degrees of freedom = 12  
 Effective number of parameters = 33

## Difference Test

Difference test for significance of specified factor loadings:

$$\text{Test statistic} = 695 \times (0.190857 - 0.016359) = 132.65 - 11.37 = 121.28$$

$$\text{df} = 24 - 12 = 12$$

$$p < .001$$

This result implies that we reject the hypothesis that the fit of these two models is equivalent, meaning we prefer the less restricted model because it fits the data significantly better.

## References

Albert, A. A. (1944a). The matrices of factor analysis. *Proceedings of the National Academy of Sciences*, 30, 90-95.

Albert, A. A. (1944b). The minimum rank of a correlation matrix. *Proceedings of the National Academy of Sciences*, 30, 144-146.

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *The American Statistician*, 21, 17-21.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, 38, 25-56.

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

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230-258.

Browne, M. W., MacCallum, R. C., Kim, C., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7, 403-421.

Carroll, J. B. (1953). An analytic solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23-38.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.

Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, 12, 289-325.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, 35, 321-332.

Cudeck, R. (1991). Noniterative factor analysis estimators with algorithms for subset and instrumental variable selection. *Journal of Educational Statistics*, 16, 35-52.

Cudeck, R. (2000). Factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of multivariate statistics and mathematical modeling* (pp. 265-296). New York: Academic Press.

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.

Everitt, B. S. (1975). Multivariate analysis: The need for data, and other problems. *British Journal of Psychiatry*, 126, 237-240.

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.

Ferguson, G. A. (1954). The concept of parsimony in factor analysis. *Psychometrika*, 19, 281-290.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39, 291-314.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1990). Common factor analysis versus principal components analysis: Some well and little known facts. *Multivariate Behavioral Research*, 25, 33-39.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Guttman, L. (1940). Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 5, 75-99.

Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19, 149-161.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, 17, 193-219.

Harman, H. H. (1960). *Modern factor analysis*. Chicago: University of Chicago Press.

Harris, C. W., & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 29, 347-362.

Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17, 65-70.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Hubbard, R., & Allan, S. J. (1987). A cautionary note on the use of principal components analysis: Supportive empirical evidence. *Sociological Methods and Research*, 16, 301-308.

Ihara, M., & Kano, Y. (1986). A new estimator of the uniqueness in factor analysis. *Psychometrika*, 51, 563-566.

Jennrich, R. I., & Sampson, P. F. (1966). Rotation for simple loadings.

*Psychometrika*, 31, 313-323.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32, 443-482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.

Kaiser, H. F. (1970). A second generation Little-Jiffy. *Psychometrika*, 35, 401-415.

Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.

Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, 33, 37-72.

MacCallum, R. C. (1983). A comparison of factor analysis programs in SPSS, BMDP, and SAS. *Psychometrika*, 48, 223-231.

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in factor analysis: Implications for theory and practice. *Psychological Bulletin*, 109, 502-511.

MacCallum, R. C., Tucker, L. R., & Briggs, N. E. (2001). An alternative perspective on parameter estimation in factor analysis and related methods. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future*. (pp. 39-57). Lincolnwood, IL: SSI.

MacCallum, R. C., Widaman, K. F., Preacher, K., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, 36, 611-637.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999) Sample size in factor analysis. *Psychological Methods*, 4, 84-99.

Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.

Mulaik, S. A. (1990). Blurring the distinction between component analysis and common factor analysis. *Multivariate Behavioral Research*, 25, 53-59.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.

Neuhaus, J. O., & Wrigley, C. (1954). The quartimax method: An analytical approach to orthogonal simple structure. *British Journal of Statistical Psychology*, 7, 81-91.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-32.

Saunders, D. R. (1953). *An analytic method for rotation to orthogonal simple structure*. Research Bulletin 53-10. Princeton, NJ: ETS.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, 106, 148-154.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Steiger, J. H. (1979). Factor indeterminacy in the 1930's and 1970's: Some interesting parallels. *Psychometrika*, 44, 157-167.

Steiger, J. H., & Lind, J. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City.

Thomson, G. H. (1956). *The factorial analysis of human ability*. Boston: Houghton-Mifflin.

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*, 421-459.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Tucker, L. R., & MacCallum, R. C. (1997). *Exploratory factor analysis*. Unpublished manuscript available at <http://quantrm2.psy.ohio-state.edu/maccallum/>.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231-251.

Vernon, P. E. (1961). *The structure of human abilities* (2nd ed.). London: Methuen.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters. *Multivariate Behavioral Research*, *28*, 263-311.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and over-extraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, *1*, 354-365.

Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, *17*, 253-269.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *17*, 253-269.