

Latent Variable Models

Fall 2003

Intro

A latent variable model can have two parts

- measurement model - Models the way that the observed variables are related to the latent variables.
- structural model - Models the way that the latent variable are related to one another.

Measurement model

- Measures the latent variables
- reduces the dimensionality of the data
- find patterns of correlations among several observed variables that are measuring the same thing
- observed variables are just a reflection of some underlying phenomena (i.e. latent variable)
- goal is to lose as little information as possible when reducing the dimensionality
- goal is to quantify how well each observed variables actually measure the latent variable

Examples of latent variables

- social economic status, quality of life, body satisfaction, depression, motivation for alcohol treatment, post traumatic stress disorder, peri menopause, parenting style, proneness to falls, family meal environment, couples relationship skills

Like observed variables, latent variables can be (hypothesized to be) continuous or categorical and if they are categorical they can be ordinal (ordered) or nominal (unordered). Depending upon what is assumed about the distribution of the latent variable and upon what kind of observed variables are used to measure them (i.e. continuous or categorical), the method for estimating the measurement model will change. Table 1.1 from BK.

Structural model

- allows researcher to translate idea about how causes are related to effects into a model
- total effect of one variable on another can be broken down into direct and indirect effects
- mediation and moderation
- Can be used with or without latent variables, that is variables of interest can be observed directly (i.e. no need for a measurement model). When there are no latent variables, it is often called path analysis.
- The general ideas of structural models will be introduced using models without latent variables, then latent variables will be added in later.

Linear Factor Analysis

1. The Model

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon} \quad (1)$$

\mathbf{x} : p-dimensional vector of continuous observed variables

\mathbf{f} : q-dimensional vector of underlying latent factors. Often called "common factors". Assume \mathbf{f} is random such that $E(\mathbf{f}) = \mathbf{0}$ and $Var(\mathbf{f}) = \boldsymbol{\Phi}$

$\boldsymbol{\epsilon}$: p-dimensional vector of random error. Often called "unique factors" or "specific factors". Assume $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var\boldsymbol{\epsilon} = \boldsymbol{\Psi}$. Element along the diagonal f $\boldsymbol{\Psi}$ often called "uniquenesses" or "specific variances"

Λ : $p \times q$ matrix of scalars called "factor loadings". This matrix describes how the observed variables \mathbf{x} are related to the latent factors \mathbf{f} .

μ : $p \times 1$ vector of scalars. Often ignored, most software assumes by default that $\mu = \mathbf{0}$ and analyze centered \mathbf{x} variables, i.e. analyze $\mathbf{x} - \bar{\mathbf{x}}$

An assumption is made for model (1) that $Cov(\mathbf{f}, \epsilon) = \mathbf{0}$. This is a very critical assumption because what it means is that the variability in \mathbf{x} can be separated into two additive parts, one coming from the common factors and one coming from the errors. That is there is no covariance between the two. Specifically,

$$Var(\mathbf{x}) = \Lambda Var(\mathbf{f}) \Lambda' + Var(\epsilon) + 2Cov(\mathbf{f}, \epsilon) \quad (2)$$

$$Var(\mathbf{x}) = \Lambda \Phi \Lambda' + \Psi. \quad (3)$$

It is assumed that equation (1) holds for each individual in the population and thus for $i = 1 \dots n$ independently sampled individuals we have

$$\mathbf{x}_i = \mu + \Lambda \mathbf{f}_i + \epsilon_i. \quad (4)$$

As an example of how to write model (1), here are the equations when $p = 5$ and $q = 2$ and it is assumed that Ψ is a diagonal matrix, for $i = 1 \dots n$ we have:

$$\begin{aligned} x_{1i} &= \mu_1 + \lambda_{11}f_{1i} + \lambda_{12}f_{2i} + \epsilon_{1i} \\ x_{2i} &= \mu_2 + \lambda_{21}f_{1i} + \lambda_{22}f_{2i} + \epsilon_{2i} \\ x_{3i} &= \mu_3 + \lambda_{31}f_{1i} + \lambda_{32}f_{2i} + \epsilon_{3i} \\ x_{4i} &= \mu_4 + \lambda_{41}f_{1i} + \lambda_{42}f_{2i} + \epsilon_{4i} \\ x_{5i} &= \mu_5 + \lambda_{51}f_{1i} + \lambda_{52}f_{2i} + \epsilon_{5i} \end{aligned}$$

or in matrix/vector notation

$$\begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ x_{5i} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \end{pmatrix} \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \end{pmatrix} \quad (5)$$

and we also must specify the variance structure of \mathbf{f} and ϵ

$$Var(\mathbf{f}_i) = Var \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} = \Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix} \quad (6)$$

$$Var(\epsilon_i) = Var \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \end{pmatrix} = \Psi = \begin{pmatrix} \psi_1 & 0 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 & 0 \\ 0 & 0 & 0 & \psi_4 & 0 \\ 0 & 0 & 0 & 0 & \psi_5 \end{pmatrix} \quad (7)$$

2. Exploratory factor analysis vs. Confirmatory factor analysis (EFA vs. CFA)

(a) EFA general purposes:

- To determine how many underlying factors are necessary to explain most of the correlations and variance in the data.
- To determine the relationship via **rotation** between each of these underlying factors with each of the observed variables in a meaningful way so that the factors can be interpreted and named.
- To weed out observed variables that do not tend to measure well the underlying factors shared by the other variables.
- To propose blocks of variables that may be subsequently be used to create a simple sum scale.
- To propose a CFA model

In EFA every element in $\mathbf{\Lambda}$ is estimated and it is assumed that $\mathbf{\Psi}$ is diagonal. Also, it is common to assume that $Var(\mathbf{f}) = \mathbf{\Phi} = \mathbf{I}$, i.e. the factors are uncorrelated with variance 1 (but this is not a necessary assumption, it is dropped when examining oblique rotations).

(b) CFA general purposes:

- To define a measurement model for the relationship between multivariate observations and underlying factors
- To test the statistical significance of factor loadings and correlations. Note this testing cannot currently be done in the EFA model. Thus one may be interested in testing whether rotated factor loadings from an EFA that look "close to zero" are, in fact, significantly different from zero or not.
- To test whether the measurement model for one group is the same as the measurement model for some other group
- As a precursor to a Structural equation model

In CFA usually several elements in $\mathbf{\Lambda}$ are fixed to zero and it is possible to consider correlated ϵ which means that $\mathbf{\Psi}$ is not necessarily diagonal. Furthermore, it is usually assumed that the factors are correlated so that no restriction is placed on $\mathbf{\Phi}$.

(c) PUT PLOTS of EFA and CFA two factor model FROM AMOS here

What are the differences in the two models?

In the EFA model, there are arrows going from every factor to every observed variable. Also there are no double headed curved arrows going from factor to factor.

In the CFA model, there are clearly a lot less arrows going from factors to observed variable and in particular each observed variable has exactly 1 arrow hitting it. Also there are curved double headed arrows going from each factor to each other factor.

(d) Discuss paper: Hurley, A., Scandura, T, Sriesheim, C. Brannick, M., Seers, A, Vandenberg, R., and Williams, L. (1997) "Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives" *Journal of organizational behavior*, 18, 667-683.

3. **What does it mean for p observed variable to have q underlying latent factors?**

PUT GEOMETRIC INTERPRETATION HERE FROM OLD NOTES

give a step-by-step about how the data was generated for XGobi and include plot

PUT EIGENVALUE EIGENVECTOR DISCUSSION HERE FROM OLD NOTES

4. **Modeling the covariance matrix - the natural choice for the linear factor analysis model**

- Consider p continuous random variables \mathbf{x} from some population, let $\mathbf{\Sigma}$ represent the true covariance matrix of \mathbf{x} in the population. That is, $Var(\mathbf{x}) = \mathbf{\Sigma}$.
- If \mathbf{x} were Normally distributed then everything about it is described simply by its mean and its covariance matrix. Thus if \mathbf{x} can be assumed Normally distributed, the natural thing to model is its mean and variance(covariance) matrix.
- Our goal is to find a parametric model $\mathbf{\Sigma}(\boldsymbol{\theta})$ that describes $\mathbf{\Sigma}$ as closely as possible.

- $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon}$ is a model such that

$$Var(\mathbf{x}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) \quad (8)$$

Here $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ are the parameters

- $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is called the "model covariance matrix"
- Given n independent samples of \mathbf{x} and given that \mathbf{x} is Normally distributed then the sufficient statistic for $\boldsymbol{\Sigma}$ is the sample covariance matrix \mathbf{S} . The $p \times p$ sample covariance matrix is

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

- Given a model (i.e. $\boldsymbol{\Sigma}(\boldsymbol{\theta})$) and the data (i.e. \mathbf{S}), the goal is to estimate $\boldsymbol{\theta}$ so that $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ is as close to \mathbf{S} as possible while still being parsimonious.
- Whether model (1) is an EFA or a CFA we will fit the model covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ to the sample covariance matrix

- OR Standardize \mathbf{x} to get \mathbf{Z} , i.e. $\mathbf{Z} = \begin{pmatrix} \frac{x_1 - \bar{x}_1}{s_1} \\ \frac{x_2 - \bar{x}_2}{s_2} \\ \vdots \\ \frac{x_p - \bar{x}_p}{s_p} \end{pmatrix}$

$$Var(\mathbf{Z}) = \boldsymbol{\rho} = \boldsymbol{\Lambda}^s \boldsymbol{\Lambda}^{s'} + \boldsymbol{\Psi}^s$$

- $\boldsymbol{\Lambda}^s$ are the "standardized factor loadings"
- We will estimate $\boldsymbol{\rho}$ using \mathbf{R} , the sample correlation matrix
- For some estimation methods, e.g. maximum likelihood method, $\boldsymbol{\Lambda}^s$ obtained by analyzing the correlation matrix is the same as rescaling the $\boldsymbol{\Lambda}$ obtained by analyzing the covariance matrix by the observed standard deviations, i.e. $\boldsymbol{\Lambda}^s = (diag(\mathbf{S}))^{-\frac{1}{2}} \boldsymbol{\Lambda}$. (See section 3.17 in BK) Note THIS IS NOT TRUE in general.

5. Estimating $\boldsymbol{\theta}$

There are many ways to estimate $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$ in the model covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. All use some sort of discrepancy function

$$F = F(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

A discrepancy function is just a rule that determines how you will find the "best" estimate for the parameters. As an analogy, in linear regression you have data y_i , $i = 1 \dots n$ and you have some model for for each observed y_i , for example $\beta_0 + \beta_1 x_i$. The goal there is to find the values for parameters β_0 and β_1 that make the model as close to the data as possible. The discrepancy function is what we use to define "close". A discrepancy function which could be used for this regression is $F(data, model) = \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_i)^2$ which is the ordinary least squares discrepancy function, but there are other discrepancy functions that could be used for example $F(data, model) = \sum_{i=1}^n w_i (y_i - \beta_0 + \beta_1 x_i)^2$ where w_i is some predetermined weight.

The common discrepancy functions used for the factor analysis model are

- Normal theory maximum likelihood
- Generalized least squares
- Weighted least squares
- Asymptotically distribution free (ADF)
- when $\boldsymbol{\Phi} = \mathbf{I}$ and $\boldsymbol{\Psi}$ is diagonal the principal factor method can be used which is very quick because it uses eigenvalue eigenvectors

We will call the estimates: $\hat{\theta} = (\hat{\Lambda}, \hat{\Psi}, \hat{\Phi})$, and $\hat{\Sigma} = \Sigma(\hat{\theta})$ will be referred to as the “fitted model covariance matrix” or “Implied covariance” (in AMOS) or “estimated covariance matrix” (in MPLUS).

6. Model identification

In general it is possible that there exists $\theta_1 \neq \theta_2$ such that $\Sigma(\theta_1) = \Sigma(\theta_2)$. This is a problem because it means that there is not just one unique solution to the discrepancy function. A name for this is “under-identification”. In EFA this is directly related to the issue of rotation. In CFA it is usually not a matter of rotation but of some parameters not being separable or simply trying to estimate more parameters than you have pieces of information to do so.

(a) General Principles of identification (pp. 5.5 in Kline is a nice general introduction)

- Model under-identification or “non-identified” models - Theoretically no matter how much data you have, there will not be a unique solution for the parameters in a non-identified model. This is always true when the number of parameters is greater than the number of unique elements in \mathbf{S} , i.e. $p(p+1)/2$. But a model can also be non-identified even if the degrees of freedom are > 0 , in this case we often rely on software to tell us the model is non-identified which it does by showing an error somewhere.
- Empirical under-identification - This occurs when, for a particular data set, it is not possible to uniquely estimate the parameters due to the some anomaly in the data, e.g. a correlation between two variables is very high and thus it is not possible to distinguish them from each other, or in a model where it is necessary for factors be correlated (in order that the model be identified), if the empirical correlation between them is very small, the model may have problems with empirical under-identification.
- Identified or “over-identified” - This means that each parameter is uniquely estimable. In other words it is possible to estimate all the parameters in the model. A special kind of identified model is the “just identified model” or “saturated model”, here the number of parameters equals the number of observations, i.e. the degrees of freedom equal zero. The model fits the data perfectly. Note there can be more than one saturated model.

7. Rotation in EFA - this is an issue related to model identification

No matter what estimation procedure, for exploratory factor analysis we get estimates that look like:

$$\hat{\Sigma} = \hat{\Lambda} \hat{\Lambda}' + \hat{\Psi}$$

We can get the exact same $\hat{\Sigma}$ by taking

$$\hat{\Sigma} = \hat{\Lambda} T T' \hat{\Lambda}' + \hat{\Psi}$$

where T is a $q \times q$ matrix such that $T T' = \text{Identity matrix}$.

Thus

$$\hat{\Sigma} = \hat{\Lambda}^* \hat{\Lambda}^{*'} + \hat{\Psi}$$

where $\hat{\Lambda}^*$ is the **orthogonally rotated** factor loading matrix, i.e. $\hat{\Lambda}^* = \hat{\Lambda} T$. BUT, There are an infinite number of matrices T that satisfy $T T' = I$.

- Why do we bother to rotate? – We want to find out which variables “stick together”
- How to choose a T ? – Find one that gives close to “simple” structure to the factor loadings
- “Simple structure” means that each observed variable only loads on one factor. In an AMOS notation, each observed variable only has one arrow going to it from a factor. Simple structure allows you to say, for example, “observed variable 1 is a direct measure of only factor 1”, instead of “variable 1 is direct measure of factor 1, 2, and 3”

- What about if simple structure is not found by orthogonal rotation?
- NOTE: implicitly orthogonal rotation keeps the assumption that $Var(\mathbf{f}) = I$, i.e. the factors are uncorrelated.
- Maybe the assumption that $Var(\mathbf{f}) = I$ should be dropped. This leads to **oblique rotations** of the factor loadings.
- We can search for

$$\hat{\Sigma} = \hat{\Lambda}^{**} \Phi \hat{\Lambda}^{**'} + \hat{\Psi}$$

so that $\hat{\Lambda}^{**}$ has “simple structure” and $\Phi \neq I$

8. Rotation continued: Determining which variables measure which factors and how to look for simple structure

- Examine the rotated factor loadings
- Rule of thumb, if the absolute value of the standardized loading is $> .3$, the variable is relevant for the particular factor.
- Do the variables that load on a given factor share some conceptual meaning?
- Examine the communality. Rule of thumb, if the communality is less than $.10$ then the observed variable can be deleted. A communality $< .10$ means that less than 10% of the variability in the observed variable is explained by all of the common factors.
 - Communalities are the diagonal elements of $\Lambda\Lambda'$ or $\Lambda^s\Lambda^{s'}$ when the factors are assumed to be uncorrelated
 - Communalities are the part of the variance of each observed variable which is due to the q underlying factors

$$\begin{aligned} x_1 &= \mu_1 + \lambda_{11}f_1 + \lambda_{12}f_2 + \epsilon_1 \\ x_2 &= \mu_2 + \lambda_{21}f_1 + \lambda_{22}f_2 + \epsilon_2 \\ x_3 &= \mu_3 + \lambda_{31}f_1 + \lambda_{32}f_2 + \epsilon_3 \\ x_4 &= \mu_4 + \lambda_{41}f_1 + \lambda_{42}f_2 + \epsilon_4 \\ x_5 &= \mu_5 + \lambda_{51}f_1 + \lambda_{52}f_2 + \epsilon_5 \end{aligned}$$

when we consider orthogonal rotations, i.e. f_1 and f_2 are uncorrelated then

$$\begin{aligned} Var(x_1) &= \lambda_{11}^2 + \lambda_{12}^2 + \psi_1 \\ Var(x_2) &= \lambda_{21}^2 + \lambda_{22}^2 + \psi_2 \\ Var(x_3) &= \lambda_{31}^2 + \lambda_{32}^2 + \psi_3 \\ Var(x_4) &= \lambda_{41}^2 + \lambda_{42}^2 + \psi_4 \\ Var(x_5) &= \lambda_{51}^2 + \lambda_{52}^2 + \psi_5 \end{aligned}$$

EXAMPLE: The communality for x_2 is $\lambda_{21}^2 + \lambda_{22}^2$. On the other hand, ψ_2 is the part of the variability in x_2 that is unique to x_2 , i.e. is not shared with other observed variables, “is not common”.

- Cudeck, R. and O’Dell, L. (1994) “Applications of standard error estimates in unrestricted factor analysis: Significance Tests for factor loadings and correlations” *Psychological Bulletin* This paper argues that the $> .3$ cutoff is too simplistic and provides a method for significance testing of rotated loadings.
- BK describe some of the work that has been done in order to come up with a way to test the significance in EFA. While the original work for standard errors in EFA were worked out for the rotation that fixes $\Lambda'\Psi\Lambda$ to be diagonal, there are now methods for calculating standard errors even for obliquely rotated loadings, in particular Ogasawara (1998) “Standard errors for rotation matrices with an application to the promax solution”. There is a specialized program available from Michael Browne’s website (at Ohio State University) that can calculate these standard errors. All of these results are based on asymptotic results, thus it is still not clear how useful they are for problems with small sample sizes.

- CROSS VALIDATION. If the sample size is large enough, a simpler way to gain an idea about the stability of the factor structure (rather than relying on rules of thumb) is to split the sample randomly into two equal parts and then fit the model to each part. If the factor structure is similar, this "tends to increase our confidence in the genuineness of the factors".

9. **Identifiability in the CFA model** Kline Section 7.5, 203-207 Table 7.1.

- Necessary conditions:
 - 1 The d.f. must be greater than zero. That is $p(p + 1)/2$ - the number of parameters being estimated must be greater than zero.
 - 2 Every factor must have a scale. This can be accomplished in one of two ways. Either the variance of the factor is fixed to be 1, or the loading of one unique observed variable for each factor is fixed to 1.
- Both methods of fixing the scale produce the exact same fit (Chi-square) value.
- If you fix a factor loading equal to 1 then the results are "unstandardized", the scale (and thus variance) of the factor is similar to that of the variable it was identified with, i.e. a one unit increase in the factor results in a one unit increase in the observed variable.
- In AMOS and Mplus no matter which observed variable has loading fixed to 1, the "standardized estimates" created from the unstandardized estimates will be the same. Note, though that the unstandardized estimates will change depending on which loading is fixed to 1.
- The standardized estimates are also obtained when the factors have had their variance fixed to 1, thus a 1 standard deviation increase in the factor results in a λ_1^s standard deviation increase in the observed variable.
- Here is an example of the model covariance for the 3 observed variables, one factor model when the first loading is fixed to 1:

$$x_1 = f + \epsilon_1 \tag{9}$$

$$x_2 = \lambda_2 f + \epsilon_2 \tag{10}$$

$$x_3 = \lambda_3 f + \epsilon_3 \tag{11}$$

$$\Sigma = Var(\mathbf{x}) = \begin{pmatrix} \phi + \psi_1 & \lambda_2 \phi & \lambda_3 \phi \\ \lambda_2 \phi & \lambda_2^2 \phi + \psi_2 & \lambda_2 \lambda_3 \phi \\ \lambda_3 \phi & \lambda_3 \lambda_2 \phi & \lambda_3^2 \phi + \psi_3 \end{pmatrix} \text{ fit this to the data } \mathbf{S} = \begin{pmatrix} s_{11}^2 & s_{12} & s_{13} \\ s_{12} & s_{22}^2 & s_{23} \\ s_{13} & s_{23} & s_{33}^2 \end{pmatrix} \tag{12}$$

- Commonly the direction given is to use unstandardized estimates when making comparisons across groups (so that way the factor variance can be allowed to vary) but otherwise interpreting standardized estimates is usually simpler.
- See Table 7.1 in Kline for some sufficient conditions for identifiability in CFA models

10. **VISPERC EXAMPLE**

- SHOW SAS OUTPUT FROM VISPERC EXAMPLE eigenvalues, screeplot Section of BK p. 58 3.14-3.15. Mention that chi-square will be discussed later Examine closely EFA results of 2 factor model first factor pattern orthogonal promax
- SHOW MPLUS OUTPUT FROM VISPERC EXAMPLE Point out what $Sig(\hat{\theta})$ is Can use $Sig(\hat{\theta})$ from Mplus as reference for results from Splus
- SHOW ROTATION PLOTS
- SHOW AMOS RESULTS FROM FITTING THE CFA MODEL IN LAB

11. **Normal distribution and the likelihood function**

Recall that maximum likelihood estimation asks us to find the μ and Σ which maximize the likelihood L (or the log likelihood).

Given n i.i.d. random vectors $\mathbf{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the log likelihood is

$$\log L = \frac{pn}{2} \log 2\pi - \frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

After a bit of matrix algebra, this can be transformed into

$$\log L = \text{constant} + \frac{n}{2} [\log |\boldsymbol{\Sigma}^{-1}| - \text{trace}(\mathbf{S}\boldsymbol{\Sigma}^{-1})] - \frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})$$

where \mathbf{S} is the sample covariance matrix. Since $\boldsymbol{\mu}$ only appears in the last term, we can maximize the likelihood with respect to $\boldsymbol{\mu}$ by minimizing the last term with respect to $\boldsymbol{\mu}$. This is clearly done when $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$

So what we are left with doing is maximizing the remaining part with respect to the parameters in $\boldsymbol{\Sigma}$, i.e. with respect to $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$. So we want to maximize the following:

$$\log L = \frac{n}{2} [\log |\boldsymbol{\Sigma}^{-1}| - \text{trace}(\mathbf{S}\boldsymbol{\Sigma}^{-1})] \quad (13)$$

This is equation (3.11) in BK. The next step is to actually perform the maximization.

12. Maximizing the Likelihood

Most of section 2.5 in BK is describing how to maximize the likelihood (13). It is not crucial that you follow the rest of this section, but I have provided a brief sketch below:

- The maximization is done by iteration.
- Needs starting values
- The iteration starts by taking the eigenvalues of the following: $(\hat{\boldsymbol{\Psi}}_{(0)}^s)^{-\frac{1}{2}} \mathbf{R} (\hat{\boldsymbol{\Psi}}_{(0)}^s)^{-\frac{1}{2}} - \mathbf{I}_{(p \times p)}$ This is implicit in the discussion of p. 47 of BK. It is the eigenvalues of this that appear on the first page of the PROC FACTOR output.
- I don't think I can provide any simple intuition about this iterative technique. Just trust that it converges to the Maximum Likelihood Estimators.

We will call the maximum likelihood estimators: $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Phi}})$, and $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ will be referred to as the “fitted model covariance matrix” or “Implied covariance” (in AMOS) or “estimated covariance matrix” (in MPLUS).

13. Chi-square Goodness of Fit

Besides giving estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, Maximum Likelihood Provides a Goodness of Fit test.

- Use Likelihood Ratio Test, i.e.,

$$\begin{aligned} -2 \log \frac{L(\text{MLE of restricted model})}{L(\text{MLE of unrestricted model})} \\ &= 2 \{ \log(L(\mathbf{S})) - \log(L(\hat{\boldsymbol{\Sigma}})) \} \\ &= n \{ \text{trace} \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{S} - \log |\hat{\boldsymbol{\Sigma}}^{-1} \mathbf{S}| - p \} \end{aligned}$$

- IF the model fits the data well, this statistic should be small!
- Its distribution is asymptotically distributed χ^2 with degrees of freedom = $(\frac{p(p+1)}{2})$ - number of unique parameters in model).
- We can determine if the the statistic is “small” enough by comparing to the χ^2 distribution and obtaining a p-value.
- In general, the Hypothesis being tested is

$H_0: \Sigma = \Sigma(\theta)$ (your model)

$H_A: \Sigma = \mathbf{S}$ the saturated model (or just-identified model)

- Since the degrees of freedom for the saturated model are 0, this means it fits the data perfectly
- So we are comparing $\Sigma(\theta)$ to a model that fits the data perfectly. Thus if it is not significantly different than the model that fits perfectly, it means it is pretty good
- Thus we are looking for the model where we DO NOT REJECT the H_0 (i.e. find a big p-value)
- Note, the chi-square test has been proven to be **asymptotically** valid even when the data is not normally distributed (Amemiya and Anderson, 1985). Note: this is an asymptotic result.
- FROM KLINE, page 209-210 Satorra Bentler correction to the Chi-square statistic when the data is non-normal. This correction is available in Mplus but not in AMOS.

14. Choosing q in the EFA model

- The chi-square test can also be used to see how well an EFA model with q factors fits the data
- (See BK section 3.16) For the EFA model the d.f. are specifically $\frac{p(p+1)}{2} - pq - p + \frac{q(q-1)}{2}$. The first term is the number of unique elements in the sample covariance matrix \mathbf{S} , then we are going to estimate pq factor loadings in $\mathbf{\Lambda}$ and p variances in $\mathbf{\Psi}$ and since we have to put restrictions in order to fix rotation we get those $\frac{q(q-1)}{2}$ d.f. back. These $\frac{q(q-1)}{2}$ d.f. correspond to fixing the $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ off diagonal elements to be zero.
- In EFA, the model being tested is

H_0 : The q factor model is correct (i.e. q factors sufficiently describe the p dimensional vector)

H_A : More factors are needed

- BK state the following: “Starting with $q = 1$, we then take successive values in turn until the fit of the model is judged to be adequate. **Viewed as a testing procedure this is not strictly valid because it does not adjust the significance levels to allow for the sequential character of the test. It rather depends on regarding the p-value of the test as a measure of the adequacy of the model**”
- Because of the tendency for the approach described just above to keep adding more and more factors (since more factors will make the model fit better), BK suggest considering the AIC Akaike’s information criterion. This is simply a penalized version of the log likelihood where models with more parameters are penalized.
- Section 3.14 in BK describes a method of choosing q (i.e. the number of underlying factors) by taking it to be the number of eigenvalues ≥ 1 . They report a paper by Fachel 1986 (actually this is Fachel’s dissertation) that showed via simulation that this method tends to overestimate the number of factors.
- Section 3.15 in BK describes the use of the “scree test”. That is, plotting the ordered eigenvalues and then looking for the elbow in the plot. q is taken to be the numbered eigenvalue where the elbow appears.

15. **Equivalent models** Two different models $\Sigma_1(\theta)$ and $\Sigma_2(\theta)$ are said to be equivalent if each model is identified and for every θ_1 there exists some θ_2 such that $\Sigma_1(\theta_1) = \Sigma_2(\theta_2)$. In other words there is no way to distinguish the two models by looking at how close $\hat{\Sigma}$ is to \mathbf{S} because $\hat{\Sigma}$ is identical in both models. Kline pp.220-224 talks about this for CFA models.

- SES and ruralness example from Homework 3, following page 221 in Kline
- give 3 variable model with one underlying factor and show equal to the simple correlation model
- This topic will show up again in path analysis and in general for any SEM.

16. Method effect and correlated errors

- Homework 1, method effect in self-esteem scale. Most likely caused by some questions being asked in the positive direction and others in the negative.

- If it is found that the factor analysis model fits much better if the error terms are allowed to correlate, this may be an indication of a lurking variable. If the errors need to be correlated this means that there is more correlation between the observed variable than can be explained by the common factors shared in the model, i.e. there is some other factor they share as well that your model is missing.
- If this additional correlation can be explained by the researcher, it is possible to then hypothesize an additional factor which the respective observed variables can load on.

17. **Fitting without Normality Assumptions** See BK 3.10-3.11 and Kline p. 209.

The ML method assumes that \mathbf{x} is distributed multivariate normal.

ADD SOME NOTES HERE ABOUT RULES OF THUMB FOR SKEW AND KURTOSIS

A general class of discrepancy functions is given by

$$F = \text{trace}\{(\mathbf{S} - \boldsymbol{\Sigma})\mathbf{V}\}^2 \quad (14)$$

where \mathbf{V} represents different ways of weighting the difference between the observed covariance matrix and the model covariance matrix.

Mplus and AMOS will allow various choices for \mathbf{V} . Unweighted least squares, weighted least squares and asymptotically distribution free.

DiStefano, Christine (2002) "The impact of categorization with confirmatory factor analysis" *Structural Equation Modeling*, 9(3), 327-346. For Homework 2. This paper is about analysis of non-normally distributed data using GLS versus ML.

MPLUS can also handle ordered categorical observed variable by creating polychoric, tetrachoric, and polyserial correlations. Discuss later.

18. **Multiple group CFA models and Nested models**

- ADD SOME NOTES HERE
- LAB EXAMPLE comparing boys and girls self esteem CFA models
- Multiple group comparison in Greenberger, E., Chen, C., Dmitrieva, J., Farruggia, S. (2003) "Item-wording and the dimensionality of the Rosenberg self-esteem scale: Do they matter?" *Personality and Individual Differences*, 35, 124-1254. Homework 3.
 - Comparison of the model across racial groups.
 - Comparison of the fit of a one factor model to the two factor model with simple structure.
- Modification indices, a nested model with just one restriction

19. **Creating scales based on a factor analysis model**

SHOW EXAMPLE with simulated data and two factors, show how factor score estimates are obtained and then also how simple sum scores are obtained.

- **Factor Score Estimates**

$E(\mathbf{f}|\mathbf{x})$,

optimal predictor of \mathbf{f}

Comparison of factor score estimates as predictors to a CFA model or a multiple regression. Wall, M.M. and Li, R. (2003) "Comparison of multiple regression to two latent variable techniques for estimation and prediction", in press at *Statistics in Medicine*. A .pdf preprint of this article is available at the class web site www.biostat.umn.edu/~melanie/Pubh5482.html. Paper read for Homework 4. ADD MORE NOTES HERE

- **Cronbach's alpha** ADD NOTES HERE

20. **Factor analysis for ordinal categorical observed variables**

- MPLUS and LISREL can handle this
- uses the "underlying variable method" rather than the direct maximum likelihood method
- Discussed in 4.10 of BK.
- ADD SOME NOTES HERE
- We will see more later