# Chapter 4 Continued: More about Factor Analysis

- One mathematical issue of interest is that there is no unique solution to the factor analysis problem.
- Mathematically speaking, an infinite number of matrices of loadings are equally good.
- Note the regression equation (in matrix form) linking the manifest variables and the latent variables:

$$\mathbf{x} = \mathbf{\Lambda f} + \mathbf{u}$$

- Consider any $k \times k$ orthogonal matrix $\mathbf{M}$.
- Then since $\mathbf{MM}' = \mathbf{I}$, we can rewrite the regression equation as:

$$\mathbf{x} = \mathbf{\Lambda MM}'\mathbf{f} + \mathbf{u} \Rightarrow \mathbf{x} = \mathbf{\Lambda}^*\mathbf{f}^* + \mathbf{u}$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda M}$ and $\mathbf{f}^* = \mathbf{M}'\mathbf{f}$.

## Nonuniqueness of the Factor Analysis Solution

- Note: Under this altered model formulation, the factors are $\mathbf{f}^* = \mathbf{M}'\mathbf{f}$ and the loadings are in the matrix $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{M}$.
- This implies that (since $\mathbf{M}\mathbf{M}' = \mathbf{I}$)

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

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{M}(\mathbf{\Lambda}\mathbf{M})' + \mathbf{\Psi}$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{M}\mathbf{M}'\mathbf{\Lambda}' + \mathbf{\Psi}$$

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

as with the original model formulation.

- Hence any of an infinite number of orthogonal matrices $\mathbf{M}$ yields another solution to the factor analysis problem.

## Ways of Solving the Nonuniqueness Problem

- Usual solution: Force $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ to be a diagonal matrix with its diagonal elements in descending order.
- This constraint resembles the PCA constraint that orders the importance of the principal components.
- The first factor will contribute the most to the shared variance of the manifest variables.
- The second factor is the factor that contributed the second-most to this shared variance, among all possible factors that are *uncorrelated* with the first factor, and so on.
- However, such a choice of constraint can lead to factors with poor interpretation.
- The resulting factors may have substantial loadings on many variables.
- Some of the resulting factors may be *bipolar*, having both positive and negative loadings, which can make interpretation difficult.

## What is a More Interpretable Solution?

- We can find an alternative factor solution simply by choosing some orthogonal matrix $\mathbf{M}$ and using the loadings $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{M}$.
- What kind of solution is easily "interpretable"?
- We would like each factor to have high loadings on only a few variables.
- We would like other loadings to be near zero.
- We would like each variable to have a high loading for only one factor.
- *Idea*: Each factor is substantially based on only a couple of variables.
- And variables don't contribute to multiple factors.

## What is a More Interpretable Solution?

- Thurstone (1931) defined *simple structure* for a $k$-factor solution:

1. Each row of $\mathbf{\Lambda}$ should contain at least one zero.
2. Each column of $\mathbf{\Lambda}$ should contain at least $k$ zeroes.
3. When $k \geq 4$, each pair of columns should contain many variables with zero loadings in each column.
4. For each pair of columns, very few variables should have nonzero loadings for both columns.

- Simple structure is difficult to obtain in practice, but represents an “ideal” factor analysis solution.

An example of a loadings matrix with (highly idealized) simple structure:

[$x$ = moderate to large loading; $o$ = small (near zero) loading]

$$\mathbf{\Lambda} = \begin{bmatrix} x & o & o & o \\ o & x & o & o \\ o & x & o & o \\ o & o & o & x \\ o & o & o & x \\ o & o & x & o \\ o & o & x & o \\ o & o & o & x \end{bmatrix}$$

## Methods of Factor Rotation

- From a geometrical standpoint, choosing a matrix $\mathbf{M}$ to alter the factor analysis solution amounts to *rotating* the variables' axes so that the factor loadings (when plotted) fall closer to those axes.
- There are two classes of rotation: *orthogonal* and *oblique*.
- Orthogonal rotations preserve the property that the factors are uncorrelated with each other, making interpretation easier.
- The factor loadings still represent correlations between the factors and the manifest variables.
- Oblique rotations allow for factors to be correlated with each other.

## Types of Orthogonal Factor Rotation

- *Varimax* rotation tries to create factors with a few large loadings and many near-zero loadings.
- Essentially, the varimax criterion yields factors with loadings whose squares are as "spread out" as possible.
- If there is a single dominant "general factor" (with high loadings on each variable), then it will tend to be obscured by the varimax rotation.
- There are ways to hold one (dominant?) factor fixed and rotate the other factors.
- Another rotation method, *quartimax*, seeks to make each variable correlate strongly with only one factor.

## Is Factor Rotation Valid?

- Some have argued that factor rotation is too subjective, allowing the investigator to produce whatever conclusion he/she wants.
- The essential solution does not change, however. The reason for rotation is simply to provide a more understandable interpretation of the solution.
- Rotation is particularly useful for maximum likelihood factor analysis.
- The initial constraint that $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ be diagonal is useful computationally, but can produce a solution that is not very interpretable.
- How to tell if factor analysis is successful? Note Johnson and Wichern's WOW criterion.

# Factor Scores

- We often wish to find (and perhaps plot) *factor scores* for each individual in the data set.
- This is conceptually more difficult than getting the principal component scores in PCA.
- In PCA, the components are defined in terms of the observed variables, but in factor analysis, the variables are defined in terms of the factors.
- If we assume multivariate normality, the conditional distribution of $\mathbf{f}$ given a data vector $\mathbf{x}$ having mean $\mathbf{0}$ is multivariate normal with mean $\mathbf{\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{x}$.
- We could predict (not estimate) factor scores by plugging in sample values, so that $\hat{\mathbf{f}} = \hat{\mathbf{\Lambda}}'\mathbf{S}^{-1}\mathbf{x}$, where $\mathbf{x}$ has been centered by subtracting off the mean vector.
- This is known as the "regression method" of obtaining factor scores. An alternative method is the "weighted least squares method."
- If there are two or three factors in our solution, we can plot the factor scores for our data set using a 2-D or 3-D scatterplot.

## Checking Model Fit

- Some diagnostic methods are available to check the fit of a factor analysis model.
- The *reproduced correlation matrix* contains correlations based on the factors we have chosen to extract.
- We can compare this to the original correlation matrix for the entire data set.
- If they are close, this is evidence of a good model fit.
- The *residual correlations* are the differences between the original correlations and the reproduced correlations — we would like most of these residuals to be near zero.

## Graphs to Diagnose Model Fit

- The R function `fact` on the course web page (code written by Dr. Brian Habing) gives fitted factor analysis output and several plots to check model fit.
- A scree plot is given to help check the correct number of factors.
- In the plot of reproduced (or predicted) correlations vs. actual correlations: If most points fall near the line, this indicates a good fit.
- A plot of residual vs. predicted correlations with no noticeable pattern indicates a good fit.
- In the histogram of the residuals: A tight distribution around zero, with no heavy tails, indicates a good fit.
- Any notable pattern in these plots indicates we should reconsider our factor analysis solution — it may not be sufficient.

# PCA vs. Factor Analysis

- Both PCA and factor analysis are exploratory dimension reduction techniques.
- Factor analysis attempts to explain correlations between observed variables, while PCA attempts to explain variances of variables.
- In PCA, changing the number of components from $m$ to $m+1$ doesn't affect the first $m$ components. In factor analysis, changing the number of factors may completely change the solution.
- Using the covariance matrix vs. the correlation matrix in (ML) factor analysis makes no difference, but in PCA using $\mathbf{S}$ compared to $\mathbf{R}$ produces different results.
- Factor analysis accounts for the specific variances, so if the specific variances are small, PCA and factor analysis will lead to similar conclusions — but if they are large, this is not so.
- Both methods will be unnecessary and not helpful if the observed variables are nearly uncorrelated.

# Confirmatory Factor Analysis

- The methods we have covered can be described as *exploratory factor analysis*.
- No previous knowledge about the number of factors, loading structure, etc., were incorporated into the model.
- In *confirmatory factor analysis*, we may have strong previous beliefs about the nature of the factors.
- The number of factors may be specified ahead of time, and specific variables may be fixed to load on one or more specific factors.
- The analysis then determines how well the data support the prior theory.
- Actually fitting the confirmatory factor analysis model is more difficult and requires specialized software.