Factor Analysis: The Common Factor Model

Measurement, Scaling, and Dimensional Analysis
2019 ICPSR Summer Program
Prof. Adam M. Enders

Housekeeping

- Lab @ 5:30 in 3411 Mason Hall
- Picnic this Saturday in Burns Park (same location as last time) from 1–5 PM
- Publishing Blalock lecture on Monday night featuring Jim Johnson (Rational Choice Theory) and Sara Mitchell (Time Series)

What is Factor Analysis?

- Suggests hypothetical variables – factors – that are immediate, common causes of observed variables
- FA is a formal model of the hypothetical variables that account for the *linear relationships* between observed variables
  - We're modeling a correlation matrix
- The properties of these latent (or hypothetical) variables are:
  - They form linearly independent sets of variables (i.e., no hypothetical component can be derived from a linear combination of the other hypothetical components)
  - The hypothetical components are of two types:
    1. Common factors, which have more than one variable with a non-zero weight associated with that factor
    2. Unique factors, which have only one non-zero weight associated with the factor

What is Factor Analysis?, cont'd

- A note on correlation and causation in FA:
  - Can only help when we have correlated data that is not manifested due to a causal relationship between two or more variables
  - We can use EFA when we have correlated data that is correlated because of a common, latent factor (no theoretical causal relationship between the observed variables)
- There are only some situations in which correlated data may be amenable to the factor model
  - $X \rightarrow Y$, X is a direct cause of Y
  - $Y \rightarrow X$, Y is a direct cause of X
  - $\xi \rightarrow X$ and $\xi \rightarrow Y$, ξ is an immediate common cause of X and Y
  - $\xi \rightarrow X$ and $\xi \rightarrow \eta$ and $\eta \rightarrow Y$. ξ is an immediate cause of X and is a cause of Y, though its effect is mediated through another unmeasured variable η

What is Factor Analysis?, cont'd

- All four of these sets of relationships will produce non-zero correlations between X and Y
- Common FA can only really deal with the third one. It cannot model causal relationships among the unobserved variables
- Common FA does not inherently handle dynamic structures (i.e., the assumption is that all variables are measured at the same time)
- At least three indicators, or pieces of information, are needed to identify a factor (i.e., three variables have a common factor as their immediate common cause)

A Little History

- Spearman (1904) defined common factors as the element(s) in common to two or more indicators
- Thurstone (1931) developed the "center of gravity" method for estimating factor loadings that worked only on the off-diagonal elements of the correlation matrix
- Hotelling (1933) proposed principal components for generating orthogonal, conditionally variance-maximizing representations of manifest variables
  - Hotelling left the ones on the main diagonal of the correlation matrix and thus retained their influence in the solution

A Little History, cont'd

- Thurstone (1935) identified 3 problems with Hotelling's solution:
  1. Leaving ones on the diagonal indicated that all of a variable's variance should be explained by the dimension
  2. Hotelling retained (and interpreted) all of the components (i.e., no dimension reduction)
  3. Hotelling did not rotate the components into a more interpretable orientation
- Guttman and later Kaiser (both writing in the mid-20th Century) suggested PCA extraction (i.e., unities on the diagonal), retain all factors with eigenvalues greater than 1 and varimax (orthogonally) rotate the solution into as simple structure as possible while maintaining orthogonality
- After this, the debate changed focus to the factor indeterminacy problem (i.e., an infinite number of equally well-fitting solutions)

A Little History, cont'd

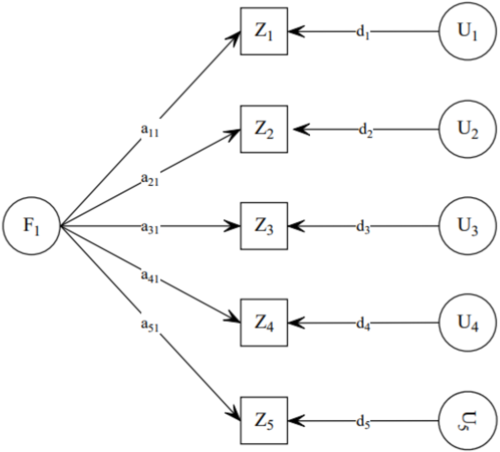
- The first substantive application of Factor Analysis was measuring ability
- Book called *The Mismeasure of Man* that details how the interpretation of latent factors as real physical quantities guided early research on intelligence
  - Basically, this psychologist named Cyril Burt "reinvented" factor analysis (stealing credit from Spearman) and used his solutions to justify ideological principles by interpreting the factors as real, physical quantities
- Lots of common misunderstandings and misperceptions:
  - Factor Analysis is a "Rumpelstiltskin" method whereby gold is spun from yarn (the indeterminacy problem)
  - Factors should not be rotated
  - Factors should only be rotated orthogonally

Different "Flavors" of Factor Analysis

- Here, we will focus on the "Common Factor Model," or what is more commonly referred to as "Exploratory Factor Analysis"
- Also "confirmatory factor analysis," which is born out of the structural equation modeling (SEM) tradition
- Why EFA?
    1. SEM is a whole other ballgame – different general perspective, different prerequisites RE estimation
    2. We have a course on SEM in the second session – covers CFA, structural equations with latent variables, latent growth curves, etc.
    3. EFA is a natural starting point to CFA, and more familiar to social scientists (at least in a general sense)
    4. Confirmatory/exploratory distinction is not really all that important

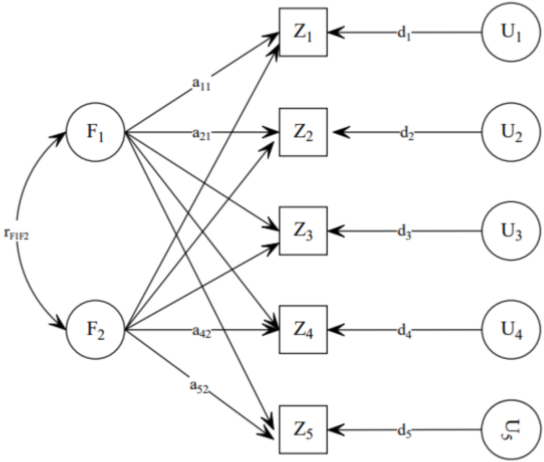
Factor Analysis via Path Diagrams

A single common factor model



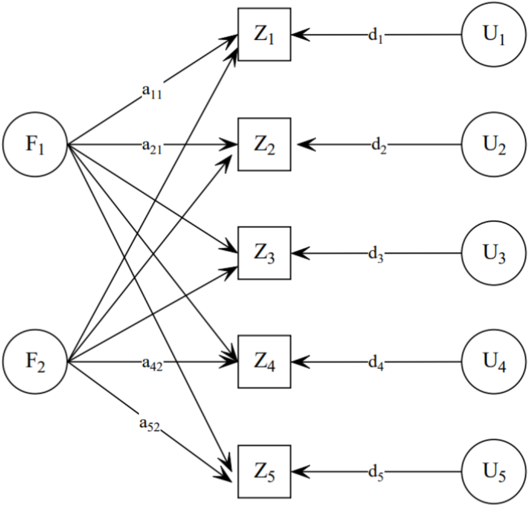
Factor Analysis via Path Diagrams

Two correlated common factors (oblique rotation)



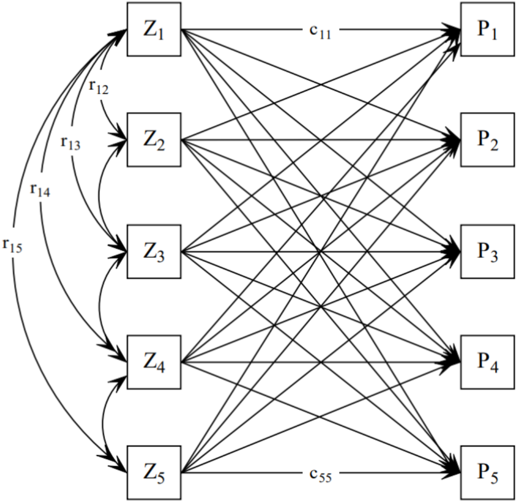
Factor Analysis via Path Diagrams

Two uncorrelated common factors (initial solution, orthogonal rotation)



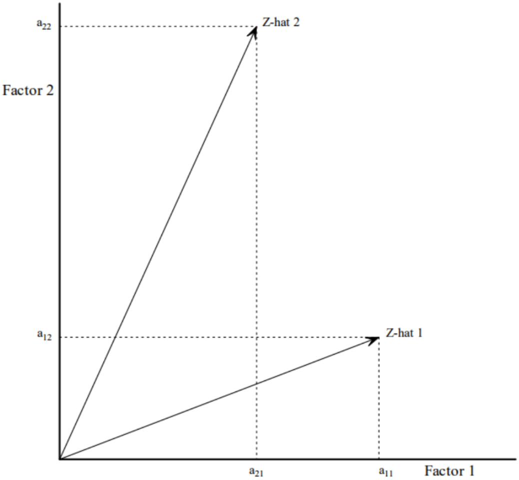
Factor Analysis via Path Diagrams

A principal components analysis (several correlations omitted for clarity)



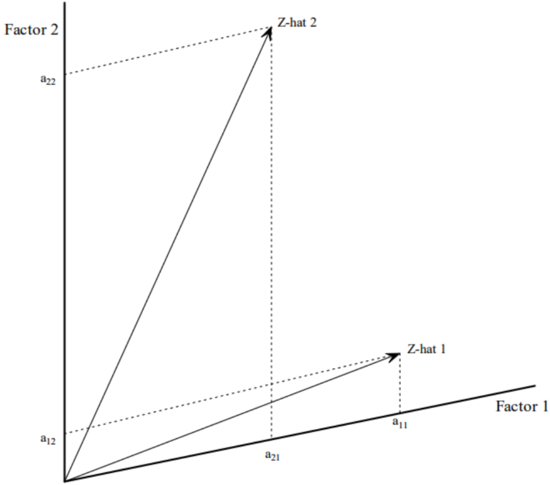
Factor Analysis via Vectors

Orthogonal (uncorrelated) factors as unit-length vectors



Factor Analysis via Vectors

Oblique (correlated) factors as unit-length vectors



Objectives of FA

- In some sense, factor analysis can be viewed as a multivariate regression, in which k observed variables $(y_1, y_2, \ldots, y_k)$ are regressed on m unobserved factors $(Z_1, Z_2, \ldots, Z_m)$
- Working from this perspective, there are six main objectives that we want to accomplish:

1. The set of regression coefficients representing the relationships between the k dependent (i.e., observed) variables and the m independent variables (i.e., the unobserved common factors)
2. The bivariate correlations between the k observed variables and the m unobserved common factors
3. The proportion of variance "explained" in each of the observed variables by the m unobserved common factors (analogous to k distinct R^2 values, one for each of the observed variables)
4. The correlations between the common factors (if any)
5. Scores for each of the n observations on each of the m common factors, for use in subsequent analyses

Fundamental Equations of Factor Analysis

The factor model can be expressed mathematically as follows:

$$
\begin{aligned}
Y_1 &= \lambda_{11} Z_1 + \cdots + \lambda_{1r} Z_r + \psi_1 \varepsilon_1 \\
Y_2 &= \lambda_{21} Z_1 + \cdots + \lambda_{2r} Z_r + \psi_2 \varepsilon_2 \\
Y_3 &= \lambda_{31} Z_1 + \cdots + \lambda_{3r} Z_r + \psi_3 \varepsilon_3 \\
&\vdots \\
Y_k &= \lambda_{k1} Z_1 + \cdots + \lambda_{kr} Z_r + \psi_k \varepsilon_k
\end{aligned}
$$

List of Terms

- $Y_1, \ldots, Y_k$ are the k observed variables
- $Z_1, \ldots, Z_r$ are the r common factors, where $r < k$
- $\lambda_{j1}, \ldots, \lambda_{jr}, j = \{1, \ldots, k\}$ are the factor pattern coefficients (loadings) that characterize the nature of the linear relationship between the common factors and the observed variables
- $\varepsilon_1, \ldots, \varepsilon_k$ are the unique factors, such that each observed variable has its own associated unique factor
- $\psi_1, \ldots, \psi_k$ (pronounced "p-sigh") are the coefficients relating the unique factors to the observed variables

Some Identifying Assumptions

- The Z_i and ε_j variables have mean 0 and unit variance (var $=$ 1)
- $r_{Z_i \varepsilon_j} = 0$
- The ε_j variables contain potentially two sources of variation:
  1. Random measurement error (that will be uncorrelated with everything)
  2. True variances that are not present in any other variables (i.e., specific variance)

Communality, Uniqueness, and Reliability

- $h_j^2 = \sum_{k=1}^{r} \lambda_{jk}^2$ is the variable's *communality*
- The amount of Y_j's variance that is shared with the other variables
- (Remember, in PCA, we implicitly assumed that we wanted to capture/explain all of the variance? Our FA model wants to explain only the *common* variance.)
- ψ_j^2 is Y_j's *unique* variance.
- Since the variance of a given observed variable Y_j is 1 (because of standardization), we can easily decompose that variance into common and unique variance:

$$h_j^2 = 1 - \psi_j^2$$
$$\psi_j^2 = 1 - h_j^2$$

Communality, Uniqueness, and Reliability

We might consider the unique variance to be composed of two parts:

$$\psi_j^2 = s_j^2 + e_j^2$$

where:

$$s_j^2 = \text{Specific (i.e., true) variance}$$
$$e_j^2 = \text{Measurement error}$$

Since we know that a variable's reliability is true variance divided by the variable's total variance, then reliability is:

$$\rho_{jj} = \frac{h_j^2 + s_j^2}{1} = h_j^2 + s_j^2$$

So, a variable's communality (its shared variance) will always be less than or equal to the variable's reliability. This puts a lower bound on reliability.

Factor Analysis via Matrix Notation

Assume :

- $\boldsymbol{Y}$ is an $n \times k$ matrix of observed variables
- $E(\boldsymbol{Y}) = \mathbf{0}$
- $E(\boldsymbol{Y}'\boldsymbol{Y}) = \boldsymbol{R}_{YY}$ is a correlation matrix with ones on the diagonal
- $\boldsymbol{Z}$ is our $n \times r$ matrix of common factors such that $E(\boldsymbol{Z}) = \mathbf{0}$ and $E(\boldsymbol{Z}'\boldsymbol{Z}) = \boldsymbol{R}_{ZZ}$ (the correlations between the latent factors)
- $\boldsymbol{E}$ is a $n \times k$ matrix of unique factors such that $E(\boldsymbol{E}) = \mathbf{0}$ and $E(\boldsymbol{E}'\boldsymbol{E}) = \boldsymbol{I}$
- $\boldsymbol{\Lambda}$ is a $r \times k$ matrix of factor pattern coefficients relating the factors to the observed variables
- $\boldsymbol{\Psi}$ is a $k \times k$ diagonal matrix with ψ_k on the diagonal giving the coefficient relating the unique factor to the observed variable

Factor Analysis via Matrix Notation

Then, the fundamental equation of Factor Analysis is:

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{E}\boldsymbol{\Psi}$$

Since $\boldsymbol{R}_{YY} = E(\boldsymbol{Y}'\boldsymbol{Y})$, we can do the following:

$$\begin{aligned}
\boldsymbol{R}_{YY} &= E((\boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{E}\boldsymbol{\Psi})'(\boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{E}\boldsymbol{\Psi})) \\
&= E((\boldsymbol{\Lambda}'\boldsymbol{Z}' + \boldsymbol{\Psi}'\boldsymbol{E}')(\boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{E}\boldsymbol{\Psi})) \\
&= E(\boldsymbol{\Lambda}'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}'\boldsymbol{Z}'\boldsymbol{E}\boldsymbol{\Psi} + \boldsymbol{\Psi}'\boldsymbol{E}'\boldsymbol{Z}\boldsymbol{\Lambda} + \boldsymbol{\Psi}'\boldsymbol{E}'\boldsymbol{E}\boldsymbol{\Psi}) \\
&= \boldsymbol{\Lambda}'E(\boldsymbol{Z}'\boldsymbol{Z})\boldsymbol{\Lambda} + \boldsymbol{\Lambda}'E(\boldsymbol{Z}'\boldsymbol{E})\boldsymbol{\Psi} + \boldsymbol{\Psi}'E(\boldsymbol{E}'\boldsymbol{Z})\boldsymbol{\Lambda} + \boldsymbol{\Psi}'E(\boldsymbol{E}'\boldsymbol{E})\boldsymbol{\Psi} \\
&= \boldsymbol{\Lambda}'\boldsymbol{R}_{ZZ}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}'\boldsymbol{R}_{ZE}\boldsymbol{\Psi} + \boldsymbol{\Psi}'\boldsymbol{R}_{EZ}\boldsymbol{\Lambda} + \boldsymbol{\Psi}'\boldsymbol{I}\boldsymbol{\Psi} \\
&= \boldsymbol{\Lambda}'\boldsymbol{R}_{ZZ}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}'\boldsymbol{R}_{ZE}\boldsymbol{\Psi} + \boldsymbol{\Psi}'\boldsymbol{R}_{EZ}\boldsymbol{\Lambda} + \boldsymbol{\Psi}^2
\end{aligned}$$

Since $\boldsymbol{R}_{ZE} = \boldsymbol{R}_{EZ} = \boldsymbol{0}$ by assumption, we get...

The Fundamental Theorem of Factor Analysis

$$\boldsymbol{R}_{YY} = \boldsymbol{\Lambda}' \boldsymbol{R}_{ZZ} \boldsymbol{\Lambda} + \boldsymbol{\Psi}^2$$

Subtracting $\boldsymbol{\Psi}^2$ (a diagonal matrix) from both sides, we get:

$$\boldsymbol{R}_{YY} - \boldsymbol{\Psi}^2 = \boldsymbol{\Lambda}' \boldsymbol{R}_{ZZ} \boldsymbol{\Lambda}$$

$\boldsymbol{R}_c = \boldsymbol{R}_{YY} - \boldsymbol{\Psi}^2$, where $\boldsymbol{R}_c$ is the reduced correlation matrix. This is what we will be inputting into the factor analysis since we are most interested in explaining shared/common variance.

Since $\boldsymbol{\Psi}^2$ (the unique factor variance) is diagonal, the off-diagonal elements of $\boldsymbol{R}_{YY}$ are unaffected, and hence are due entirely to the common factors

Factor Extraction

- We basically have a regression problem, except we only know the DV (observed Y_k's)
- Can't estimate with OLS regression because, of course, we don't know values of any of the variables on the right-hand side of the equation
- If $E(Y_i) = 0$ and $Var(Y_i) = 1$; AND, $E(Z_i) = 0$ and $Var(Z_i) = 1$, then $\lambda_{11} = r_{Y_1 Z_1}$
  - That is, the λ's (factor pattern coefficients) are the correlations between the observed variables and the common factors
  - Furthermore, $h_1^2 = \lambda_{11}^2 = r_{Y_1 Z_1}^2$, the variance in Y explained by Z, which is the communality

Estimation, cont'd

- Factor pattern coefficients: coefficients relating indicators to factors ("loadings")
- Factor structure coefficients: correlations between indicators and factors
  - When the factors are orthogonal, the factor pattern and structure coefficients are the same
  - When factors are correlated, the structure matrix contains the correlations and the pattern contains the loadings
- Many ways to estimate the common factor model, but the two most popular are probably:
  - Maximum likelihood
  - Iterated principal axis factoring
- We will focus on the latter since it it more intuitive and doesn't have convergence issues MLE is occasionally plagued by
- That said, results should be substantively identical

Iterated Principal Axis Factoring

- Start with a correlation matrix of variables one's interested in
  - Problem: we have 1's on the diagonal
  - This means that we're modeling the total variance of each item, not the shared variance (this is what PCA does)
  - We need estimates of communality, h_j^2 – the variance shared between each variable and the rest
- We provide an initial estimate of communality using the squared multiple correlation between a variable and all the rest
  - We regress each variable on the rest and use the R^2 as the estimate of communality
  - This is an underestimate, but using 1's is a (gross, in most cases) over estimate
  - This is the reduced correlation matrix, $\boldsymbol{R}_c$, from before

Iterated Principal Axis Factoring, cont'd

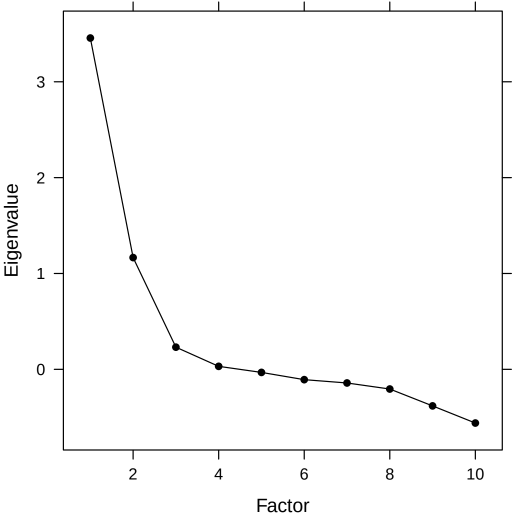
- Next, we submit our estimate of the reduced correlation matrix to an eigendecomposition
- Calculate our estimates of the factor loadings, $\mathbf{\Lambda} = \mathbf{VD}$
- Reproduce the correlation matrix, $\mathbf{R}_c^* = \mathbf{\Lambda\Lambda}'$
- Replace the initial (SMC) communality estimates with the new ones and repeat the procedure
- Once the change in communalities is less than some pre-specified criterion (usually 0.001), we stop iteration and produce results
- Results include:
  - Factor pattern coefficients (loadings)
  - Factor structure coefficients (correlations between variables and latent factors)
  - Communalities
  - Uniquenesses
  - Residuals

How Many Factors?

- Like with PCA, we can look at the proportion of variance explained by each factor in the initial, unrotated solution
- We can also examine a scree plot of the eigenvalues against the associated factor and look for the "elbow" (most common)
- Some also use "parallel analysis"
  - Compares the eigenvalues of factors of the observed data with those of a random data matrix of the same size as the original
  - First, data of the same size as the original data matrix is simulated and a correlation matrix is produced
  - Second, a factor analysis is conducted and the eigenvalues from the simulated data are retained
  - Finally, the difference in eigenvalues for each factor are compared – if there's a difference, one should retain that factor
  - NOTE: parallel analysis quite sensitive to sample size...this leads to a probable overestimation of the factors that should be retained

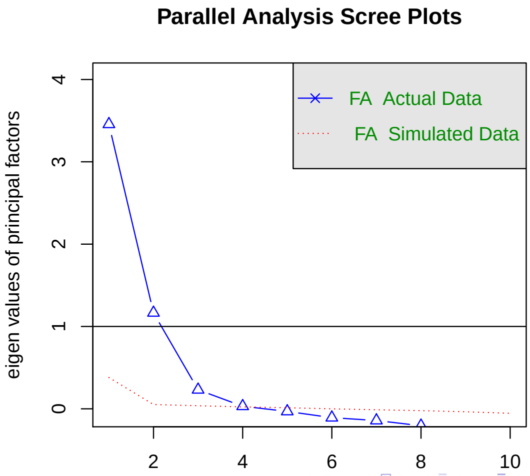
Scree Plot

```
scree(cor(groups), pc=FALSE)
```



Parallel Analysis

```
> fa.parallel(cor(groups), n.obs = nrow(groups), fa = "fa")
Parallel analysis suggests that the number of factors =  4
```



Factor Rotation

- Basic idea is that there are some views of the data that are more useful than others
- Imagine looking at a table in two dimensions – there are lots of different views that give you different insights into the nature of the table
- Rotation is the process of projecting the variable vectors into the factor space in a way that makes the factors more interpretable
- Called rotating because we do this by actually rotating the axes
- EFA is unique only up to a rotation
  - There exist infinite pairs of rotated factors such that all have exactly identical fit since

Factor Rotation, cont'd

- Critics suggest that this property of rotation allows researchers to project their preconceived notions onto the data – they're "p-hacking," in today's terminology
- But, practitioners aren't devising their own rotational strategies to make the analysis conform to their expectations – this criticism is very weak
- Rather rotation is about aiding our substantive interpretation of the latent factors, which can become difficult when there are many variables and factors
- When we rotate, the proportion of variance explained, model residuals, and communality/uniqueness estimates remain unchanged
- All we're doing is taking a different "view" of the data, which means that loadings will change

Factor Rotation, cont'd

- Thurstone: recommends considering "simple structure" (parsimony)
  1. Each factor should effect as few variables as possible
  2. Each variable should be explained by as few factors as possible
- With positively correlated variables in a two factor solution, the unrotated solution will generally show two clusters of variable vectors (one in each of two quadrants)
  - The unrotated solutions is almost never going to be a simple structure solution
- Best way to rotate: find the groups of variable vectors and run the axes through them
- Want to find a rotation that maximizes the variability within the columns of the factor pattern matrix (i.e., push the higher coefficients toward 1 and the lower coefficients toward 0)

Factor Rotation, cont'd

- Once a rotation matrix of cosines is produced, the factor loadings are transformed by post-multiplication by that matrix to produce the new loadings
- During the rotation phase of the analysis, we might choose to depart from orthogonality toward correlated (oblique) factors
- Orthogonal vs. Oblique
  - Neither is "better" than the other
  - Should the latent variables be correlated? Then it makes sense for the factors to be rotated obliquely
  - Do you think the latent variables are uncorrelated (probably the less likely scenario, really)? Then go with orthogonal

Factor Rotation, cont'd

- Orthogonal and oblique (factors are correlated) rotations
- Kaiser's varimax rotation (orthogonal)
  - For two factors, the rotation matrix $\mathbf{T}$ will rotate the two factors θ radians in a counterclockwise direction
  - Maximizes variance in the columns of $\boldsymbol{\Lambda}$
  - Tries to get the factor axes as close as possible to running through the vector clusters while still respecting orthogonality
- Promax rotation (oblique)
  - Makes a target matrix $\mathbf{B}$, such that the elements are $b_{ij} = \frac{\lambda_{ij}^{n+1}}{\lambda_{ij}}$
  - Then, finds a transformation matrix $\mathbf{T_1} = (\boldsymbol{\Lambda}_\mathbf{v}'\boldsymbol{\Lambda}_\mathbf{v})^{-\mathbf{1}}\boldsymbol{\Lambda}_\mathbf{v}\mathbf{B}$, where $\boldsymbol{\Lambda}_v$ is the varimax rotation
  - With oblique rotation, the axes can run right through the middle of the variable vector clusters since it's OK for the factor axes to be correlated
  - Potential to provide "simpler" structure than orthogonal rotations

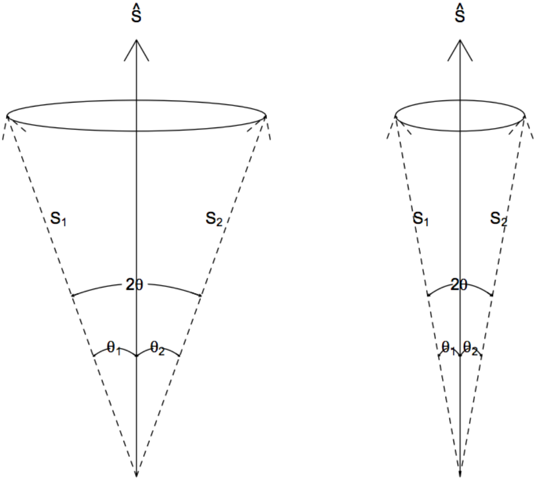
Rotation, cont'd

- Practical advice:
  - Don't spend much time looking at unrotated solution – this solution is only good for assessing dimensionality (i.e., how many factors you probably want to retain)
  - ALWAYS ROTATE!!!!
  - Compare the factor pattern matrix for the varimax and promax solutions – are they even that different?
  - Use the solution that allows you to give the simplest and most theoretically appropriate meaning or interpretation to the solution

Factor Scores

- We might want to estimate factor scores for the row objects
  1. They represent a parsimonious summary of the original data possibly useful in subsequent analyses
  2. They are likely to be more reliable than the observed variable values.
  3. The factor score is a "pure" measure of a latent variable, while an observed value may be ambiguous because we do not know what combination of latent variables may be represented by that observed value.
- They are *estimates* of the common part of the variables and are not identical to the factors themselves
- Unfortunately, factor scores are **indeterminant**
  - More information to estimate than we have
  - Means there are infinite equally good estimates of factor scores
  - Exception is when communalities are 1 and uniquesnesses are 0 – i.e., doing a PCA

Indeterminacy (from Revelle 2019)



Factor score estimate $\hat{S}$ is equally representative of factor S_1 and S_2

Factor Scores, cont'd

- Lots of ways to estimate factor scores...we will use the regression-based "Thurstone method"
- Want to come up with a matrix of estimated values of the latent variables, $\hat{\mathbf{Z}}$

$$\begin{aligned}\hat{\mathbf{Z}} &= \mathbf{R_{ZZ}}\boldsymbol{\Lambda}\mathbf{R_{YY}^{-1}}\mathbf{Y} \\ &= \mathbf{R_{ZY}}\mathbf{R_{YY}^{-1}}\mathbf{Y}\end{aligned}$$

- $\mathbf{P}^2 =$ diag$(\mathbf{R_{ZZ}}\mathbf{R_{YY}^{-1}}\mathbf{R_{ZY}})$, where $\mathbf{P}^2$ is the squared multiple correlation of the observed variables and the factors
- In PCA, our "hat-like" values are calculated: $\mathbf{C} = \mathbf{XA}$
  - We have no "hat" over the $\mathbf{C}$ in PCA - this is because we are not *estimating* in PCA, but we are in EFA

Steps in an Exploratory Factor Analysis

1. Estimate initial model, not constraining the number of factors, so that the appropriate number of factors to retain can be discerned
   - Usually practitioners check a scree plot of the eigenvalues against the number of factors retaining the number of factors that comes before the "elbow"
2. Re-estimate factor model with the desired number of factors
3. Rotate factors
4. Substantively interpret factors by examining patterns in factor loadings
5. If using in subsequent analyses, compute factor scores for row objects

Difference between PCA and EFA

- People confuse the techniques because the mathematics behind the models are somewhat similar
- Both do aim to reduce the dimensionality of a multivariate data set, in some way
- PCA tries to capture variation, EFA tries to explain it
- Both are pointless if the observed variables are almost uncorrelated
- But, PCA and EFA are different "models" (PCA isn't really a model)
    - In PCA, the components exist only in relation to the observed data: $\mathbf{C} = \mathbf{X}\mathbf{A}$
    - In EFA, the latent constructs exist in their own right: $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Lambda} + \mathbf{E}\boldsymbol{\Psi}$

Difference between PCA and EFA, cont'd

- Interpretation:
  - In PCA we can interpret insofar as we can see what variables are related to the components
  - We do not really understand meaning of the components the way that we do the factors from PCA
- In PCA we are not making hypotheses about the causal structure between some latent variables and some observed indicators - this is for EFA
  - EFA has an underlying theoretical model, PCA doesn't
- If the number of retained components is increased, say from m to $m+1$, the first m components are unchanged. This is not the case in factor analysis, where there can be substantial changes in all factors if the number of factors is changed

Difference between PCA and EFA, cont'd

- If there is a perfect simple structure solution in EFA, then $\mathbf{R_{yy}} \leftrightarrow \mathbf{\Lambda}'\mathbf{R_{ZZ}}\mathbf{\Lambda}$
- In PCA, $\mathbf{R_{yy}} \rightarrow \mathbf{A}'\mathbf{A}$
- The actual principal components are interpreted as controlling for the variation that exists in the multivariate data on which the PCA was performed
  - If a two-dimensional simple structure exits, PCA will not recover that while EFA will (or, at least, can)
  - In fact, PCA is not designed to do this at all - its designed to soak up as much variation in a set of multivariate data as possible
- All that being said, the PCs and factors on the same data will usually be very highly correlated

Non-Interval Level Data

- Perhaps your variables aren't amenable to the standard, Pearson product-moment correlation
  - That is, you have ordinal or dichotomous data (can't do much with nominal)
- As with PCA, we have (at least) two options:
  1. Use correlations designed to decipher linear relationships between ordinal (polychoric) or dichotomous (tetrachoric, point-biserial) data
  2. "Linearize" relationships by optimally scaling the original variables
- Could also move to a (multidimensional) IRT model, since they are designed for categorical data and nonlinear relationships between observed indicator and latent variable
- As we saw with PCA, oftentimes ordinal variables behave in a linear fashion and we don't need to worry too much

Looking Back to IRT

- A key assumption of all IRT models is unidimensionality
- Could compute correlation matrix and do PCA or EFA
  - BUT...we don't expect variables to be related to each other or the latent factor in a linear way
  - Therefore, this isn't necessarily appropriate, especially with dichotomous input variables
- Four good options:
  - Nonlinear (categorical) PCA via "princals" routine/function
  - EFA of tetrachoric correlation matrix
  - "Linearize" variables using the "aspect" package
  - Compare 1- and 2-dimensional IRT models using "mirt" package
- We can actually estimate a 2PL IRT model in the factor analysis framework
  - Use tetrachoric correlations as input data
  - Transform loadings into discrimination parameters