



THE UNIVERSITY
of EDINBURGH

WEEK 4

Exploratory Factor Analysis 1

Data Analysis for Psychology in R 3

dapR3 Team

Department of Psychology
The University of Edinburgh

Learning Objectives

1. Understand the conceptual difference between PCA and EFA
2. Understand the strengths and weaknesses of different methods to estimate the optimal number of factors
3. Understand the purpose of factor rotation and the difference between orthogonal and oblique rotation.
4. Run and interpret EFA analyses in R.

Part 1: Introduction to EFA

Part 2: Estimation & Number of factors problem

Part 3: Factor rotation

Part 4: Example and interpretation

Real friends don't let friends do PCA.

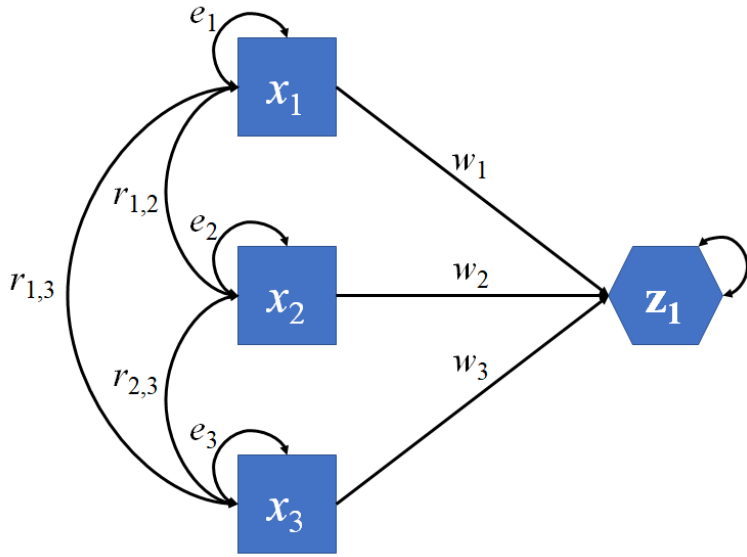
W. Revelle, 25 October 2020

Why are we reducing?

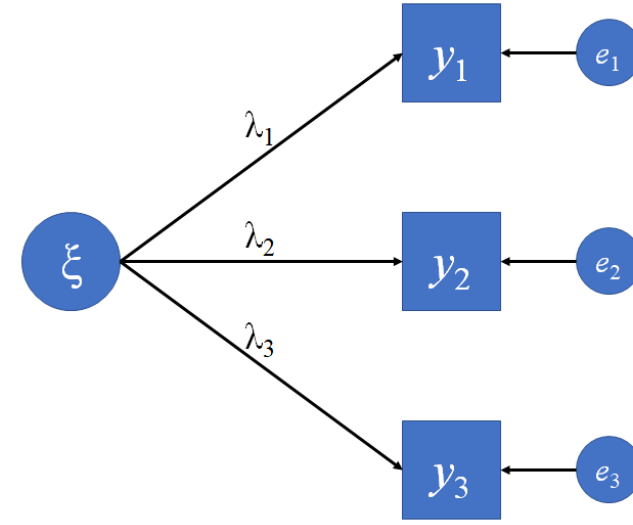
- Why are your variables correlated?
 - Agnostic/don't care
 - **Believe there are underlying "causes" of these correlations**
- What are your goals?
 - Just reduce the number of variables
 - **Reduce your variables and learn about/model their underlying (latent) causes**
- It is possible to use FA for pure reduction.

Latent variables

- One of many features that distinguish factor analysis and principal components analysis
- Key concept of psychometrics (factor analysis is a part)
- Theorized common cause (e.g., cognitive ability) of responses to a set of variables
 - Explain correlations between measured variables
 - Held to be true
 - No direct test of this theory



$$\mathbf{z} = x_1 w_1 + x_2 w_2 + x_3 w_3$$



$$y_1 = \lambda_1 \xi + e_1$$

$$y_2 = \lambda_2 \xi + e_2$$

$$y_3 = \lambda_3 \xi + e_3$$

$$\text{cov}(\xi, e_j) = 0$$

PCA versus EFA

- PCA
 - The observed measures (x_1, x_2, x_3) are independent variables
 - The component (\mathbf{z}) is the dependent variable
 - Explains as much variance in the measures (x_1, x_2, x_3) as possible
 - Components are determinate
- EFA
 - The observed measures (y_1, y_2, y_3) are dependent variables
 - The factor (ξ), is the independent variable
 - Models the relationships between variables ($r_{y_1, y_2}, r_{y_1, y_3}, r_{y_2, y_3}$)
 - Factors are *indeterminate*

Modeling the data

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Item 1 | 1.00 | | | | | | | |
| Item 2 | 0.60 | 1.00 | | | | | | |
| Item 3 | 0.55 | 0.61 | 1.00 | | | | | |
| Item 4 | 0.45 | 0.48 | 0.71 | 1.00 | | | | |
| Item 5 | 0.10 | 0.01 | 0.04 | 0.13 | 1.00 | | | |
| Item 6 | 0.05 | 0.00 | 0.08 | 0.20 | 0.52 | 1.00 | | |
| Item 7 | 0.14 | 0.02 | 0.11 | 0.14 | 0.76 | 0.51 | 1.00 | |
| Item 8 | 0.07 | 0.11 | 0.13 | 0.04 | 0.68 | 0.54 | 0.48 | 1.00 |

Modeling the data

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 | Item 8 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Item 1 | 1.00 | | | | | | | |
| Item 2 | 0.60 | 1.00 | | | | | | |
| Item 3 | 0.55 | 0.61 | 1.00 | | | | | |
| Item 4 | 0.45 | 0.48 | 0.71 | 1.00 | | | | |
| Item 5 | 0.10 | 0.01 | 0.04 | 0.13 | 1.00 | | | |
| Item 6 | 0.05 | 0.00 | 0.08 | 0.20 | 0.52 | 1.00 | | |
| Item 7 | 0.14 | 0.02 | 0.11 | 0.14 | 0.76 | 0.51 | 1.00 | |
| Item 8 | 0.07 | 0.11 | 0.13 | 0.04 | 0.68 | 0.54 | 0.48 | 1.00 |

Modeling the data

- EFA tries to explain patterns of correlations
 - If, for our three items, the model (factor or ξ) is good, it will explain their interrelationships
 - Read the dots (\cdot) as "given" or "controlling for"

$$\rho(y_1, y_2 \cdot \xi) = \text{corr}(e_1, e_2) = 0$$

$$\rho(y_1, y_3 \cdot \xi) = \text{corr}(e_1, e_3) = 0$$

$$\rho(y_2, y_3 \cdot \xi) = \text{corr}(e_2, e_3) = 0$$

Modeling the data

- Factor analysis has to distinguish between the true and unique variance
 - True variance
 - Variance common to an item and at least one other item
 - Variance specific to an item that is not shared with any other items
 - Unique variance
 - Variance specific to an item that is not shared with any other items
 - Error variance

$$\text{var}(\text{total}) = \text{var}(\text{common}) + \text{var}(\text{specific}) + \text{var}(\text{error})$$

The general factor model equation

$$\Sigma = \Lambda\Phi\Lambda' + \Psi$$

Σ : A $p \times p$ observed covariance matrix (from data)

Λ : A $p \times m$ matrix of factor loadings (relates the m factors to the p items)

Φ : An $m \times m$ matrix of correlations between factors ("goes away" with orthogonal factors)

Ψ : A diagonal matrix with p elements indicating unique (error) variance for each item

Assumptions

- Another way that factor analysis resembles regression
 - The residuals/error terms (e) should be uncorrelated (it's a diagonal matrix, remember!)
 - The residuals/errors should not correlate with factor
 - Relationships between items and factors should be linear, although there are models that can account for nonlinear relationships

Practical Steps

1. Check the appropriateness of the data and decide of the appropriate estimator.
2. Decide which methods to use to select a number of factors.
3. Decide conceptually whether to apply rotation and how to do so.
4. Decide on the criteria to assess and modify your solution.
5. Run the analysis.
6. Evaluate the solution (apply 4)
7. Select a final solution and interpret the model, labeling the factors.
8. Report your results.

End of Part 1

Part 1: Introduction to EFA

Part 2: Estimation & Number of factors problem

Part 3: Factor rotation

Part 4: Example and interpretation

Estimation

- For PCA, we discussed the use of the eigendecomposition.
 - This is not an estimation method, it is simply a calculation
- As we have a model for the data in factor analysis, we need to estimate the model parameters
 - primarily here the factor loading's.

A brief pause for factor loadings

- When we run an FA, the main element of our output are the factor loading's and factor correlations (see rotation)
- Factor loading's, like PCA loading's, show the relationship of each measured variable to each factor.
 - They range between -1.00 and 1.00
 - Larger absolute values = stronger relationship between measured variable and factor
- We interpret our factor models by the pattern and size of these loading's.
 - **Primary loading's**: refer to the factor on which a measured variable has it's highest loading
 - **Cross-loading's**: refer to all other factor loading's for a given measured variable

Factor loadings

```
##
## Loadings:
##      MR1      MR2
## A1      -0.405
## A2       0.677
## A3       0.760
## A4  0.145  0.439
## A5       0.602
## C1  0.571
## C2  0.637
## C3  0.542
## C4 -0.649
## C5 -0.562
##
##              MR1   MR2
## SS loadings    1.791 1.764
## Proportion Var 0.179 0.176
## Cumulative Var 0.179 0.355
```

Estimation & Communalities

- The most efficient (i.e., least computationally intensive) way to factor analyze data is to start by estimating communalities
 - Communalities are estimates of how much true variance any variable has
 - Indicate how much variance in an item is explained by other variables, or factors
 - They appear in the diagonal of your correlation matrix
- Estimating communalities is difficult because population communalities are unknown
 - Range from 0 (no shared variance) to 1 (all variance is shared)
 - Occasionally estimates will be ≥ 1 (called a 'Heywood Case')
 - Methods often are iterative and "mechanical" as a result

Principal axis factoring

- Principal factors with squared multiple correlation (SMC)
 1. Compute initial communalities from SMCs, which are multiple correlations of each item regressed on all $p - 1$ other variables
 2. Once we have these reasonable lower bounds, we substitute the 1s in the diagonal of our correlation matrix with the SMCs derived in step 1
 3. Obtain the factor loading matrix using the eigenvalues and eigenvectors of the matrix obtained in the step 2
- Some versions of principal axis factor use an iterative approach in which they replace the diagonal with the communalities obtained in step 3, and then repeat step 3, and so on, a set number of times

Method of minimum residuals

- This is an iterative approach and the default of the **fa** procedure
 1. Starts with some other solution, e.g., PCA or principal axes, extracting a set number of factors
 2. Adjusts loading's of all factors on each variable so as to minimize the residual correlations for that variable
- MINRES doesn't "try" to estimate communalities
- If you apply principal axis factoring to the original correlation matrix with a diagonal of communalities derived from step 2, you get the same factors as in the method of minimum residuals

Maximum likelihood estimation

- Uses a general iterative procedure for estimating parameters that we have previously discussed.
 - Assume a distribution for your data, e.g., a normal distribution
 - Your covariance matrix contains information related to the parameters, that is, factor loading's, uniqueness, and correlations between factors
 - The procedure works to find values for these parameters that maximize the likelihood of obtaining the covariance matrix
- Method offers the advantage of providing numerous "fit" statistics that you can use to evaluate how good your model is compared to alternative models
- Assumes a distribution (usually the multivariate normal)
 - The methods I described before do not
 - Which approach would be the most 'robust' to deviations from normality in your data?

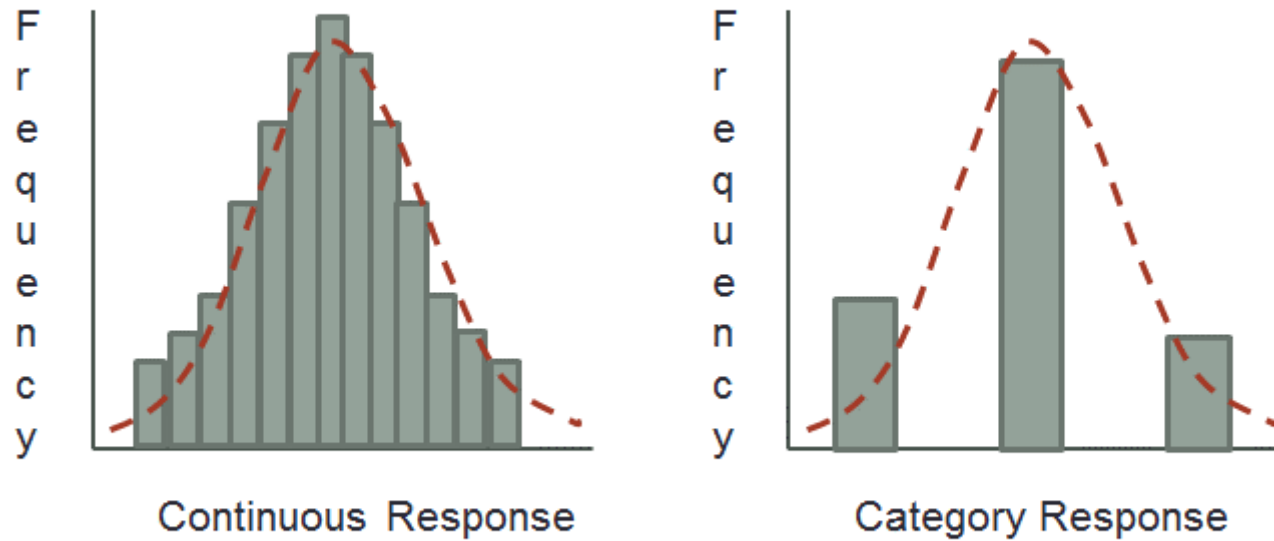
ML con's

- The issue is that for big analyses, sometimes it is not possible to find values for factor loadings that = MLE estimates.
 - Referred to as non-convergence (you may see warnings)
- Also may produce solutions with impossible values
 - Factor loadings > 1.00 (Heywood cases), thus negative residuals.
 - Factor correlations > 1.00

Non-continuous data

- Don't mistake response options for how the construct is likely to be distributed
 - I can easily get some life into a rather dull party? Yes/No
 - I have committed: 1. No crimes; 2. A minor crime; 3. A major crime
- Most constructs we seek to measure by questionnaire are assumed to be continuously distributed
 - It's thus *usually* okay to treat the data as if they are, too
 - The exception is for maximum likelihood factor analysis
 - Or if the observed distribution is very skewed

Non-continuous data



class:center

Non-continuous data

- If we are concerned and the construct is normally distributed, we can conduct our analysis on a matrix of tetra/polychoric correlations
- If the construct is not normally distributed, you can conduct a factor analysis that allows for these kinds of variables

Choosing an estimator

- The best option, as with many statistical models, is ML.
- If ML solutions fail to converge, principal axis is a simple approach which typically yields reliable results.
- If concerns over the distribution of variables, use PAF on the polychoric correlations.

Number of factors

- We have discussed the methods for deciding on the number of factors in the context of PCA.
- Recall we have 4 tools:
 - Variance explained
 - Scree plots
 - MAP
 - Parallel Analysis (FA or PCA)
- For FA, we generally want a slightly more nuanced approach than pure variance:
 - Use them all to provide a range of plausible number of factors
 - Treat MAP as a minimum
 - PA as a maximum
 - Explore all solutions in this range and select the one that yields the best numerically and theoretically.
- We will go through this process in later video.

Factor indeterminacy

- One tricky thing about factor analysis is that, when there are two or more factors, you run into the problem of *factor indeterminacy*.
- This occurs because factors are inferred from the pattern of correlations
 - Compare that to PCA where components are products of the data
- Factor indeterminacy means that there are an infinite number of pairs of factor loading's and factor score matrices which will fit the data **equally well**, and are thus **indistinguishable** by any numeric criteria
- There is no **unique solution** to the factor problem
- And this is also in part why the theoretical coherence of the models plays a much bigger role in FA than PCA.

End of Part 2

Part 1: Introduction to EFA

Part 2: Estimation & Number of factors problem

Part 3: Factor rotation

Part 4: Example and interpretation

Factor rotation: what and why?

- Factor solutions can sometimes be complex to interpret.
 - the pattern of the factor loading's is not clear.
 - The difference between the primary and cross-loading's is small
- Factor rotation is an approach to clarifying these relationships.
- Rotation aims to maximize the relationship of a measured item with a factor.
 - That is, make the primary loading big and cross-loading's small.

Analytic rotation

- Orthogonal
 - Includes varimax and quartimax rotations
 - Axes at right angles; correlations between components are zero
 - In example, I used a varimax rotation

Analytic rotation

- Orthogonal
 - Includes varimax and quartimax rotations
 - Axes at right angles; correlations between components are zero
 - In example, I used a varimax rotation
- Oblique
 - Includes promax and oblimin rotations
 - Axes are not at right angles; correlations between components are not zero
 - What happens when we apply an oblimin "rotation"?

The impact of rotation

Original correlations

The impact of rotation

The impact of rotation

Analytic rotation

- Rotation is about finding the most interpretable set of components
- There is not a "unique" solution to this problem
 - Rotations are mathematically indistinguishable
 - All you're doing is changing the reference point
 - This problem is known as "rotational indeterminacy"
- "Simple structure" was proposed by Thurstone as the criterion
 - Different means of rotation try to achieve this differently
 - I won't get into the maths!

Simple structure

- Adapted from Sass and Schmitt (2011):
 1. Each variable (row) should have at least one zero loading
 2. Each component (column) should have same number of zero's as there are components
 3. Every pair of components (columns) should have several variables which load on one component, but not the other
 4. Whenever more than four components are extracted, each pair of components (columns) should have a large proportion of variables which do not load on either component
 5. Every pair of components should have few variables which load on both components

How do I choose which rotation?

- Easy, my recommendation is always to choose oblique.
- Why?
 - It is very unlikely factors have correlations of 0
 - If they are close to zero, this is allowed within oblique rotation
 - The whole approach is exploratory, and the constraint is unnecessary.
- However, there is a catch...

Interpretation and oblique rotation

- When we have an obliquely rotated solution, we need to draw a distinction between the **pattern** and **structure** matrix.
 - Pattern Matrix: matrix of regression weights (loading's) from factors to variables.
 - Structure Matrix: matrix of correlations between factors and variables.
- When we orthogonally rotate, the pattern and structure matrix are the same.
- When we obliquely rotate, the structure matrix is the pattern matrix multiplied by the factor correlations.
- In most practical situations, this does not impact what we do, but it is important to highlight the distinction.

End of Part 3

Part 1: Introduction to EFA

Part 2: Estimation & Number of factors problem

Part 3: Factor rotation

Part 4: Example and interpretation

BFI data in `psych()`

- Factor analysis yields a lot of output.
 - It is tricky to get this on slides.
- So let's attempt to run an analysis in R itself.

End