The Common Factor Model

PSY 600K

Dr. Alyssa Gibbons

February 22, 2018

Agenda

- The big picture of factor analysis:
  - Homogeneity/heterogeneity (or, how many factors?).
  - DeVellis' 4 purposes of factor analysis.
  - The formal (general) common factor model.
    - Loadings
    - Uniquenesses
  - A few important properties of the common factor model.

Homogeneity/Heterogeneity

- We keep saying that tests should be **homogeneous**.
  - All items measuring one thing in common.
- But no test is **perfectly** homogeneous.
  - Some pairs of items have more in common than others.
  - Homogeneity is really a matter of degree.
- Factor analysis is all about determining **how** homogeneous or heterogeneous your items are.
  - Is one common factor enough? Two? Three?
  - Balancing parsimony with accuracy.

R & M's Example (3.1), Simplified

1.	1					
2.	.65	1				
3.	.66	.67	1			
4.	.16	.15	.13	1		
5.	.17	.14	.15	.53	1	
6.	.13	.16	.16	.52	.50	1

- All of these items are positively correlated…
- … but some are more correlated than others.
- Does it make more sense to describe this test as measuring one thing or two?

Exploratory vs. Confirmatory

- We can start with a hypothesis about how many dimensions we should have and what they are, or we can start with a set of items and no hypothesis.
- The underlying math is a little different depending on where we start.
  - We'll highlight these differences over the next couple of weeks.
- Most of what we'll say **today** applies to both.

DeVellis: 4 Uses of Factor Analysis

- 1. Condense information.
  - E.g., describe someone in terms of 5 personality traits rather than 100 adjectives.
- 2. Determine how many latent variables underlie a set of items.
  - This sounds like #1, but can be a little bit different…
  - We can condense information without proposing a common cause or latent variable .
    - **Principal components analysis** – NOT FA.

DeVellis: 4 Uses of Factor Analysis

- 3. Define/describe the meaning of those latent variables.
  - Use the content of highly related items to "identify" important factors.
  - Is this justifiable? Why or why not?
- 4. Evaluate the quality of individual items.
  - Some items tell us more about the underlying construct(s) than others.
  - FA allows us to assess how much information an item gives us about a latent factor.

A Formal Model

- Every item response is driven by some number of underlying factors, plus some random error.
  - $X_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + u_1$
  - $X_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + u_2$
- fs are the underlying factors – true values are unknown.
- as are like regression coefficients – they tell us how much each factor influences X_i.
  - $X_1 = .85f_1 + .15f_2 + u_1$
  - $X_2 = .20f_1 + .90f_2 + u_2$
- us are the residual – whatever part of X_i is not determined by the common factors.

How Many Factors?

- $X_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + u_1$
- R & M state that the number of factors (m) is less than the number of items (p).
- The number of factors is actually a choice we make.
  - Single factor model: $m = 1$.
  - Fully unrestricted EFA starts with $m = p$.
  - In CFA, choice of m is based on theory.
- Our goal is to choose the number of factors that gets us "close enough" to the observed data.
  - The more factors we have, the better our model will fit.
  - The fewer factors we have, the simpler our explanation.

Factor Loadings

- *a* coefficients (more commonly called λs) tell us how good the item is.
  - Difficult to interpret unless standardized, but bigger is better!
  - If the factors are uncorrelated, λ_{ij} = the correlation between item *i* and common factor *j*.
  - Note that if we have multiple factors, an item may be "good" on one factor and "poor" on another.

Factor Loadings

- If factors are uncorrelated, the correlation between two items equals the sum of the products of their factor loadings on all the factors:
  - $\rho_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jm}a_{km}$
  - For just one factor, $\rho_{jk} = a_{j1}a_{k1}$
  - An extension of DeVellis' path diagram stuff!
- We can work backwards from this equation to estimate the factor loadings.
  - We can also account for correlated factors – the equation is just more complex.

Uniquenesses

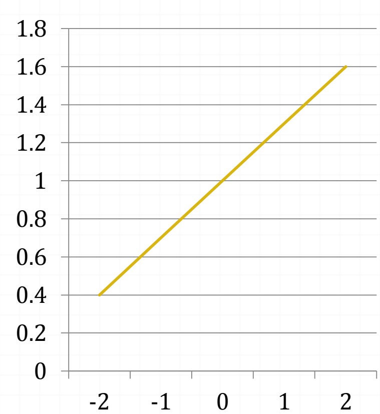
- The *us*, (more commonly ψs), are the part of the item response that is not explained by any common factor in the model.
- By definition, *us* are not related to any of the common factors nor to any other *u*.
  - Called the "unique part" or "uniqueness."
- The size of *u* is not particularly interpretable.
  - The size of *a* is much more important.

Conditional Independence

- Imagine a group of individuals whose scores on ALL of the common factors are exactly the same:
  - $X_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + u_1$
  - $X_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + u_2$
- Everything except the *us* is now constant across all members of the group.
  - And we know the *us* are uncorrelated across items.
  - So X_1 and X_2 are **uncorrelated** in this group.
  - **All** of the relationship between X_1 and X_2 is accounted for by the common factors.
  - This is called **conditional independence**, and it is an important property of many psychometric models (not just factor analysis).

Linear Relationship

- In this model, X_1 and f are related in a linear fashion:
  - $X_1 = a_{11}f_1 + u_1$
- Most of the time, this is not a problem.
  - But at extreme values of *f*, the model predicts impossible values of *X*.
- We'll come back to this when we discuss IRT.



Questions?

For next time:
Exploratory Factor Analysis
Read: DeVellis pp. 125-132 AND R & M 3.6 – 3.10
6th Reading Response