Issues in CFA

PSY 600K
Dr. Alyssa Gibbons
March 22, 2018

Agenda

- Theoretical background for the single factor model
  - Identifiability
  - Standardized loadings
- Homogeneity & omega
- Multiple factor models
  - Basic
  - Hierarchical

The Single Factor Model

- Developed by Spearman.
- Another way to write it:
  - $X_j = \mu_j + \lambda_j F + E_j$
- Break it down:
  - μ_j = item mean
    - Items vary in difficulty.
    - We often omit the mean from our models – that doesn't change the other values, we're just not usually concerned with the means in a CFA. We *can* model them explicitly if we choose to.
  - λ_j = factor loading
    - Some items are more sensitive to the common factor than others.
  - E_j = uniqueness
    - Each item has its own amount of unique variance.

Important Properties

- If the single factor model holds, then:
  - The covariance between two items = the product of their factor loadings.
    - $\sigma_{jk} = \lambda_j \lambda_k$
  - The variance of an item = the square of its factor loading plus the variance of the unique part.
    - $\sigma_{jj} = \lambda^2{}_j + \psi^2{}_j$
    - Don't square the uniqueness again!
- If we have at least 3 items, we can solve these equations to find the factor loadings and uniquenesses from the item covariances.

Identifiability

- That system of equations gives us a unique solution AS LONG AS:
  - We set the factor variance = 1 OR
  - We fix one factor loading = 1
- Either is acceptable in practice.
  - The construct is theoretical, we can scale it however we want.
  - Factor loadings need to be standardized anyway before they can be interpreted in absolute terms.
- This means we don't need rotation in CFA.

Standardized Factor Loadings

- Generally, we perform CFAs based on the covariance matrix.
  - Mplus, for example, does this by default.
- This is for most purposes a good thing.
- However, it means that the values for our factor loadings, etc. are difficult to interpret.
  - Just like covariances!
- We can get standardized factor loadings (and other parameters) by analyzing the correlation matrix instead.
  - Or requesting a standardized solution in Mplus.
- When we do that, our factor loadings are now on a scale from 0 to 1.
  - By convention, we like them to be above .30.
  - Mplus uses standard errors to test "significance" of factor loadings – standardized or not.

Homogeneity

- Single factor (**homogeneous**) scales are often desirable. Why?
- At the very least, individual factor scales ought to be homogeneous.
  - Otherwise, they measure more than one factor!
- Recall that Cronbach's alpha is **not** a good indicator of homogeneity.
  - In fact, R & M argue that alpha is not really a measure of **reliability** (% of variance due to true score) at all.

Omega

- If we can divide each item into true score and error:
  - $X_j = \mu_j + \lambda_j F + E_j$
- And our test score is the sum of all the X_j:
  - $Y = \mu_1 + \lambda_1 F + E_1 + \mu_2 + \lambda_2 F + E_2 \ldots$
  - $Y = \sum \mu_j + (\sum \lambda_j)F + \sum E_j$
- We can use our formula for the variance of a composite to get:
  - $\sigma_Y = Var[(\lambda_1 + \lambda_2 + \ldots + \lambda_j)F] + Var(E_1 + E_2 + \ldots + E_j)$
- And then our reliability formula is:
  - $\dfrac{Var[(\lambda_1 + \lambda_2 + \ldots + \lambda_j)F}{Var[(\lambda_1 + \lambda_2 + \ldots + \lambda_j)F] + Var(E_1 + E_2 + \ldots + E_j)}$

Omega, cont.

- This reduces to:
  - $$\frac{(\lambda_1 + \lambda_2 + \ldots + \lambda_j)^2}{(\lambda_1 + \lambda_2 + \ldots + \lambda_j)^2 + (\psi_1 + \psi_2 + \ldots + \psi_j)}$$
  - Sum the factor loadings, then square them.
  - Divide by (the squared sum of the factor loadings + the sum of the unique variances).
- This yields a reliability coefficient based on the factor analysis information.
- McDonald (1999) calls it *omega*.

Benefits of Omega

- For R & M, omega *is* reliability.
  - Direct assessment of the ratio of common variance to total variance.
  - Avoids the potential problems with correlated errors, true-score equivalence, etc. that plague alpha.
- Omega *does* give you information about homogeneity, because you have to test homogeneity in order to calculate omega.
  - Need to assess model fit *first* – omega is meaningless if the data don't fit a single-factor model.

Multiple Factor Models

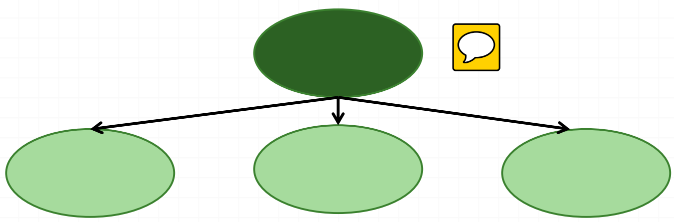
- We can extend the logic of CFA to models with more than one factor.
- Typically, we expect items to have an *independent clusters basis*.
  - Each item loads onto only one factor.
  - $X_1 = a_{11} f_1 + 0 f_2 + u_1$
  - $X_2 = a_{12} f_1 + 0 f_2 + u_2$
  - $X_3 = a_{13} f_1 + 0 f_2 + u_3$
  - $X_4 = 0 f_1 + a_{21} f_2 + u_4$
  - $X_5 = 0 f_1 + a_{22} f_2 + u_5$
  - $X_6 = 0 f_1 + a_{23} f_2 + u_6$
- This does not mean the factors are uncorrelated!

Comparing Multiple Factor Models

- As we discussed last time… the more complex model always fits better than the simpler model.
  - Unless your more complex model is badly misspecified.
- The question is *how much better*.
  - Use the chi-square difference test.

Hierarchical Models

- Sometimes, we might find evidence for multiple factors but our multiple factors are highly correlated.
- Sometimes, it makes sense to think of our factors as facets of an overarching construct rather than totally distinct constructs.



Hierarchical Models

- We can specify our factor model to reflect this structure:
  - $X_1 = a_{11}G + a_{21}f_1 + \ 0f_2 + u_1$
  - $X_2 = a_{12}G + a_{21}f_1 + \ 0f_2 + u_2$
  - $X_3 = a_{13}G + a_{21}f_1 + \ 0f_2 + u_3$
  - $X_4 = a_{14}G + \ 0f_1 + a_{31}f_2 + u_4$
  - $X_5 = a_{15}G + \ 0f_1 + a_{32}f_2 + u_5$
  - $X_6 = a_{16}G + \ 0f_1 + a_{33}f_2 + u_6$
- But there’s a catch…

Hierarchical Models and Fit

- The fit of a hierarchical model will be **identical** to the fit of a non-hierarchical model with correlated factors.
  - As long as the subfactors are the same, of course (same items on each factor).
  - Why?
- In a correlated-factors model, the relationships among factors are accounted for by the correlations. In a hierarchical model, they are accounted for by the general factor.
  - Two ways to describe the same relationship.
  - In fact, we can estimate those general factor loadings from the non-hierarchical model:
    - Loading on G = loading on F x $\sqrt{\text{factor correlation}}$

Hierarchical Models and Fit

- So how do you choose?
- Judgment!
  - Do all of the items have reasonably high (> .30) standardized loadings on the general factor?
  - Do all of the items have reasonably high (> .30) standardized loadings on their respective subfactors?
  - Which factor loadings are higher?
  - What are you using the scale for? Is a general score more valuable, or is it important to distinguish subdimensions?

Questions?

For next time: Classical Item Analysis
Read: McDonald (1999) Ch. 11 on Canvas

Lab Friday: CFAs in Mplus
Midterm due tonight at midnight!