Confirmatory Factor Analysis

PSY 600K

Dr. Alyssa Gibbons

March 20, 2018

Agenda

- Survey is live!
- How CFA is different from EFA.
- Testing & comparing CFA models.
  - Model fit
    - The logic of fit indices.
    - Common fit indices and how to use them.
    - Using discrepancies / residuals to understand fit.
  - Model comparisons.
    - Free & fixed parameters.
    - Nested models
    - $\Delta\ \chi^2$
    - Testing CTT models

CFA and EFA

- In EFA, we did not specify any hypotheses about how the items should relate or how many underlying factors there should be.
  - We let the empirical relationships drive our judgments about which items were related.
- In CFA, we specify how many factors there should be and which items should be related to each factor.
  - We let prior information (theory, previous data) tell us which items should belong where.
  - The CFA model then tests how well our hypothesized model fits the data.

More technically…

- EFA is an *unrestricted* factor model.
  - We set no rules about which items belong to which factor, how many items load on each factor, how many factors an item loads on.
  - $X_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + u_1$
- CFA is a *restricted* model.
  - We are going to constrain some – actually a lot – of the possible parameters in this model to equal zero.
  - $X_1 = a_{11}f_1 + 0f_2 + \cdots + 0f_m + u_1$

Tradeoffs

- EFA models will always fit better than CFA models!
  - Allowing more parameters to vary produces better fit.
- But CFA models are more parsimonious.
  - Usually, each item loads on only one factor.
  - Much easier to interpret (and already based on some amount of theory).
- CFA resolves the indeterminacy problems of EFA.
- CFA is a stricter and cleaner test of our hypotheses.
  - Reduces capitalization on chance.

EFA then CFA?

- Many (e.g., R & M) recommend using an EFA followed by a CFA in scale development.
- If you are really uncertain about your factor structure, this may be reasonable…
- BUT you MUST have a separate sample!
  - Otherwise your CFA will fit too well… the EFA capitalizes on chance.
  - The point is to replicate your structure.
- If you already have a reasonable theory about your items, there's no real need for the EFA step.
  - Often just introduces confusion.

So why *not* do an EFA?

- The premise of EFA is that we can "discover" the structure of the data.
  - But if we wrote the items… we put that structure there!
- If you cannot even venture a hypothesis about how many factors your construct has, and which items measure each one, then you should probably think more about your construct before you try to measure it.
  - It's OK to have **competing** hypotheses… but these are easier to test in CFA.
- EFA presumes you've identified all of the relevant items.
  - Counterexamples: Early Big Five studies, Thurstone's study of mental abilities.
  - But that's not typical of most EFA studies today.

Testing CFA Models

- We evaluate the quality CFA models based on their fit to the data.
- Models, by definition, are approximations. We want the most parsimonious model that still does a good job of approximating the actual data.
- Goodness of fit = how close is our model to the real data?
  - Technically: How close is the covariance matrix we get from our estimated parameters to the covariance matrix from the actual data?

Goodness of Fit

- How do we quantify this?
  - We can calculate a chi-square statistic comparing the predicted and observed covariance matrices.
  - A “significant” result indicates **misfit** – these matrices are not the same.
- But there’s a problem: χ^2 is **extremely** sensitive to sample size.
  - With a large enough sample, even small deviations produce a “significant” χ^2.
  - This is because no **model** is ever **true**! It’s an approximation…
  - McDonald & Ho (2002) report that only 5/41 SEM papers over a 3-year span had nonsignificant chi-squares.

Relative Fit Indices

- **Relative** fit indices compare the fit of our model to a null model.
  - Typically, the null is that all items are completely uncorrelated.
- There are a bunch of relative goodness-of-fit indices out there.
  - GFI, CFI, TLI, NFI, NNFI
  - Pretty much all of them are slight variations on this fundamental principle.
  - Different indices correct differently for bias, model complexity, etc.
  - **No** real consensus on which one is the best… and they don’t always agree.

Recommendations

- **CFI:** Comparative Fit Index (Bentler, 1990; McDonald & Marsh, 1990).
- **TLI:** Tucker-Lewis Index (Tucker & Lewis, 1975)
- Why? These are **unbiased** by sample size.
- To interpret:
  - Scaled from 0 to 1, with higher values indicating better fit.
  - Different guidelines for “good” and “acceptable” fit depending on who you ask.
    - > .90 is usually “acceptable” and > .95 is usually “good.”
- Also: **RMSEA:** Root Mean Squared Error of Approximation
  - Function of χ^2 and *df.*
  - “Good” is < .05; “acceptable” is either < .08 or < .10.

Even so…

- Even these indices may not always agree.
- For example, RMSEA tends to be inflated in models with few degrees of freedom.
- What should you do?
  - Report **multiple indices** (χ^2, CFI, TLI, RMSEA) so that the reader can judge whether the overall pattern of indices supports your conclusion (don’t cherry-pick the best index!).
  - Examine and report the discrepancies!

What's a discrepancy?

- If you literally subtract the observed correlation matrix from the predicted correlation matrix, the resulting matrix is called the standardized **discrepancy** or **residual** matrix:

- $$\begin{bmatrix} 1 & .35 & .38 \\ .35 & 1 & .40 \\ .38 & .40 & 1 \end{bmatrix} - \begin{bmatrix} 1 & .34 & .39 \\ .34 & 1 & .28 \\ .39 & .28 & 1 \end{bmatrix} = \begin{bmatrix} 0 & .01 & .01 \\ .01 & 0 & .12 \\ -.01 & .12 & 0 \end{bmatrix}$$

Interpreting Discrepancies

- Large discrepancies (absolute value > .10) tell you which specific **pairs of items** are causing the misfit in your model (McDonald, 1999).
  - Items that are more or less correlated than they ought to be.
  - Global fit indices (e.g., CFI) can hide this type of misfit.
- Examining your discrepancy matrix can help you identify and resolve misfit. McDonald & Ho (2002) recommend reporting the discrepancy matrix to help the reader judge for themselves.

Competing Models

- We said before that we don't just want **any** model that fits acceptably… we want the **most parsimonious** model that fits acceptably.
- More parsimonious = estimating fewer parameters.
  - Parameters we can estimate:
    - Factor loadings
    - Uniquenesses
    - Correlations between factors
- We can choose to **estimate** (free) or **constrain** (fix) every possible parameter.

Free & Fixed Parameters

- A complex model:
  - $X_1 = a_{11}f_1 + a_{12}f_2 + a_{13}f_3 + u_1$
  - $X_2 = a_{21}f_1 + a_{22}f_2 + a_{23}f_3 + u_2$
  - $X_3 = a_{31}f_1 + a_{32}f_2 + a_{33}f_3 + u_3$
  - All factors are correlated: $r_{12}, r_{13}, r_{23}.$
  - This model is **saturated:** we are estimating every parameter we can possibly estimate.
- A constrained version:
  - $X_1 = a_{11}f_1 + 0f_2 + 0f_3 + u_1$
  - $X_2 = a_{21}f_1 + 0f_2 + 0f_3 + u_2$
  - $X_3 = a_{31}f_1 + 0f_2 + 0f_3 + u_3$

Nested Models

- When you have 2 models, one of which can be written as a constrained version of the other, they are called **nested** models.
- You can statistically compare these models to determine whether adding the constraints "significantly" worsens the fit of the model.
  - The constrained model should always fit worse than the unconstrained model.

χ^2 Difference Test for Nested Models

- Formal test for the difference between χ^2 values for two nested models:
  - $\chi^2_{constrained} - \chi^2_{unconstrained} = \Delta\,\chi^2$
  - $df_{constrained} - df_{unconstrained} = \Delta\, df$
  - $\Delta\,\chi^2$ is distributed as a χ^2 with $\Delta\, df$ degrees of freedom.
- If this value is significant, we conclude that including the constraints makes the model worse (i.e., that the less constrained model is better).

Other Types of Constraints

- Another example – what new constraint have I added?
  - $X_1 = a_{11}f_1 + 0f_2 + 0f_3 + u_1$
  - $X_2 = a_{11}f_1 + 0f_2 + 0f_3 + u_2$
  - $X_3 = a_{11}f_1 + 0f_2 + 0f_3 + u_3$
- We can constrain parameters to be equal to one another.
- Special types of CTT models:
  - True-Score Equivalent Items = constrain factor loadings to be equal.
  - Parallel Items = constrain factor loadings **and** uniquenesses to be equal.
  - These are nested within our basic factor model – we can use the $\Delta\ \chi^2$ test to determine whether these more specific models hold.

Very Technical Issues

- Discrete data:
  - Most of the time, our Likert scales approximate continuous variables reasonably well.
  - When we have truly non-continuous data, we need more complex estimation methods.
  - R & M provide considerable (good) detail on this; this is beyond the scope of this class.
    - Mplus handles categorical data very well.

Questions?

For next time: Issues in CFA
Read: R & M 7.5 – 7.6

Remember that the midterm is due Thursday at midnight!