Model fit in CFA

PSYn5440 – Introduction to Factor Analysis

Week 13

CFA Review

- Think of CFA as a succession of three steps:
  - 1) Specify a model
  - 2) Fit the model to data (estimate the free parameters)
  - 3) Evaluate the result

CFA Review

1) Specify a model

- We know what this is – either specify the model through specification of the model matrices or through path diagrams

CFA Review

2) Fit the model to data

- Obtain estimates of the free parameters that minimize the discrepancy function

- I haven't mentioned one thing yet, but it's very important – do **not** carry out an EFA with a blind rotation to decide on the number of factors and the position of zero loadings and then fit an analogous CFA model using the **same data**.

CFA Review

2) Fit the model to data

- That would be capitalizing on chance – what you would be effectively doing is “confirming” a solution by CFA that was suggested by EFA.
- First of all, that’s kind of running in circles, second of all, it invalidates statistical tests and standard errors obtained by fitting the confirmatory model

CFA Review

2) Fit the model to data

- What you can do is exploring the factor structure using one half of the data (split randomly) and then “cross-validating” your hypothesized model using the other half of the data.
- This is an example of a good practice. You want to formulate a model that holds for the population, so being able to replicate the model on different data tells you good things about your model.

CFA Review

3) Evaluate the result

- That’s what we’ll be talking about today.

- Generally:

  - A) Evaluate the parameter estimates
  - B) Evaluate the overall model fit

Evaluating the result

- First, make sure the solution looks reasonable enough.
- Did the software produce any warnings?
- Examine the parameter estimates – the point estimates and the standard errors.

- Are the point estimates within their logical bounds? Look for correlations greater than 1, variances less than or equal to 0. These are signs that things are breaking down.

Evaluating the result

- Such “broken down” things are called *inadmissible* or *improper* solutions. You should be super-careful when interpreting such solutions, as *something* is most likely wrong in your data or with your model. Treat this as a warning and try to figure out what’s wrong.

Evaluating the result

- Even if the point estimates look reasonable, do they make sense to you?

- Were you expecting a positive correlation where a negative correlation is?
- Were you expecting a high loading where you got a small one?
- Were you expecting a negative loading on an MV, but got a positive one?

Evaluating the result

- Standard errors.
- Standard error is the estimate of the standard deviation of the parameter estimate over repeated sampling.

- How much would a parameter estimate vary over different samples?
- High value = unstable estimate
- Low value = stable estimate

Evaluating the result

- Treat the standard errors like you would treat them with any other parameter estimates – say, in linear regression.
- You can use the standard error of a point estimate to compute a t-statistic [point estimate / its standard error] to test difference from zero.
- Typically, the absolute value of these t-statistics will be much, much larger than 2

Evaluating the result

- Large standard errors can be an indication of trouble.
- If all standard errors are large, then your N is probably too small
- If only some standard errors are large while most are small, then you probably have some other problem that is more specific – investigate the respective MVs or LVs that are involved.

Model fit

- We’ll cover only a couple of fit indices, but there are many many maaaaaaaaaaaaaaaany more.

- Good fit doesn’t mean your model is “true” or “correct” or “best”. There are probably other models that fit just as well or even better (models with more free parameters, for instance)

- Does your model fit “well enough”, though?

Model fit

- As in EFA, one can perform a test of perfect fit using the likelihood ratio test statistic. But you know better than that.

- Traditionally, the χ^2/ degrees of freedom ratio was also used to assess fit. The problem is that there are no agreed-upon convention regarding its value. Some frequently used fit indices build upon this ratio (such as RMSEA or TLI) and so there is zero reason not to use them instead of this ratio.

Model fit

- We already covered RMSEA (Root Mean Square Error of Approximation) and TLI (Tucker-Lewis Index) before. They can be used in the same way as in EFA, and they have identical meaning.

- TLI is sometimes called NNFI (Non-normed Fit Index) in CFA context.

Model fit

- **CFI (Comparative Fit Index) is very similar in logic to the TLI:**

$$\frac{(\chi_0^2 - df_0) - (\chi_m^2 - df_m)}{(\chi_0^2 - df_0)}$$

- **This index, again, pits the proposed model (*m*) against the null model (*0*).**
- **CFI and TLI are highly correlated**
- **Same suggested "cut-offs" as for TLI (>.95 awesome, >.90 great)**

Model fit

- GFI (Goodness-of-fit index) and AGFI (Adjusted Goodness-of-fit index) are affected by sample size and have been shunned by the community. Don’t use them, don’t report them, ignore them.

Model fit

- Residual-based indices:

- The residual matrix is a good source of information on model fit (i.e., how well does the model reproduce the observed correlation / covariance matrix).

- Some fit indices directly work with the elements in the residual matrix.

Model fit

- RMSR (Root mean squared residual) – a zero indicates perfect fit

- SRMR (Standardized root mean [squared] residual) – again, a zero indicates perfect fit. Typically a value of less than .08 is considered good fit.

- Keep in mind that most residual-based fit indices incorporate no penalty for model complexity.