Fitting the Common Factor Model II

PSY544 – Introduction to Factor Analysis

Week 8

Maximum likelihood estimation

- In the preceding lectures, we covered the least-squares approach to estimating the model parameters (model matrices).

- An important property of the techniques previously covered is that they do not make any distributional assumptions about the manifest variables.

- Now, we will briefly cover another technique – maximum likelihood (ML) estimation.

Maximum likelihood estimation

- Maximum likelihood (ML) estimation does make distributional assumptions.

- These assumptions are about the normality of data. In other words, ML estimation assumes the MVs are normally distributed.

- ML estimation is not only used in FA, but pretty widely in statistics overall, so it comes in handy to have an idea about how it works. Right? ☺

Maximum likelihood estimation

- ML estimation, in general, can be described as follows:

- 1) You assume you have a random sample from some well-defined population.
- 2) You assume that the distribution of manifest variables has some particular form (in our case, multivariate normality)

- The likelihood function can be defined as follows:

L = Likelihood function = function(data, parameters)

Maximum likelihood estimation

- The likelihood function indicates the likelihood of the data, given model parameters.

- The principle goes as follows: Given the data we have, we obtain the values of parameters that maximize the likelihood function. These values of the parameters are the maximum likelihood estimates of the real (unknown) parameters.

- Let me illustrate with aliens.

Maximum likelihood estimation

- I'll spare you most of the math here, the important thing is that you understand MLE (maximum likelihood estimation) conceptually

- The actual values of the likelihood function tend to be very small, which can give computers hard time (due to rounding errors) and can be grossly ineffective

- For these reasons, we work with a function that is inversely related to the likelihood function – a function based on **-2 x log(likelihood)** (so we're really looking for the minimum of this function which corresponds to the ML maximum)

Maximum likelihood estimation

- So, the goal of ML estimation is to find such $\widehat{\Lambda}$ and $\widehat{D_\psi}$ so that the value of the -2 x log-likelihood function is minimized

- We can run into troubles with **local** minima / maxima – let me explain on board

- This function is again a discrepancy function – it is always larger than or equal to 0 and is zero if and only if the model-implied correlation matrix equals the sample correlation matrix

Maximum likelihood estimation

- The logic of ML estimation is very similar to that of (iterative) OLS:

- 1) Initial estimates of $\widehat{D}_\psi$ are obtained (by SMCs or other means)

- 2) A maximum likelihood estimate of Λ is obtained using eigenvalue-eigenvector decomposition, conditional on the estimated $\widehat{D}_\psi$

- 3) A model-implied reduced correlation matrix is obtained, which completes the first iteration

- 4) New iteration is initiated with most recent estimates of $\widehat{D}_\psi$

- 5) Iterations continue until convergence is achieved

Maximum likelihood estimation

- The convergence criterion is usually defined as the change in the discrepancy function values of two successive iterations. When the difference in these values drops below some pre-defined point, the iterations cease.

- Heywood cases can occur just like in the iterative least squares case.

Maximum likelihood estimation

- After the converge, the **likelihood ratio test statistic** is computed by multiplying the value of the discrepancy function by (N – 1).

- This statistic is used in assessing model fit and testing hypotheses about model fit (we'll get there shortly)

Summary

- We have described three different methods for fitting the common factor model to sample data:

- Principal factors with prior communality estimates (noniterative)

- Ordinary least squares (iterative principal factors)

- Maximum likelihood

- These are *methods* for fitting the *model* to data. Many more methods exist, but OLS and ML are commonly available.

Evaluating model fit

- Trying to fit a model to data is all nice and neat, but what is it good for without a way to evaluate how well the model fits?

- Models have an ultimate purpose – explanation, description and understanding (or prediction, but that's not very hot in psychology) through simplification.

- Models represent an attempt to simplify reality as much as we can so that we can understand it better. Simplify too much, and your model will not successfully represent reality. Simplify too little, and your model becomes unwieldy and begins to defy its purpose.

Evaluating model fit

- Models are **never true**, by definition (they're models!)

- Anyway, when evaluating model fit, we ask ourselves a bunch of questions

- How appropriate is the model for the data?

- How well does the model reconstruct the observed data?

- …

Test of perfect fit

- One of the earliest and still very much used ways to assess model fit

- It is a direct hypothesis test of our model

- $H_0: \boldsymbol{\Sigma} = \boldsymbol{\Lambda\Lambda'} + \boldsymbol{D_\psi}$ -- in other words, the model fits perfectly in the population with m factors

- $H_1: \boldsymbol{\Sigma}$ has no particular structure

Test of perfect fit

- Note that the null hypothesis H_0 has a different role from the one that is usual in research. We actually don't want to reject this null hypothesis – rejecting the null means rejecting the model. Failing to reject the null means that the model is plausible.

Test of perfect fit

- Under the null hypothesis, the likelihood ratio test statistic $(N - 1)F_{ML}$ [the value of the discrepancy function F_{ML} multiplied by the sample size minus one] has a χ^2 distribution with degrees of freedom equal to:

$$df = \frac{1}{2}p(p+1) - \left\{p + pm - \frac{1}{2}m(m-1)\right\}$$

$$= \frac{(p-m)^2 - (p+m)}{2}$$

DF intermezzo

$$df = \frac{1}{2}p(p+1) - \left\{p + pm - \frac{1}{2}m(m-1)\right\} = \frac{(p-m)^2 - (p+m)}{2}$$

- The number of degrees of freedom follows the same logic as usual – it is the number of unique data points at our disposal minus the number of model parameters that we estimate.

- In our case, the number of unique data points is the number of distinct elements in the covariance matrix, $\frac{1}{2}p(p+1)$ (the number of elements in the diagonal and one of the lower / upper triangles off the diagonal, since the matrix is symmetric)

DF intermezzo

$$df = \frac{1}{2}p(p+1) - \left\{ p + pm - \frac{1}{2}m(m-1) \right\} = \frac{(p-m)^2 - (p+m)}{2}$$

- The number of model parameters we are estimating is $\left\{ p + pm - \frac{1}{2}m(m-1) \right\}$:

  - pm factor loadings (p manifest variables times m factors)
  - p unique variances (one for each MV)
  - minus a part that corrects for rotational indeterminacy

DF intermezzo

- So, we have $\frac{1}{2}p(p+1)$ data points (distinct elements in covariance / correlation matrix), they represent „bits" of information for our model

- We spend those bits of information to „pay" for estimation of model parameters. The difference between the number of data points available and the parameters to be estimated is degrees of freedom (df)

- Using the same data, simpler models will have more degrees of freedom and complex models will have fewer.

Test of perfect fit

- Aaaaaanyway, back to the test of perfect fit:

- Under the null hypothesis, the likelihood ratio test statistic $(N-1)F_{ML}$ [the value of the discrepancy function F_{ML} multiplied by the sample size minus one] has a χ^2 distribution with degrees of freedom equal to:

$$df = \frac{1}{2}p(p+1) - \left\{ p + pm - \frac{1}{2}m(m-1) \right\} = \frac{(p-m)^2 - (p+m)}{2}$$

- ...so we can **test** the null hypothesis.

Test of perfect fit

- If $(N - 1)F_{ML}$ is significant, we reject the null hypothesis. In other words, we reject the model with *m* factors.

- If $(N - 1)F_{ML}$ is NOT significant, we failed to reject the null hypothesis. In other words, we failed to reject the notion that the model with *m* factors fits perfectly in the population.

- Seems like a nice, clear, beautiful way to assess the fit of our model, don't you think?

Test of perfect fit

- It has one tiny problem. It doesn't make sense.

- The problem is – did we ever believe H_0 to be true? We **know** it isn't.

- We are building a model – an approximation to reality. What sense does it make to think that an approximation to reality will be perfect? It's called an approximation for a reason! What's the point of testing something we know isn't true?

Test of perfect fit

- Moreover, the likelihood ratio test statistic is defined as $(N - 1)F_{ML}$ which follows the chi-square distribution if the model is correct.

- Even if the discrepancy (F_{ML}) is small, as N increases, so will the test statistic. A large enough N will inevitably result in a significant likelihood ratio test statistic, thus rejecting the model. With N large enough, even a well-fitting parsimonious model will be rejected.

- The only thing the test basically tells us is whether sample size is large enough to reject the null hypothesis.

Test of perfect fit

- We hope for a model that fits well-enough, not for a model that fits perfectly (if it does fit perfectly, something is probably wrong – like the model being overly complex and not parsimonious)

- However, folks still use this test and you will see it in almost EVERY paper that contains a factor analysis model. I myself still put it in papers, just because I don't want to deal with nagging reviewers.

- Be wiser than most FA users and don't pay attention to this test.

RMSEA

- RMSEA = Root Mean Square Error of Approximation
- Steiger & Lind, 1980; Browne & Cudeck, 1992
- The sample value of the discrepancy function, $\hat{F}$, is known. However, the population value, F_0, is not – we would only know what it is if we would fit a model to the population correlation or covariance matrix

- RMSEA = $\sqrt{\dfrac{F_0}{d}}$, where d is the degrees of freedom
- Measure of discrepancy per degree of freedom

RMSEA

- RMSEA = $\sqrt{\dfrac{F_0}{d}}$, where d is the degrees of freedom
- This is a population statistic, but we don't have population data. We would like to estimate RMSEA in a sample.

- It turns out that the best estimate of F_0 is the following:

$$\hat{F}_0 = \hat{F} - \frac{d}{N-1}$$

- However, if it turns out to be less than zero, the estimate is then exactly zero.

RMSEA

- So we substitute the estimate into the formula for RMSEA to obtain a point estimate of RMSEA: $\sqrt{\dfrac{\hat{F}_0}{d}}$

RMSEA

- Browne & Cudeck, 1992, provide the following guidelines for interpretation of RMSEA:

- < .05 -- close fit

- .05 - .08 -- good fit

- .08 - .10 -- acceptable fit

- > .10 -- unacceptable fit

- These numbers are guidelines, they should NOT be used as cutoffs (which is what everyone does, of course)

RMSEA

- Note that the formula for RMSEA contains degrees of freedom. RMSEA "prefers" simpler model over more complex models.

- Generally speaking, if the value of the discrepancy function would be the same for two models, one with $m = 2$ factors and one with $m = 3$ factors, RMSEA would favor the simpler model ($m = 2$ factors).

Confidence intervals for RMSEA

- One cool thing about RMSEA is that we know its theoretical distribution and thus we can calculate confidence intervals around the point estimate.

- The confidence intervals are provided by some software (CEFA, R…)

- Use the confidence intervals! They give you information that the point estimate does not (e.g., uncertainty about the point estimate)

- The point estimate might be, say, .06, but you might get different CIs:
  - (0.00; 0.14) – we don't know if the model fits great or not so much
  - (0.05; 0.07) – we know the model probably fits well

ML vs OLS

- We've learned about both ML and OLS. Remember, both are just different *methods* for fitting the *same model* to data.

- In other words, both are methods for estimating the model parameters. Typically, the estimates will not differ too much.

- Both are optimal in their own definition of optimality.