Vectors of Mind IV

All roads lead to CFA

Contents

- Deeper dive into restricted FA
  - Restrictions
  - Model identification
- Confirming and exploring with FA
- Model fit!
  - Local fit
  - Global fit
  - Alternative models
  - Other checks
- Advanced confirmatory methods in FA
  - Target rotations and GORIC
- Sins of FA + reporting standards

Unrestricted FA

- The model:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\phi\Lambda}' + \boldsymbol{D_\Psi}$$

- The goal:
  Find the best solution to reproduce $\boldsymbol{\Sigma}$ given pre-specified number of factors (i.e., order of $\boldsymbol{\Lambda}$ and $\boldsymbol{\phi}$)

- Restrictions:
  None, we just need to select one solution out of the infinite possibilities due to rotational indeterminancy

Restricted FA

- The model:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\phi}\boldsymbol{\Lambda}' + \boldsymbol{D_\Psi}$$

- The goal:
  Find the best solution to reproduce $\boldsymbol{\Sigma}$ given pre-specified $\boldsymbol{\Lambda}, \boldsymbol{\phi}, \boldsymbol{D_\Psi}$

- Restrictions:
  At least enough to achieve model identification (i.e., one unique solution)
  At most pre-specifying all of the parameters!

Unrestricted vs. restricted

- The underlying model is the same
- The difference is only in the way we restrict the values in the model matrices

Technical specifics of unrestricted FA

- Restrictions
- Correlated errors (residual covariances)
- Estimation
- Identification

Restrictions in restricted FA

- Elements in model matrices Λ, ϕ, D_Ψ can be:
- **Free** = freely estimated, unconstrained
- **Fixed** = pre-specified to a specific value (i.e., two factors are correlated r = .2)
- **Constrained** = e.g., an *equality constraint* (i.e., all the factor loadings must be equal)

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ \lambda_2 & 0 \\ \lambda_3 & 0 \\ \lambda_4 & 0 \\ \lambda_5 & 0 \\ 0 & \lambda_6 \\ 0 & \lambda_7 \\ 0 & \lambda_8 \\ 0 & \lambda_9 \\ 0 & \lambda_{10} \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$D_{\Psi} = \begin{bmatrix} \xi_1 & \xi_{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \xi_{21} & \xi_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \xi_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \xi_4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \xi_5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \xi_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \xi_7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \xi_{10} \end{bmatrix}$$

Restrictions in restricted FA

- Now you can model scenarios like:
  - All items have the same loading (tau equivalence)
  - A precisly hypothesized latent structure (we know the sizes of lambdas in advance)
  - Invariance (same factor loadings between two groups)

Residual covariances

- Since we can play around with the elements of any model matrix, we can free up the off-diagonal elements of $\boldsymbol{D_\Psi}$!

- Thanks to that, you can model the covariances between unique factors

- In other words, you can ask for a model that assumes that:

After controlling for the effect of the common factor, there is still some association left between a pair of items

Residual covariances

- Now you can model scenarios like:
  - Some items have similar format while others don't
  - Passing one item helps you pass another because you learn something other don't
  - Some items measure a specific facet of a construct
  - Some items overlap contextually

Estimation in restricted FA

- We need to estimate all parameters at once, so we can no longer rely on the eigen-decomposition of the reduced correlation matrix

- Hence, we can't use the principal factors approach but rather something from the LS or ML family

- But you already know these well!

Model identification

- Fancy term for *one unique solution*

- Model is *not identified* if you can find multiple solutions

- You already know this problem from high school maths when you were solving equations that had infinitely many solutions

Model identification rules

1. Latent variables must be **scaled** (either one loading per factor needs to be fixed to 1, or the variance of each factor needs to be fixed to 1)

2. At least zero **degrees of freedom**

3. At least **3 MVs per factor**

2. and 3. assume the simplest possible model, i.e., no residual covariances and/or crossloadings

Model identification vs. convergence

- Models that are not identified do not converge

- Recall the n-dimensional mountain you are climbing to find the parameter estimates

- If there are multiple peaks of the same height, the algorithm would just keep moving from one to the other, unable to decide when to stop

Differences between the two FAs

- …are only technical!
- Your goal determines whether you call it exploratory or confirmatory
- You use exploration when you want to use the data to generate hypotheses
  - „What latent structure might be behind these data?"
- You use confirmation when you want to test a hypothesis
  - „Which of these latent structures might be the closest to the data-generating process?"

Surprising uses of FAs

- You might want to model a complex latent structure full o cross-loadings –> restricted FA would fail -> do confirmatory unrestricted FA!

- You might want to tune a model to inform your theory-building, e.g., through residual associations -> you can't use unrestricted FA -> do exploratory restricted FA!

2 FAs, 2 approaches

	Restricted FA	**Unrestricted FA**
Confirmatory approach	Testing a hypothesized model	Target rotation
Exploratory approach	Tuning a model using modification indices / residual matrix	Blind rotation

Model fit

Model fit in one slide

- We observed some correlations between items
- Our model implies some correlations between items
- The difference = the misfit
- You can either look at all the differences (= local fit) or aggregate the differences in some way (= global fit)
- Fit indices like RMSEA are global fit measures
- Local fit is severly underused in papers
- Report both, use both

Model fit – a tutorial paper

- Assessing model fit is like building a court case for / against your model. You get some evidence and need to interpret it in relation to each other.

- All this is summarized in Anna's and mine upcoming tutorial paper
- Its almost done version shall be posted into Study Materials

- It's called: *Abandon All Thumbs Ye Who Model*

Observed vs. expected

- Remember that we are doing all this to reproduce a variance-covariance matrix:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\phi\Lambda'} + \boldsymbol{D_\Psi}$$

- But the relationship is imperfect as we don't know the true values of anything:

$$\boldsymbol{\Sigma} \approx \boldsymbol{\widehat{\Lambda}\widehat{\phi}\widehat{\Lambda}'} + \boldsymbol{\widehat{D_\Psi}}$$

Observed vs. expected

- But if we reconstruct the variance-covariance matrix from our estimates:

$$\widehat{\Sigma} = \widehat{\Lambda}\widehat{\phi}\widehat{\Lambda}' + \widehat{D_\Psi}$$

- We get the **model-implied variance-covariance matrix**

Observed vs. expected

- …and the distance between the observed Σ and the model-implied $\hat{\Sigma}$ variance-covariance matrix *is* model fit

- All the fit indices you know are simply different takes on this distance

- Henceforth, we shall cover some of them

Residual matrix

- Not to be confused with $\widehat{D_{\Psi}}$, the residual matrix is simply $\widehat{\Sigma} - \Sigma$

- In other words, for each item pair it tells you how off the model is from the observed variance-covariance matrix

- So we can start right here!

- Looking at the residual matrix to assess model fit is called **local fit testing**

Residual matrix – how to find it

- In R, use lavaan::lavResiduals(model)
- In JASP, tick „Misfit plot" in the Plots menu

	x1	x2	x3	x4	x5	x6	x7	x8	x9
x1	0								
x2	.2	0							
x3	.34	.29	0						
x4	0	.03	.03	0					
x5	.07	.05	.11	.02	0				
x6	.01	.01	.01	.01	.02	0			
x7	.01	.12	.03	.02	.05	.03	0		
x8	.14	.05	.14	.06	.03	.02	.45	0	
x9	.26	.14	.26	.05	.03	.04	.29	.39	0

Residual matrix – how to interpret it

- Most software gives you correlations in the residual matrix (although you can ask for covariances as well)
- **But these correlations should be interpreted as partial correlations**
- **Partial correlations are generally smaller than standard ones**
- So you should consider residuals around ca .25 to be noteworthy while .40 and more are huge
- A good model should have some local misfit distributed evenly

Local fit testing – recap

- Residual matrix shows how much the model misrepresents the relationship of each item pair

- Residuals imply dependence of the item pair that remains even after controlling for their common cause

- **ALWAYS consider and report the residual matrix**

- The elements of the residual matrix are *partial correlations*

- You want *some* misfit that is distributed *evenly*

SRMR

- The residual matrix can be easily summarized into one number
- That's the **Standardized Root Mean Residual** (SRMR)
- In other words, SRMR is the average correlation residual of your model:

$$SRMR = \sqrt{\frac{\sum(Expected - Observed)^2}{Number\ of\ elements}}$$

SRMR

- Upsides:
  - Easy to interpret
  - Not influenced by anything but the residuals

- Downsides:
  - May be inflated in small samples due to sampling error
  - Not influenced by anything but the residuals
  - Like all averages, inflated by one huge residual

Bottom line: Always report, use along with others

Chi squared

- Where do these come from? The discrepancy function!

- Remember when we covered estimation and said that it's basically about defining some quantity and trying to minimize it

- That's called **the discrepancy function**

Chi squared

- Each estimation technique defines „discrepancy" differently, hence you get different expressions for the discrepancy function you are trying to minimize

- For ML, it's:

Chi squared

- This function has a certain value that doesn't have much meaning
- But!

Chi squared

- This function has a certain value that doesn't have much meaning

- But!

- If you multiply the ML version by N-1 you get a statistic that's approximately chi-squared distributed with the degrees of freedom of your model

Test of perfect fit

- Testing (i.e., getting the p-value) of this chi-square gets you the **test of perfect fit**

- The null hypothesis tested is that the model fits perfectly in population so **you want a high p-value**

- **Unfortunately, this is basically just a test of the sample size**

- Bottom line: **Report, but do not rely on, unless p-value high.**

Degrees of freedom

1. Measure of your courage

2. The expected value of a chi-squared distribution

3. The difference between the number of observed pieces of information and the number of parameters

Degrees of freedom

- In FA, we model the variance-covariance matrix so we have the information from the diagonal (the variances) and from one of the triangles (the covariances)

- So the number of pieces of information is:

$$df_{max} = k(k+1)/2$$

where k is the number of MVs

Degrees of freedom

- Each freely estimated parameter costs us one degree of freedom, so:

$$df = df_{max} - number\ of\ estimated\ parameters$$

- Every freely estimated factor loading, factor correlation or unique (co)variance costs one df

Degrees of freedom

- Why are they a measure of your courage?

- Because each freely estimated parameter brings your model closer to the data

- Maximum parameters freely estimated => df = 0

- No parameters freely estimated => the most stringent test of your theory

Degrees of freedom

- Dfs represent the potential for misfit

- If you have high df model that fits well, the model is great!

- If you have low df model that fits well, meh!

Degrees of freedom

- Imagine there are two persons and you know one of their IQ scores:

$$X1 = 105$$

What values can X2 (the IQ scores of the second person) have?

Degrees of freedom

- Imagine there are two persons and you know one of their IQ scores:

$$X1 = 105$$

What values can X2 (the IQ scores of the second person) have?

Any value!

Degrees of freedom

- Now, imagine there are, again, two persons and you know one of their IQ scores:

$$X1 = 105$$

And the mean of their IQs:

$$M = 95$$

What values can X2 (the IQ scores of the second person) have?

Degrees of freedom

- Now, imagine there are, again, two persons and you know one of their IQ scores:

$$X1 = 105$$

And the mean of their IQs:

$$M = 100$$

What values can X2 (the IQ scores of the second person) have?

Only X2 = 95! Otherwise, their mean cannot be 100 or X1 cannot be 105

Degrees of freedom

- One could say that with N = 2, we only have one degree of freedom after we compute the mean because we lose one piece of information by estimating the mean

- That's why we do the N-1 correction so often (e.g., when computing variance), because since we already use the information from the mean, one of the data points is determined by it

Parsimony ratio

- Dividing the number of *df*s of your model with the maximum number of *df*s gives you the **parsimony ratio**
- It tells you the percentage of *df*s you retained

$$PR = \frac{df}{df_{max}}$$
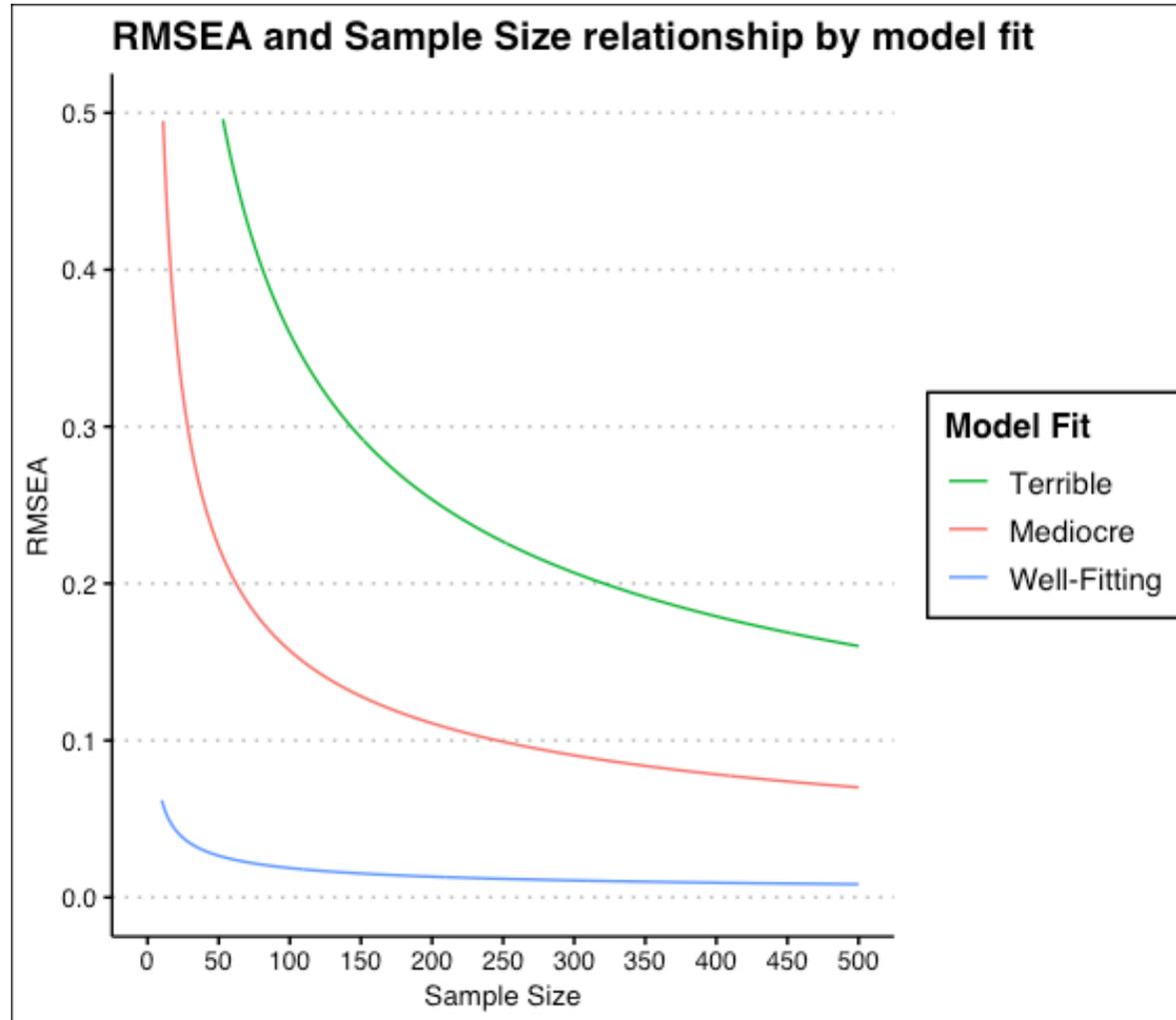
RMSEA

- Root Mean Square Error of Approximation

$$RMSEA = \sqrt{\frac{\chi^2 - df}{[df(N-1)]}}$$

- Basically, you look at how far the chi-squared is from its expected value and standardize the difference
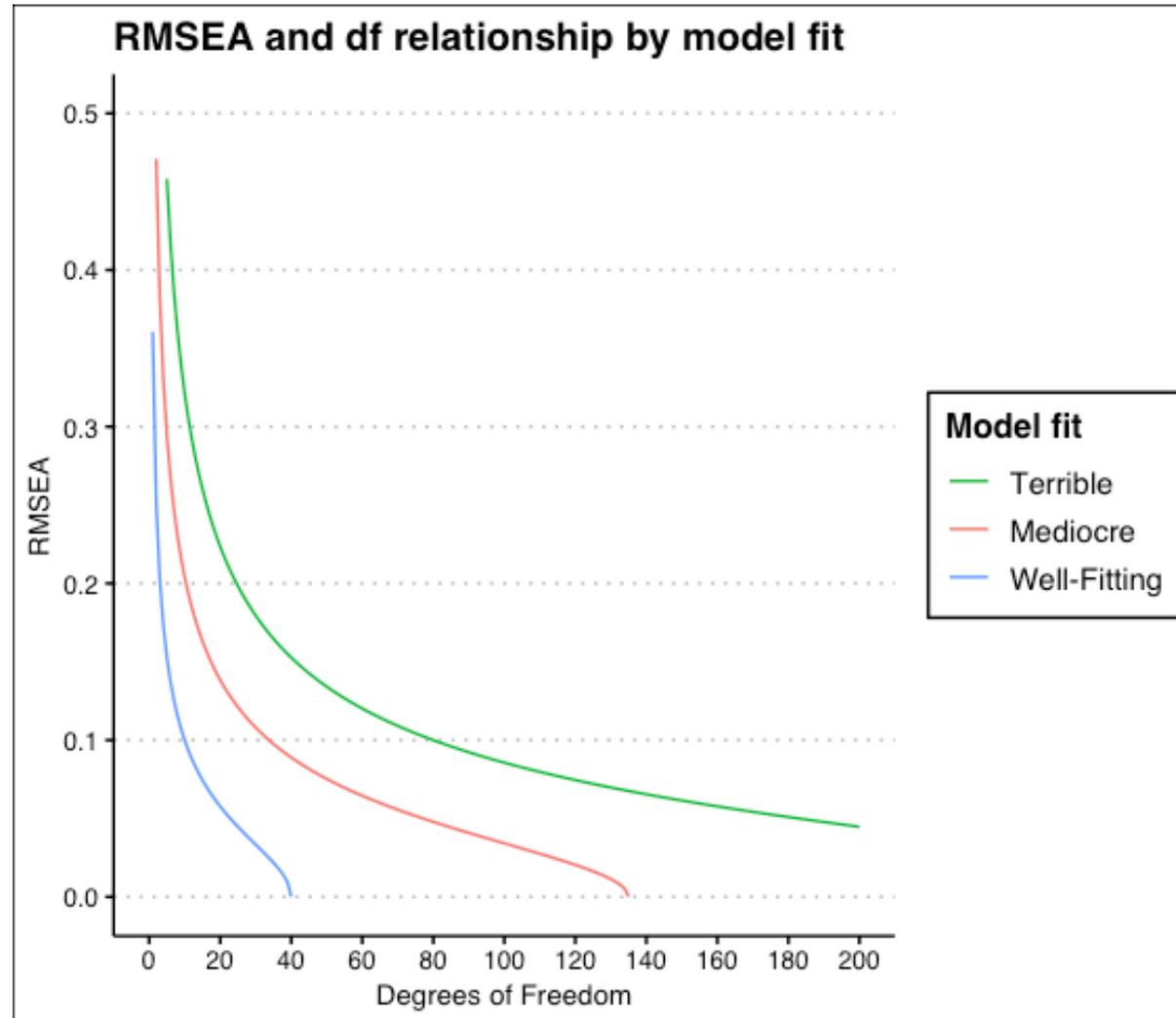- Consequently, RMSEA cares about **misfit, parsimony and sample size**

RMSEA
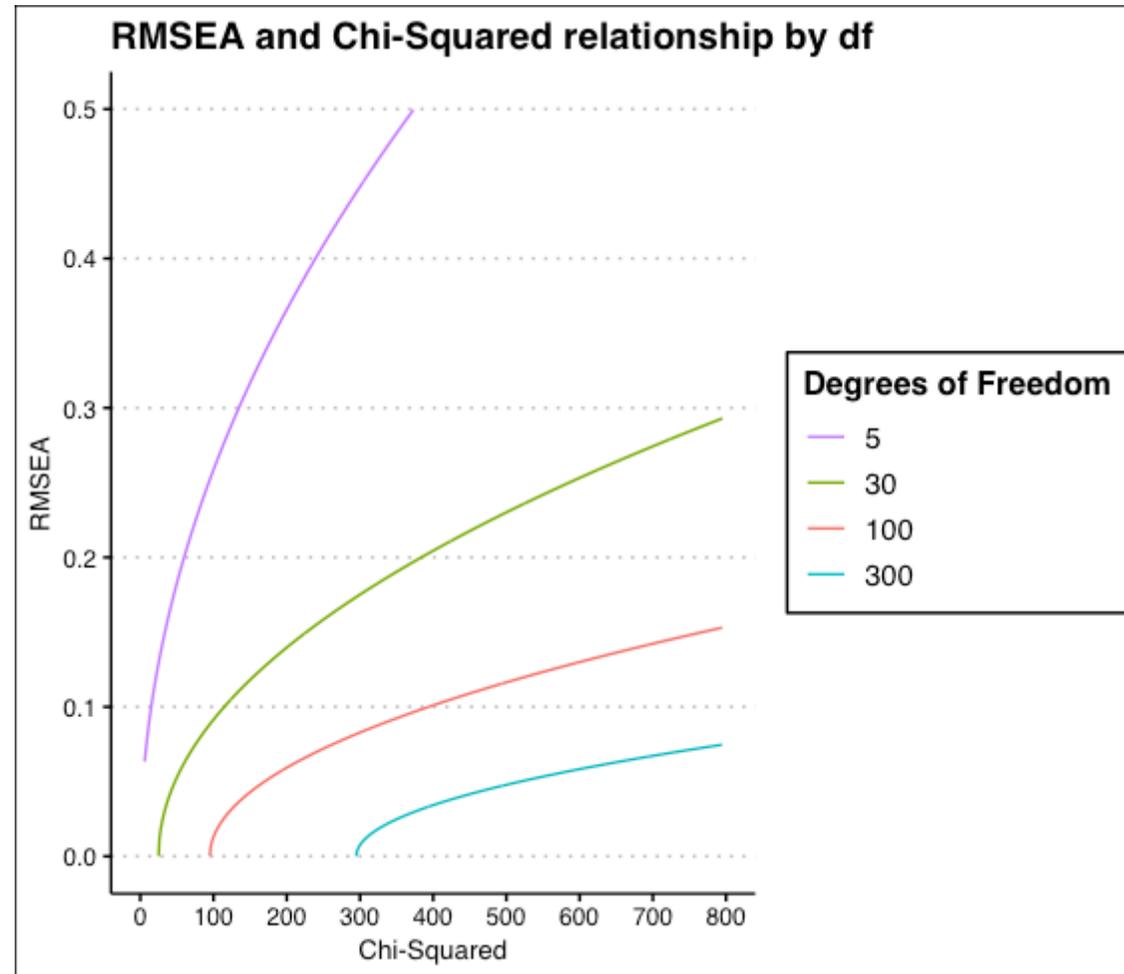
RMSEA

RMSEA

RMSEA – recap

- Consolidates more information about the model into one number
- Can compute confidence intervals!
- Gold standard of the industry
- May fail when df and N too low

- Bottom line: **Always report (with a CI) but interpret in relation to df and N**

TLI

- "Incremental" index of fit
- Compares your model to the best possible model and places it on a scale from the worst to the best model

$$TLI = \frac{\dfrac{\chi^2_{null}}{df_{null}} - \dfrac{\chi^2}{df}}{\dfrac{\chi^2_{null}}{df_{null}} - 1}$$

TLI

- Or figuratively:

$$TLI = \frac{worst\ possible\ model - proposed\ model}{worst\ possible\ model - best\ possible\ model}$$

TLI

- But when the null model is alright, TLI is unreliable
- That's why it's a good idea to calculate the RMSEA for the null model and checking if it's high enough (= the null is bad)
- Kenny says that null RMSEA = .158 implies that TLI can never by high

TLI – recap

- TLI tells you how close your model is to the perfect model
- It doesn't work if the scale from worst to best is broken, i.e., when the null model fits well
- You should report RMSEA of the null model to make sure this isn't the case

- Bottom line: **Use TLI and report it. But be wary of the null model.**

Model comparison

- Your model is only good if it beats its alternatives
- For example: If you want to support a unidimensional model, you need to show it's not too bad compared to the three-dimensional model

- It's always going to be worse (because of dfs!) but maybe not by a lot

Model comparison

- When comparing models against each other, you can:

1. Judge their absolute fit, i.e., everything we covered so far, and

2. Compare them directly using AIC or LRT

Likelihood Ratio Test

- Gives you a significance test of the chi-square difference between to models
- In other words: Is the reduction in misfit (chisquare) worth this loss of dfs?

- In R, it's the *anova* function

AIC

- An „information criterion" that considers chi-square along with parsimony (degrees of freedom)
- Interpretation makes sense only when comparing models with same variables and data
- It's better to use the corrected version AICc

Other checks

- Are the estimates within reasonable bounds? (variances, correlations)
- Do the estimates make sense? (negative loading on a reverse coded item)
- Are standard errors low? High? Or a mixed bag?

Model fit workflow

1. Note the degrees of freedom
2. Glance over the estimates to detect weird values (e.g., negative variances)
3. Check out SRMR to get an idea of the residual matrix
4. Compare SRMR to RMSEA and TLI – do they tell the same story?
5. Inspect the residual matrix
6. Pit the model against alternatives
7. Make your judgement!

Advanced confirmatory techniques

Blind rotations

- Taking advantage of the rotational indeterminancy is called **rotation**
- Algebraically, it means choosing an orthogonal transformation matrix **T** that satisfies certain criteria that help achieve **simple structure**
- These can be general (or **blind**):
  - Maximize the sum of the variances of squared loadings (that's varimax!)
  - Like varimax but correlated factors (that's oblimin!)

Target rotations

- …or you can tailor them specifically for your dataset

- Then they are called **target** or, poetically, **Procrustean**

- As I understand it, you can either:
  - rotate to a pre-specified simple structure or
  - rotate to another loading matrix (so you can compare them)

- So you can actually do a confirmatory analysis with the „exploratory" factor analysis, wow!

quartimin	oblique	
targetT	orthogonal	target rotation
targetQ	oblique	target rotation
pstT	orthogonal	partially specified target rotation
pstQ	oblique	partially specified target rotation
oblimax	oblique	
entropy	orthogonal	minimum entropy
quartimax	orthogonal	
varimax	orthogonal	
simplimax	oblique	
bentlerT	orthogonal	Bentler's invariant pattern simplicity criterion
bentlerQ	oblique	Bentler's invariant pattern simplicity criterion
tandemI	orthogonal	Tandem Criterion
tandemII	orthogonal	Tandem Criterion
geominT	orthogonal	
geominQ	oblique	
cfT	orthogonal	Crawford-Ferguson family
cfQ	oblique	Crawford-Ferguson family
infomaxT	orthogonal	
infomaxQ	oblique	
mccammon	orthogonal	McCammon minimum entropy ratio
bifactorT	orthogonal	Jennrich and Bentler bifactor rotation
bifactorQ	oblique	Jennrich and Bentler biquartimin rotation

GPArotation package

GORIC

Seven Deadly Sins of FA

Sin I: Misreporting

- Many papers omit or even wrongly report:
    - Chi-squared statistics
    - Degrees of freedom
    - Residual matrix
    - Alternative model comparison
    - Confidence intervals for RMSEA
    - Estimation method
    - Missing data treatment
    - Model specification
    - Estimates (i.e., the contents of the model matrices)

Sin I: Misreporting

- How to overcome this sin?
  - When you read a paper, get into the habit of checking degrees of freedom by hand (it's easy!)
  - Doubt papers that omit important information, especially multiple pieces are missing
  - In your own work, be as transparent as possible
  - But on the other hand, identify what's important so that you don't flood the reader with technical jargon

Sin I example – my diploma thesis

Sin II: Skipping model fit assessment

- Many practitioners don't spend enough time assessing model fit in the manner we discussed before

- You should really look at all the evidence including the residual matrix and reconcile discrepancies between them before you declare your model well-fitting

Sin III: Messing up the theory

- This sin includes:
  - Doing PCA instead of FA
  - Doing FA instead of formative models
  - Doing FA instead of network models
  - Correctly choosing FA but omitting any reasoning for that

- All this stems from insufficient attention paid to the data-generating process
- And partly from the lack of statistical education
  - If all you know is FA, you will always use FA
  - If you don't understand FA, you will use default options (like PCA in SPSS)

Sin IV: Abusing modification indices

Sin V: Exploring and confirming the same data

Sin VI: Not doing sensitivity analyses

Reporting standards

- Unrestricted FA:
  https://journals.sagepub.com/doi/full/10.1177/0095798418771807

- Restricted FA: https://www.researchgate.net/profile/Rebecca-Purc-Stephenson/publication/24187223_Reporting_Practices_in_Confirmatory_Factor_Analysis_An_Overview_and_Some_Recommendations/links/02e7e524c2d7dda3f1000000/Reporting-Practices-in-Confirmatory-Factor-Analysis-A