

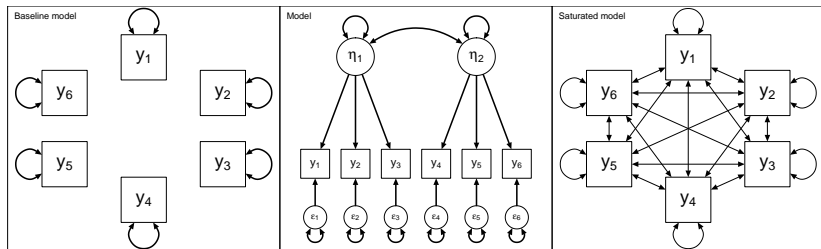
SEM 1: Confirmatory Factor Analysis

Week 2 - Fit indices

Sacha Epskamp

April 3, 2020

Fit indices



To assess fit, model can be compared to *baseline model* or *saturated model*.

- ▶ Baseline model: a model in which no items covary
- ▶ Saturated model: A perfectly fitting model ($T = 0$; $DF = 0$)

The χ^2 test compares the model (T_M) to the saturated model (should fit about the same). Many **fit indices** compare the model to the baseline model instead (T_B ; should fit much worse than tested model).

RMSEA

Root Mean Square Error of Approximation

$$RMSEA = \sqrt{\frac{T_M - DF_M}{(nDF_M)}}$$

```
Tm <- fitMeasures(fit)[['chisq']]
DFm <- fitMeasures(fit)['df']
sqrt((Tm - DFm)/(n * DFm))

##           df
## 0.09212148

fitMeasures(fit)[['rmsea']]

## [1] 0.09212148
```

RMSEA

RMSEA is a measure of absolute fit (no comparison model). It measures the amount of misfit per degrees of freedom. Smaller values indicate better fit.

Proposed benchmarks from a selection of papers:

- ▶ $< .05$ “very good fit” or “close fit”
- ▶ $.05 - .08$ “good fit” or “fair fit”
- ▶ $.08 - .1$ “mediocre fit” or “good”!
- ▶ $.05 - .08$ “good fit” or “fair fit”
- ▶ $> .10$ “poor or unacceptable”

RMSEA

RMSEA is one of the only fit indices for which the asymptotic sampling distribution is known, so we can make confidence intervals and conduct hypothesis tests about its population value.

- ▶ Test of Exact Fit
 - ▶ H_0 : RMSEA = 0 in the population
 - ▶ Equivalent to the significance test on the chi-square statistic
- ▶ Test of Close Fit (MacCallum et al., 1996)
 - ▶ H_0 : Null hypothesis: RMSEA < RMSEA_{good} in the population.
 - ▶ RMSEA_{good} is some acceptable value of RMSEA (in lavaan: 0.05)
- ▶ Test of Not-Close Fit (MacCallum et al., 1996)
 - ▶ H_0 : Null hypothesis: RMSEA > RMSEA_{bad} in the population.
 - ▶ RMSEA_{bad} is some unacceptable value of RMSEA (e.g., 0.08)

RMSEA test of close fit

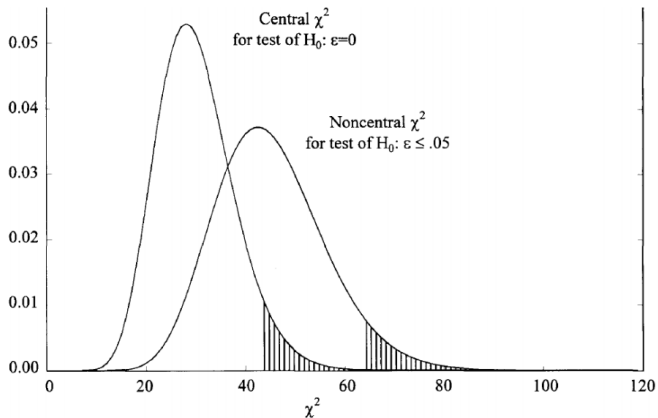


Figure 1. Illustration of difference in critical values between central and noncentral χ^2 distributions.

MacCallum, Browne, & Sugawara (1996)

RMSEA test of close fit

Given some boundary RMSEA_{good} , compute non-centrality parameter:

$$\lambda_c = \text{RMSEA}_{good}^2 \times n \times \text{DF}_M$$

Compute a one-tailed test with T , now using non-central distribution $\chi_2(\text{DF}_M, \lambda_c)$.

```
lambda_c <- 0.05^2 * n * DFm
pchisq(Tm, DFm, lambda_c, lower.tail = FALSE)

## [1] 0.0006612368

fitMeasures(fit)['rmsea.pvalue']

## rmsea.pvalue
## 0.0006612368
```

We reject the hypothesis the model fits well (RMSEA smaller than 0.05), this is **not good** (but also not very bad, as RMSEA around 0.05 – 0.07 is still acceptable)!

RMSEA test of not-close fit

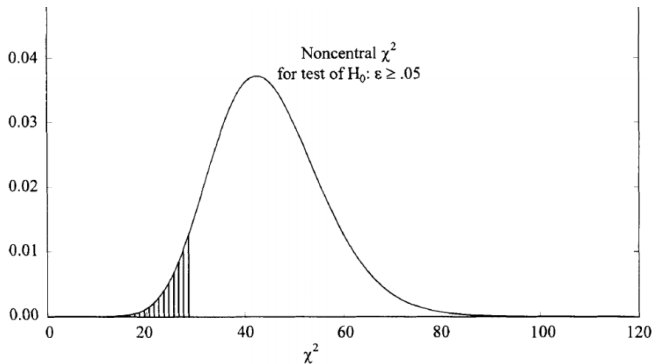


Figure 2. Illustration of critical value of noncentral χ^2 distribution for testing hypothesis of not-close fit.

MacCallum, Browne, & Sugawara (1996)

RMSEA test of not-close fit

Given some boundary RMSEA_{bad} , compute non-centrality parameter:

$$\lambda_c = \text{RMSEA}_{bad}^2 \times n \times \text{DF}_M$$

Compute a one-tailed test with T , using the **lower tail** of non-central distribution $\chi_2(\text{DF}_M, \lambda_c)$.

```
lambda_c <- 0.08^2 * n * DFm
pchisq(Tm, DFm, lambda_c, lower.tail = TRUE)

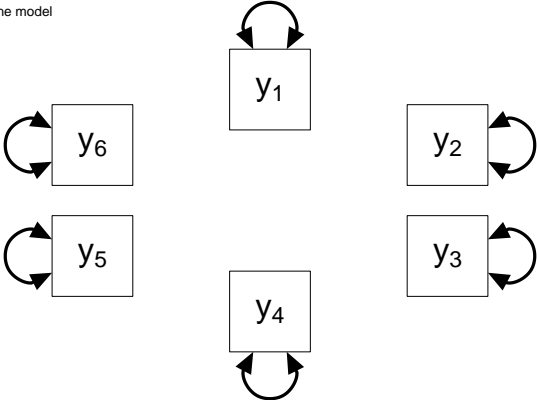
## [1] 0.8395529
```

We can **not** reject the hypothesis the model fits poorly (RMSEA larger than 0.08), this is **bad**!

Incremental fit indices

Incremental fit indices compare the model to the baseline model, which only estimates variances:

Baseline model



Incremental fit indices

```
Tb <- fitMeasures(fit)['baseline.chisq']  
Tb  
  
## baseline.chisq  
##      918.8516  
  
DFb <- fitMeasures(fit)['baseline.df']  
DFb  
  
## baseline.df  
##      36
```

How much better does the model fit than the worst possible model?



Incremental fit indices

T_M , df_M : Model test statistic and DF; T_B , df_B : Baseline model test statistic and DF.

$$NFI = \frac{T_B - T_M}{T_B}$$

% change in the test statistic, ranges from 0-1. Tends to be higher with larger N.

Normed Fit Index,
Bentler & Bonnett (1980)

$$TLI = \frac{T_B - \frac{df_B}{df_M} T_M}{T_B}$$

NFI + reward for parsimonious models. "non-normed" = can take values higher than 1.

Tucker-Lewis Index (TLI),
Tucker & Lewis (1973)
Non-Normed Fit Index (NNFI),
Bentler & Bonnett (1980)

$$RFI = \frac{\frac{T_B}{df_B} - \frac{T_M}{df_M}}{T_B / df_B}$$

% change in the test statistic relative to its df

Relative Fit Index
(Bollen, 1986)

$$IFI = \frac{T_B - T_M}{T_B - df_M}$$

IFI is less sensitive to sample size than NFI, NNFI, RFI

Incremental Fit Index
(Bollen, 1989)

(higher is better)

Incremental fit indices

T_M , df_M : Model test statistic and DF; T_B , df_B : Baseline model test statistic and DF.

$$RNI = \frac{(T_B - df_B) - (T_M - df_M)}{T_B - df_B}$$

Nonnormed (can exceed 1)

Relative Noncentrality Index,
McDonald & Marsh, 1990

$$CFI = 1 - \frac{T_M - df_M}{T_B - df_B}$$

Constrained to 1 if $df_M > T_M$

Comparative Fit Index,
Bentler, 1990

In Practice:

CFI, TLI/NNFI are most commonly reported incremental fit indices. .95 is often used as a cutoff rule-of-thumb for “good fit”, and .90 for “acceptable fit” though these cutoffs do not have much empirical support

Goodness of Fit (GFI) and Adjusted GFI

- ▶ GFI and AGFI are analogous to R^2 and adjusted R^2 in regression
 - ▶ R^2 estimates the proportion of variance in Y that is accounted for by the regression model.
 - ▶ GFI estimates the proportion of variance in the sample covariance matrix \mathbf{S} that is accounted for by the model structure Σ
- ▶ Both GFI and AGFI can take values from 0 to 1; higher is better.
- ▶ Rule of thumb: $> .90$ is acceptable fit.
- ▶ GFI and AGFI tend to be underestimated in small samples.

SRMR

- ▶ The largest residual correlation, or list of 5 largest residuals, is very useful for identifying why/how the model does not fit.
- ▶ Only look at these if the test of exact fit is significant - if not, the residuals are within the range of sampling error and are most likely noise.
- ▶ SRMR is the average of the squared values in the residual correlation matrix. It has been suggested that SRMR should be less than .05 or definitely less than .08. It is less informative than just looking for the biggest residuals!

Residuals

```
residuals(fit)$cov
```

```
##      x1      x2      x3      x4      x5      x6      x7      x8      x9
## x1  0.000
## x2 -0.041  0.000
## x3 -0.010  0.124  0.000
## x4  0.097 -0.017 -0.090  0.000
## x5 -0.014 -0.040 -0.219  0.008  0.000
## x6  0.077  0.038 -0.032 -0.012  0.005  0.000
## x7 -0.177 -0.242 -0.103  0.046 -0.050 -0.017  0.000
## x8 -0.046 -0.062 -0.013 -0.079 -0.047 -0.024  0.082  0.000
## x9  0.175  0.087  0.167  0.056  0.086  0.062 -0.042 -0.032  0.000
```


Strategies for Assessing Fit

- ▶ Always report the chi-square test statistic. You may argue that it is too sensitive to minor misspecifications because you have a large sample size, but report it anyway!
- ▶ Report several indices (e.g., RMSEA, CFI, RNI, TLI)
- ▶ The RMSEA tests of close and not-close fit can be a good index of power (i.e., if neither are significant, you may lack power to detect misspecifications)
- ▶ Try to make a holistic judgment based on a set of fit indices
- ▶ It goes without saying but... don't cherry pick the indices that make your model look good :)