# Evaluation of model fit

# Outline

- Some basics
- Significance testing
- Descriptive indices
- Other aspects of model evaluation

# Some basics

- After having estimated the indices and parameters it needs to be checked whether the model explains the data well.
- There are a number of fit indices (statistics) that are provided by the software programs for latent variable modeling.
- It is „a number“ of fit indices because there is no individual index that meets all the demands to fit evaluation.

# Some basics

- The indices are subdivided in …

| Significance tests | Descriptive indices … | | |
|---|---|---|---|
| | … referring to the structure of the complete model | … performing a comparison of models | … evaluating model parsimony |
| 1. $\chi 2$ statistic | 3. RMSEA | 6. NFI | 8. CFI/GFI |
| 2. $\chi 2$ difference test | 4. RMR | 7. NNFI | 9. PGFI/PNFI |
| 3. CFI difference test | 5. SRMR | 8. CFI/GFI | 10. AIC |

# Outline

- Some basics
- **Significance testing**
- Descriptive statistics
- Other aspects of model evaluation

# Significance testing

- $\chi^2$ **test statistic**

Check of the null hypothesis $H_o$: $S = \Sigma(\theta)$ Empirical matrix

Against the alternative hypothesis $H_1$: $S \neq \Sigma(\theta)$ Model-implied matrix

Please consider the changed logic of goodness-of-fit tests (Gof tests)!

→ Since it is a Gof test, the investigations aims at an insignificant $\chi^2$- value!

→ This means that this research interest opposes the research interest of inferential statistics!

- i.e. a p( $\chi^2$) that signifies a good model fit must be larger than .05.

# Significance testing

- $\chi^2$ **test statistic**

Check of the null hypothesis H$_o$: $S = \sum(\theta)$

Against the alternative hypothesis H$_1$: $S \neq \sum(\theta)$

→ Since it is a Gof test, the investigations aims at an insignificant $\chi^2$- value!

→ This means that research interest opposes the research interest of inferential statistics!

- Problems:

→ Distribution: … should be multivariate normal

→ Sample: an increasing sample size leads to increasing $\chi^2$ - values!

(problem) in very large samples $\chi^2$ is always significant!

→ Model complexity: adding more parameter to a model causes a decrease of the $\chi^2$- Wert!

# Significance testing

- $\chi^2$ **test statistic**

Check of the null hypothesis H$_o$: $S = \sum(\theta)$

Against the alternative hypothesis H$_1$: $S \neq \sum(\theta)$

→ Since it is a Gof test, the investigations aims at an insignificant $\chi^2$- value!

→ This means that research interest opposes the research interest of inferential statistics!

- Further information:

→ Because of the dependency on sample size, the probability of the $\chi^2$- value is **rarely used** in the evaluation of a model.

→ The $\chi^2$- value is an important ingredient of many descriptive statistics!

→ It is used as „normed $\chi^2$" genutzt: **normed** $\chi^2 = \chi^2$ **/ df**

with normed $\chi^2 \leq 2$ is expected for N about 200

$\leq 3$ is expected for $N \leq 500$

# Significance testing

## $\chi^2$ difference test

Checks the difference between two ***nested*** models for significance :

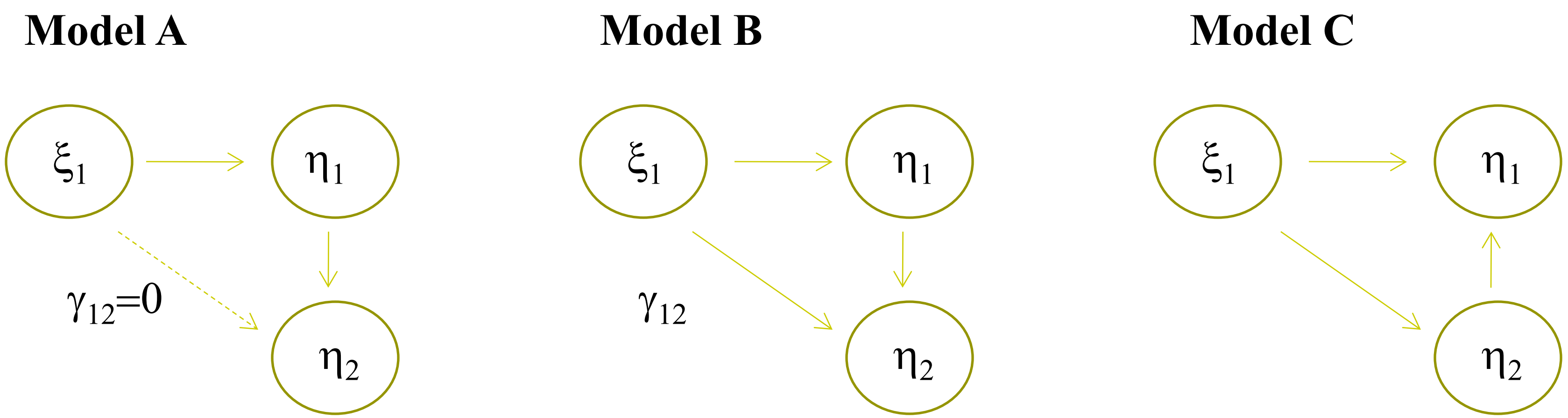
Modell A $\longrightarrow \chi^2{}_{\mathrm{A}}(df_{\mathrm{A}})$

Modell B $\longrightarrow \chi^2{}_{\mathrm{B}}(df_{\mathrm{B}})$

$$\chi^2_{diff}\ \left(df_{diff}\right) = \chi^2_A\left(df_A\right) - \chi^2_B\left(df_B\right)$$

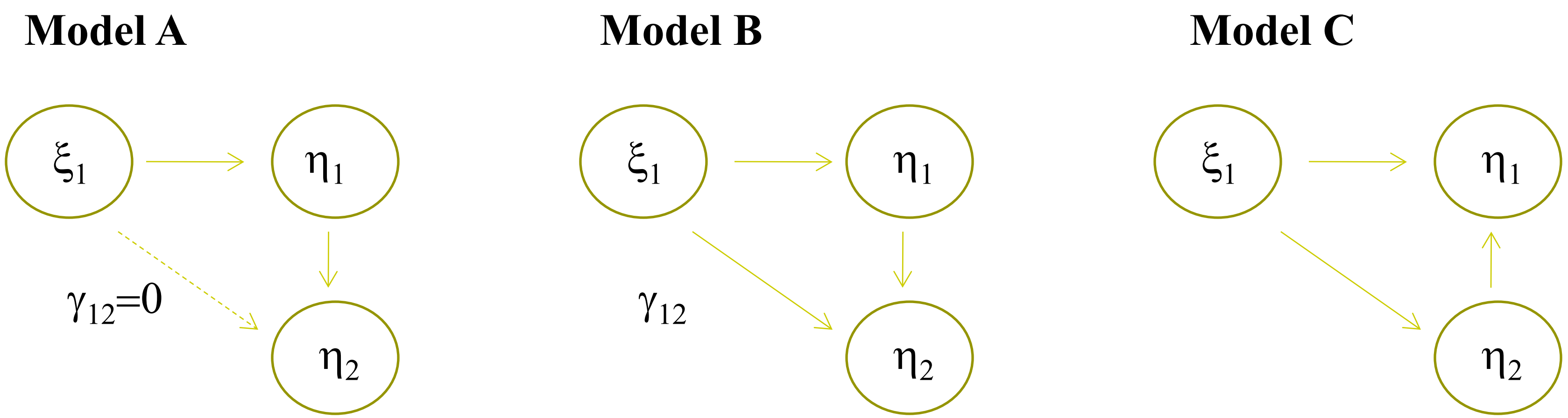
# Significance testing

## **Nesting of models:**



# Significance testing

## Nesting of models:




Model A and model B are nested since there is one more link or one less link:

- Model B is complete
- Model A is constrained

# Significance testing

## $\chi^2$ difference test

Checks the difference between two or more nested models:

$$\chi^2_{diff}\ (df_{diff}) = \chi^2_A\,(df_A) - \chi^2_B(df_B)$$

Model A = simpler

Model B = more complex

**Rule:**

Is the $\chi^2$ difference significant → model B is preferred

Is the $\chi^2$ difference not significant → model A is preferred

**Simplicity principle!**

# Significance testing

$\chi^2$ **difference test**

- … was a very important and often used test in the past
- … is no more used so often since simulation studies showed that it is not very accurate.
- … is **only** accepted for comparing **nested models**

# Significance testing

## CFI difference test

Checks the difference between any pair of models for significance :

$$\text{CFI difference} = \text{CFI}_A - \text{CFI}_B$$

Modell A $\longrightarrow$ $\text{CFI}_A$

Modell B $\longrightarrow$ $\text{CFI}_B$

- The exact error probability ***cannot*** be determined! It is only possible to find out whether the probability of the CFI difference is larger or smaller $p = .05$. The cutoff of the critical difference is **.01**.
- The meaning of CFI is explained in the section on descriptive statistics

Advantage: simulation studies showed that the .05 boundary is more often correct than in the $\chi^2$ difference test

# Outline

- Some basics
- Significance testing
- **Descriptive indices**
- Other aspects of model evaluation

# Descriptive indices

**In most descriptive indices the results are interpreted in considering three categories :**

→ „**good**“ model fit

→ „**acceptable**“ model fit

→ „**bad**“ model fit

# Descriptive indices

**Descriptive evaluation of the complete structure**

***Root Mean Square Residual (RMR) and Standardized RMR (SRMR)***

These „badness of fit measures“ are based on „fitted residuals“: S- Σ(θ)

→ desirable is a small RMR

→ „fitted residuals“ including RMR shows dependency on the scale of the data

Therefore, the standardized version that is **SRMR** is preferred.

→ good fit is indicated by: SRMR ≤ .08

→ acceptable fit is indicated by: 0.08 < SRMR ≤ 1.00

# Descriptive indices

**Descriptive evaluation of the complete structure**

***Root Mean Square Residual (RMR) and Standardized RMR (SRMR)***

These „dadness of fit measures“ are based on „fitted residuals“: S- $\Sigma(\theta)$

→ desirable is a small RMR

…. because of the better interpretability virtually only SRMR is used

→ good fit is indicated by: SRMR $\leq .08$

***Root Mean Square Error of Approximation (RMSEA)***

Indicates how well $\Sigma(\theta)$ approximates S

→ RMSEA should be as small as possible. Good model fit means:

$0 \leq$ RMSEA $\leq .05$ / **.06** ;

.05 / .06 $\leq$ RMSEA $\leq$ .08 indicates acceptable fit.

It was found to be rather independent of sample size!

# Descriptive indices

**Descriptive evaluation of model structure using comparisons of models**

**Comparative-Fit-Index (CFI) - important index!**

- Compares the specified model with the independence model (uncor.)
- Takes into consideration the degrees of freedom

$$\text{CFI} = 1 - \left[\frac{Model\ chisquare - Model\ df}{Independence\ model\ chisquare - Independence\ model\ df}\right]$$

# Descriptive indices

**Descriptive evaluation of model structure using comparisons of models**

**Comparative-Fit-Index (CFI) - important index!**

- Compares the specified model with the independence model (uncor.)
- Takes into consideration the degrees of freedom
- Should be at least larger than 0.8;
  Values of larger than 0.9 are recommended;
  Range between 0 and 1
  $.95 \leq CFI \leq 1.00$ = good model fit
  $.90 \leq CFI < .95$ = acceptable model fit

- Advantages
  - Independence of sample size
  - Small standard error
  - Reflects mis-specification of the model
  - Robust regarding the violation of the normal distribution assumption

# Descriptive indices

**Descriptive evaluation of model structure using comparisons of models**

**Comparative-Fit-Index (CFI) - important index!**

- Compares the specified model with the independence model (uncor.)
- Takes into consideration the degrees of freedom
- Should be at least larger than 0.8;

  Values of larger than 0.9 are recommended;

  Range between 0 and 1

  $.95 \leq \text{CFI} \leq 1.00$ = good model fit

  $.90 \leq \text{CFI} < .95$ = acceptable model fit

- Based on the work by Cheung et al. (2002) a CFI difference $\geq .01$ can be considered as **indication of a model fit difference** at the 5% level.

Advantage: it is independent of a data distribution and of whether the models are nested or not nested!

# Descriptive indices

**Descriptive evaluation of model structure using comparisons of models**

***Normed Fit Index (NFI) & Goodness of Fit Index (GFI)***

… describe model fit in comparing the specified model with a „baseline model“

**NFI**: Comparison with „null model“

→ all variables are not correlated (… are not allowed to …)

**GFI:** Comparison with another kind of a „null model“:

→ all parameter are fixed to zero

This means that it is checked how much better the specified model fits than a „bad“ baseline model (NFI) respectively an unspecified model

→ the larger the fit index, the better the model fit: $\geq .95$

- NFI rarely indicates a good model fit and, therefore, is rarely used.
- The recommendation is „avoid it“

# Descriptive indices

**Descriptive evaluation of model structure using comparisons of models**

**Other fit indices:**

***Nonnormed Fit Index (NNFI) & Adjusted GFI (AGFI)***

They are achieved by kind of scaling of NFI respectively GFI

→ the integration of the degrees of freedom disadvantages more complex and less restrained models

A note. *NNFI is equivalent to TLI* (another fit index)

# Descriptive indices

**Descriptive evaluation of the parsimony of models**

*Parsimony Normed Fit Index (PNFI) &*

*Parsimony Goodness-of-Fit Index (PGFI)*

These indices favor simplicity of the structure of a model

They are achieved by weighting *NFI (PNFI) respectively GFI (PGFI) using the degrees of freedom*

→ more complex lead to smaller values indicating worse model fit

# Descriptive indices

**Descriptive evaluation of the parsimony of models**

**AIC =** Akaike Information Criterion

- Used for comparing two non-nested models
- High complexity of models is penalized
- The model with the smaller AIC is preferred
- But - there is no cut-off

(furthermore, there are BIC and CAIC)

# Outline

- Some basics
- Significance testing
- Descriptive indies
- **Other aspects of model evaluation**

# Other aspects of model evaluation

**The quality of special aspects of a model**

→ These aspects are mostly considered in the case of a bad model fit

**- Evaluation of the standardized residuals:**

- ***must always be larger than zero***
- ***must be clearly smaller than* 1** (otherwise the model does not explain variance)

(Comment: occassionally a model shows good model fit although it explains almost no variance → this model must be rejected)

# Other aspects of model evaluation

**The quality of special aspects of a model**

→ These aspects are mostly considered in the case of a bad model fit

**- Evaluation of the standardized residuals:**

- ***must always be larger than zero***
- ***must be clearly smaller than* 1** (otherwise the model does not explain variance)

(Comment: occassionally a model shows good model fit although it explains almost no variance → this model must be rejected)

**- Evaluation of Q plot:**

Graphical illustration of the standardized residuals using a two-dimensional plot

# Other aspects of model evaluation

**The quality of special aspects of a model**

→ **Evaluation of parameter estimates by t values (Wald test):**

$$t_i = \frac{\pi_i}{s_i}$$

$\pi_i$ = estimated parameter *i* of non-standardized solution

$s_i$ = standard error of estimation of parameter *i*

Is the estimated value significantly different from zero?

→ Yes, if the *t* value is larger than 1.96

# Recommendation

Among the recommended indices are …

…. **RMSEA**

…. **SRMR**

…. **CFI**

..... **NNFI**

..... (AIC)

Also the report of $\chi^2$ is recommended although it is rarely used for the evaluation.

.

***Note.***

**- There is considerable overlap of information provided by some fit statistics!**

**- Therefore, recently the consideration of only a few especially important statistics is recommended when reporting empirical results!**

# QUESTIONS REGARDING COURSE UNIT 11

- What are nested models?
- What is the critical difference according to the CFI difference test?
- What are the categories for the fit evaluation when using the descriptive fit indices?
- The use of which fit statistics is recommended?

# Literature:

**DiStefano, C. (2016). Examining fit with structural equation models. In K. Schweizer, & C. DiStefano (Eds.), *Principles and Methods of Test Construction* (pp. 166-193). Göttingen: Hogrefe Publishing.**

Kline, R. B. (2011). *Principles and practices of structural equation modeling* (3rd edition) (Chapter 1: Introduction). New York, NY: The Guilford Press.