

GESIS Spring Seminar 23

Comparative Social Research with Multi-Group SEM

Daniel Seddig

Eldad Davidov

Peter Schmidt

Yannick Diehl

Day 1 - 27.02.2023

Table of Contents

| | |
|--|----------|
| Packages | 1 |
| Data | 1 |
| Confirmatory measurement analysis (CFA) | 2 |
| Procedure | 2 |
| CFA model of Universalism | 3 |

Packages

```
# basic tools
library(tidyverse)

# checking data
library(MVN)
library(correlation)
library(performance)
library(parameters)
library(insight)

# building the models
library(lavaan)

# cheking the models
library(semTools)

# model visualization
library(semPlot)
library(semptools)
```

Data

```
ESS07 <- read.csv("../Data/ESS07.csv")
```

Confirmatory measurement analysis (CFA)

Procedure

Step 1: Translation of the theoretical model into a measurement model

- We need a model that is able to conceptualize the relationship between manifest indicators and latent variables

Step 2: Checking whether it is a reflexive or formative construct

- Reflective measurement: indicators are “effects” of the latent variable
- Formative measurement: latent variable is “caused” or composed by the manifest indicators

Step 3: Specification of the model (Choose a method to assign a scale to the latent variable)

- Reference indicator method: factor loading of one indicator is fixed to 1.0
- Fixed factor method: the factor variance is fixed to 1.0

Step 4: Checking the model assumptions (ML estimation requirements)

- Data are continuous and multivariate normal (case of non-normal data standard errors and chi-square-statistics should be adjusted - e.g., robust ML = MLR)
- Sample size is sufficiently large (ratio of number of cases and number of parameters - N:q rule => 20:1 or 10:1)

Step 5: Checking whether the model is identified

- Complexity of a CFA models is limited by the total number of observations (p) and the number of unknown model parameters (q)
- Degrees of freedom must be at least zero: $df = p - q \geq 0$
- j = number of indicators
- $p = j * (j + 1) / 2$
- q = counting of the parameters to be calculated (factor loading, residual error, factor variance)
- $df = 0$ => model is identified and saturated / $df > 0$ => model is overidentified

Step 6: Checking the fit measurements

| Index | Typ | Theoretical range | Cut-off | N sensitive | Penalty for complexity |
|--------|---------|-------------------|-----------------------------|----------------|------------------------|
| X2/df | badness | ≥ 0 | < 5 | yes | no |
| CFI | godness | 0.0 – 1.0 | ≥ 0.95 (≥ 0.90) | no | yes |
| TLI | godness | 0.0 – 1.0* | / | no | yes |
| SRMR | badness | ≥ 0 | < 0.08 | yes | no |
| RMSEA | badness | ≥ 0 | ≤ 0.05 (≤ 0.08) | yes to small N | yes |
| PCLOSE | badness | 0.0 – 1.0 | ≥ 0.95 | yes | / |

*negative values indicate extremely misspecified model; when exceeds 1, model is extremely well-fitting

Step 7: Checking for misspecification

- What causes model misfit? => Indicator choice, Factor choice, Violations of assumptions (e.g., multivariate normality) and Causal structure (e.g., restrictions)
- How to diagnose model misfit? => Parameter estimates (Heywood cases?), Residual matrices (i.e., differences between observed and estimated covariances), Modification indices (approximation of the reduction of chi-square if a single constrained parameter is freely estimated)

Step 8: Identifying the mean structure (optional)

- Fixing the intercept of the reference indicator to zero
- Fixing the latent mean to zero

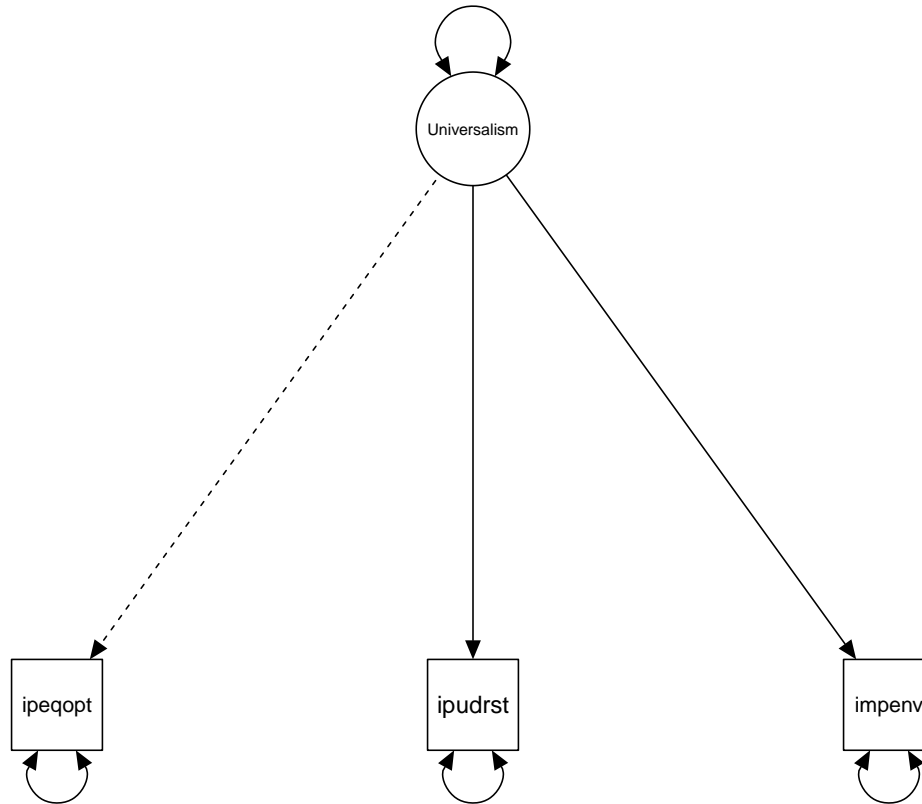
CFA model of Universalism

Step 1

?!

Step 2

Indicators are “effects” of the latent variable and therefore we specify a reflective measurement model



Step 3

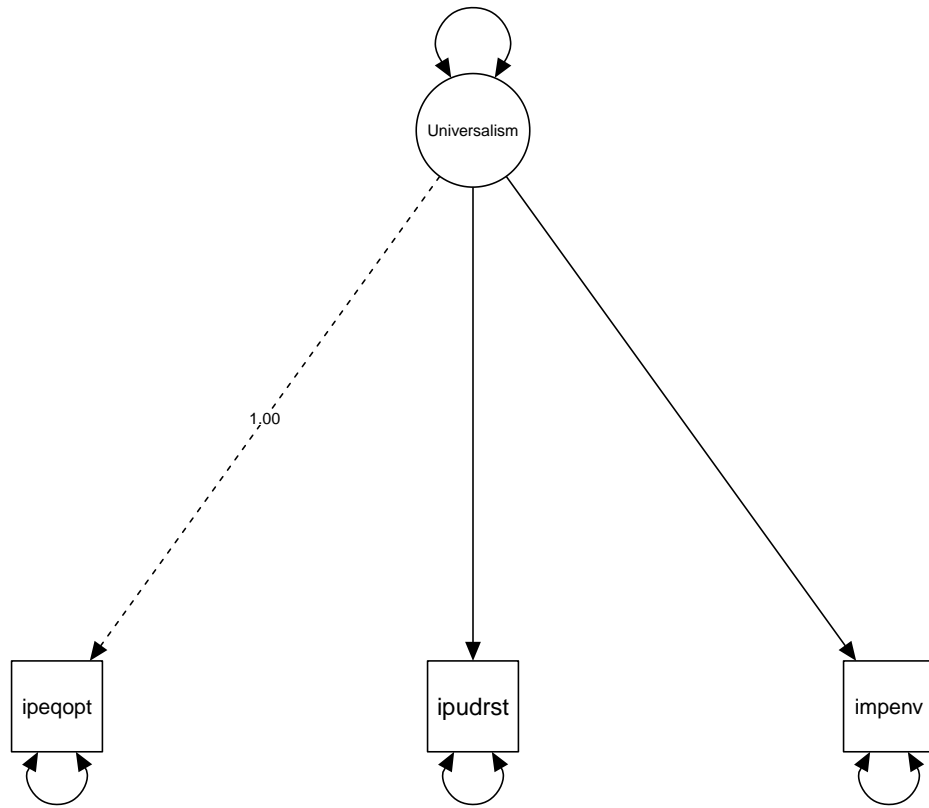
The syntax commands for specifying a lavaan model:

<https://lavaan.ugent.be/tutorial/syntax1.html>

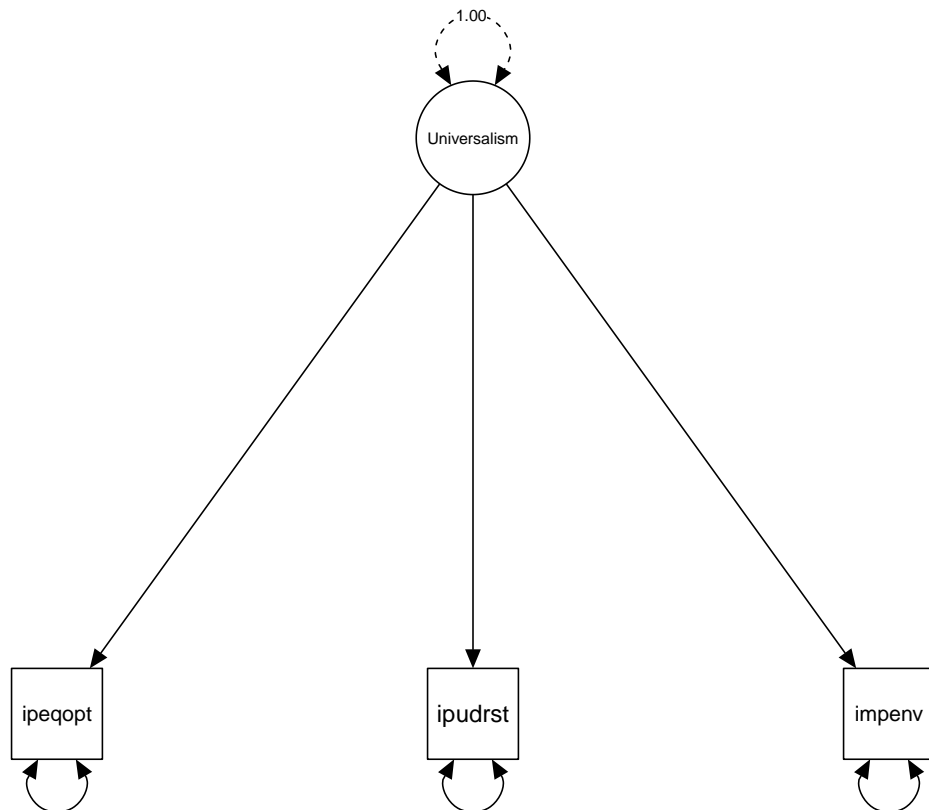
| Formula type | Operator | Mnemonic |
|----------------------------|-----------------|--------------------|
| latent variable definition | <code>=~</code> | is measured by |
| regression | <code>~</code> | is regressed on |
| (residual) (co)variance | <code>~~</code> | is correlated with |
| intercept | <code>~1</code> | intercept |

```
universalism <- "uni =~ ipeqopt + ipudrst + impenv"
```

Lavaan automatically uses the Reference indicator method, so one indicator’s factual loading is fixed at 1. Accordingly, the variance of the latent factor is estimated.



In Lavaan, however, it is also very easy to use the fixed factor method, i.e. to fix the variance of the latent factor to 1. This allows all factor loadings to be freely estimated.



This requires the use of labels, i.e. the labeling of individual parameters in order to be able to control them directly. The syntax commands for labeling a lavaan model can be found here: <https://lavaan.ugent.be/tutorial/syntax2.html>. We just need to understand how parameters are fixed in Lavaan and how they can be calculated freely.

Reference indicator method:

```
universalism <- "uni =~ 1*ipeqopt + ipudrst + impenv"
```

Fixed factor method:

```
universalism <- "
# specification of the measurement model
uni =~ NA*ipeqopt + ipudrst + impenv

# specification of the variance
uni ~~ 1*uni
"
```

Freely estimate the factor loading of the first indicator by NA*ipeqopt and fix the variance of the latent factor by 1*uni. The variance of a variable is specified as covariance with itself, hence $uni \sim uni$. In the end we get $uni \sim 1*uni$

Step 4

Now let's get an overview of the properties of our variables. First, let's look at the descriptive statistics of our items to gain knowledge of missing values, central tendency properties, variability, and distribution.

```
ESS07 %>%
  select(ipeqopt, ipudrst, impenv) %>%
  mvn() %>%
  # optional for better display
  export_table(table_width = 1, digits = 3)
```

| Test | | HZ | p value | MVN |
|---------------|--|---------|---------|--------|
| Henze-Zirkler | | 133.042 | | 0 NO |

| Test | | Variable | Statistic | p value | Normality |
|------------------|--|----------|-----------|----------|-------------|
| Anderson-Darling | | ipeqopt | | 362.6880 | <0.001 NO |
| Anderson-Darling | | ipudrst | | 332.9049 | <0.001 NO |
| Anderson-Darling | | impenv | | 322.7889 | <0.001 NO |

| n | Mean | Std.Dev | Median | Min | Max | 25th | 75th | Skew | Kurtosis |
|------|-------|---------|--------|-----|-----|------|------|--------|----------|
| 5844 | 4.945 | 1.040 | 5 | 1 | 6 | 4 | 6 | -1.142 | 1.295 |
| 5844 | 4.717 | 1.027 | 5 | 1 | 6 | 4 | 5 | -0.937 | 0.892 |
| 5844 | 4.878 | 1.025 | 5 | 1 | 6 | 4 | 6 | -0.933 | 0.732 |