

Assignment 4

SEM 1: Confirmatory Factor Analysis

Please hand in a .pdf file containing your report and a .R containing your codes or screenshots of every Jasp analysis. The deadline of this assignment is Friday May 4 23:59.

Assignment

Model nesting

An important concept in model comparison is *nesting*, as the χ^2 difference test is only valid when comparing two nested models.

Question 1 (1 point) Explain nesting and what it means that model 1 is nested in model 2. Which of the models will then be more parsimonious (simpler)?

We can test in Lavaan if two models are nested as follows:

```
library("lavaan")

# One CFA model:
model1 <- '
eta_1 =~ y1 + y2 + y3
eta_2 =~ y4 + y5 + y6
'

# Another CFA model:
model2 <- '
eta_1 =~ y1 + y2 + y3 + y4
eta_2 =~ y4 + y5 + y6
'

# Is model A nested in model B?
# Step 1: fit model A to some dataset (here I simulate data):
data <- simulateData(model1)
fit1 <- cfa(model1, data)

# step 2: fit model 2 on the MODEL IMPLIED cov matrix (Sigma) from model 1 with some arbitrary sample size
Sigma1 <- inspect(fit1, "sigma")
fit2 <- cfa(model2, sample.cov = Sigma1, sample.nobs = 100000)

# If model 1 is nested in model 2, chi-square = 0 and p = 1!
fit2

## lavaan (0.5-23.1097) converged normally after 35 iterations
##
##   Number of observations              100000
##
##   Estimator                          ML
##   Minimum Function Test Statistic    0.000
##   Degrees of freedom                  7
##   P-value (Chi-square)                1.000
```

Question 2 (1 point) Explain why we would expect model 2 to return perfect fit ($T = 0$) based on Σ obtained from model 1, if model 1 is nested in model 2.

Given the following CFA models (syntax assumes `cfa()` is used):

	Model 1	Model 2
Comparison A	$\eta_1 \sim y_1 + y_2$ $\eta_2 \sim y_3 + y_4$ $\eta_3 \sim y_5 + y_6$	$\eta_1 \sim y_1 + y_2 + y_5 + y_6$ $\eta_2 \sim y_3 + y_4$
Comparison B	$\eta_1 \sim a*y_1 + a*y_2 + a*y_3$	$\eta_1 \sim y_1 + y_2 + y_3$
Comparison C	$\eta_1 \sim y_1 + y_2 + y_3 + y_4$ $y_3 \sim y_4$	$\eta_1 \sim y_1 + y_2$ $\eta_2 \sim y_3 + y_4$
Comparison D	$\eta_1 \sim y_1 + y_2 + y_3 + y_4$ $\eta_2 \sim y_4 + y_5 + y_6$	$\eta_1 \sim y_1 + y_2 + y_3$ $\eta_2 \sim y_3 + y_4 + y_5 + y_6$

Question 3 (2 points) For each comparison listed above (A, B, C, and D), first describe conceptually if you would expect model 1 to be nested in model 2, model 2 to be nested in model 1, model 1 and 2 to both be nested in each-other (equivalent), or neither of the models to be nested in the other. Next, investigate the nesting relationships using Lavaan. It is possible that you expected a different nesting relationship than you eventually observe. Note: some of the cases may lead to warnings/failed estimations in some simulations, just simulate a different dataset and try again. ■

Higher order factor models

The following code in R gives you the covariance matrix of an $N = 500$ simulated example:

```
covMat <- structure(c(1.32, 0.53, 0.54, 0.16, 0.04, 0.1, 0.02, 0.03, 0.15,
0.53, 0.9, 0.53, 0.08, 0.04, 0.14, 0.08, 0.06, 0.05, 0.54, 0.53,
1.19, 0.11, 0, 0.08, 0.03, 0, 0.05, 0.16, 0.08, 0.11, 1.96, 0.87,
0.79, 0.09, 0, 0, 0.04, 0.04, 0, 0.87, 1.38, 0.75, 0.1, 0.02,
-0.03, 0.1, 0.14, 0.08, 0.79, 0.75, 1.31, 0.15, 0.08, -0.02,
0.02, 0.08, 0.03, 0.09, 0.1, 0.15, 1.06, 0.42, 0.37, 0.03, 0.06,
0, 0, 0.02, 0.08, 0.42, 0.94, 0.36, 0.15, 0.05, 0.05, 0, -0.03,
-0.02, 0.37, 0.36, 0.97), .Dim = c(9L, 9L), .Dimnames = list(
c("y1", "y2", "y3", "y4", "y5", "y6", "y7", "y8", "y9"),
c("y1", "y2", "y3", "y4", "y5", "y6", "y7", "y8", "y9")))
```

We assume three factors underlying these variables:

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ \lambda_{31} & 0 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & \lambda_{52} & 0 \\ 0 & \lambda_{62} & 0 \\ 0 & 0 & \lambda_{73} \\ 0 & 0 & \lambda_{83} \\ 0 & 0 & \lambda_{93} \end{bmatrix}$$

Suppose the items represent personality inventory items and the latent traits are aimed to represent the dark triad personality traits: Machiavellianism, Narcissism and Psychopathy.

Question 4 (1 point) Fit the CFA model to this dataset and judge the fit. ■

Question 5 (1 point) Fit a *higher order* factor model, in which one general factor causes covariation between the three latent variables. Compare the fit to the CFA model. Can you explain your results? ■

Question 6 (1 point) Fit a *bifactor order* factor model. This is a factor model in which (a) one general factor is added that causes *all* indicators *directly*, and (b) all latent variables (personality traits and the general factor) are *uncorrelated*. ■

Question 7 (1 point) Compare your fitted models and describe which model you would select based (a) on model comparisons and (b) on theory. ■

Exploratory Factor Analysis

Repeat the codes from week 2 in R (replace ... with your student number, or your birthday (YYYYMMDD) if you do not have a student number):

```
library("install.packages")
library("psych")
studentNumber <- ...
set.seed(studentNumber)
data(bfi)
noNA <- na.omit(bfi)
Data <- noNA[sample(1:nrow(noNA), 500), 1:25]
write.csv(Data, paste0("bfi_", studentNumber, ".csv"), row.names = FALSE)
```

This code created an object `Data`, which is also stored as a csv file in your working directory. This should be exactly the same data you used before. Note, some of you noted this dataset is *not* the actual Big Five Inventory but rather the IPIP (see `?bfi`).

Question 8 (2 points) Perform an exploratory factor analyses on this dataset using the same number of latent variables you used in assignment 2, and compare the obtained factor loadings to your final model from week 2. Take rotation into account, and note that the order of latent variables may differ from the order of latent variables in the CFA model. ■

Ordinal data

Question 9 (1 points) Describe variables that can be assumed binary or ordinally distributed (2-4 answering categories), but on which it might not make plausibly sense to compute polychoric correlations. ■