Assignment 1

SEM 1: Confirmatory Factor Analysis

Please hand in a `.pdf` file containing your answers and, if applicable, a `.R`, `.Rmd` (compile to PDF) or `Rnw` file containing your codes. The deadline of this assignment is Tuesday April 9 13:00.

Conceptual questions

Question 1 (1 point) In a model with one latent variable and three indicators, F_{ML} is always exactly 0 in the maximum likelihood solution. Can you explain why? ■

Question 2 (1 point) Are the following statements true or false?

- A nonzero residual covariance implies a violation of local independence.
- Both the first factor loading and the variance of the latent variable need to be set to 1 to identify a model.
- Assuming the data is multivariate normally distributed implies that all relationships between variables are assumed *linear*.

■

Practical questions

In this assignment we will make use of the general CFA framework:

$$\begin{aligned}
\boldsymbol{y}_i &= \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \\
\boldsymbol{y} &\sim N(\boldsymbol{0}, \boldsymbol{\Sigma}) \\
\boldsymbol{\eta} &\sim N(\boldsymbol{0}, \boldsymbol{\Psi}) \\
\boldsymbol{\varepsilon} &\sim N(\boldsymbol{0}, \boldsymbol{\Theta}),
\end{aligned}$$

First, let:

$$\begin{aligned}
\boldsymbol{\Lambda} &= \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix} \\
\boldsymbol{\Theta} &= \begin{bmatrix} \theta_{11} & & & \\ 0 & \theta_{22} & & \\ 0 & 0 & \theta_{33} & \\ 0 & 0 & 0 & \theta_{44} \end{bmatrix} \\
\boldsymbol{\Sigma} &= \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Theta}
\end{aligned}$$

Question 3 (1 point) Draw a path diagram for this model. ■

Question 4 (1 point) Derive an expression for each element of $\boldsymbol{\Sigma}$. ■

Question 5 (1 point) How many degrees of freedom does this model have? Is it identified? ■

Suppose a dataset of $n = 50$ subjects leads to the following *unbiased* estimate $\boldsymbol{S}$ (obtained with the `cov()` function in R):

$$\boldsymbol{S}^* = \begin{bmatrix} 0.72 & 0.30 & 0.11 & 0.28 \\ 0.30 & 1.10 & 0.21 & 0.58 \\ 0.11 & 0.21 & 1.14 & 0.30 \\ 0.28 & 0.58 & 0.30 & 1.77 \end{bmatrix}$$

Question 6 (1 point) Transform $\boldsymbol{S}^*$ to the Maximum Likelihood sample variance–covariance matrix $\boldsymbol{S}$. ▪

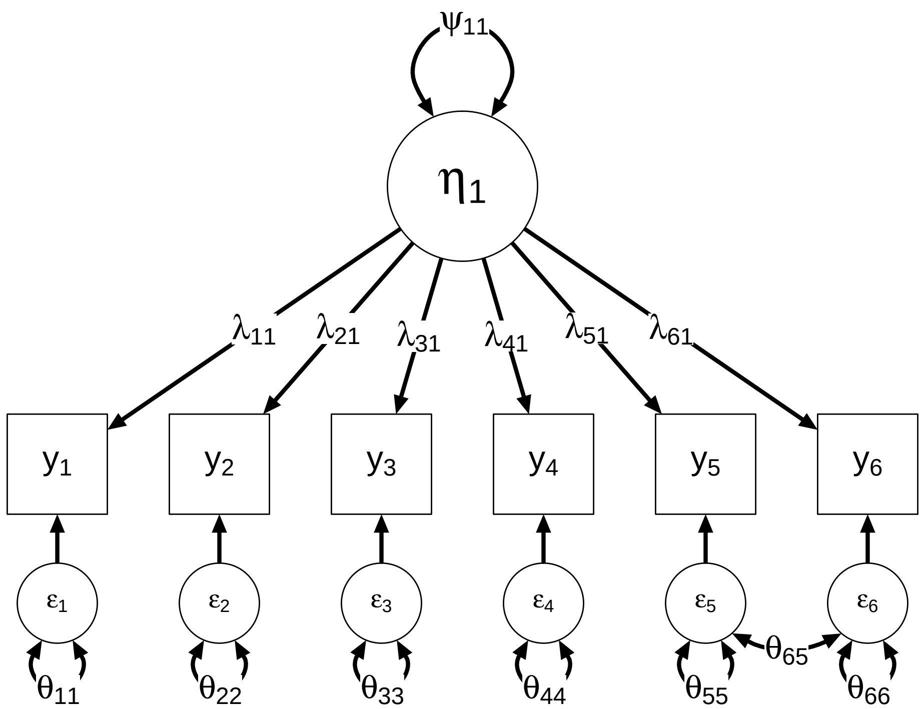
Consider the following R code (it is ok if you do not understand every step), which makes use of an object `S` that you defined in the previous exercise:

```
# Form a fit function given S:
f <- function(x, S){
  Lambda <- matrix(x[1:4])
  Theta <- diag(x[5:8])
  Sigma <- Lambda %*% t(Lambda) + Theta
  sum(diag(S %*% solve(Sigma))) - log(det(S %*% solve(Sigma))) - ncol(Theta)
}

# Form start values (from the matrices above)
start <- rep(0.5,8)

# Optimize fit function:
optres <- nlminb(start, f, S = S)

# Obtain estimates for lambda:
Lambda <- matrix(optres$par[1:4])
print(round(Lambda,2))

# Obtain estimates for theta:
Theta <- diag(optres$par[5:8])
print(round(Theta,2))
```

Question 7 (1 point) Can you explain *conceptually* (in one sentence) what this R code does (1 point)? ▪

Question 8 (1 point) Use the estimated parameter matrices to compute $\boldsymbol{\Sigma}$, and compare it to $\boldsymbol{S}$. ▪

Given the following path diagram:



Question 9 (2 points) Write down $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ and give the number of degrees of freedom. Is the model identified? Why (not)? ▪