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.

Questions

Suppose we observe the following data of three variables:

$$\boldsymbol{Y} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 2 \\ 3 & 5 & 4 \\ 1 & 0 & 0 \\ 4 & 5 & 5 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \\ 2 & 0 & 3 \\ 2 & 1 & 3 \\ 4 & 5 & 4 \end{bmatrix}.$$

To save some work, you can obtain this data in R as follows:

```
Y <- matrix(c(0, 2, 3, 1, 4, 3, 2, 2, 2, 4, 0, 0, 5, 0, 5, 0, 1,
0, 1, 5, 0, 2, 4, 0, 5, 1, 0, 3, 3, 4), 10, 3)
```

Question 1 (1 point) Compute the maximum-likelihood estimate $\boldsymbol{S}$ using any software you want. ■

Now we will use the general CFA framework:

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

With the following matrices:

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1.0 \\ 2.0 \\ 1.5 \end{bmatrix}, \boldsymbol{\Theta} = \begin{bmatrix} 0.5 & & \\ 0.0 & 1.2 & \\ 0.0 & 0.0 & 1.0 \end{bmatrix}$$

(note that upper-triangular elements of $\boldsymbol{\Theta}$ are also zero, but not shown for notational clarity because $\boldsymbol{\Theta}$ is symmetrical). Suppose the implied variance–covariance matrix equals:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Theta}$$

Question 2 (2 points) Compute $\boldsymbol{\Sigma}$ by hand and verify your work using any software. Does your result resemble $\boldsymbol{S}$? Do you think these are good parameter estimates? ■

Question 3 (1 point) What is the value of the implied variance, $\boldsymbol{\Psi}$, of the single latent variable in this model? To what form of *scaling* does this value relate? ■

Question 4 (1 point) What are (1) the model-implied *correlation* (not covariance) between y_1 and y_3, and (2) the observed correlation between these two variables? ■

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

```r
# Form a fit function given S:
f <- function(x, S){
  Lambda <- matrix(x[1:3])
  Theta <- diag(x[4:6])
  Sigma <- Lambda %*% t(Lambda) + Theta
  sum(diag(S %*% solve(Sigma))) - log(det(S %*% solve(Sigma))) - 3
}

# Form start values (from the matrices above)
start <- c(1,2,1.5,0.5,1.2,1)

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

# Obtain estimates:
Lambda <- matrix(optres$par[1:3])
print(round(Lambda,3))

##        [,1]
## [1,] 1.033
## [2,] 1.927
## [3,] 1.588

Theta <- diag(optres$par[4:6])
print(round(Theta,3))

##        [,1]  [,2]  [,3]
## [1,] 0.343 0.000 0.000
## [2,] 0.000 1.097 0.000
## [3,] 0.000 0.000 0.638
```

Question 5 (1 point) Can you explain *conceptually* (in one sentence) what this R code does? Use a term from the lecture! ■

Question 6 (1 point) My code gives the following objects:

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1.033 \\ 1.927 \\ 1.588 \end{bmatrix}, \boldsymbol{\Theta} = \begin{bmatrix} 0.343 & & \\ 0.0 & 1.097 & \\ 0.0 & 0.0 & 0.638 \end{bmatrix}$$

Use these objects to form $\boldsymbol{\Sigma}$, and compare your result to the answer of Question 1. Explain why you would or would not expect this result. ■

Now let:

$$\mathbf{\Lambda} = \begin{bmatrix} 0.7 & 0 \\ 0.7 & 0 \\ 0.7 & 0 \\ 0 & 0.7 \\ 0 & 0.7 \\ 0 & 0.7 \end{bmatrix}$$

$$\mathbf{\Psi} = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$$

$$\mathbf{\Theta} = \begin{bmatrix} 0.51 & & & & & \\ 0 & 0.51 & & & & \\ 0 & 0 & 0.51 & & & \\ 0 & 0 & 0 & 0.51 & & \\ 0 & 0 & 0 & 0 & 0.51 & \\ 0 & 0 & 0 & 0 & 0 & 0.51 \end{bmatrix}$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^\top + \mathbf{\Theta}$$

Question 7 (1 point) Compute $\mathbf{\Sigma}$.

Now let:

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{\Psi} = \begin{bmatrix} 0.490 & 0.147 \\ 0.147 & 0.490 \end{bmatrix}$$

$$\mathbf{\Theta} = \begin{bmatrix} 0.51 & & & & & \\ 0 & 0.51 & & & & \\ 0 & 0 & 0.51 & & & \\ 0 & 0 & 0 & 0.51 & & \\ 0 & 0 & 0 & 0 & 0.51 & \\ 0 & 0 & 0 & 0 & 0 & 0.51 \end{bmatrix}$$

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Psi}\mathbf{\Lambda}^\top + \mathbf{\Theta}$$

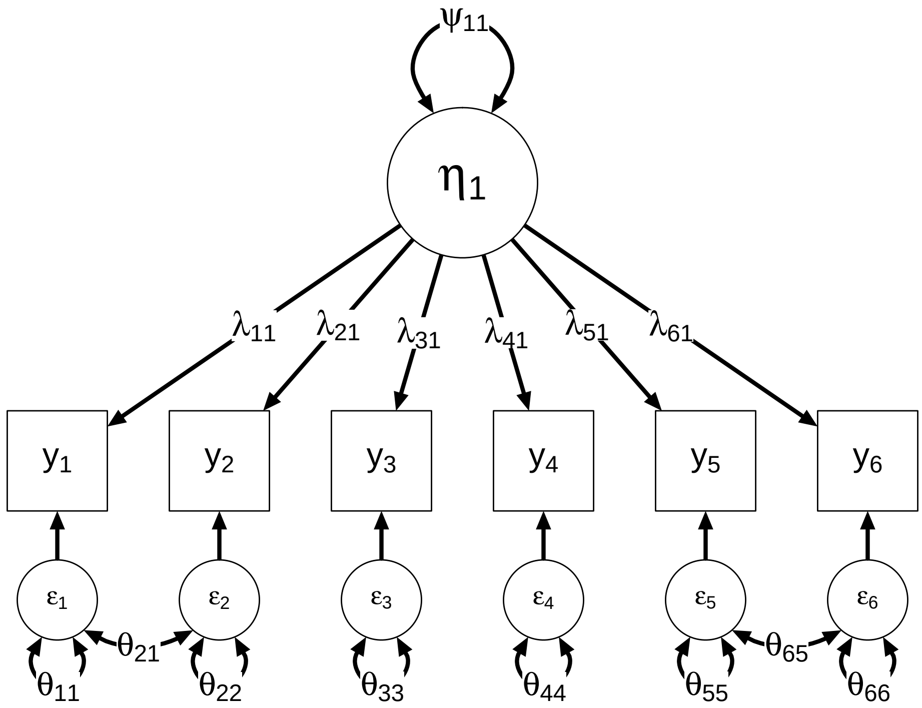
Question 8 (1 point) Recompute $\mathbf{\Sigma}$ and compare your answer to the previous question. Can you explain the result?

Suppose we observe:

$$\boldsymbol{S} = \begin{bmatrix} 1.03 & & & & & \\ 0.49 & 1.02 & & & & \\ 0.52 & 0.49 & 1.05 & & & \\ 0.63 & 0.14 & 0.15 & 0.97 & & \\ 0.10 & 0.18 & 0.13 & 0.45 & 0.97 & \\ 0.15 & 0.18 & 0.13 & 0.49 & 0.50 & 1.02 \end{bmatrix}$$

Question 9 (2 points) Compare your computed $\mathbf{\Sigma}$ to the observed $\boldsymbol{S}$. Do you think the model fits well? Is there a parameter you can think of to add to improve model fit?

Given the following path diagram:



Question 10 (2 points) Write down $\mathbf{\Lambda}$, $\mathbf{\Theta}$ and $\mathbf{\Psi}$ and give the number of degrees of freedom. Is the model identified? Why (not)?