Psychometrics and SEM Data Assignment II 2024-2025

Psychometrics and SEM

This assignment will be graded individually. Therefore, you are obliged to complete this assignment individually. Although you can of course make use of external sources, the work you submit should be your own.

Dataset

Download the SPSS system file "TOTAAL.SAV" containing all data (contains more than we need) from Brightspace. Load it into R as follows:

```
library("foreign")
dat <- read.spss("TOTAAL.SAV", to.data.frame = TRUE)
```

The data come from a large-scale survey on the social integration of young adults. The survey was administered both in 1987 and 1991. In 1987 the subjects where from 17 through 27 years of age. There are responses of 1775 subjects on psychological tests and other variables.

In this assignment, we focus on three 'Life Satisfaction' indicators assessed in 1987 and 1991.

The following code provides information about the meaning of the variables and values and can be copy-pasted and used to select the indicators:

```
Satisfaction87 <- c(
  ##  Likert scale response: 1 (helemaal niet van toepassing / not at all applicable)
  ##  through 7 (volledig van toepassing / completely applicable)

  "S341",  #  In most respects my life is ideal
  "S342",  #     My life conditions are exellent
  "S343"  #     Overall I am satisfied with my life
  )

Satisfaction91 <- c(
  ##  Likert scale response: --- (helemaal niet van toepassing / not at all applicable)
  ##  through +++ (volledig van toepassing / completely applicable)

  "SS281",  #     In most respects my life is ideal
  "SS282",  #    My life conditions are exellent
  "SS283"  #    Overall I am satisfied with my life
)
```

The character vectors may be used to extract a subset of variables. For example, to select only the 1987 items:

```
dat[,Satisfaction87]
```

Create a subset of the data containing both 1987 and 1991 'Life Satisfaction' items. Name this subset `dat1`.

For this assignment, we assume the items are numerical responses (instead of ordered-categorical). Recode the variables as numeric using the following code:

```
dat1 <- sapply(dat[, c(Satisfaction87, Satisfaction91)], as.numeric)
```

As always, some data are missing. Use the default `lavaan` behavior of list-wise deletion for rows with missing data. Although sub-optimal, for the purposes of the current Assignment it will not be problematic. **The treatment of missing data should be specified in your report.**

Report requirements

The Data Assignments are meant to provide hands-on experience in analyzing real data as they come in practice. Write a comprehensive report, making use of what you have learned. In the report, all questions should be answered, but you do not have to indicate the question numbers anywhere in text. Originality is appreciated.

A good data-analysis report consists of four main sections: Introduction, Methods, Results, and Discussion.

The report should be concise but contain the following information under the appropriate headings:

- **Introduction.** What empirical questions are to be answered by the analysis?
- **Method.** How are these questions answered? This section has the following subsections:
  - Dataset. Shortly describe the sample and the variables.
  - Statistical analysis. Describe the statistical models used.
- **Results.** What are the answers to the research questions? Are expectations, if any, supported by the data? Give the results of the analysis in tables and figures and describe each of them in the running text (focus on patterns and possible striking deviations, no need to literally repeat results from tables or figures in text). Although good tables and figures should be self-explanatory so that, with their captions, they can stand apart from the text, never give tables or figures that are not discussed in text and do not discuss results that are not given numerically. Short numerical results such as the results of statistical tests can be given in the running text if readability is not hampered by them. Never give raw computer output.
- **Discussion.** Give a short summary of the results and discuss main limitations of the analyses, shortcomings, unforeseen issues and to what extend the research questions are answered.

Above all, let the report be clean, clear, and concise.

The terms 'test', 'scale' and 'test score' are used interchangeably in research. From a psychometric perspective, it is often best to refer to the 'test score', because reliability and validity are properties of a test score with respect to a certain use, they are not properties of a test, in and of itself.

Grading

Your report will be evaluated in terms of completeness, technical adequacy of data-analysis, reasoning,interpretation of results, and adequacy of tables and graphs.

The Results section will be most important in determining your grade. For the purposes of the current assignment, the Introduction, Method and Discussion sections may be kept very short (e.g., 5-30 sentences each).

What to hand in?

Hand in two files on Brightspace:

1) Your report. This could be a .docx or .pdf file, or such (and should not contain code!)
2) The R code you used to generate the reported results. This could be an .R, .Rmd file, or such.

You can use Markdown, Quarto, Sweave and such, and hand in the compiled .pdf file as well as the source.You may also write a separate R script and write your report in a separate text processing app.

Before turning in the report, make sure that your code can be run and your results replicated from scratch. For example, if you refer to an external file in your script, make sure you refer to the file name in the code (not the complete location of the file) and the file is available in the current working directory.

Longitudinal Confirmatory Factor Analysis

1. Fit a two-factor model for the 1987 and 1991 'Life satisfaction' items (see Figure 1). Report and interpret the estimated factor loadings and factor covariance.

In the above model, the 1987 and 1991 factors correlate. In addition, the 1987 and 1991 indicators of the same items are assumed to covary. We want to check whether 'Life Satisfaction' improves or deteriorates from 1987 to 1991.

2. Is the model above identified? Why or why not?
3. Test longitudinal invariance across 1987 and 1991:
   i. Model 1: configural invariance
   ii. Model 2: weak factorial invariance / metric invariance
   iii. Model 3: strong factorial invariance / scalar invariance

For each of the above models,

- Report and interpret both exact (χ^2) and approximate (RMSEA and CFI) model fit for each step.
- Report and interpret difference in model fit for each step (compare $\Delta\chi^2$, AIC, and BIC values).

4. Does strong factorial/scalar invariance hold? If scalar invariance does not hold, which model is your final model?
5. Can you make any conclusions about whether 'Life Satisfaction' improves or deteriorates from 1987 to 1991? Why or why not? If you can make conclusions, what conclusions can you make?
6. Plot your final model using the `semPaths()` function from the `semPlot` package.

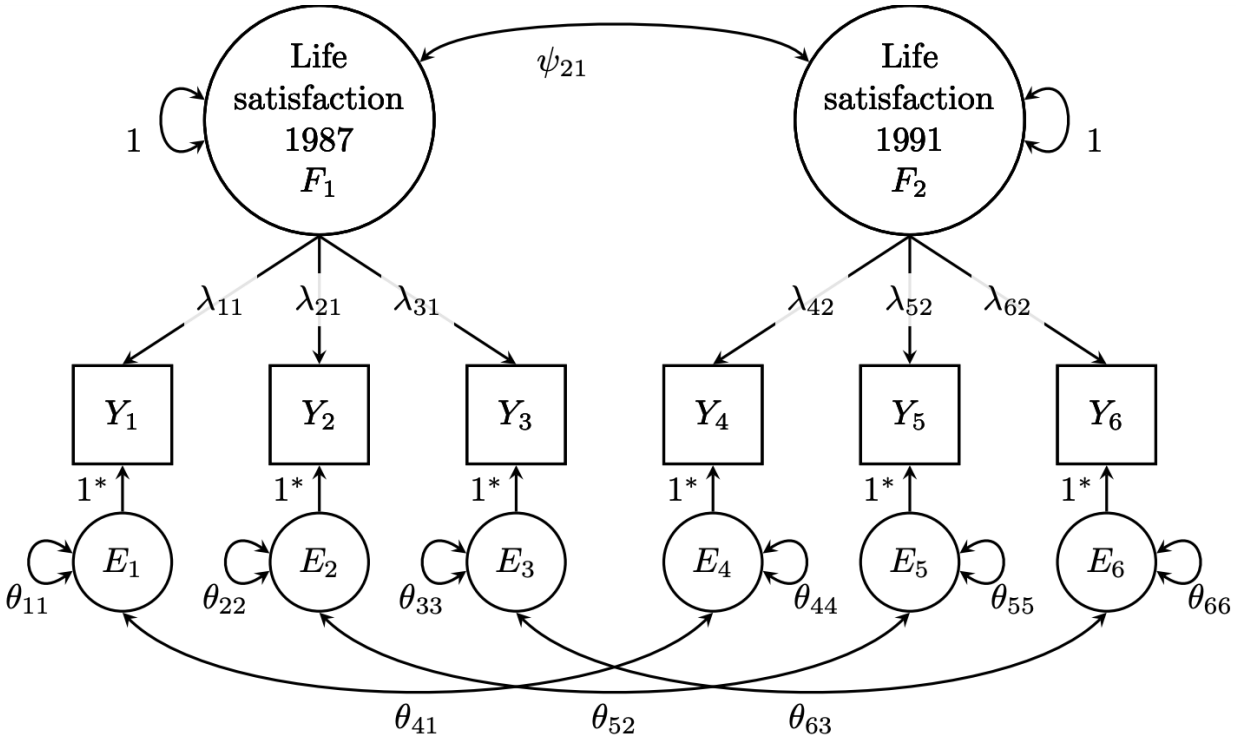

Figure 1: Longitudinal CFA for 'Life Satisfaction'