Psychometrics and SEM Data Assignment I 2024-2025

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. We focus on the following variables assessed in 1987:

- Impulsivity: Item scores on a test to measure the propensity to react fast and sometimes recklessly.
- General psychological health: Item scores on a test to measure the degree to which the subject suffers from minor psychiatric disorders.
- Gender: Whether the subject identifies as a man or a woman.
- Study delay: How many years the person is behind in his/her study.
- Age: Age in years.

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

```
Impulsivity87 <- c(

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

  "S255",  ##  If I have to make a decision in a confused situation, I prefer to use quite some time
  "S264",  ##  My reactions to unexpected situations are rather careful
  "S269",  ##  I feel comfortable in situations that need quick action
  "S274",  ##  I usually make decisions only after contemplating all arguments for or against
  "S279",  ##  I rather work fast with a higher chance of errors than slow but faultless
  "S292"   ##  If I want to do something important, I usually contemplate all possible consequences
           ##  because I absolutely don't want to blunder.
)
```

```
General.psychological.health87 <- c(

  ## 1: more / better than usual,
  ## 2: the same as usual,
  ## 3: less / less well than usual,
  ## 4: much less than usual.

  ## Or reversed if the item is counter indicative:

  ## 1: not at all
  ## 2: not more than usual
  ## 3: a bit more than usual
  ## 4: much more than usual

  "S327",  ##  Can you concentrate on the things you're busy with lately?
  "S328",  ##  Have you lost sleep over worries lately?
  "S329",  ##  Do you have the feeling that your activities are useful and meaningful?
  "S330",  ##  Do you feel able to make decisions lately?
  "S331",  ##  Do you have the feeling that you are constantly under pressure lately?
  "S332",  ##  Do you have the feeling that you can't manage your problems lately?
  "S333",  ##  Do you have pleasure in your daily activities lately?
  "S335",  ##  Did you feel unhappy and dejected lately?
  "S336",  ##  Did you lose self-confidence lately?
  "S337",  ##  Did you consider yourself a worthless person lately?
  "S338"   ##  On the whole, did you feel reasonably happy lately?
)

Gender <- c("M4") ## Identifies as 0 (woman) or 1 (man)

Study.delay87 <- c("M232") ## Years (0 through ...)

Age87 <- c("M37") ## Years (17 through 27)
```

Note that a mistake seems to have been made in labeling of the levels of the Impuslivity items. While the 5th level in the dataset was labeled as "completely applicable", this should have been the 7th level. We assume the order of the factor levels is correct, and that we thus do not need to adjust anything.

The character vectors may be used to access a set of variables. For example, to select only the items in the 1987 Impulsivity test:

```
dat[ , Impulsivity87]
```

As always, some data are missing. The treatment of missing data should be specified in your report. You may remove rows with missing data for the variables relevant for your analysis (i.e., listwise deletion). Although suboptimal, for the purposes of the current Assignment it will not be problematic.

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. E.g., 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.

Part I: Classical test theory

For this part of the assignment, we assume the items are numerical responses (instead of ordered-categorical). If necessary, you can recode variables as numeric. For example, to for example recode general psychological health as numeric:

```
sapply(dat[ , General.psychological.health87], as.numeric)
```

Include the following analyses in your report:

1. Give descriptive statistics for item and test scores (i.e. simple sum of item scores) of both tests. Make sure to give a short description of general patterns and, if relevant, possible deviations from those patterns.
2. Compute Cronbach-Guttman's α for both test scores, and compute the 90% confidence intervals of these α's using a bootstrapping approach, e.g. using package **boot**.
3. For both tests, estimate a one-factor (congeneric) model where the factor has a variance of one. Report the relevant parameter estimates in a table. From these estimates, compute ω for both tests. Compare their values with the α's.
4. Do any of the items lower the reliability of the test scores?
5. Compute the standard error of measurement for both test scores. Use it to compute the length of the 90% confidence interval for each of the test scores.
6. How many parallel items have to be added to (removed from) each of the tests to obtain a test score with reliability (α) of 0.80? Depending on whether items need to be added or removed, formulate an item to add, or suggest an item to remove for each test.
7. Compute the correlation between the test scores, and between the true scores on the two tests.

Part II: Item response theory

For this part of the assignment, we assume the items are ordered-categorical responses. However, package **ltm** accepts numerical item responses all the same. Still, for estimation purposes, it can be helpful to reverse-code items within the same test or scale to have the same direction. The easiest way to do this is with numeric variables, so we can subtract the item responses from (the highest level + 1). E.g.:

```
library("knitr")
num_imp_dat <- sapply(dat[ , Impulsivity87], as.numeric)
kable(cov(num_imp_dat, use = "pairwise"), digits = 2L)
```

	S255	S264	S269	S274	S279	S292
S255	3.21	0.60	-0.78	0.79	-0.45	0.67
S264	0.60	2.71	-0.47	0.79	-0.53	0.72
S269	-0.78	-0.47	2.50	-0.26	0.63	-0.19
S274	0.79	0.79	-0.26	2.76	-0.66	1.14
S279	-0.45	-0.53	0.63	-0.66	3.41	-0.55
S292	0.67	0.72	-0.19	1.14	-0.55	2.96

```
num_imp_dat[ , "S279"] <- 8L - num_imp_dat[ , "S279"]
kable(cov(num_imp_dat, use = "pairwise"), digits = 2L) ## covs for S279 now in right direction
```

	S255	S264	S269	S274	S279	S292
S255	3.21	0.60	-0.78	0.79	0.45	0.67
S264	0.60	2.71	-0.47	0.79	0.53	0.72
S269	-0.78	-0.47	2.50	-0.26	-0.63	-0.19
S274	0.79	0.79	-0.26	2.76	0.66	1.14
S279	0.45	0.53	-0.63	0.66	3.41	0.55
S292	0.67	0.72	-0.19	1.14	0.55	2.96

(Note that is relatively easy to create nice tables with function `knitr::kable`, some relabeling of columns and/or rows to obtain a more informative table may be needed though.)

Include the following analyses in your report:

1. Which item response model best fits the data of each test?
2. Report and discuss the item parameters for each test.
3. Plot and interpret the test information function for each test.
4. Compare the fit of the chosen IRT model with that of a model where the responses are completely independent. What can you conclude?
5. Many behavioral scientists use the linear factor model on discrete item responses although a logistic model is more appropriate from a statistical point of view. Given the test information functions of both models and standard normal density function of the latent variable, does this make much difference for your primary test in practice? Hint: Eyeball a graph wherein these functions are plotted.
6. Check whether the items of your primary test are biased with respect to a categorical background variable, such as gender, study delay or age (the latter two are not binary in these data, but for this exercise you can dichotomize them using e.g., a median split, or any other meaningful dichotomous assignment of your choice).
7. In research, where the interest might be in modeling the effects or associations of impulsivity or general psychological health with other variables of interest in large groups of individuals, would you prefer using the one-factor model, the IRT model or the sum scores for relating these constructs to other predictors of interest? Give the two main reasons for your choice. In (clinical) practice, where test results would be reported back to individual patients or used in decision making, would you prefer using the one-factor model, the IRT model or the sum scores? Again, give the two main reasons for your choice (which need not but could overlap with the previous two reasons).