**Lab Day 1: Regression**

Today’s lab meeting consists of three exercises, below you can find a short description for each.

* All the data and other files for these exercises can be found at the LLL platform.

Make sure to unzip the files. The folder containing these files will be your working directory.

* Solutions to the exercises can be found in the Solutions folder.

We provided R scripts for doing each of the exercises with lavaan.

In the subfolders, you can find the files for performing the analyses in SPSS or Mplus.

Short description of exercises:

* The first exercise of today makes an explicit comparison on how certain analyses look in the software you are used to (e.g., SPSS, Excel, Mplus or SAS/STATA) and how the corresponding analyses look in R when using the lavaan package. The aim is to understand the similarities and differences between software packages. Most exercises include SPSS syntax and Mplus setup. Feel free to use the windows interface in SPSS, or your own software (Excel, or SAS/STATA).
* The second exercise is aimed at doing a multivariate multiple regression analysis yourself using lavaan.
* The third exercise is aimed at doing a path analysis yourself using lavaan. Additionally, it will make you familiar with the ‘technical output’.

***Exercise 1: Multiple regression***

The data for this exercise is taken from Van de Schoot et al (2010)[[1]](#footnote-1). The study examined the understanding of anti-social behaviour and its association with popularity and sociometric status in a sample of at-risk adolescents from diverse ethnic backgrounds (n = 1491, average age 14.7 years). Both overt and covert types of anti-social behaviour were used to distinguish subgroups. For the current exercise, you will carry out a regression analysis where you want to predict levels of socially desirable answering patterns of adolescents (*sw*) using the predictors overt (*overt*) and covert antisocial behaviour (*covert*).

Before you will carry out this regression analysis using lavaan, you will first conduct the analysis in a program of your own choice (SPSS, Mplus, SAS, STATA). By first conducting the analysis in another program, you can later check whether the results of lavaan are comparable. If they are the same, you (most certainly) know you specified everything correctly.

In the folder ‘Data files’, you can find four times the corresponding data set, each with a different extension: **popular\_regr.sav, popular\_regr.xlsx, popular\_regr\_missingsrecoded.dat,** and **popular\_regr.txt**. Use the one that fits the software you are used to work with.

In the subfolder ‘Ex. 1’ in the Solutions folder, you can find the files to perform the analyses in this exercise for SPSS and for Mplus.

Note: There are missing data (depending on the file you use, these are codes as system missing, -999, or -99 & -999).

***1a1.*** Ask for descriptive statistics and run a correlation analysis between the variables of interest (*sw*, *overt* and *covert*).

*What do you think of the correlations with respect to significance, direction, and magnitude?*

***1a2.*** Run a (multiple) regression analysis (dependent variable (DV) is *sw*; independent variables (IVs) are *covert* and *overt*); also check the standardized results.

***1b.*** Do the same as in 1a1 and 1a2 using lavaan in R.

*Interpret the output. What are your conclusions?*

***Exercise 2: Multivariate regression***

In this exercise, we make use of a part (the Child Development Supplement) of the Panel Study of Income Dynamics (PSID)[[2]](#footnote-2). The goal is to investigate whether three DVs, Applied problems (*APst02*), Behavioral problems (*Problems*), and Self-esteem (*Selfesteem*), can be predicted from the following two IV’s: digit span (*DS02)* and letters words (*LWst02*).

Use the data file **CDSsummerschool.sav** or **CDSsummerschool.txt**.

***2a.***

It is a good exercise to draw the model you plan to estimate for yourself.

Be very precise in how you draw the correlations.

***2b***. Request for descriptive statistics and correlations using R.

Check if the descriptive statistics obtained in another software program are comparable. Notably, the corresponding Mplus in- and output files can be found in the subfolder ‘…\Solutions\Ex. 2\Mplus’.

***2c***. Create lavaan code to analyze the research question (running a multivariate regression).

*Interpret the output. What are your conclusions?*

***Exercise 3: Path analysis***

The data for this exercise is about corporal punishment[[3]](#footnote-3), which can be defined as the deliberate infliction of pain as retribution for an offence, or for the purpose of disciplining or reforming a wrongdoer or to change an undesirable attitude or behavior. Here, we are interested in how corporal punishment influences children’s psychological maladjustment. The data come from 175 children between the ages of 8 and 18. The Physical Punishment Questionnaire (PPQ) was used to measure the level of physical punishment that was experienced.

In this exercise, we focus on predicting psychological *maladjustment* (higher score implies more problems) by perceived *rejection* (e.g., my mother does not really love me; my mother ignores me as long as I do nothing to bother her; my mother goes out of her way to hurt my feelings). Moreover, *rejection* is predicted by perceived *harshness* (0 = never punished physically in any way; 16 = punished more than 12 times a week, very hard) and perceived *justness* (2 = very unfair and almost never deserved; 8 = very fair and almost always deserved).

The data consists of a covariance matrix taken from a published paper and can be found in the file **CorPun.dat** or **CorPun.txt***.* Hence, besides analyzing a data set you collected yourself, in lavaan (and Mplus) it is possible to base your analyses on a covariance or correlation table (and this is the reason why reviewers always ask you to include it):

lower <- scan("CorPun.txt")

CovMx <- getCov(lower, names = c("harsh", "just", "reject", "maladj"))

**3a.** Make a drawing of the statistical model about corporal punishment and write down which parameters (e.g., regression paths, covariances, residuals) you expect that are estimated, and number these parameters.

**3b.** Write your lavaan code for this model based on the drawing you made in **3a** and run it.

Note that since you are using a covariance matrix as the input file, you should indicate this also in the input file, as well as the number of observations:

sample.cov = <name covariance matrix>,

sample.nobs = 175

**3c.** Ask for the technical output to inspect which parameters are estimated. For a lavaan fit object, use the function ‘lavInspect()’. Note that this comparable to asking for the TECH1 output in Mplus.

In this output, you can find six matrices (nu, lambda, theta, alpha, beta, and psi) with numbers summing up to the total number of parameters that are estimated. This way you can check whether you analyzed the model you wanted (and you will discover this is not always the case).

*Which of the parameters you listed in exercise* ***3a*** *should be in which matrix?*

For example: the regression coefficient between *rejection* and *harshness* should be in the Beta matrix. Write this down for each parameter.

*Did you estimate all the parameters that you expected?* *If not, what should be changed to the input?*

Notice that you can also answer this question by means of a plot: e.g., using lavaanPlot() from the lavaanPlot package or graph\_sem() from the tidysem package. This will show you all (means, intercepts,) regression paths, (residual) variances, and covariances. Notably, lavaanPlot() will not provide you with information about the means and intercepts (i.e., the nu & alpha matrices) nor the residual (co)variances.

Note:

A first check is to see whether ‘Number of model parameters’ equals the number of elements in your variance-covariance matrix and if available the means. You can count the number of elements in your variance-covariance matrix:

number of variables \* (number of variables +1) / 2.

If you also have means, ‘Number of model parameters’ should equal the number of elements in your variance-covariance matrix plus the number of variables.

Thus, for our four variables without means, the calculation is: 4\*5/2=10.

If we would have means, the number of sample statistics would be 14 (i.e., 4\*5/2 +4).

*Write down the chi-square statistic and its degrees of freedom from the model fit information. Also inspect the model results.*

1. Van de Schoot, R., Van der Velden, F., Boom, J., & Brugman, D. (2010). Can at-risk young adolescents be popular and anti-social? Sociometric status groups, anti-social behavior, gender and ethnic background. Journal of Adolescence, 33(5), 583–592. doi:10.1016/j.adolescence.2009.12.004 [↑](#footnote-ref-1)
2. Produced and distributed by the Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI [↑](#footnote-ref-2)
3. Rohner, R. P., Bourque, S. L.,&Elordi, C. A. (1996). Children's perceptions of corporal punishment, caretaker acceptance, and psychological adjustment in a poor, biracial southern community. *Journal of Marriage and the Family, 58*, 842-852. [↑](#footnote-ref-3)