Computer Lab Introduction

Exercises Structural Equation Modeling in M*plus*.

The data files used are: moodschl, TPBdata, grant

EXERCISE 1 (CFA):
(Moodschl)

Take the file **moodschl.sav**. This is an SPSS file that holds the data on a measurement of students' mood just before class. To measure the mood, a semantic differential was used, asking the students to answer 3 evaluation items (e.g. 'My mood right now is bad … good') and 3 activation items (e.g., Right now I feel passive … active) on a nine-point scale. For all items a high score indicates a high evaluation/activity. The 3rd semantic differential dimension 'Potency' is not used because it does not apply well to mood states. There are 2 additional variables: student gender, and whether the class was a regular class, or whether the students were to take a test in this class, both dichotomous, scored 0=no, 1=yes). Before you can analyze these data in M*plus*, they must be saved on a file as an ASCII (text) file. Before that, check if there are missing values that are defined using the SPSS default dot. If there are, recode these to minus nine (-9). You can use the M*plus* language generator to set up the part of the input that defines the data and variables. The model must be specified by hand.

1. Set up a 1-factor and a 2-factore model to test formally using a chi-square difference test whether the factor structure indeed involves two latent factors, or whether one mood factor might be sufficient.
2. If you decide on two factors, test their covariance, and if that is significant interpret their correlation to decide how strongly the two constructs are related.
3. Use the gender and 'tested' variables as independent variables to predict the latent variables(s) describing the students' mood. What are your conclusions?

Some variations:

1) The default estimator for a general SEM model is Maximum Likelihood (ML). If you check the data in SPSS, you will find that the 6 scale items do not have a normal distribution. Although the sample size is probably large enough to overcome this, it is better to use robust statistics in this case. Change the estimator to either MLR (robust ML) or MLMV (Mean & Variance corrected ML), run model 3 above and observe the difference.

EXERCISE 2 (Path Model):
(TPBData)

A popular theory in psychology to explain social behavior is the Theory of Planned Behavior (TPB) of Ajzen and Fishbein (sometimes called the Ajzen-Fishbein model). This states that *behavior* is predicted by *behavioral intention*, which is in turn predicted by the *attitude toward the behavior* and the *subjective norm about the behavior*. Ajzen and Fisbein originally proposed a complex way to measure attitude and norm, but in practice researchers usually measure the attitude by including a number of attitude questions, and the norm by including a number of questions about the perceived social norms (e.g., how many of your friends would do…). The intention is usually measured by a single question on a 7 or 9-point scale and the behavior is either a dichotomous outcome (yes/no) or a frequency (how often do you…).

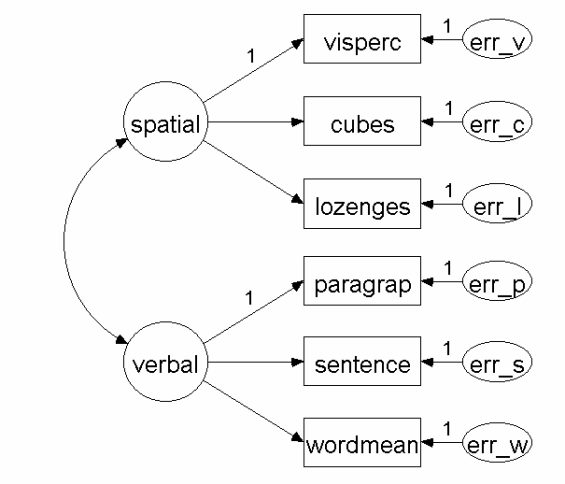
The TPB data are an example data set after Reinecke (1998). The behavior under investigation is condom use by 16-24 year adolescents. The dependent variable 'condom use' is measured on a 5-point frequency scale (How often do you…), and the behavioral intention on a similar 5-point scale (In general do you intend to…). There are three attitude items about condom use (e.g., using a condom is awkward), each measured on a 5-point scale, and three normative items (e.g., I think most of my friends would use…), also measured on a 5-point scale. Before you can analyze these data in M*plus*, they must be saved on a file as an ASCII (text) file. Before that, check if there are missing values that are defined using the SPSS default dot. If there are, recode these to minus nine (-9). For this analysis, you may ignore the *control* variables that are also in the SPSS file. You can use the M*plus* language generator to set up the part of the input that defines the data and variables. The model itself must be specified by hand. It is convenient to recode the one negatively worded item so for all attitude items a high score means a positive attitude. This is already done in the SPSS file PCBdata.

1. Set up a CFA model with attitude and norms as latent variable to test the measurement model separately.
2. Set up the full TPB model with attitude and norms as latent variables, and analyze it. Interpret the results. Since we use the demo version of Mplus we can use only 6 dependent variables, and therefore attit3 and norm3 must be removed from the measurement model. How much variance does the model explain?
3. The TPB model forbids a direct path between attitude and actual behavior; the effect should be mediated totally by the behavioral intention. Test if this path is indeed absent. If the path is indeed absent, all effects of attitude and norms on behavior are indirect.

(Reference: Reinecke, J. (1998). *Sexual- und Verhütungsverhalten 16- bis 24 jähriger Jugendlicher und junger Erwachsener*. Köln, Bundeszentrale fúr gesundheitliche Aufklärung.)

EXERCISE 3 (Multigroup):
(Grant)

The file grant.sav holds the data for the confirmatory factor analysis on six psychological tests used in the lecture (with the verbal & spatial latent factor). The model is given by the following path diagram.



1. Specify this two-factor model. Since these are data for both boys and girls, the results will not be identical to the results in the handouts from the lecture. Mplus will automatically add all constraints needed to make the measurement model equivalent across the two groups. Inspect the output and locate the constraints Mplus automatically imposes. Does the model fit? Are there indications that any of the imposed constraints fail to fit well? (Hint: inspect modification indices for this).
2. Free all constraints imposed by Mplus. Estimate the model again, and carry out a chi-square difference test to see if the imposed constraints fit well (they should, otherwise comparing the groups becomes difficult if not impossible).
3. Specify a model with latent factor means constrained at zero and variances constrained at one in the first group, and unconstrained in the second group. Mplus will automatically constrain all corresponding loadings and intercepts to be equal. Estimate the parameters of this model, and inspect the output to identify the constraints. Next, in the model with the measurement invariance constraints in place, constrain the factor means of the boys and girls to be equal, and carry out a formal chi-square difference test. Are the means of boys and girls equal? And how about the variances?

Variations:

Testing the measurement invariance constraints is often done in 2 steps, first releasing intercepts and next releasing both intercepts and loadings. This allows a better view at what is wrong when the constrained model fits significantly worse than the unconstrained model.

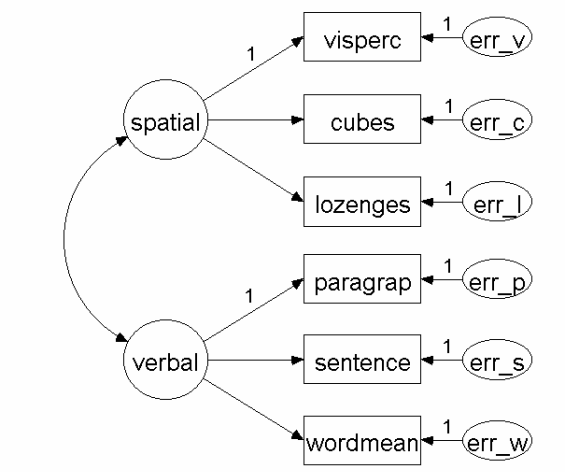
Computer Lab Intermediate

Exercises Structural Equation Modeling in M*plus*.

The data files used are: grant (various types) and GPA

EXERCISE 1 (Multigroup CFA with nonnormal variables):
(Grant)

The file grant.sav holds the data for the confirmatory factor analysis on six psychological tests used in the lecture (with the verbal & spatial latent factor). The model is given by the following path diagram.



The file GrantCat has the Grant data categorized as dichotomous variables, GrantMis has normal data with (MAR) missing data for 2 variables, and GrantCatMis has dichotomous data including missing data.

1. Specify this two-factor model for the original Grant file. Copy the resulting parameter estimates and fit indices to a Word file, under the heading *Continuous Normal*. Check if the hypothesized two-factor model is confirmed. Check if the covariance between the two intelligence factors is significant.
2. Do the same with the categorized variables in GrantCat, with the incomplete data in GrantMis, and with the categorized and incomplete data in GrantCatMis. Put the parameter estimates and fit indices in the Word file with appropriate headings. Compare the results with the original continuous normal results.
3. It is instructive to repeat step 2 with variables treated as continuous normal and using listwise deletion. The results should deviate quite much from the continuous normal results obtained in step 1.
4. Use the GrantCatMis file. Specify the analysis as multiple group model with grouping variable *gender* and including a mean structure. Specify the variables as categorical. Rely on Mplus to estimate this model with appropriate measurement equivalence constraints in place. Check the output to find out what constraints are imposed.
5. Test in the original Grand data and in the GrantCatMis file if the latent factor means for boys and girls are the same.

EXERCISE 2 (Latent Growth Model / LGM):
(GPA)

The file GPA.sav holds the data for the longitudinal GPA example.

1. Set up a basic LGM (no predictors) for the GPA data. Test if the trend is linear or nonlinear. Use the Mplus command | (vertical slash) for this.
2. Use high school GPA gender as a predictor for the intercept and slope.

Extra:

3. Do the same on the Categorized, Incomplete and Categorized + Incomplete data in the files GPACat, GPAMis, & GPACatMis. Observe the different constraints that Mplus adds to the models (provided the | 'vertical slash' notation is used!).

EXERCISE 3 (Regression on Count Data):
(Migrations.sav)

The file Migrations.sav contains data on migrations from school leacers. The variables are: Case ID, number of migrations since leaving school, number of years since leaving school, educational level (high score indicates low educational level).

1. The dependent variable *nmigrat* is a count variable. Set up a model to predict the number of migrations from the education and exposure, using a linear model
2. Specify the dependent variable as a COUNT variable, and inspect what changes this causes in the Mplus output
3. Add an inflation part to the model. Inspect the log-likelihood from the model with and without inflation. Any conclusions?
4. Finally, use exposure and educational level to model the inflation part. Conclusions?

Computer Lab Advanced

Exercises Structural Equation Modeling in M*plus*.

EXERCISE 1 Latent Class Analysis / LCA):
(alcohol, diabetes)

The file alcohol.sav holds the data for the alcohol example.

1. To run the alcohol example in the demo version of Mplus, you need to select 6 items from the available 9. Use the information in the graph (in the handouts) to select 6 items that appear to distinguish well between the three classes. Then use USEVARIABLES in the command setup to select them.
2. Since we now have less information, it may not be necessary / possible to specify 3 latent classes. Run a model with 2 and a model with 3 latent classes, and decide which model fits the data best. Read the output carefully to make sure you understand it and are able to interpret the results.

The file diabetes.sav holds 3 biomedical markers for diabetes; glucose, insulin, and sspg (Steady-State Plasma Glucose, a different glucose indicator). There is also a variable 'diagnose' which contains the clinical diagnosis.

1. Theory states that there should be 3 classes (no diabetes, type I and type II). Set up a LCA model to investigate this hypothesis. Note: the three variables have widely different scales. This causes problems that are solved by using DEFINE to rescale the variables using: DEFINE: glucose = glucose/10; insulin = insulin/100; spg = sspg/10;

EXERCISE 2 Multilevel CFA:
(FamIQData)

The file FamIQData.sav holds the data on 6 intelligence tests for children in 60 families. Replicate the multilevel analysis reported in the handouts. Try ESTIMATION=MUML and compare the results to the (default) MLR estimation.

EXERCISE 3 Mediation:

The TPB model used on Day 1 contains several indirect paths. Use the Mplus MODEL INDIRECT command to estimate the indirect effects. Use bootstrapping to obtain p-values and a bootstrapped 95% confidence interval. Are the indirect effects significant?