Mplus homework M&S – 1

Below are the Mplus homework exercises. Do not forget to make the analytical homework as well!

Exercise 1.

Consider Example 5.1 from Mplus: Carefully read the explanation in the Mplus User's Guide (http://www.statmodel.com/ugexcerpts.shtml, and go to Chapter 5). TIP: To run this model, make sure the **inp-file** (i.e., the input file) and the **data file** are in the same folder.

a. How many sample statistics are there? Indicate how you obtained this number.
A: sample statistics: covariance matrix has 6*7 / 2 = 21 unique elements… and the mean vector has 6 elements → 21 + 6 = 27 sample statistics in total.

b. How many free parameters are there? Describe them (i.e, how many factor loadings etc.)
A: free parameters = parameters to be estimated:

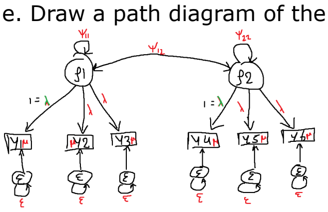
6 factor loadings (but 2 will be fixed to 1, so only 4)
6 residual variances
2 variances of the factors
1 covariance between the factors
6 means

→ 4 + 6 + 2 + 1 + 6 = 19

c. Hence, how many df are there?
A: degrees of freedom = sample statistics – free parameters = 27 – 19 = 8

d. Which defaults does Mplus use when running this model?
A: unit loading identification (i.e., setting the first of the factor loadings for each factor to 1 instead of setting the variances of the factors to 1, which is unit variance identification.

e. Draw a path diagram of the model, and indicate where the free parameters are.



Red = free parameter

f. What do the model matrices $\boldsymbol{\Lambda}$ (=Lambda), $\boldsymbol{\Psi}$ (=Psi), $\boldsymbol{\Theta}$ (=Theta), ν (=nu) and α (= alpha) look like?
A:
Lambda = two columns of 6 values (first column: first value 1, last 3 values are 0 and last column: first 3 values are 0, fourth value 4)
Psi = a 2 by 2 symmetric matrix
Theta = a 6 by 6 diagonal matrix
Nu = a column of 6 values (the means – as intercepts)
Alpha = a two by one matrix that contains the means of the two latent variables (by default set to 0)

g. Run the model (by pressing the button RUN) and describe the model fit. What can you conclude about this model?
A:
Chi-square = non-significant, i.e., good model fit
RMSEA = lower than .05, i.e., good model fit
CFI / TLI = both higher .95, i.e., good model fit
SRMR = lower than .08, i.e., good model fit
→ The model fits very well! (Chi-square for baseline is ignored, only used to compute SRMR)

h. Is there a need to consider Modification Indices? Explain your answer.
A: Absolutely not! The model already fits very well, and modification indices is basically an exploratory approach to see how model fit could be improved *in future research*.

i. Consider the parameter estimates: Are they all significant? If a parameter is non-significant, what does that mean?
A: All but the intercepts and covariance between the two factors are significant. If a parameter is non-significant, it means that the parameter does not add to the model fit.

j. Adjust your model in light of the previous question. TIP: Check the options for the Model Command in the *Mplus User's Guide* (see pdf version the website of Mplus; chapter 17); this provides information on how to fix a specific parameter to a certain value, (e.g., @1 to fix at the value 1 or @0 to fix at the value 0), and how to refer to covariances and means in the model. How did you adjust your model?
A: Because the covariance is not significant, we probably should not allow the factors to covary and instead run an orthogonal factor analysis. Furthermore, the intercepts don't seem to matter so we should fix these to zero / not estimate the mean structure.
However, not allowing the factors to covary, means that the whole covariance matrix will be 0 (so none of the items from different factors would theoretically be allowed to covary then either) - a very 'courageous' statement to make that might not ever hold in reality.

k. Run the adjusted model. How many parameters does it have? How can you compare this model to the previous model? Compare the models and report your findings.
A: Now it has 12 free parameters (as we removed 1 covariance and 6 intercepts: 19 – 7 = 12). We can compare this model to the previous model through information criteria.
If the models are nested (which they are, the second model is a special case of the first model), we could also use a chi-square difference test.

Comparing the models:
Log likelihood values are equal for both values, however, the second model uses 7 less free parameters. Hence, the AIC of the second model is 9851.218 – 9837.557 = 13.66 lower (with 7 df, this is very significant!), so we should keep the more parsimonious model.

Chi-square difference test:
DeltaChi^2 = 6.6 - 3.9 = 2.7, with df = 7… this is not significant (*p*-value = .91). Hence, we should keep the more parsimonious model.

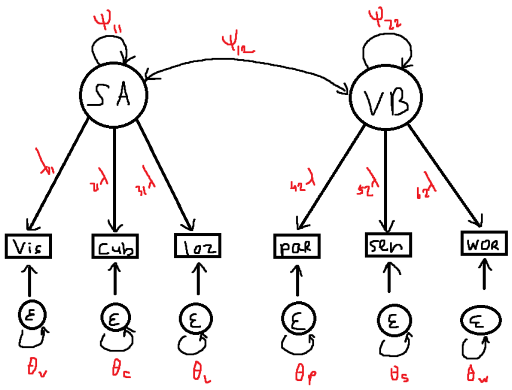
Exercise 2

Consider the SPSS file Grnt_mal.sav. This file contains the scores of 72 boys on 6 measures:

visperc	Visual perception scores
cubes	Test of spatial visualization
lozenges	Test of spatial orientation
paragraph	Paragraph comprehension score
sentence	Sentence completion score
wordmean	Word meaning test score

It is assumed that these 6 variables measure two factors, that is Spatial Ability, which is measured by the first three variables, and Verbal Ability, which is measured by the last three variables.

a. Draw the path diagram that represents this model.



b. Write down the model in regression equations.

$$\begin{bmatrix} vis_i \\ cub_i \\ loz_i \\ par_i \\ sen_i \\ wor_i \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} [SA_i \quad VA_i] + \begin{bmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \\ \epsilon_{6i} \end{bmatrix}$$

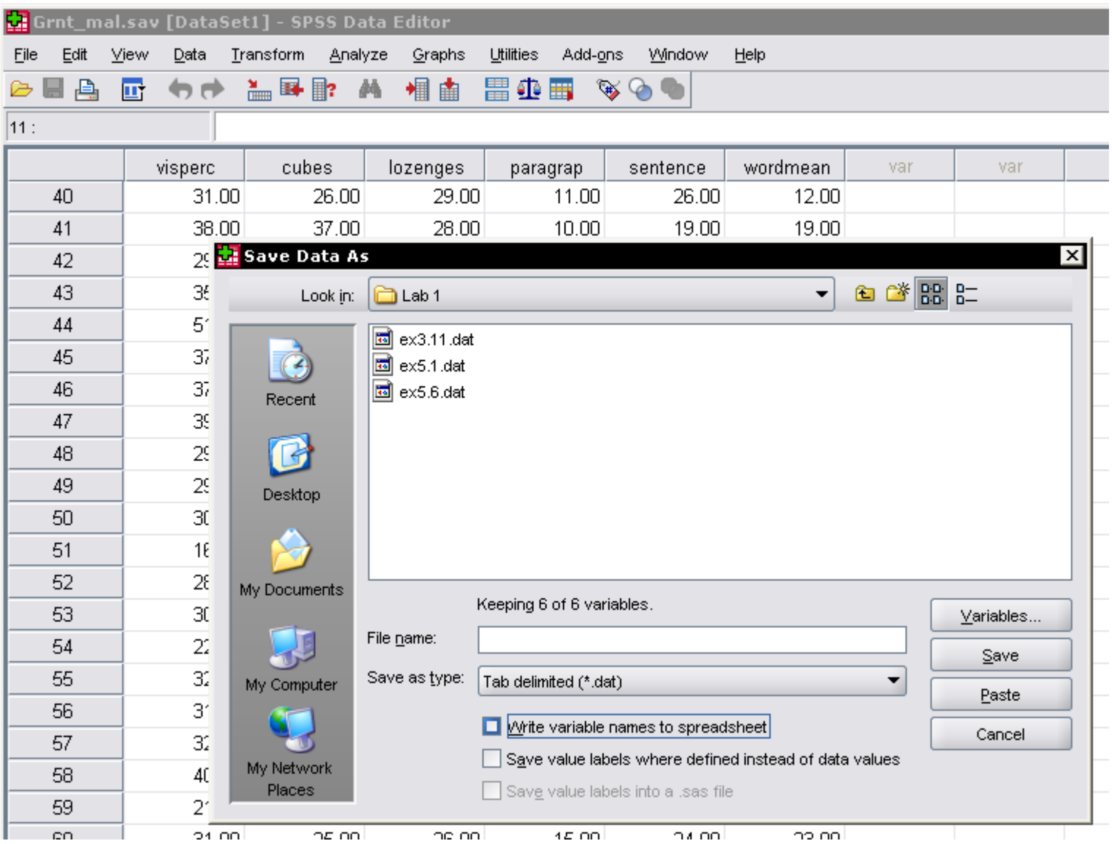
I forgot the intercepts, so there should be a column of 6 nu's in there. Also I didn't state factor loadings as 1 where necessary.

c. How many observed statistics are there (which ones)? How many parameters are there (which ones)? So how many df does this model have?
A: Sample statistics: means: 6 + unique covariance matrix elements:(6 * 7 / 2) = 21
➔ Total: 6 + 21 = 27
Free parameters:

- 4 factor loadings (2 fixed to 1)
- 6 residuals
- 6 means (if not specified otherwise)
- 2 factor variances
- 1 factor covariance (if not specified otherwise)

➔ Total: 16 + 3 = 19
Degrees of freedom: 27 – 19 = 8 (unless otherwise specified)

d. Save the file in SPSS as a tab delimited dat-file (see below). Make sure that you do not save the variable names (uncheck the box). Save the data in the folder you use to run your Mplus models from.



e. Specify the two-factor model in Mplus and run this model. Report on the model fit. Is there a need to modify the model? Explain your answer.
A: Model specification:

```
TITLE:  Exercise 2C - Lab 1
DATA:  FILE IS Grnt_mal.dat;
VARIABLE:       NAMES ARE y1-y6;
```

```
MODEL:
  f1 BY y1-y3;
        f2 BY y4-y6;
```

Model fit:
Chi^2 p-value = 0.3640
RMSEA = 0.036
CFI = 0.995
TLI = 0.990
SRMR = 0.041
➔ Very good model fit!

Is there a need to modify the model?
A: None at all! The model fits very well to the data.
However, something to take into account is that the sample size is quite small and hence we do not have a lot of power to reject the null hypothesis (i.e., to reject the two factor model))

f. Based on the parameter estimates, would you say Spatial Ability and Verbal Ability are strongly related? Explain your answer.
A: They most certainly are, the parameter estimate for the covariance between the factors was 6.840 (SE = 2.543) with a p-value of 0.007. In other words, they are significantly related to one another. However, we do not know the strength of the relationship (for this we need the correlation).

g. Run the model again including:

```
OUTPUT: TECH1 STDYX TECH4;
```

to obtain the standardized results. Is this helpful with respect to the previous question? Explain your answer.
A: It is helpful! Now, we can see that the estimated correlation (and other standardized results) between the two factors is 0.650, which denotes a strong relationship between two variables.

h. As explained in the lecture, each latent variable (or factor) needs to be scaled. Scaling can be done by fixing a factor loading to 1, and this is what Mplus does by default. Compare the following ways of specifying the two factor model:

Method 1:
```
MODEL:
        SA BY visperc cubes lozenges;
        VA BY paragrap sentence wordmean;
```

Method 2:
```
MODEL:
       SA BY cubes visperc lozenges;
        VA BY paragrap sentence wordmean;
```

Are these models different or the same? Explain your answer.
A: They are similar, but not completely. This is because only the first variable has its factor loadings scaled. And in both models the first variables are different, in other words, there is other scaling which can affect the point estimate results of factor loadings (but model fit and residual variances will remain the same).

i. You can also scale a factor using the factor variance, rather than one of the factor loadings. This implies you have to overrule the default of fixing the first factor loading to 1, which you can do by saying:
```
MODEL:
        SA BY visperc* cubes lozenges;
        VA BY paragraph* sentence wordmean;
```

(the asterix implies this parameter—factor loading here—will be estimated freely. In addition, you need to fix the variances of the factors to a positive number (for instance 1), which you can do by adding this to the MODEl statement:

```
SA@1;
VA@1;
```

Run this model, and discuss the results.
A: This model again has the exact same model fit (i.e., statistically identical) but the parameterization is different.

j. What can you say about the covariance between the factors when using this model specification?
A: This is now equal to a correlation and can be interpreted as such.

k. The one-factor model is nested under the two-factor model. You can specify a one-factor model in Mplus, and use a chi-square difference test to compare the one-factor model to the two-factor model. Alternatively, you could constrain a parameter in the previous model such that it becomes equivalent to the one-factor model. Which parameter is this, and to what value should it be constrained?
A: If you scale the covariance between the two factors to 1, you are implying that it is a one factor model (as they are now exactly the same variable).

You can check whether this is true by comparing the two factor model with covariance = 1 to a one factor model, if they have the exact same chi-square value, you know that the models are nested in one another. (Or use the analysis syntax in Mplus)

l. Use both approaches and compare the results (model fit should be the same). Also perform a chi-square difference test to determine whether this constraint can be imposed. Explain your conclusion.
A: The constraint should not be imposed as the chi-square is significant. In other words, we should use the less parsimonious model (i.e., not fix covariance to 1 but letting it vary) —> i.e., it is better a two factor model than a one factor model.

Exercise 3

Consider Example 5.6 from Mplus. This is a higher order factor model, where the four first-order factors are indicators of the second-order factor. Such higher-order factor models are for instance used in intelligence research, where diverse abilities (verbal ability, spatial ability, working memory) are modeled as first order factors, which are in turn indicators of a second order factor which is interpreted as general intelligence.

a. How many sample statistics are there (which ones)?
A: Sample statistics = 12 + 12 * 13 / 2 = 12 + 78 = 90

b. Draw a path diagram of the model used in exercise 5.6, and indicate where the free parameters are.
A: It'll look like a pyramid. Important to note is that the first order factors will also now have residual variances (denoted by Zeta)!

c. How many free parameters are there (which ones)? So how many df are there?
A:
Free parameters =
- 8 factor loadings in items (4 fixed to 1)
- 3 factor loadings in first order factors (1 fixed to 1)
- 4 residual variances in first order factor
- 12 residual variances in second order factors
- 12 intercepts (mean structure)
- 1 variance in second order factor

—> Summed up: 40 free parameters

DF = Sample stat - Free par = 90 - 40 = 50 df

d. If you run the model and ask for the TECH1 output, this is what you get:

```
TECHNICAL 1 OUTPUT


     PARAMETER SPECIFICATION


           NU
              Y1            Y2            Y3            Y4            Y5
              ________      ________      ________      ________      ________
      1           1             2             3             4             5


           NU
              Y6            Y7            Y8            Y9            Y10
              ________      ________      ________      ________      ________
      1           6             7             8             9            10


           NU
```

	Y11	Y12
1	11	12

LAMBDA

	F1	F2	F3	F4	F5
Y1	0	0	0	0	0
Y2	13	0	0	0	0
Y3	14	0	0	0	0
Y4	0	0	0	0	0
Y5	0	15	0	0	0
Y6	0	16	0	0	0
Y7	0	0	0	0	0
Y8	0	0	17	0	0
Y9	0	0	18	0	0
Y10	0	0	0	0	0
Y11	0	0	0	19	0
Y12	0	0	0	20	0

THETA

	Y1	Y2	Y3	Y4	Y5
Y1	21				
Y2	0	22			
Y3	0	0	23		
Y4	0	0	0	24	
Y5	0	0	0	0	25
Y6	0	0	0	0	0
Y7	0	0	0	0	0
Y8	0	0	0	0	0
Y9	0	0	0	0	0
Y10	0	0	0	0	0
Y11	0	0	0	0	0
Y12	0	0	0	0	0

THETA

	Y6	Y7	Y8	Y9	Y10
Y6	26				
Y7	0	27			
Y8	0	0	28		
Y9	0	0	0	29	
Y10	0	0	0	0	30
Y11	0	0	0	0	0
Y12	0	0	0	0	0

THETA

	Y11	Y12
Y11	31	
Y12	0	32

ALPHA

	F1	F2	F3	F4	F5
1	0	0	0	0	0

BETA

	F1	F2	F3	F4	F5
F1	0	0	0	0	0
F2	0	0	0	0	33
F3	0	0	0	0	34
F4	0	0	0	0	35
F5	0	0	0	0	0

PSI

	F1	F2	F3	F4	F5

F1	36				
F2	0	37			
F3	0	0	38		
F4	0	0	0	39	
F5	0	0	0	0	40

Can you relate the parameters in these matrices to the parameters in the path diagram you drew? Indicate what is new.

A: In this output you can see exactly where the free parameters that are estimated are! So you see 12 thetas for residual variances for items, 12 nu's for intercepts for items, 8 factor loadings in the item, 3 factor loadings (beta - NOT lambda, because it is relating a latent variable to another latent variable) for the second order factor, Psi: 5 variances in the factors (first four are now the residual variances of the first order factors! Last one is still the variance of the second order factor)

e. If we are interested in the covariance matrix or correlation matrix of the first-order factors, how can we get these?

A: You can get these by in the output argument stating TECH4;

Appendix

RCode for the regression equation for question 2b:

```
$$ \begin{bmatrix}
vis_i \\
cub_i \\
loz_i \\
par_i \\
sen_i \\
wor_i
\end{bmatrix}
= \begin{bmatrix}
\lambda_{11} & 0 \\
\lambda_{21} & 0 \\
\lambda_{31} & 0 \\
0 & \lambda_{42} \\
0 & \lambda_{52} \\
0 & \lambda_{62} \\
\end{bmatrix}
\begin{bmatrix}
SA_i & VA_i
\end{bmatrix} +
\begin{bmatrix}
\epsilon_{1i} \\
\epsilon_{2i} \\
\epsilon_{3i} \\
\epsilon_{4i} \\
\epsilon_{5i} \\
\epsilon_{6i} \\
\end{bmatrix}
$$
```