Practical sessions: Structural Equation Modeling in M*plus*.

EXERCISE 1 (EFA): Thurstone's study of mental abilities
(THUR.dat; script ThurstoneCFA.inp)

**Note: if you only have the demo version of Mplus, modify the syntax to include only first 6 variable, and attempt to extract 1 and 2 factors

This is a classic study of mental abilities by Thurstone. We have 9 subtests (continuous variables) measuring 3 mental abilities:
Subtest 1 – subtest 3 measure Verbal Ability
Subtest 4 – subtest 6 measure Word Fluency
Subtest 7 – subtest 9 measure Reasoning Ability
The objective is to run an exploratory factor analysis with the Maximum Likelihood method.
We do not have individual data, instead we have a correlation matrix in file THUR.dat, and we know that the data was obtained from N=215 individuals.

Before you can analyze this matrix in M*plus*, the file must be declared in Mplus syntax. Declare its type, and how many observations it was based on. Describe the variables to use in the analysis.
Now we are ready to specify the analysis.

1. Specify EFA with 1, 2 and 3 factors.
2. Specify rotation, in the first run orthogonal, and then oblique.
3. Ask for TYPE=PLOT2 in PLOT command (it will include a scree plot), and for RESIDUALS in the OUTPUT command.
4. Run the script. Examine the scree plot, and then assess chi-square statistics for 1-factor, 2-factor and 3-factor models. How many factors do you think are necessary to explain the observed correlations?
5. Examine residuals for all models and interpret the results; repeat for both orthogonal and oblique run.
6. If you decide on three factors, examine and interpret their correlations to assess how strongly the three constructs are related.

EXERCISE 2 (CFA): Thurstone's study of mental abilities
(THUR.dat; script ThurstoneCFA.inp)

**Note: if you only have the demo version of Mplus, modify the syntax to include only first 6 variable, and attempt to fit 2 factors

We used this study for EFA.
The objective here is to test a confirmatory factor analytic model with the Maximum Likelihood method. Same as before, we have a correlation matrix in file THUR.dat, and we know that the data was obtained from N=215 individuals.

Before you can analyze this matrix in M*plus*, the file must be declared in Mplus syntax. Declare its type, and how many observations it was based on.
Now we are ready to specify the model.

1. Specify 3 latent factors with 3 indicators each (use BY statements).

2. Specify a model with uncorrelated factors.
3. Ask for RESIDUALS in the OUTPUT command.
4. Run the script. Examine the output; assess chi-square statistics and standard errors.
5. Examine residuals and interpret the results.
6. Now specify correlated factors and repeat steps 4-5.
7. Examine and interpret factor correlations.

EXERCISE 3 (Regression for continuous data): Relationships between body measurements
(Body.dat; script BodyPartsRegression.inp)

The file Body.dat contains data on two body measurements (height and weight) for 507 individuals. The variables are: age (in years), weight (in kg), height (in cm) and gender (coded 1 for males and 0 for females). We will try to predict weight from the other 3 variables.

1. Describe the data file and the variables. Set up a model regressing weight on height, age and gender. What are your predictions for relationships?
2. Inspect the Mplus output for regression parameter estimates. Interpret results.
3. Request the sample statistics output, and the standardized output. Interpret results.

EXERCISE 4 (Path analysis): Mediated relationships between body measurements
(Body.dat; script BodyPartsRegression.inp)

The file Body.dat contains data on two body measurements (height and weight) for 507 individuals. The variables are: age (in years), weight (in kg), height (in cm) and gender (coded 1 for males and 0 for females). This time, we explore the relationship between gender and weight mediated by height.

1. Describe the data file and the variables.
2. First, test a model where weight depends on age and height only, and height depends on gender, i.e. there is only a mediated relationship between weight and gender. What are your expectations for this model, based on the results of the previous exercise? Inspect the output and interpret the results.
3. Now include a direct path from gender to weight. Inspect the output and interpret the results.

EXERCISE 5 (Logistic Regression + DIF): Predicting success on ability item from the test score and group membership
(AbilityDIF.dat; script AbilityRegressionLog.inp)

The file AbilityDIF.dat contains ability test data for 1000 individuals, including responses to 20 items (coded 1=correct, 0=incorrect), and group membership (coded 0=majority, 1=minority). First, e will try to predict probability of correct response to item 10 based on the ability score for the whole test. Next, we will examine if the group membership adds to this prediction.

1. Describe the data file and the variables. Declare i10 variable as CATEGORICAL.

2. Use DEFINE command to compute a new variable *ability*, which is the sum of all item responses excluding the item of interest – item 10 (this sum score represents the number of correct answers on the test). Add this newly created variable to the USEVARIABLES list.
3. Specify ML estimator (with categorical dependent variable, it will use the logistic link by default).
4. First, set up a model regressing *i10* on *ability* score and *group*, fixing the regression path on group to 0. Take notes of log likelihood and number of free parameters. Inspect the Mplus output for regression parameter estimates, and odds ratio. Interpret results.
5. Next, modify the model by freeing the regression path on group. Take notes of log likelihood and number of free parameters. Inspect the Mplus output for regression parameter estimates. Interpret results.
6. Which group has an advantage on this item? Look at the odds ratio for the group variable, and interpret. What does the direct path from group variable to the item response mean?

EXERCISE 6 (Factor analysis with covariate - MIMIC model): Inductive reasoning
(IndReasoning.dat; script IndReasoningCFAcovariate.inp)

The file IndReasoning.dat holds the data on five testlets (a, b, c, d and e) measuring one common factor – inductive reasoning ability. The test was in English. The last variable *nat_eng* indicates whether the respondent was native English speaker (coded 1) or not (coded 2). There are few missing values for the grouping variable (coded .), for which we will exclude whole cases setting LISTWISE=ON in the DATA section.

1. Specify the variables for analysis, and code for missing values.
2. Specify a one-factor model, and regress the ability factor on grouping variable.
3. Examine the results and interpret the regression parameter. Is there difference between native and non-native speakers? Who does better?
4. Now add a direct regression path from *nat_eng* to testlet *d*. Examine the results. Is this path significant? Which group does better on the testlet? What effect did this have on the overall difference in ability?

EXERCISE 7 (Factor analysis with covariate - MIMIC model): Inductive reasoning
(IndReasoning.dat; script IndReasoningCFAmultigroup.inp)

The file IndReasoning.dat holds the data on five testlets (a, b, c, d and e) measuring one common factor – inductive reasoning ability. The test was in English. The last variable *nat_eng* indicates whether the respondent was native English speaker (coded 1) or not (coded 2). There are few missing values for the grouping variable (coded .), for which we will exclude whole cases setting LISTWISE=ON in the DATA section.

1. Specify the variables for analysis, and symbol used for missing values.
2. Specify a one-factor model, and regress the ability factor on grouping variable = *nat_eng*. 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 any of the

imposed constraints too restrictive for the model? (Hint: inspect modification indices for this).

3. Free the constraints with the highest modification index. Estimate the model again, and carry out a chi-square difference test to see if the released constraint improves the fit significantly.
4. Specify a model with latent factor variances constrained to be the same across groups. Does this model fit?
5. Now fix both group means at zero. Does this model fit? Interpret the overall result of the analysis.

Variation:

The testlet scores were compiled from 3 binary items each, and may therefore be treated as categorical. Repeat the analysis above declaring testlet scores categorical.

EXERCISE 8 (Latent Class Analysis): Macready and Dayton's Mastery model (Mastery.dat; script LCAmastery.inp)

Here we will test Macready and Dayton's Mastery model of performance. Four test items selected at random from a domain of items testing mastery in the multiplication of a two-digit number by a three- or four-digit number. Responses to items are coded as 0=fail, 1=pass. Based on their scores on the items, N=142 respondents are expected to belong to one of the two groups: Masters and Non-Masters.

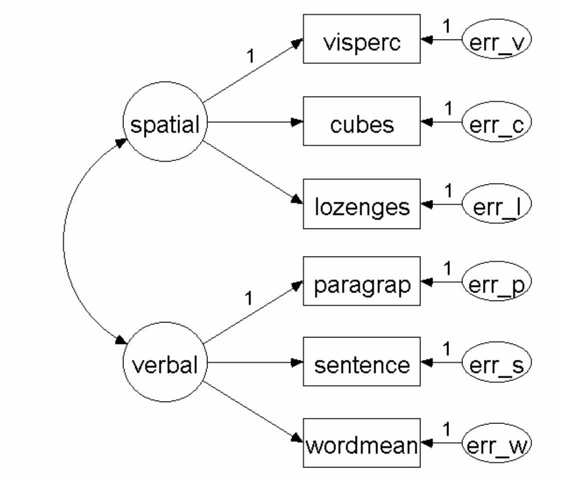
1. Specify the data file and the variables. In this file, item responses are not separated by blanks so FORMAT statement must be used.
2. Specify 2 latent classes. Examine the model fit and the model parameters (unconditional and conditional probabilities).
3. Request the plot of conditional probabilities of item responses, and examine them.
4. Request to save the estimates of class membership and probability for each individual. Examine the saved file.

Variation:

Try to fit the latent trait model with one common factor (CFA model with categorical variables, or IRT model) to the Mastery data. The correct script is in IRTmastery.inp. How would you interpret this model and its results?

OPTIONAL EXERCISE (Target rotation): Cognitive ability data
(Holzinger.dat; script HolzingerEFAtarget.inp)

The file Holzinger.dat holds the data on six cognitive ability tests (with the verbal & spatial latent factor). The model is given by the following diagram.



There is a gender variable in the data file (last variable), coded 1=girl, and 2=boy.

1) Describe the data and the variables for analysis.
2) First run the EFA including boys only (use USEOBS for that). Assess model fit – 2-factor model should fit data well. Pick two factor loadings that are close to 0.
3) Now select girls only, and perform EFA target rotation (do not use EFA command, use MODEL command) with the target loadings equal to the two loadings you picked from the boys' analysis.
4) Compare other loadings between the boys and girls – are they similar?