

Introduction to Mplus Syntax

For much more documentation, examples, and help, go to: <http://www.statmodel.com/index.shtml>

- Basic rules about Mplus and its syntax and output:
 - Two kinds of files, both of which are just text files (so you can open them in notepad or word at home):
 - **.inp** is for input (syntax), and **.out** is for output (the input syntax will re-appear at the top)
 - Can have 90 characters in each syntax line, otherwise Mplus will get mad (check **column #** in bottom right)
 - Mplus commands are **blue** and end in **colons**; subcommands are **black** and end in **semi-colons**
 - Comments are indicated with **!** at the beginning of each so that the text that follows will turn **green**. There are no block comments, so you have to repeat the **!** for each line
 - There is no way of selecting parts of code to run one-at-a-time – the entire file is run each time. So if you want to only run certain models, you can comment out the others via the **!** for each line
 - The terms **IS**, **ARE**, and **=** in defining suboptions (see below) are interchangeable
 - You can abbreviate variables using the **ALL** keyword or using a dash for contiguous names (e.g., var1-var10 indicates var1, var2, var3.... var10)
- Getting data into Mplus:
 - There is no provision for viewing data within Mplus. For this reason, although it is possible to perform variable transformations within Mplus (e.g., centering, stacking/unstacking data), it will be easier to make sure this is done correctly if you do so in the native file (e.g., SPSS or SAS) in which you can see the data.
 - You will need to write in the syntax the names of all variables in the data, so it will be more convenient if you limit your dataset for importing to just those variables you will need for your analyses. Mplus does not handle string variables or dates well in my experience, so remove those variables especially.
 - **All variable names must be 8 characters or less.** Mplus does not know what the variables were called within the original data, so you can call them whatever you want in Mplus.
 - Beware of missing data! Although Mplus accepts “blank” as a missing data indicator, this tends not to work as well as defined missing data codes (e.g., -999). It will be easiest if all variables have the same missing data code.
 - Mplus only accepts tab-delimited files (**.dat**), fixed-format text files (**.dat**), or comma-separated-values files (**.csv**). Of these, **.csv** is nice because it opens by default in excel so that you can check it as needed. If this fails, you can try the other formats, with fixed-format as the last resort.
 - Once you have taken care of these issues, to save data from SPSS .sav to a .csv for use in Mplus:
 - Through the windows: FILE → SAVE AS → change “Save as type” box to “comma delimited (*.csv)” and UNCHECK the box that says “Write variable names to spreadsheet”. Mplus does NOT read variable names from the file, and will give you an error message if they appear on the first line instead of data.
 - In SAS, you can use PROC IMPORT, in which DATA tells it which SAS file to export, OUTFILE lists the path and name of the new .csv file, REPLACE means it will be replaced if a file already exists with that name, and PUTNAMES=NO tells it not to write the names to the top of the .csv file.

```
PROC EXPORT DATA=work.mydata OUTFILE= "F:\folder\mydata.csv"  
DBMS=CSV REPLACE; PUTNAMES=NO; RUN;
```
 - When you begin an Mplus analysis, the VERY FIRST THING YOU SHOULD DO is make sure your data were imported correctly. You can use TYPE IS BASIC analysis option or the SAMPSTAT output option to do so (see example). It will provide means, variances, and covariances for your variables. These should match those of your original SPSS or SAS data exactly— unless you have missing data, in which case they will not be exact. But if they are not even close, then chances are it is not reading your data correctly yet. Verify that you have listed the variable names in the order in which they appear in the data and that you didn't forget to list any variables.

- Mplus commands:
 - **TITLE:** This is optional, but it will print at the top of your output (just like in SPSS/SAS)
 - **DATA:** This is NOT optional, and here you tell it where your data are and their format. It has the following subcommands:
 - **FILE** gives the name of the data file, plus path if the syntax is not stored in the same place as the data:
 - FILE IS mydata.csv; or FILE IS F:\OtherFolder\mydata.csv;
 - **FORMAT** tells it whether you have delimited (free) or fixed-format data
 - If fixed-format, you need to give the formatting statement, for example if you had one ID variable with 4 characters and no decimals, followed by 24 variables with 4 characters, 2 of which were decimals:
 - FORMAT IS free; or if fixed: FORMAT IS F4.0 24F4.2;
 - **TYPE** tells it whether you have individual data (default) or summary data (e.g., covariance matrix)
 - TYPE IS individual; or TYPE IS fullcov;
 - **VARIABLE:** This is NOT optional, and here you tell it what your variables are called and which ones you are using for what purpose:
 - **NAMES ARE** is where you list ALL the variables in the DATASET
 - **USEVARIABLES ARE** is where you list JUST the variables in the MODEL
 - **USEOBSERVATIONS ARE** is how you tell it to select only certain cases, for example:
 - USEOBSERVATIONS ARE gender EQ 1;
 - **MISSING ARE** is how you tell it how missing data are indicated, for example:
 - MISSING ARE ALL (-999);
 - **IDVARIABLE IS** tells it what variable indicates which case is which – if you add this option, then any outputted dataset you ask for will have this variable included, which is necessary if you want to merge any outputted data from Mplus into the original data
 - IDVARIABLE IS PersonID;
 - **AUXILIARY ARE** is how you tell it to keep the other variables not currently in the model in any outputted datasets, or to use those variables in the missingness model
 - AUXILIARY ARE gender age group;
 - In addition, the following options are specific to multilevel models in Mplus:
 - CLUSTER IS tells it how level-2 units are organized (e.g., PersonID, GroupID)
 - WITHIN is used to identify predictors to be included ONLY at level 1
 - BETWEEN is used to identify predictors to be included ONLY at level 2
 - If a variable is measured at level 1 and has variance at both levels for which you want to estimate level-1 residual and level-2 random intercept variances, then do NOT include it on WITHIN
 - **DEFINE:** This is optional, and is how you make new variables out of existing ones. You then need to list any newly created variables on the USEVARIABLES list after the existing ones.
 - For example, to create a new predictor centered at 30:
 - DEFINE: c_pred = pred – 30;
 - To recode a predictor, EQ is used to evaluate an equality, = is used to transform data:
 - DEFINE: IF oldvar EQ 1 THEN newvar=0; IF oldvar EQ 2 THEN newvar=1;

- **ANALYSIS:** This is not usually optional, and is how you set any estimation options. These will be model-specific for the most part, and include suboptions for TYPE, ESTIMATOR, and so forth.
 - Here's how you get the descriptive statistics to check your data: TYPE IS BASIC;
- **OUTPUT:** This is where you request specific pieces of output not provided by default, such as standardized coefficients, analysis of results, and sample statistics for checking the data. This will vary by model.
- **SAVEDATA:** This is how you ask for outputted datasets, such as factor scores or model parameters.
- **PLOT:** There are many kinds of plots available in Mplus. To get all of them you could ever get for any model:
 - PLOT: TYPE IS PLOT1 PLOT2 PLOT3;
- **MODEL:** This is model-specific and based on three key words:
 - **BY** is to define latent variables, for example: factor BY var1-var10;
 - **ON** is to define regression slopes, for example: dv ON pred1 pred2 pred3;
 - **WITH** is to define covariances, for example: dv1 WITH dv2;
 - In addition, free (estimated) parameters are indicated with a * and fixed parameters with @
 - **XWITH** is used to specify interactions involving latent variables:
 - Interaction between two latent variables: F1F2int | F1 XWITH F2;
 - Interaction between one latent and one observed variable: F1x1int | F1 XWITH x1;
 - To define a random slope in the level-1 part of a multilevel model:
 - Slopenam | dv ON pred; where "slopenam" is your own name that gets used at level 2
 - Mplus knows based on your model whether each variable is a predictor or an outcome. Variables that are only predictors can have estimated means and variances. Variables that are outcomes (even if they are also predictors) can have estimates intercepts and residual variances. Either way, the syntax to refer to these parameters looks the same.
 - To refer to intercepts or means, use []:
 - For example, estimate means/intercepts for var1 and var2 only: [var1* var2* var3@0];
 - To refer to residual variances or variances, just list the variable:
 - For example, estimate residual/variances for var1 and var2 only: var1* var2* var3@0;
 - For example, to estimate a covariance between var1 and var2, regardless of whether they are both predictors, both outcomes, or a mix of the two: var1 WITH var2*;
 - To estimate all possible covariances of var1-var4: var1 var2 var3 var4 WITH var1 var2 var3 var4;