Getting Started with Mplus
CSDA Spring 2015 Brown Bag Workshop

Jeffrey Napierala
jnapieral@albany.edu
CSDA & Dept. of Sociology

What you will learn:

- Some basic information about the program:
  - How you can access it; what it can do; how it is designed; where to go for more info.
  - How to write basic code
  - How to get your data into Mplus
    - I will work through a simple example.
  - What basic analysis programs and output looks like and how to read and interpret them.
About Mplus

- It is a statistical software program designed specifically for analyses with latent variables. It can handle continuous, censored, nominal, ordinal, and discrete variables with a number of different estimators.
  - Is a powerful statistical package, but has limited data management capabilities. For this reason, it is often used in conjunction with another package.
  - Developed and managed by UCLA Statistician Bengt Muthen and Dr. Linda Muthen. First released in 1998.

About Mplus: Access

- The program is installed on CSDA Lab computer #8 in BA B-18 (just across the hall).
  - It is the Windows version with a full “combination add-on” and the Diagrammer program.
- If you are interested in your own copy… it is expensive: $600 for base program, but up to $900 for both add-ons. It is about 1/3rd of that price for students. There is a demo version, but you are very limited in the number of variables you can use (more info at www.statmodel.com).
  - Note: They only supply updates and support for one year; you need to pay for extended updates and support, but can apparently still use the software. From what I have seen they are very responsive (less than 24 hours) in answering questions on the discussion boards!
About Mplus: Statistical framework

- It is based around an integrated statistical framework that includes continuous and categorical latent variables:
  - Observed variables are in squares and latent variables are in circles.
  - x: “background” variables
  - y: continuous and censored outcome variables
  - u: binary, count, ordinal, or nominal outcomes
  - f: continuous latent variables
  - c: categorical latent variables

---

About Mplus: Statistical framework

- Statistical Concepts Captured By Latent Variables:
  - Continuous (latent variables)
    - Measurement errors
    - Factors
    - Random effects
    - Frailties, liabilities
    - Variance components
    - Missing data
  - Categorical (latent variables)
    - Latent classes
    - Clusters
    - Finite mixtures
    - Missing data

Source: Statmodel.com

Source: Muthen and Muthen (2008):
About Mplus: Some capabilities

- Exploratory and confirmatory factor analysis
- Structural equation modeling
- Latent class analysis
- Mixture modeling
- Growth modeling (in SEM framework)
- Multilevel analysis
- Item response theory analysis
- Longitudinal mixture modeling (hidden Markov, latent transition analysis, latent class growth analysis, growth mixture analysis)
- Survival analysis (continuous- and discrete-time)
- Bayesian analysis
- Monte Carlo simulation
- Complex survey data analysis (clustering, stratification, replicate weights...)
- Methods for missing data under MCAR, MAR, and NMAR and with multiple imputation
- Bootstrapped standard errors and confidence intervals for most models; robust s.e.'s for all models

About Mplus: some capabilities

- Features for continuous latent variables:
  - Single or multiple group analysis (analysis with same structure on multiple groups e.g. EFA)
  - Latent variable interactions and non-linear factor analysis using maximum likelihood
  - Random slopes (unobserved heterogeneity in effects)
  - Individually-varying times of observations
  - Linear and non-linear parameter constraints
  - Indirect effects including specific paths
  - Maximum likelihood estimation for all outcomes types
  - Wald chi-square test of parameter equalities
  - Factor scores and plausible values for latent variables
About Mplus: some capabilities

- Features for categorical latent variables:
  - Analysis with between-level categorical latent variables
  - Tests to identify possible covariates not included in the analysis that influence the categorical latent variables
  - Tests of equality of means across latent classes on variables not included in the analysis
  - Plausible values for latent classes

About Mplus: resources

- www.statmodel.com - Official Website with documentation and other recourses: a discussion board, training videos, example analyses, the user’s manual (a good place to start)...
- www.ats.ucla.edu/stat/mplus/ - UCLA’s Mplus web resources. Videos form seminars, example analyses, comparisons across statistical packages, etc.
Programming language and input programs

- Language is based around 10 commands:
  - TITLE - type whatever you want here and it shows up at the top of the output file. Useful for keeping organized.
  - DATA (required) - info about the dataset
  - VARIABLE (required) - info on the variables in the dataset; an ordered list of variables in the dataset; which ones you want to use in the analysis; the classification of the DV...
  - DEFINE - recode and create new variables here
  - ANALYSIS - specify technical details of the analysis: the type of analysis; the estimator to use
  - MODEL - define the model here: multiple equations; correlations between variables; between and within effect models...
  - OUTPUT - request additional output such as standardized coefficients and c.i.’s
  - SAVEDATA - save data and results
  - PLOT - get graphical displays
  - MONTECARLO - for simulations

“Input program” = .do file or .sas syntax file (run in batch mode)

- Mplus runs all code from input programs; it is not possible to run code line-by-line as in R, Stata and SAS. The input programs are run through the Mplus engine and an output file is generated (and an diagram if applicable).
- Can write code yourself or it can be generated (and modified if needed) using some tools included with the software
  - Language Generator - A “wizard” that automatically generates code after you to enter in info on your data and analysis using dropdown menus, checkboxes etc.
    - Does not appear to be updated recently (evidently since version 2; they are on 7.3 now)
    - Cannot access all of the features
  - Diagrammer - Allows you to generate code by making diagrams using drawing tools and to view diagrams of the code you have written.
    - Note: this just generates partial code (for the “model command”), you have to do the rest.
- I suggest finding a worked example (there are a lot of them on their website, for example) and work from there.
Programming language and input programs

- Some things to know when writing code:
  - Commands can be written in any order.
  - Commands must begin on a new line and end with a colon.
  - Command options are separated by semicolons (there can be more than one on a line).
  - The language is not case sensitive.
  - Comments start with an explanation point (everything on the line afterwards is ignored).
  - Commands and Command Options can be shortened to 4 or more letters.
  - Lines cannot have more than 90 characters (can happen with heavily commented input files and long variables lists); variables are limited to 8 characters.
  - You must save the input program before running it every time.
    - Tip: Create multiple input programs as you make major changes to keep track of work you have done.

Getting your data into Mplus

- Data files for Mplus are just plain ASCII text files. Can be space, comma, tab delimited (“free format”) or “fixed format”—data that is organized by column number. Can also enter data manually.
  - It must be numeric data.
  - Remember: variables names must be 8 characters or less. The program will not run correctly if you try to put longer ones in.
  - Can be correlational data. You need to specify this.
  - Note: many data files will have variables names/labels in the first row(s) of the file, which will generate an error in Mplus. Variable names are specified in the input program.
  - Tip: get rid of any variables you are not using before importing to Mplus.

- ...most problems generate errors and are easy to fix, but always a good idea to compare descriptives before and after importing.
Getting your data into Mplus

There are many ways to read your data into Mplus:

- Use Stattransfer software (available in BA B-18 on the same machine with Mplus) - seems to work ok, but you still may need additional preparation (be careful with missing and character values).
- Write your own input program (it is relatively easy).
- Use a function in your main stat package to write the input program for you (R and Stata).

Getting your data into Mplus: a simple program

Title: ! Notice the colon after the command here and below.
Reading a data file
Data:
File is datafile.csv; ! "=" or "are" can be used in place of "is"; it will figure out your delimiter.
Format is (6F9.0) ! When a path is not given it uses the directory of the input file.
Variable:
Names are id y1 y2 x1 x2; ! The names of all the variables in the file in the order they appear.
Usevariables are y1 x1 x2; ! Save time by selecting a subset of variables (ordering doesn’t matter here)
Missing are (-9999); ! Declare missing values codes if needed. Can declare separate codes for var.
Analysis:
Type = basic; ! The basic analysis provides means, a variance-covariance matrix, and a correlation matrix.
! To use listwise deletion use LISTWISE=ON in the DATA command.
Getting your data into Mplus: a simple program

- Simple (DIY) example in Mplus using the previous program and a comma delimited file.
- Data is from the NLSY.

Getting your data into Mplus: Stata

- Use the “stata2mplus” package (see www.ats.ucla.edu/stat/stata/faq/stata2mplus.htm). Creates a data (.dat) file and an input (.inp) file for Mplus from your Stata (.dta) dataset.
  Sample code:

  Findit stata2mplus
  stata2mplus using C:\Users\jn836464\Desktop\datafile, replace

  This will convert the datafile.dta file and will replace any existing files with the same name.
- Missing values will be automatically converted to -9999 unless the “missing()” command is used to change the value (this is done to smooth the transition to Mplus, which does not read character values). You would want to do this (change the missing value code) if a variable might take on that value.
- You can also select which variables to convert from your Stata data file with “use(varlist...)”.
Getting your data into Mplus: SAS

- SAS: Recode missing values to a unique numeric code and use the export wizard (you will need remove the variable names from the first row—just cut and paste the names into Mplus). Example syntax (see www.ats.ucla.edu/stat/mplus/faq/sas2mplus.htm):

```
data sample;
  set yourdata;
  array allvars _numeric_; *all numeric variables in the array
  do over allvars;
    if missing(allvars) then allvars = -9999; *recode missing cases to -9999
  end;
  proc export data=sample outfile='c:\datafile.dat' dbms=dlm replace;
```

Getting your data into Mplus: R

- One method—use write.table and save it to a file. Something like will work (be aware of your missing values):

```
write.table(data.frame, file="C:/dataset.csv", row.names=FALSE, col.names=TRUE, sep="", quote=FALSE)
```

- Check into the "MplusAutomation" package by Michael Hallquist (see www.statmodel.com/usingmplusvar.shtml). The function “prepareMplusData” will create a .dat file and write syntax to import the data in Mplus.
Data analysis: regressions

- Will walk you through some simple analyses
  - Linear regression (with ML estimation)
  - Probit and logistic regression
  - Path analysis

- http://www.ats.ucla.edu/stat/mplus/seminars/IntroMplus_CFA/