An Introduction to SEM in Mplus

CRMDA Staff[1]

[1]Ben Kite, Aaron Boulton, Terry Jorgensen, Paul Johnson, and others

2019

CENTER FOR
RESEARCH METHODS
& DATA ANALYSIS

**College of Liberal Arts
& Sciences**

Outline

KU

Goals

- Provide an introduction to Mplus (Muthén & Muthén, 2017)
- Companion notes "Mplus_semexamples" is a detailed review of our standard collection.
- Provide Mplus tips and suggestions
- Provide additional sources for information

KU

Outline

KU

What is Mplus?

- Created by Linda K. Muthén and Bengt O. Muthén
- Now in 8th version!
    - Version 8 April 2017
    - Version 7 September 2012
    - Version 6 April 2010
    - Version 5 November 2007
    - Version 4 February 2006
    - Version 3 March 2004
    - Version 2 February 2001
    - Version 1 November 1998

What is Mplus?

- Statistical modeling program
  - Wide choice of models, estimators, algorithms
  - Cross-sectional & longitudinal
  - Single & multi-level models
  - Observed & latent variables
  - Continuous, categorical, binary, censored, count, or combinations of these
  - Missing data
  - Monte Carlo simulation
  - Latent Classes

Some Advantages

- Has been the "state of the art" for more than 1 decade
- Compact syntax (few lines of code)
- Powerful for complex modeling, if Mplus authors have interest in the particular model you want to estimate
- Good user's guide with lots of examples
- Customer support for registered users
- Website has a discussion board open to the public
  http://www.statmodel.com

Some Disadvantages

- Expensive \$\$\$
    - Student Pricing \$195-350
    - University Pricing \$595-895
    - Commercial/Non-Profit/Gov. \$695-1095
- Syntax based. Write Code, Run Models!
- Details can be tricky for new (*or experienced*) users

KU

Outline

Three Steps

1. Get data in a text file
2. Write an Mplus input file that uses the data
3. Run Mplus with the input file to process the input data file.

Mplus Input File

- Can use any "flat text" editor and write a syntax file.
  - Mplus for Windows provides a syntax editor, not much different from Windows Notepad
- Vital: line length limit is 80 characters!
  - Variable names: 8 letters or less, no spaces, only use letters & numbers.
- Suggested to save the file with suffix "*.inp".
- **REMEMBER**: This is a flat text file, not an MS Word Document. OK to edit it with Notepad++, Sublime, Emacs, Eclipse, or any other "programmers file editor".

KU

Running An Mplus file

- In Mplus for Windows, open the *.inp file.
- Click "Run" the run button
    - Creates output in same directory as your inp file
- Mac or Linux: `mplus` is a command line program. In a terminal,

```
$ mplus yourfile.inp
```

- Creates

```
yourfile.out
```

in same directory

KU

Running in the Terminal Looks like this

```
$ mplus cfa−PhysHealth−1factor−ML.inp

    Mplus VERSION 8 (Linux)
    MUTHEN & MUTHEN

    Running input file 'cfa−PhysHealth−1factor−ML.inp'...

    Dimensions of integration:                    1
    Total number of integration points:          15

    ITER    LOGLIKELIHOOD       ABS CHANGE     REL CHANGE    ALGO    TIME    TOTAL TIME
     1   −0.92564910D+05      0.0000000      0.0000000      EM      0.08      0.1
     2   −0.91073391D+05   1491.5194977      0.0161132      EM      0.12      0.2
     3   −0.90853027D+05    220.3636945      0.0024196      EM      0.16      0.4
     4   −0.90807821D+05     45.2062046      0.0004976      EM      0.16      0.5
     5   −0.90791424D+05     16.3971344      0.0001806      EM      0.16      0.7
     6   −0.90784617D+05      6.8068919      0.0000750      EM      0.17      0.9
     7   −0.90781679D+05      2.9384770      0.0000324      EM      0.16      1.0
     8   −0.90780381D+05      1.2974772      0.0000143      EM      0.16      1.2
     9   −0.90779797D+05      0.5837423      0.0000064      EM      0.16      1.3
    10   −0.90779529D+05      0.2679598      0.0000030      EM      0.16      1.5
    11   −0.90779403D+05      0.1260571      0.0000014      EM      0.16      1.7
    12   −0.90779342D+05      0.0611618      0.0000007      EM      0.17      1.8
    13   −0.90779311D+05      0.0308225      0.0000003      EM      0.16      2.0
    14   −0.90779295D+05      0.0162304      0.0000002      EM      0.16      2.2
    15   −0.90779272D+05      0.0230502      0.0000003      FS      0.13      2.3
    16   −0.90779272D+05     −0.0001924      0.0000000      EM      0.33      2.6
     Parameters        6        2         24        16        40
     Derivatives    0.68D−05   0.46D−05   0.42D−05   0.41D−05   0.35D−05
     Beginning Time:  12:05:31
```

Running in the Terminal Looks like this ...

```
        Ending Time:   12:05:35
        Elapsed Time:  00:00:04

    Refer to 'cfa−PhysHealth−1factor−ML.out' for warning(s).
```

Outline

KU

Main Mplus Commands

These are ALWAYS in beginning of a line. They initiate a code block

- Only DATA & VARIABLE commands are required for all analyses

      TITLE:
       DATA: (required)
   VARIABLE: (required)
   ANALYSIS:
      MODEL:
     OUTPUT:
      DEFINE:
  SAVEDATA:
        PLOT:

They are often written in all uppercase to make it easy to read, but it is not required. Mplus is not case sensitive

KU

Basic Syntax Conventions

- ! Can be used to make comments
- End statements with a semicolon;
- 80 character lines!

KU

More Syntax Conventions

- Dash can be used to select consecutive variables
  - var1-var5 same as writing out var1 var2 var3 var4 var5
- Not case sensitive
  - i.e. not picky about upper/lower case letters
- Commands can be in any order
- Commands start on a new line and are followed by a colon
  - (e.g. DATA:)
- Only DATA & VARIABLE commands are required for all analyses
- No more than 90 characters wide

TITLE:

- Not required
- Provide a title for the analysis (for your own use, not part of the model)
- e.g.

```
TITLE: My Analysis of Happiness
```

KU

DATA: file input must be text

- Data should be a text file
  - Use .csv (comma delimined) or .dat (tab delimited) files
- No importer for Excel, Stata, SPSS, etc
- All procedures can utilize
  - Raw "individual level" observation data
- Some can work without individual level data.

KU

Data Files

- Numeric only (NO character variables, NO variable names)
- Maximum of 500 variables
- You need a numeric code for missing values.
    - E.g., -999
    - You cannot have any empty cells!
- No funny characters! No non-ASCII characters, so only numbers, or commas (tabs) for separators.

KU

DATA: file name

- "FILE is" specifies where the data file can be found
- Full Path

```
DATA:
FILE is "C:/mydocuments/happiness.dat";
```

- Relative path within project folder

```
FILE is "data/happiness.dat";
```

- Relative path outside project folder

```
FILE is "../data/happiness.dat";
```

- File name in current working directory (same directory as syntax input file).

KU

DATA: file name ...

```
FILE is happiness.dat ;
```

Note quotation marks are not required if no special path characters (spaces)
need protection

DATA command components

- Default data type is individual-level information.
- Some procedures can run without individual level data, can fit models with
  - covariance matrix (or correlation matrix and standard deviations)
  - Reminder of days when this was called "covariance structure analysis"
- If not using individual-level data, must specify

```
TYPE IS
```

is a keyword that can be followed by "INDIVIDUAL", "COVARIANCE", "CORRELATION", "MEANS" and some others

- Group qualifier

```
NGROUPS=
```

can declare the number of groups for which information is provided. Needed if there are several covariance matrices.

KU

VARIABLE:

- Information about the variables
  Order they appear in set
  No more than 8 characters in variable names

```
VARIABLE:
  NAMES ARE id age sex marital happines;
  USEVARIABLES ARE age sex marital happines;
  CATEGORICAL ARE sex marital; !(not always
    needed)
  MISSING ARE (-999);
```

- **Remember: Error if line longer than 80 characters!**

KU

ANALYSIS:

- Technical details of analysis (not always needed)

e.g.

```
TYPE = MIXTURE ;
```

or

```
TYPE IS TWOLEVEL RANDOM ;
BOOTSTRAP = 500;
```

Double-check the data

- To check data import, request Summary Statistics
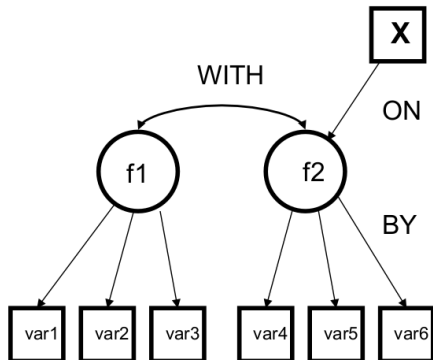
```
ANALYSIS:
  type=basic;
```

- Or in the OUTPUT section, request data datails

KU

MODEL:

Describes the model to be estimated

e.g., in a factor analysis

```
f1 BY var1 var2 var3 ;
f2 by var4-var6 ;
f1 WITH f2 ;
f2 ON X ;
```

MODEL: BY, WITH, ON

- BY – indicator (e.g. factor 1 is indicated BY var1-var3)
- WITH - correlate
- ON - regress DV on IV (order matters)

```
Y on X;
```

- To prevent estimation of a pathway

```
Var1 WITH Var2 @0;
```
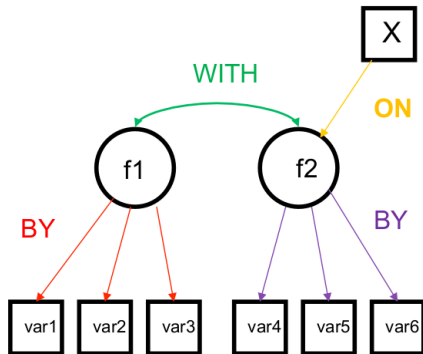
KU

MODEL: with a little color

Describes the model to be estimated

e.g., in a factor analysis

```
f1 BY var1 var2 var3;
f2 by var4-var6;
f1 WITH f2;
f2 ON X;
```

MODEL: A few details

- MPLUS will use its own rules to estimate or fix parameters.
- Can request changes
    - " `*` " is used to free a parameter estimate
    - " `*.5` " indicates parameter is free and .5 is the starting value of the calculator
    - " `@1` " fixes parameter to a specified value, such as "1" in this case.
- Can also use a parenthesized letter (a) for each parameter that is assumed to have the same value.

KU

OUTPUT:

- Mplus will, by display, print results that the authors believe are minimally necessary.
- OUTPUT can be used to gather additional information
- Common Requests

1. Modification Indices

```
OUTPUT:
  MODINDICES;
```

   - Requests modification indices (indicators of whether a coefficient that is not currently estimated might be important to take into consideration).
   - Use ($\#$) to have indices only $\geq \#$ reported (e.g. MODINDICES(10))

2. More technical details. Put these OUTPUT requests:
   1. STANDARDIZED Request standardized estimates
   2. TECH1 technical details for troubleshooting

3. There are many many other OUTPUT options, see Ch 18 in Mplus User Guide.

KU

Outline

Troubleshooting 1

- Read the *.out file. Error message will usually point at first syntax error.
- Check warnings as well

```
Warning Message
e.g. Data set contains cases with missing on all
    variables.
These cases were not included in the analysis.
Number of cases with missing on all variables: 34
```

- Common syntax errors
  - lines too long
  - variable names too long
- Review the MODEL code. Did you ask for something impossible?
  - Draw a path diagram for the MODEL you want, then compare with syntax

KU

Troubleshooting 2

With unexpected results (not errors)

- Check data details, find out if Mplus thinks you gave it the data that you intended. Output has hints about
  - Number of observations
  - Number of variables
- Eyeball your data file for funny patterns
- Check your missing codes

Troubleshooting 3

COMMON Warning Message:

```
The residual covariance matrix (Theta) is not
   positive definite. This could indicate a
   negative variance/residual variance for an
   observed variable, a correlation greater or
   equal to 1 between two observed variables or a
   linear dependency among more than 2 observed
   variables.
PROBLEM INVOLVING VARIABLE VAR8
Check the technical output (e.g. tech1)
```

Troubleshooting 4

Exceeds iterations & did not converge.

Fixes:

- Insert start values close to your best guess for parameters (can start with what the output gives you)
  e.g.,

```
y ON x*0.1;
```

- Increase the number of iterations

```
ANALYSIS: ITERATIONS = 10,000;
```

  (default is 1,000)

- Build a model piece-by-piece
  - your way up to more complex model

KU

Troubleshooting 5

Model is unidentified/under identified

Fixes

- Estimate fewer parameters
- Try fixing one or more parameters to 0 or 1 or to be equal to something else in the model
    - e.g. in a longitudinal model, fix residuals to be equal across all time points.

Additional Help/Resources

- Mplus Website http://statmodel.com
  - Mplus Language Summary (User Guide Chapter20.pdf)
  - Full Copy of User's Guide is downloadable PDF of User's guide and
  - example code with data.
  - Support forum for users
- KU CRMDA Guides (http://crmda.ku.edu/guides)

KU

Outline

The Olde Time Feeling

- People who used Microsoft DOS in the 1990s may have the "happy deja vu" feeling in Mplus
- Limitations that we left behind
  - limitations on the number of letters in words
  - line length of input files
    are returned to prominence
- Input and output are raw unstructured ASCII text.
  - Using Mplus is rather like writing letters to Mplus
    - And (frequently) getting answers.

What are the Compensating Benefits?

- The Mplus development team as at the cutting edge of SEM statistical modeling.
- They develop new methods, write their articles, and then incorporate new routines in their Mplus package.
- Especially in the case of structural equations with categorical variables, Mplus is widely considered to be the state of the art.

KU

What are the Concerns?

- Expensive, difficult to use, tedious output framework
- Closed source methods & Non-reproducability.
  - Even when there are technical documents to explain methods of calculation, our opinion is that details are obscured (sometimes omitted entirely) (possibly to preserve commercial advantage of the Mplus corporation).

KU

References

Muthén, L. K. & Muthén, B. O. (2017). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén, 8th edition.

KU

Session

```
sessionInfo ()
```

```
R version 3.6.0 (2019 -04 -26)
Platform : x86_64 -pc -linux -gnu (64 -bit)
Running under : Ubuntu 19.04

Matrix products : default
BLAS:    /usr/lib/x86_64 -linux -gnu/atlas/libblas.so.3.10.3
LAPACK : /usr/lib/x86_64 -linux -gnu/atlas/liblapack.so.3.10.3

locale :
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
     LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8     LC_MONETARY=en_US.UTF-8
     LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C                LC_ADDRESS=C
[10] LC_TELEPHONE=C             LC_MEASUREMENT=en_US.UTF-8
     LC_IDENTIFICATION=C

attached base packages :
[1] stats      graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached ):
[1] compiler_3.6.0 tools_3.6.0
```

KU