PSY 5939: Longitudinal Data Analysis

Contents

1 Introduction

1.1 What is it?

1.1.1 What is Mplus?

Mplus is a very general structural equation model (SEM) program that can estimate models involving:

- Observed variables

- Latent variables

- Clustering

- Repeated measures

- Categorical outcomes

- Multiple groups

- Monte Carlo simulations

- Maximum likelihood & multiple imputation for missing data

- and more. . .

1.1.2 Availability

Official website: www.statmodel.com

- A demo version of Mplus is available via this website

- Full and demo versions run on Windows, Mac OS, and Linux

Psychology: You should have access to the full version of Mplus through your mentor / lab

Outside of Psychology: You may have access through your department. If not, use the Demo version. You can complete the growth models homework assignment in the Demo version.

2 Getting data into Mplus

2.1 Missing data

2.1.1 Missing data

If there are missing values in the dataset, they must be assigned a **numerical value** (by you)

Cannot use SPSS and SAS default of "." or R default of "NA"

You will tell Mplus what the missing value is

- You want the missing value you assign to be very obvious

- Use a value that is **far** out of the range of possible responses

  - I usually use –999 or -9999

2.1.2 Assign missing values - SPSS

```
recode VAR1 VAR2 VAR3 VAR4 (SYSMIS = -999).
execute.
```

The default for a missing value in SPSS is "."

This will recode all missing values to –999

Do this for all variables that will go in the Mplus dataset

2.1.3 Assign missing values - SAS

```
data new;
set old;
if VAR1 = "." then VAR1 = -999;
if VAR2 = "." then VAR2 = -999;
if VAR3 = "." then VAR3 = -999;
```

```
if VAR4 = "." then VAR4 = -999;
run;
```

The default for a missing value in SAS is "."

This will recode all missing values to –999

Do this for all variables that will go in the Mplus dataset

2.1.4 Assign missing values - R

```
# defines matrix with TRUE where data is NA
is_na <- is.na(mydata)
# replaces any NA value with -999
mydata[is_na==TRUE] <- -999
```

The default for a missing value in R is "NA"

This will recode all missing values to –999

Do this for all variables that will go in the Mplus dataset

2.2 Getting data from other software

2.2.1 Output data

Mplus cannot directly read files from SAS or SPSS

Output data from SAS, SPSS, or R into a .txt, .dat, or .csv file

- Mplus cannot read a .txt or .dat file that is more than 80 characters wide

- If using .txt or .dat, you'll need to put them on multiple lines

- No problems (that I know of) with .csv files

2.2.2 Getting data out of SPSS

The **best** method I have found is to save your data as a .csv (comma separated values) file using the point-and-click interface

- This works regardless of how many variables you have / how many columns wide your data is

You can tell SPSS to omit variable names across the top of the file (because Mplus doesn't want them in the data file)

- Or (**better**) include the variable names, then open the data file in Excel and **cut and paste them** into the Mplus program
  - You need a list of the variables in the Mplus program anyway, so best to copy and paste them instead of typing them all

2.2.3 Getting data out of SAS

Save data to .csv file using `proc export`

```
proc export data=your_data dbms=csv
outfile="path\file_name.csv"
replace;
run;
```

where "path" is the path to where you're saving the file
and "file_name.csv" is the name of the file you're saving

2.2.4 Getting data out of R

Save data to .csv file using `write.csv`

```
write.csv(your_data,
"path\file_name.csv",
row.names = FALSE)
```

where "path" is the path to where you're saving the file
and "file_name.csv" is the name of the file you're saving

3 Structure of the program

3.1 Overview

3.1.1 The program

Mplus **only uses syntax** to create models

- Diagrammer works for Windows computers only
- There are point-and-click features, but they are only advanced features that are used **after** a syntax-driven analysis is complete

When you open "Mplus Editor" there is only a blank window

- You type your program in that window and save the file (`filename.inp`)

To run the analysis, click the "RUN" button at the top

- Results for the analysis are saved in the same folder as your input file and opened by the program (`filename.out`)

One analysis / model per input file

- You don't run a bunch of analyses in a single file like R, SPSS, SAS

3.1.2 Structure of an Mplus program - part 1

Title

- Name for your analysis – description is good

Data

- Where is the file I will be using?

Variable

- What are the variables called?

- If not using all variables, which variables am I using?

- Define missing values in your dataset

Define

- Define interaction terms or other transformations

3.1.3 Structure of an Mplus program - part 2

Analysis

- What type of analysis?

- Clustering, logistic, EFA, mixture, etc.

Model

- Here's the analysis I'm doing

Output

- Any extra info – modification indices, parameter numbering, output something to a file, etc.

3.2 Title

3.2.1 Title

```
TITLE:
This is the title for this analysis, it can be as long and descriptive as you want
```

For each section in the program, you name the section, followed by a **colon (:)** and then give the commands for that section

Most commands need to end with a **semicolon (;)** but the TITLE section does not

3.3 Data

3.3.1 Data

```
DATA:
file is C:\path\mydata.dat;
```

If your data file is saved in the same folder as your input file, you don't need to supply the path, just the name of the data file

If your dataset is actually a correlation or covariance matrix instead of raw individual data, add the line:

```
type = correlation;
```

or

```
type = covariance;
```

3.4 Variable

3.4.1 Variable

```
VARIABLE:
Names are VAR1 VAR2 VAR3 VAR4 VAR5 VAR6 subid gender;
!  Names of the variables in the order they appear;
Usevariables are VAR1 VAR2 VAR3 VAR4 subid gender;
!  Which variables you will use in your model;
Missing are all(-999);
!  Tell Mplus how you coded missing values;
```

The exclamation point (!) indicates that the rest of the line is a **comment** that is not read by Mplus

- So you can leave notes for yourself about what you did or how your program works

3.4.2 Define

```
DEFINE:
VAR1xVAR2 = VAR1 * VAR2;
VAR1sq = VAR1 * VAR1;
```

VERY IMPORTANT!

If you define an interaction or other term here, that variable name must also go in the USEVARIABLES statement, **after all other variables**

- If you do not put it in the USEVARIABLES list, or do not put it **after** the other varaibles, you will get an error

3.5 Analysis

3.5.1 Analysis

```
ANALYSIS:
type is general complex;
```

GENERAL includes regression, path analysis, CFA, etc.

COMPLEX is used for clustering and also to get chi-square tests

- Note: latent growth models we will do in Mplus are not conceptualized as mixed / multilevel models, so there is no clustering; we're using COMPLEX for the chi-square test function

Mplus 5 and later default to maximum likelihood for missing data

- Prior versions require `missing h1` to use ML

3.6 Model

3.6.1 Model

```
MODEL:
!  This is where your model statements go;
!  You describe your analysis model here;
!  We will talk about this in a minute;
```

3.7 Output

3.7.1 Output

```
OUTPUT:
sampstat modindices(3) standardized;
```

`sampstat`

- Sample statistics (means, variances, etc.)
- Good for checking that your data was correctly read in

`modindices(3)`

- All modification indices for your model that are greater than 3 (default is 10)

`standardized`

- Standardized solution / regression coefficients

We won't use these last two commands in this class, but they are useful in general

There are a lot of other options you can request: see Mplus User's Guide

4 MODEL section: Relationships

4.1 Model

4.1.1 Relationships between variables

ON = define a regression relationship

`Y ON X;`

Regression of Y on X

4.1.2 Relationships between variables

WITH = define a correlation

`X1 WITH X2;`

Correlation (/covariance) between X1 and X2

4.1.3 Relationships between variables
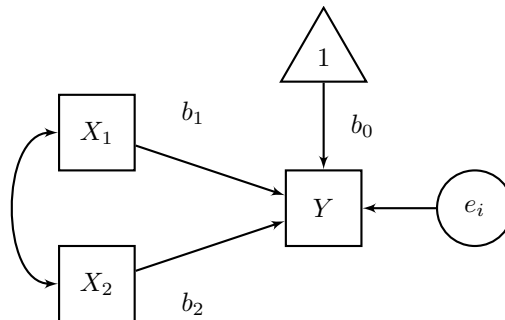
BY = define a latent variable

`Factor1 BY X1 X2 X3 X4;`

Defines a latent variable `Factor1`

- `Factor1` *is not in the dataset*
- We measure it with `X1`, `X2`, `X3`, and `X4` (which ARE in the dataset)

4.2 Prediction

4.2.1 ON = "regress on" or "predicted by"
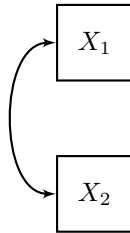

4.2.2 ON = "regress on" or "predicted by"

`Y on X1 X2;`

This command only asks **specifically** for the regression coefficients, b_1 and b_2

But Mplus also gives you the intercept (b_0) and the residual variance (e_i)

4.3 Correlation

4.3.1 WITH = "correlated with"

4.3.2 WITH = "correlated with"

`X1 with X2;`

Several variables can be included for a correlation matrix

e.g., `X1 X2 X3 with X4;`

- Does this only include `X4` with each of the other variables or the whole 4x4 matrix of correlations?
    - Mplus is inconsistent about including correlations in a model
    - Sometimes you will need to explicitly include the correlation
    - It's a good idea to VERY explicitly define your model

4.4 Measurement

4.4.1 BY = "measured by"
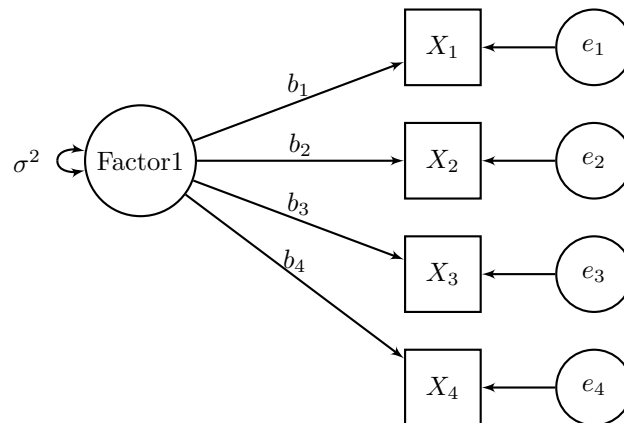
4.4.2 BY = "measured by"

`Factor1 by X1 X2 X3 X4;`

This command produces a complete factor model including:

– loadings (b_1, b_2, b_3, b_4)
- residuals on the observed variables (e_1, e_2, e_3, e_4)
- variance of the factor (σ^2)

Note that the *default* in a factor model is to fix the first loading to 1, but you can change that

- If this last bit doesn't make sense to you right now, don't worry, I'll explain it more very soon

4.5   Means and variances

4.5.1   Means and variances

Mplus is a little inconsistent with including correlations

The same thing is true with **some variances** and especially **means**

You will sometimes need to **explicitly** include means and/or variances in your model

To explicitly include **means** of variables:

```
[X1 X2 X3 X4];
```

To explicitly include **variances** of variables:

```
X1 X2 X3 X4;
```

4.6   Fixing and freeing parameters

4.6.1   Fixing and freeing parameters

You may want to fix a parameter at a certain value or otherwise change the defaults in a model

- **@ fixes** a parameter *at a value you specify*
  - Fix the loading of `X1` on `Factor1` at a value of 1:
  - This is the default for factor analysis if you don't specify anything

  ```
  Factor1 by X1@1 X2 X3 X4;
  ```

- *** frees** a parameter so that *it is estimated*
  - Fix the loading of `X3` on `Factor1` at a value of 1, and free all other loadings:

  ```
  Factor1 by X1* X2* X3@1 X4*;
  ```

4.6.2   Factor model options - example

Minimum for a factor model:

```
Factor1 by X1 X2 X3 X4;
```

Default in Mplus is to fix the loading of the first item to 1 (for identification of the model)

You could explicity show that in your model:

```
Factor1 by X1@1 X2* X3* X4*;
```

Or instead fix the loading of the second item:
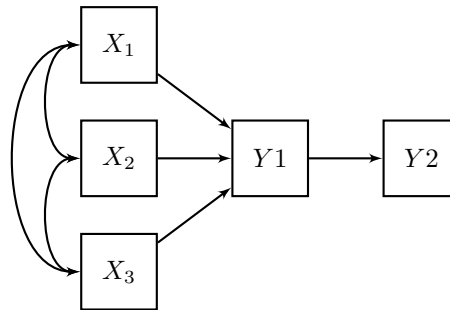
```
Factor1 by X1* X2@1 X3* X4*;
```

Or instead fix the variance of the factor to 1:

```
Factor1 by X1* X2* X3* X4*;
Factor1@1;
```

5 Examples

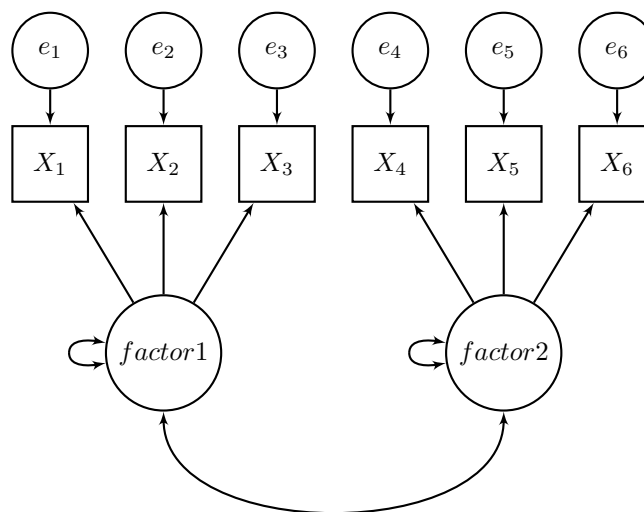## 5.1 Path model

5.1.1 Path model with observed variables

5.1.2 Path model with observed variables - syntax

```
MODEL:

!  paths between variables;
Y1 on X1 X2 X3;
Y2 on Y1;

!  means of all variable;
[Y1 Y2 X1 X2 X3];

!  variances of all variables;
Y1 Y2 X1 X2 X3;

!  correlations between predictors;
X1 with X2 X3;
X2 with X3;
```

5.2 Factor model

5.2.1 Two factor model

5.2.2 Two factor model - syntax

```
MODEL:
```

```
!  define factors;
factor1 by X1 X2 X3;
factor2 by X4 X5 X6;

!  factor correlations;
factor1 with factor2;
```

Remember that the default is to fix the first loading of each factor to 1 – you can always change that and identify the model a different way

6   More information

6.1   Other resources

6.1.1   Mplus guides

Mplus Survival Guide - on the github page for the course

- Excellent quick guide

- Lists all the sections in a program and many useful options

- Good for looking at when you need a refresher on something

Mplus user manual - on the Mplus website

- Very detailed information about all aspects of the program

- Includes numerous examples