

[PSY202B] Statistical Modeling in Psychological
Sciences

Ihnwhi Heo, M.Sc.

Spring 2024

Contents

1	Introduction	5
2	Introduction to Mplus	7
2.1	What is it? Why called Mplus?	7
2.2	Syntax-based programming	7
2.3	Some tips when programming	8
2.4	Some tips about model command particularly	8
2.5	Example: Multiple linear regression using maximum likelihood estimation	8
2.6	Additional materials	10
3	Path Analysis	11
3.1	Research scenario	11
3.2	Main questions	11
3.3	Bonus questions	13
4	Confirmatory Factor Analysis	15
4.1	Research scenario	15
4.2	Main questions	15
4.3	Bonus questions	16
5	Structural Equation Modeling	17
5.1	Research scenario	17
5.2	Main questions	17
5.3	Bonus questions	19

Chapter 1

Introduction

Hi everyone! I'm Ihnwhi.

It is my great pleasure to be your guest lecturer for PSY202B. The theme of my lecture is statistical modeling in psychological sciences.

An essential aspect of psychological research is statistical modeling based on substantive theories. It is thus important to use statistical software for accurate modeling to reach the research conclusion. I will briefly introduce Mplus (Part 1) and HLM (Part 2) and walk you around several analytic techniques using statistical software. This GitBook is your guide such that you can easily access code for Mplus and HLM practicals.

Are you ready? Let's get it on!

Chapter 2

Introduction to Mplus

2.1 What is it? Why called Mplus?

Mplus is a statistical modeling program that provides researchers with a flexible tool to analyze data

- Many models: regression, path analysis, factor analysis, SEM, MLM, longitudinal models, mixture model, mediation/moderation
- Many data: cross-sectional, longitudinal, single-/multilevel, observed/latent, incomplete
- Many variables: continuous, dichotomous, categorical, count
- Many estimator: maximum likelihood, weighted least squares, Bayesian

2.2 Syntax-based programming

- Commands and subcommands (<https://www.statmodel.com/language.html>)
- Examples of commands? (<https://www.youtube.com/watch?v=XeRRtdmu23k>)
 - We will be ‘mostly’ using TITLE, DATA, VARIABLE, ANALYSIS, MODEL, OUTPUT commands
 - But we will also be often using DEFINE, SAVEDATA, PLOT, MONTECARLO commands

2.3 Some tips when programming

1. Comments can be added with exclamation marks (!)
 2. Commands should end with colon (:), and subcommands should end with semicolon (;)
 3. Syntax is not case sensitive
 4. Data should consist of numeric values, with no variable names
 5. Data and Mplus input file should be in the same directory (like an R working directory)
- Otherwise, be sure to specify the correct directory

2.4 Some tips about model command particularly

1. Start with a path diagram
2. Think of it as specifying model parameters
3. Care to the degrees of freedom (DF)

2.5 Example: Multiple linear regression using maximum likelihood estimation

2.5.1 Model syntax

```
! Title command
TITLE: Predicting album sales using ML multiple regression

! Data command
DATA:
    ! When data and input file are in the same working directory
    FILE IS Album Sales.csv; ! Subcommands should end with ;

    ! When data and input file are in the different working directory
    ! FILE IS c:\desktop\different folder\Album Sales.csv;

! Variable command
```


2.5. EXAMPLE: MULTIPLE LINEAR REGRESSION USING MAXIMUM LIKELIHOOD ESTIMATION⁹

```
VARIABLE:
    ! Column names (i.e., ALL variable names)
    NAMES ARE adverts sales airplay attract;

    ! Variables that will be used in our analysis
    USEVARIABLES ARE adverts sales airplay;

! Analysis command
ANALYSIS:
    ESTIMATOR IS ML; ! This is the default

! Model command
MODEL:
    ! Let's predict sales using adverts and airplay
    ! We regress sales on adverts and airplay
    sales ON adverts airplay;

    ! If you do not specify variances of and covariances between predictors
    ! degrees of freedom (DF) are not correct
    ! Variances of exogenous variable
    adverts airplay;
    ! Covariances between exogenous variable
    adverts WITH airplay;

! Output command
OUTPUT:
    TECH1 SAMPSTAT STDYX;
    ! TECH1 to understand which parameters are being estimated
    ! SAMPSTAT to check sample descriptive statistics
    ! STDYX to standardize Y (i.e., DV) and X (i.e., IV)
```

2.5.2 Part of the output file

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
SALES ON					
ADVERTS		0.087	0.007	12.082	0.000
AIRPLAY		3.589	0.285	12.608	0.000
ADVERTS WITH					

AIRPLAY	604.061	421.412	1.433	0.152
Means				
ADVERTS	614.412	34.255	17.936	0.000
AIRPLAY	27.500	0.865	31.777	0.000

2.6 Additional materials

1. Official website at <https://www.statmodel.com/>
2. User's guide and examples at <https://www.statmodel.com/ugexcerpts.shtml> → Highly recommended!
3. Mplus YouTube channel at <https://www.youtube.com/c/MplusVideos>
4. QuantFish YouTube channel at <https://www.youtube.com/c/QuantFish>
5. Tutorials by Prof. Rens van de Schoot and his students at <https://www.rensvandeschoot.com/tutorials/>

Chapter 3

Path Analysis

3.1 Research scenario

A team of researchers (Charlie, Emily, Aleksandr, Madelin, and Annabella) is interested in understanding the impact of anxiety and distress on hostile behavior. They are thus to conduct a path analysis to examine the interrelationships between the variables. According to the substantive theory, depression (**depress**) is predicted by anxiety (**anxiety**), and hostile behavior (**hostile**) is predicted by depression and distress. So here, depression is a mediating variable between anxiety and hostility. Two exogenous variables are anxiety and distress. Use the `mmpi.csv` dataset.

3.2 Main questions

1. Draw a path diagram.
2. Write Mplus syntax.

```
! Annotate what you are doing in this line
title: Path analysis

data:
! Annotate what you are doing in this line
file is mmpi.csv;

variable:
! Annotate what you are doing in this line
names are subid source age sex race slpneed slpget anxiety depress
```

```
hostile posafect senseek totacctp totsavoid totici tottrust totmach
totpower totsms aggress impulse harm epixtra epineuro totrathus
totfaith totcyc totsd tothypo avoid distress;

! Annotate what you are doing in this line
usevariables are anxiety depress hostile distress;

analysis:
! Annotate what you are doing in this line
estimator = ML;

model:
! Annotate what you are doing in this line
depress on anxiety;

! Annotate what you are doing in this line
hostile on depress;

! Annotate what you are doing in this line
hostile on distress;

! Annotate what you are doing in this line
anxiety;

! Annotate what you are doing in this line
distress;

! Annotate what you are doing in this line
anxiety with distress;

output:
! Annotate what you are doing in this line
TECH1;

! Annotate what you are doing in this line
stdyx;

! Annotate what you are doing in this line
modindices (all 3.84);
```

3. Run the analysis and interpret the results.

3.3 Bonus questions

1. Check the model fit. Are there any possibilities of improving model fit? Explore such possibilities using modification indices.
2. Calculate the degrees of freedom by hand. Compare your result with that of Mplus.

Chapter 4

Confirmatory Factor Analysis

4.1 Research scenario

A team of researchers (Amish, Mercedes, Kavya, and Katey) is interested in constructing a psychometric theory on being manipulative to others. Our substantive theory suggests that sensation seeking (**senseek**), Machiavellianism (**totmach**), powerlessness (**totpower**), social monitoring (**totsms**), and faith (**totfaith**) measure the one underlying latent construct: being manipulative. We are thus to conduct a confirmatory factor analysis that measures manipulativeness by the five measures. Use the `mmpi.csv` dataset.

4.2 Main questions

1. Draw a path diagram.
2. Write Mplus syntax.

```
! Annotate what you are doing in this line
title: Confirmatory factor analysis

data:
! Annotate what you are doing in this line
file is mmpi.csv;

variable:
! Annotate what you are doing in this line
```

```

names are subid source age sex race slpneed slpget anxiety depress
hostile posafect senseek totacctp totsavoid totici tottrust totmach
totpower totsms aggress impulse harm epixtra epineuro totrathus
totfaith totcyc totsd tothypo avoid distress;

! Annotate what you are doing in this line
usevariables are senseek totmach totpower totsms totfaith;

analysis:
! Annotate what you are doing in this line
estimator = ML;

model:
! Annotate what you are doing in this line
manipulativeness BY senseek totmach totpower totsms totfaith;

output:
! Annotate what you are doing in this line
TECH1;

! Annotate what you are doing in this line
stdyx;

```

3. Run the analysis and interpret the results.

4.3 Bonus questions

1. By default, Mplus fixes the first factor loading to 1 for model identification. What if we want to free the first factor loading and instead scale the variance of the factor?
2. Calculate the degrees of freedom by hand. Compare your result with that of Mplus.

Chapter 5

Structural Equation Modeling

5.1 Research scenario

Based on the findings on the impact of anxiety and distress (Coward et al., 2024) and manipulative behavior (Patel et al., 2024), Ihnwhi is interested in modeling the interrelationships between manipulateness, anxiety, and distress. In particular, the research aim is to predict being manipulative using anxiety (**anxiety**) and distress (**distress**) via structural equation modeling. For the sake of convenience, you can only use the three following variables to measure manipulateness: sensation seeking (**senseek**), Machiavellianism (**totmach**), and social monitoring (**totsms**).

5.2 Main questions

1. Draw a path diagram.
2. Write Mplus syntax.

```
! Annotate what you are doing in this line
title: Structural equation modeling

data:
! Annotate what you are doing in this line
file is mmpi.csv;

variable:
```

```

! Annotate what you are doing in this line
names are subid source age sex race slpneed slpget anxiety depress
hostile posafect senseek totacctp totsavoid totici tottrust totmach
totpower totsms aggress impulse harm epixtra epineuro totrathus
totfaith totcyc totsd tothypo avoid distress;

! Annotate what you are doing in this line
usevariables are anxiety senseek totmach totsms distress;

analysis:
! Annotate what you are doing in this line
estimator = ML;

model:
! Annotate what you are doing in this line
manipulativeness BY senseek totmach totsms;

! Annotate what you are doing in this line
manipulativeness on anxiety;

! Annotate what you are doing in this line
manipulativeness on distress;

! Annotate what you are doing in this line
anxiety;

! Annotate what you are doing in this line
distress;

! Annotate what you are doing in this line
anxiety with distress;

output:
! Annotate what you are doing in this line
TECH1;

! Annotate what you are doing in this line
stdyx;

! Annotate what you are doing in this line
modindices (all 3.84);

```

3. Run the analysis and interpret the results.

5.3 Bonus questions

1. Check the model fit. Are there any possibilities of improving model fit? Explore such possibilities using modification indices.
2. Can you interpret the TECH1 output given that the model formulation in Mplus is based on the LISREL “all-y” notation?