[PSY202A] Statistical Modeling in Psychological Sciences

Ihnwhi Heo, M.Sc.

Fall 2024

Contents

1 Introduction **5**

2 Introduction to R: Part 1 **7**
2.1 What is it? Why called Mplus? 7
2.2 Syntax-based programming 7
2.3 Some tips when programming 7
2.4 Some tips about model command particularly 8
2.5 Example: Multiple linear regression using maximum likelihood estimation 8
2.6 Additional materials 9

3 Introduction to R: Part 2 **11**
3.1 What is it? Why called Mplus? 11
3.2 Syntax-based programming 11
3.3 Some tips when programming 11
3.4 Some tips about model command particularly 12
3.5 Example: Multiple linear regression using maximum likelihood estimation 12
3.6 Additional materials 13

4 Summarizing Data **15**
4.1 What is it? Why called Mplus? 15
4.2 Syntax-based programming 15
4.3 Some tips when programming 15
4.4 Some tips about model command particularly 16
4.5 Example: Multiple linear regression using maximum likelihood estimation 16
4.6 Additional materials 17

5 Regression and ANOVA **19**
5.1 What is it? Why called Mplus? 19
5.2 Syntax-based programming 19
5.3 Some tips when programming 19

5.4 Some tips about model command particularly 20
5.5 Example: Multiple linear regression using maximum likelihood estimation 20
5.6 Additional materials 21

Chapter 1

Introduction

Hi everyone! I'm Ihnwhi.

It is my great pleasure to be your guest lecturer for PSY202A.

Statistical modeling is a key component in conducting research in the psychological sciences. While many statistical toolkits are available to researchers, R is arguably one of the most useful free and open-source statistical software programs. It offers a dizzying array of analytic options to answer your important and exciting research questions. In the upcoming four lab sessions, you will be introduced to R and become familiar with its capabilities. These sessions are designed to help you get acquainted with the fundamentals of R and learn how to use it wisely as the next generation of psychologists. You will then be guided through summarizing and analyzing data using R.

Are you ready? Let's get it on!

Chapter 2

Introduction to R: Part 1

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
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

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    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

Introduction to R: Part 2

3.1 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

3.2 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

3.3 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

3.4 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)

3.5 Example: Multiple linear regression using maximum likelihood estimation

3.5.1 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
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)
```

3.5.2 Part of the output file

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    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
```

3.6 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 4

Summarizing Data

4.1 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

4.2 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

4.3 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

4.4 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)

4.5 Example: Multiple linear regression using maximum likelihood estimation

4.5.1 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
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)
```

4.5.2 Part of the output file

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    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
```

4.6 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 5

Regression and ANOVA

5.1 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

5.2 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

5.3 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

5.4 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)

5.5 Example: Multiple linear regression using maximum likelihood estimation

5.5.1 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
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)
```

5.5.2 Part of the output file

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    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
```

5.6 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/