Integrative Analysis of Longitudinal Studies on Aging

Introduction to Mplus statistical software and command language

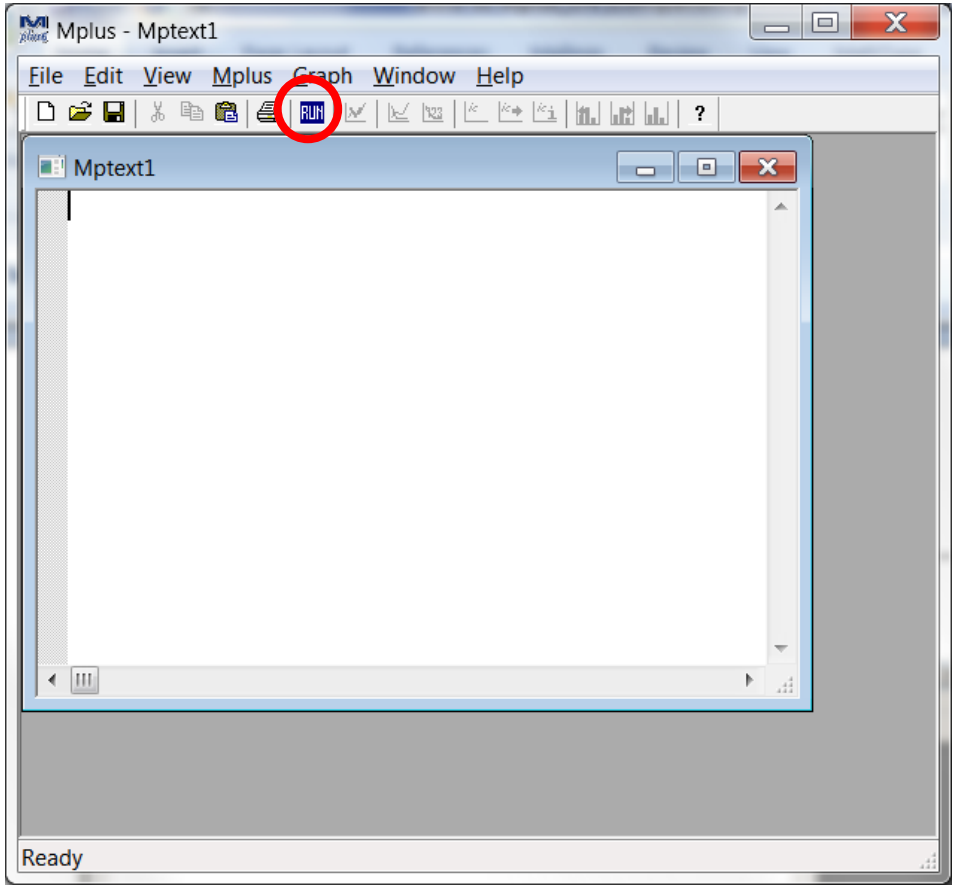
The Integrative Analysis of Longitudinal Studies of Aging (IALSA) research network is supported by a grant from the National Institutes of Health: 1P01AG043362; 1R01AG026453 and Canadian Institutes of Health Research: 200910MPA Canada-UK Aging Initiative.

About Mplus

- Statistical modeling program
- Muthén & Muthén
- Useful website https://www.statmodel.com/
- Numerous analysis capabilities:
  - EFA, SEM, , MLM, IRT, Growth modeling, Growth mixture modeling , Survival analysis, Bayesian analysis, Monte Carlo simulation...
- continuous, binary, ordinal, nominal

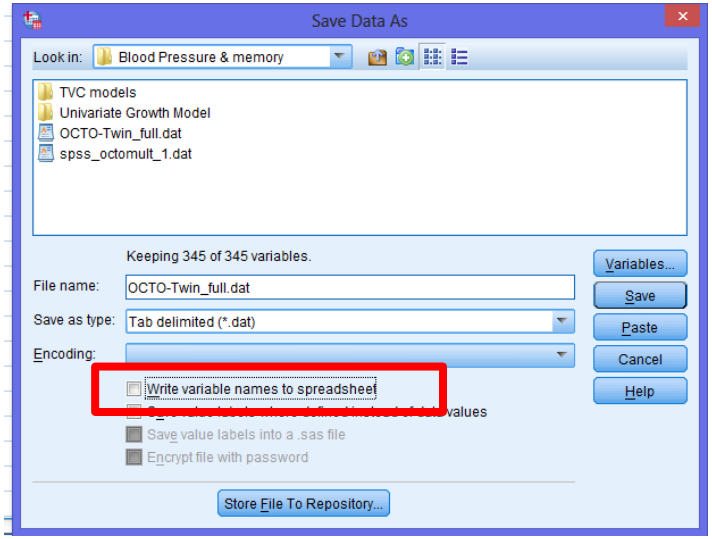
About Mplus

- Syntax-based program (not point & click)
- Input file (.int)
  - Syntax
- Output file (.out)
  - Results
- Each analysis needs a separate input file



From SPSS to Mplus

- Mplus does not open ".sav" files (SPSS files)
- Convert ".sav" to text file (.dat)
  - Open SPSS file
  - Save as type: tab delimited (*.dat) is my preference
  - Do not forget to UNCHECK the box "write variable names to spreadsheet"
  - Click Save
- ***But before converting to .dat file....



Missing Values in SPSS

- Change "." to a numeric value (e.g -9999)
  - Make sure it does not overlap with a real value
- I do not recommend dealing with missing values in SPSS
- Remember you made the change in SPSS

What .dat file should look like

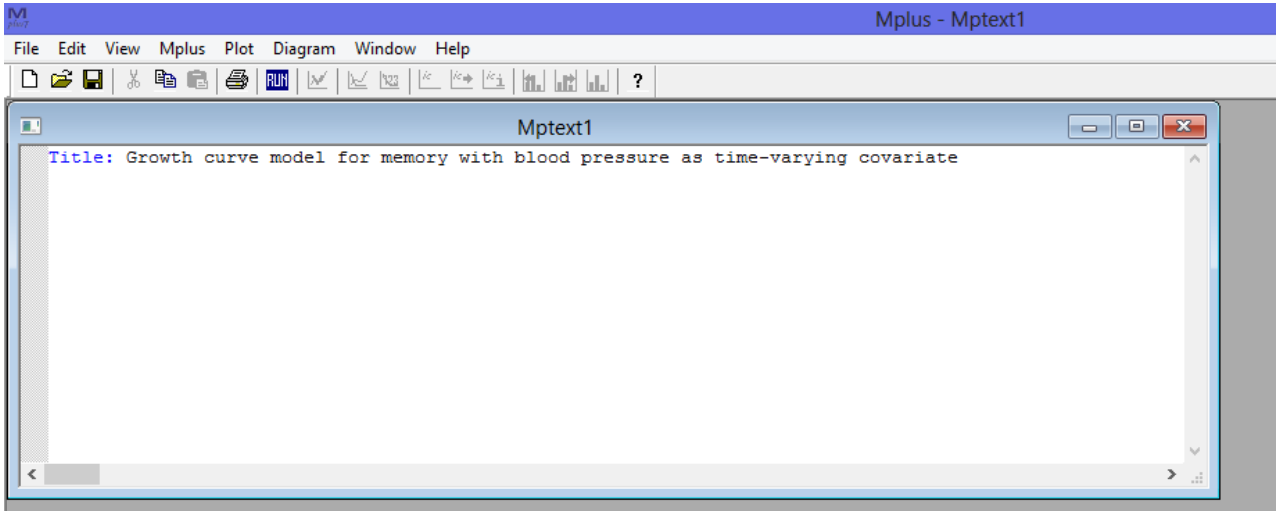
```
1     28    1     2     0     6     2     2     6     2
      5     0     -9999 1     91.660274   -.4109589   -9999
      -9999 -9999 0     -9999 -9999 -9999 -9999 -.4109589
      -9999 -9999 0     1     1     1     1     -9999 -9999
      -9999 0     -9999 -9999 -9999 -9999 91.249315   -9999
      -9999 1     -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 1     -9999 -9999 -9999 -9999
      -9999 -9999 -9999 1     -9999 -9999 -9999 -9999 1
      -9999 -9999 0     -9999 -9999 -9999 -9999 0     -9999
      -9999 0     -9999 -9999 -9999 -9999 0     -9999 -9999
      0     -9999 -9999 -9999 -9999 0     -9999 -9999 -9999
      -9999 -9999 -9999 -9999 0     -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999
      -9999 -9999 -9999 2     -9999 -9999 -9999 -9999 -9999
```

Input Files

- Commands followed by ":"
- Options separated by ";"
- Line length < 90 characters
- Variable names ≤ 8 characters
- ! Add comments excluded from analysis
- Commands can be shortened to 4 letters
- Upper or lower case

Mplus Language – Title

- Title:
  - Not a required command
  - Describe analysis
  - Help you differentiate one input/output from another



Mplus Language - Data

- DATA: FILE IS spss_octomult_1.dat;
- Where data is located
  - (path required if not in same folder as input file)
- Type of data
- Format of data (Fixed vs Free)

Type of Data

- Individual = raw data (default)
    - Rows = observations; Columns = variables
- Summary = correlation/covariance matrix
    - Means (first row)
    - Standard deviations (second row)
    - Correlation matrix/Covariance matrix
  – Mplus input
    - Type = means std corr/cova;
    - NOBSERVATIONS = 500;

Type of Data: Example of input file

Data:

TYPE IS CORRELATION MEANS STDEVIATIONS;

- Example of Data file

6 7 4 8 4

4 1 2 1 1

1.0

.75 1.0

.85 .56 1.0

.56 .25 .75 1.0

.26 .86 .53 .56 1.0

Wide vs Long

- Wide
  - Multivariate
  - Flat
  - A variable for each time point
  - SEM

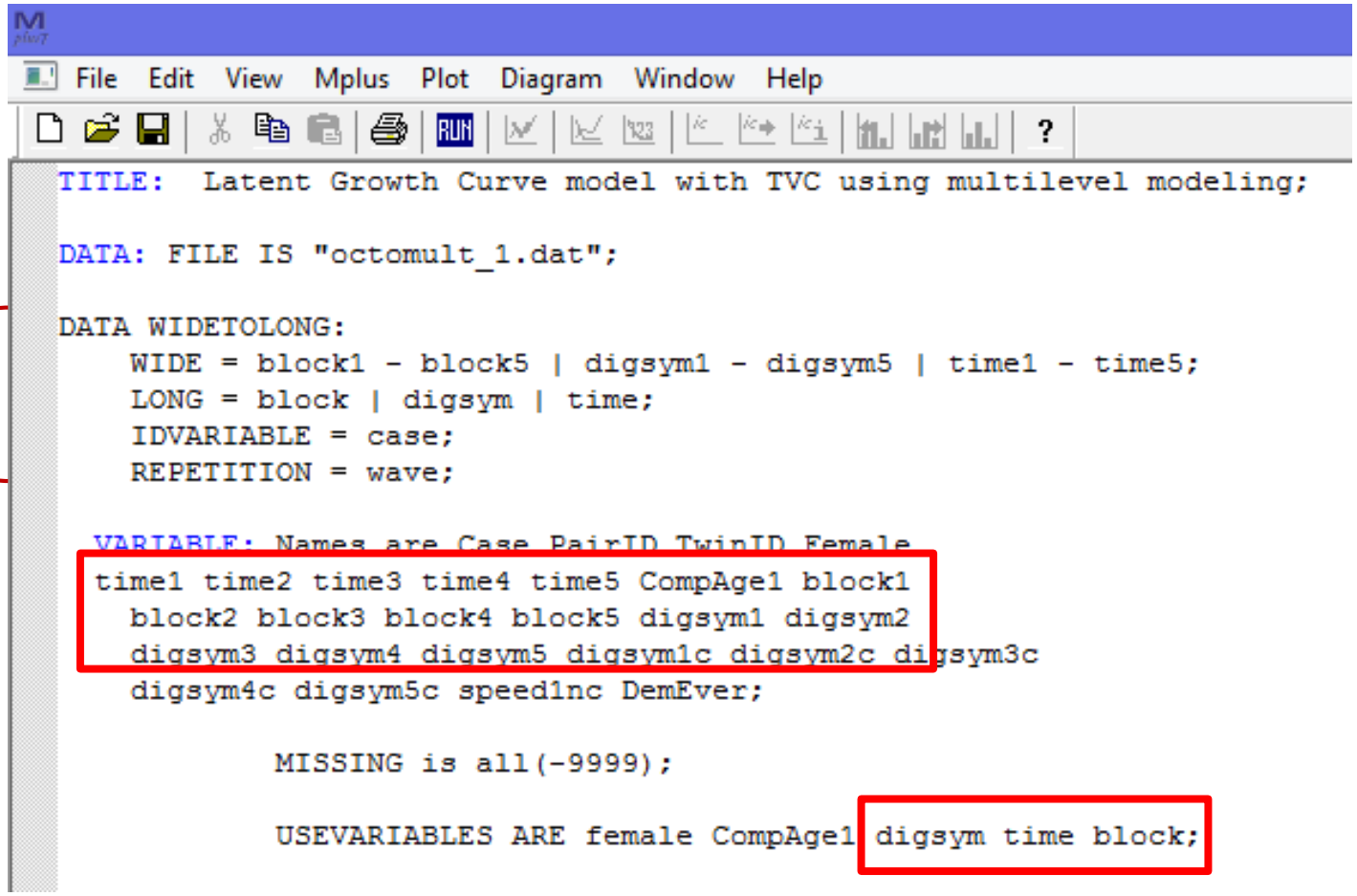
ID	mem1	mem2	mem3	cov
1	5	7	6	0
2	9	10	9	1
3	8	8	.	1

- Long
  - Univariate
  - Stacked
  - MLM

ID	mem	cov
1	5	0
1	7	0
1	6	0
2	9	1
2	10	1
2	9	1
3	8	1
3	8	1
3	.	1

Rearranging data in Mplus

- Wide to Long command



Rearranging data in Mplus

- Long to Wide

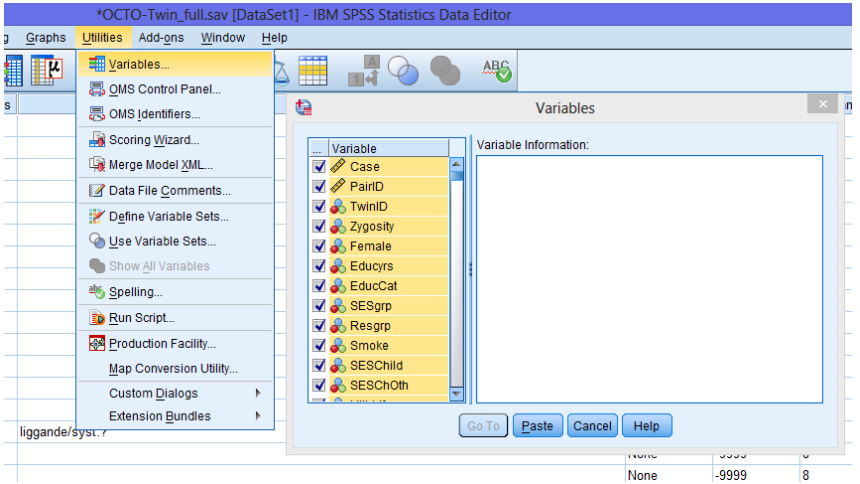
Example for presentation L

File Edit Format View Help

```
TITLE:  Latent Growth Curve model with TVC using multilevel modeling;

DATA: FILE IS "octomult_1.dat";

DATA LONGTOWIDE:
    LONG = block | digsym | time;
    WIDE = block1 - block5 | digsym1 - digsym5 | time1 - time5;
    IDVARIABLE = case;
    REPETITION = time (0, 2, 4, 6, 8);

  VARIABLE: Names are Case PairID TwinID Female
  time CompAge1 block digsym;

            MISSING is all(-9999);

            USEVARIABLES ARE female CompAge1
            block1 block2 block3 block4 block4
            digsym1 digsym2 digsym3 digsym4 digsym5
            time1 time2 time3 time4 time5;
```

Mplus language - Variable

- VARIABLE:
- NAMES ARE
  - All variables in data set
  - Keep in order
  - Block1-block5 = block1 block2 block3 block4 block5
  - In SPSS
    - Click Utilities
    - Click Variables
    - Highlight all variables
    - Click paste
    - Go to syntax window & copy



Mplus language – Variable (cont.)

- USEOBSERVATIONS ARE/SUBPOPULATION IS
  - To select observations
  - USEOBSERVATIONS ARE DemEver EQ 0;
  - Equal (EQ, ==), Not Equal (NE, /=), Greater than or equal to (GE, >=), Less than or equal to (LE, <=), greater than (GT, >), less than (LT, <), AND, OR, NOT
- USEVARIABLE ARE
  - Variables included in analysis
  - New variables in DEFINE must be included

Mplus language – Variable (cont.)

- MISSING ARE
- CATEGORICAL ARE
- TSCORES ARE
  - Individually varying times of observation
  - Use in conjunction with TYPE=RANDOM
- CLUSTER IS
  - Variables with clustering information (e.g. twins)
  - Use in conjunction with TYPE=COMPLEX

Mplus Language – Define

- DEFINE:
  - Transform variables and create new variables
    - age65c = age1-65; ! Centre variables;
    - timesq = time*time; ! Create product terms;
    - y = cdwrstd; ! “Template shortcut”;

Note: created variables must be listed last in USEVAR

Mplus Language – ANALYSIS

- ANALYSIS:
- Type = General } Default
- Use Type = GENERAL in conjunction with
  - BASIC
    - Sample stats and descriptive info
  - RANDOM
    - Random intercepts and slopes
    - Tscores in SEM
  - COMPLEX
    - Takes into account non independence of information
    - E.g. Twin data

Mplus Language – ANALYSIS (cont.)

- ESTIMATOR
  - Default depends on type of analysis & scale of DV
  - ML = Maximum Likelihood
  - MLR = Maximum Likelihood with mean adjusted chi-square test that are robust to non-normality and non-independence of observations
  - WLSMV = Weighted least square estimates

Mplus Language - MODEL

- Describe you model
  - BY: Defines latent variables
  - ON: Regressed on
    - i ON Mage Meducyrs female;
  - PON: Regressed on for paired relationships
    - mmse1-mmse5 PON sbp1 – sbp5;
  - WITH: Relationships
  - PWITH: Paired relationships
  - List variables;: Variances & residual variances

Mplus Language - MODEL

- (number);: Constrain parameters to be equal
  - Y1 ON X1 (1);
  - Y2 ON X2 (1);
  - Y3 ON X3 (1);
- |: Names and defines random variable
  - (e.g. i s | mmse1-mmse5 AT time1-time5;)
- AT: measured at
- MODEL CONSTRAINT: Specifies new model constraints
  - NEW: Assign label
    - E.g. Mediation model

Example of Model Command

MODEL:

i s q| mmse1@0 mmse2@2 mmse3@4 mmse4@6 mmse5@8;

i s q WITH i s q;

i ON Mage Meducyrs female; !control for covariates on intercept

s ON Mage Meducyrs female; !control for covariates on slope

MODEL CONSTRAINT:

NEW(indb1 indw1);

indw1 = wa1*wb1;

indb1 = ba1*bb1;

Mplus Language - Output

- For additional output not default
- Default:
  - INPUT INSTRUCTIONS: Restates input setup
  - *** WARNING
  - SUMMARY OF ANALYSIS

SUMMARY OF ANALYSIS

```
Number of groups                                                 1
Number of observations                                         390

Number of dependent variables                                   10
Number of independent variables                                  3
Number of continuous latent variables                            3

Observed dependent variables

  Continuous
   DBP1        DBP2        DBP3        DBP4        DBP5        MMSE1
   MMSE2       MMSE3       MMSE4       MMSE5

Observed independent variables
   FEMALE      MAGE        MEDUCYRS

Continuous latent variables
   IDBP        I           S

Variables with special functions

  Cluster variable      PAIRID
  Weight variable       _WEIGHT

  Time scores
   TIME1       TIME2       TIME3       TIME4       TIME5


Estimator                                                      MLR
Information matrix                                        OBSERVED
Maximum number of iterations                                  100
Convergence criterion                                  0.100D-05
Maximum number of EM iterations                               500
Convergence criteria for the EM algorithm
  Loglikelihood change                                 0.100D-02
  Relative loglikelihood change                        0.100D-05
  Derivative                                           0.100D-03
Minimum variance                                       0.100D-03
Maximum number of steepest descent iterations                  20
```

Mplus Language – Output (cont.)

OUTPUT: Sampstat;

SAMPLE STATISTICS

ESTIMATED SAMPLE STATISTICS

Means

	Y1	Y2	Y3	Y4	Y5
1	4.585	4.368	4.267	4.022	3.868

Covariances

	Y1	Y2	Y3	Y4	Y5
Y1	4.932				
Y2	2.665	4.645			
Y3	2.611	2.716	4.785		
Y4	2.270	2.309	2.437	4.363	
Y5	2.448	2.459	2.539	2.721	4.702

Correlations

	Y1	Y2	Y3	Y4	Y5
Y1	1.000				
Y2	0.557	1.000			
Y3	0.538	0.576	1.000		
Y4	0.489	0.513	0.533	1.000	
Y5	0.508	0.526	0.535	0.601	1.000

Requesting Additional Output

OUTPUT:

- PATTERNS
- STANDARDIZED
- MODINDICES
- TECH1-14;

What data are missing?

OUTPUT: SAMPSTAT;

PROPORTION OF DATA PRESENT

Covariance Coverage

	Y1	Y2	Y3	Y4	Y5
Y1	1.000				
Y2	0.868	0.868			
Y3	0.785	0.757	0.785		
Y4	0.728	0.700	0.701	0.728	
Y5	0.674	0.647	0.644	0.647	0.674

What Data are missing?

OUTPUT: Patterns;

SUMMARY OF DATA

Number of missing data patterns 16

SUMMARY OF MISSING DATA PATTERNS

MISSING DATA PATTERNS (x = not missing)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Y2	x	x				x		x	x	x	x			x		
Y3	x			x	x	x	x	x		x			x			
Y4	x			x		x	x				x	x		x		x
Y5	x						x		x	x	x	x	x		x	

MISSING DATA PATTERN FREQUENCIES

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
1	10885	7	295	13	34
2	1505	8	1116	14	100
3	1629	9	131	15	70
4	74	10	252	16	46
5	107	11	254		
6	1222	12	90		

Mplus Language – Savedata

SAVEDATA:

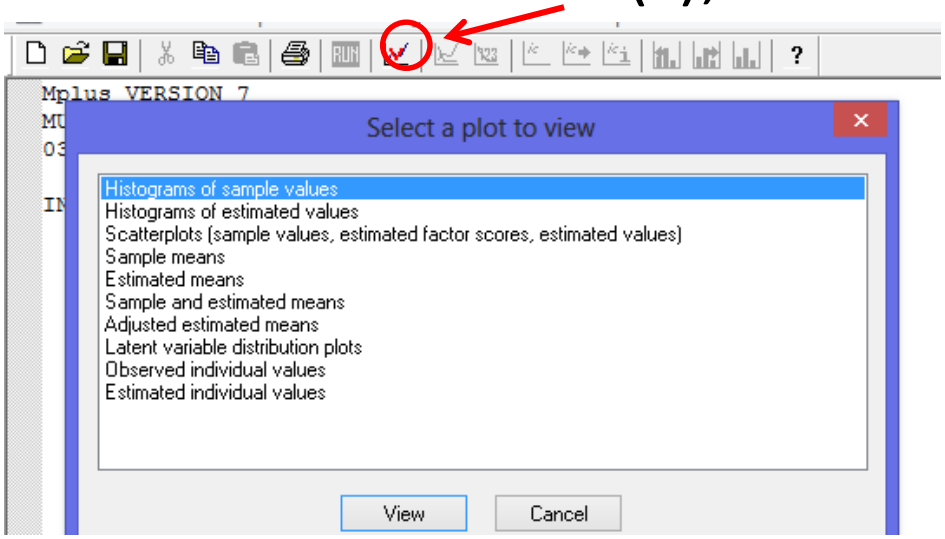
- File is: octonew.dat;
  - e.g. To save wide to long data
  - Missing values default *
    - MISSFLAG = -9999; to change default
  - AUXILIARY
    - Save variables not part of the analysis

Mplus Language – Savedata (cont.)

- SAMPLE
  - Save sample statistics (e.g. correlation/covariance matrix)
  - SAMPLE IS covariancematrix.dat;
- MFILE
  - Merging data files
  - MFILE IS C:\Blood Pressure & memory\OCTO-Twin_full.dat;

Plot Command

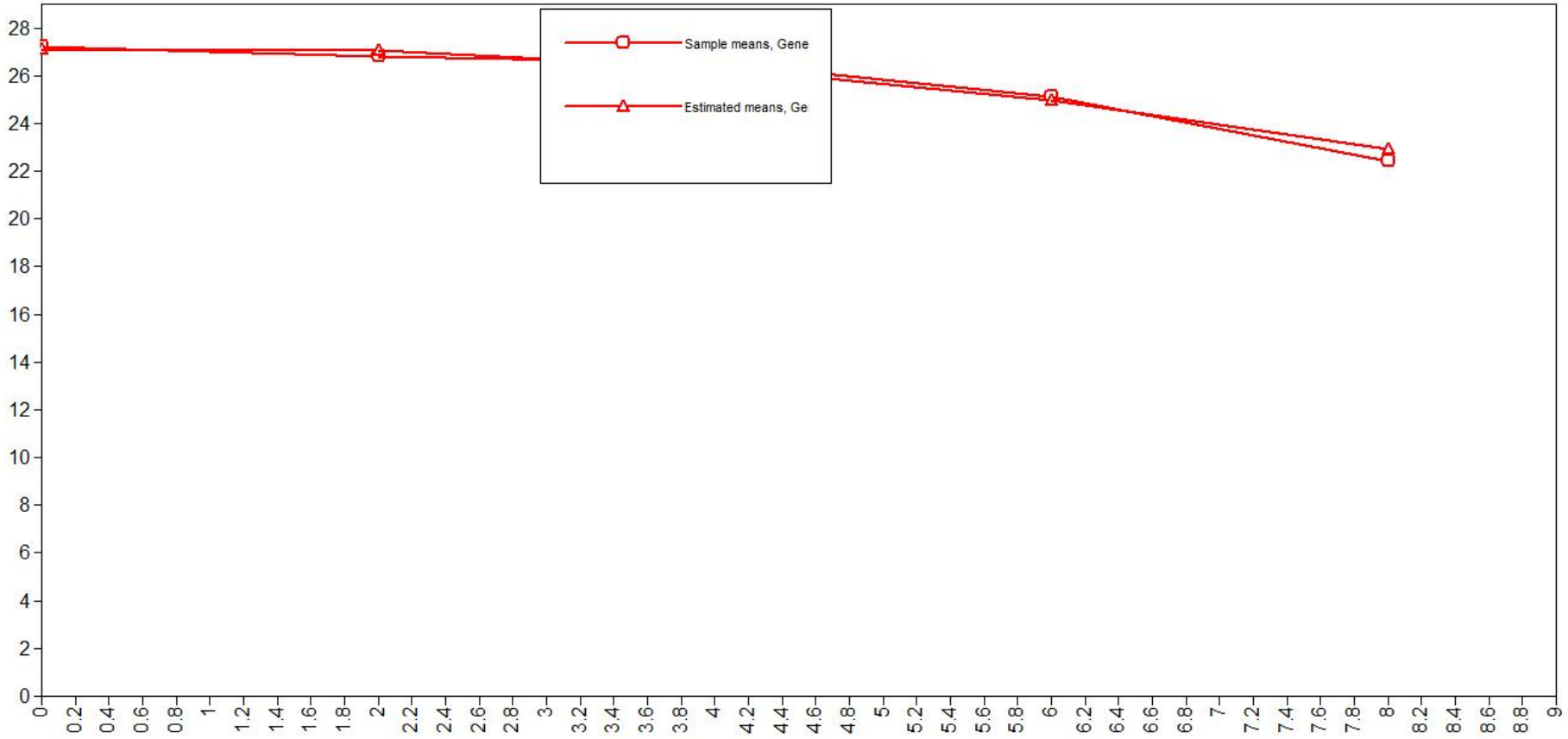
- Check for errors, outliers
  - PLOT: Type=PLOT3;
  - Series=y1 y2 y3 y4 y5 (*);
    - * =y1 (0) y2 (2) y3 (4)…;
  - Series = mmse1-mmse5(s);



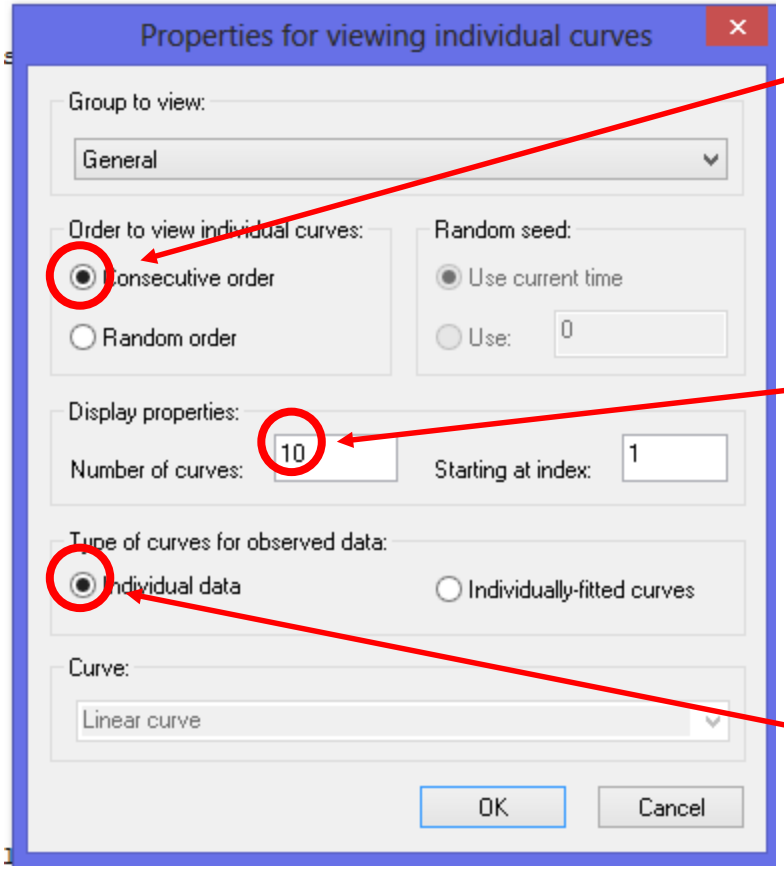
Plot Command

- Sample Means
  - Plot of observed mean as a function of time
- Estimated Means
  - Mean for model-implied growth curve
- Sample and estimated means
  - Both observed and mean model-implied on one graph
- Observed individual values
  - Observed individual growth curves for all individuals in sample

Sample & estimated means



Observed individual values

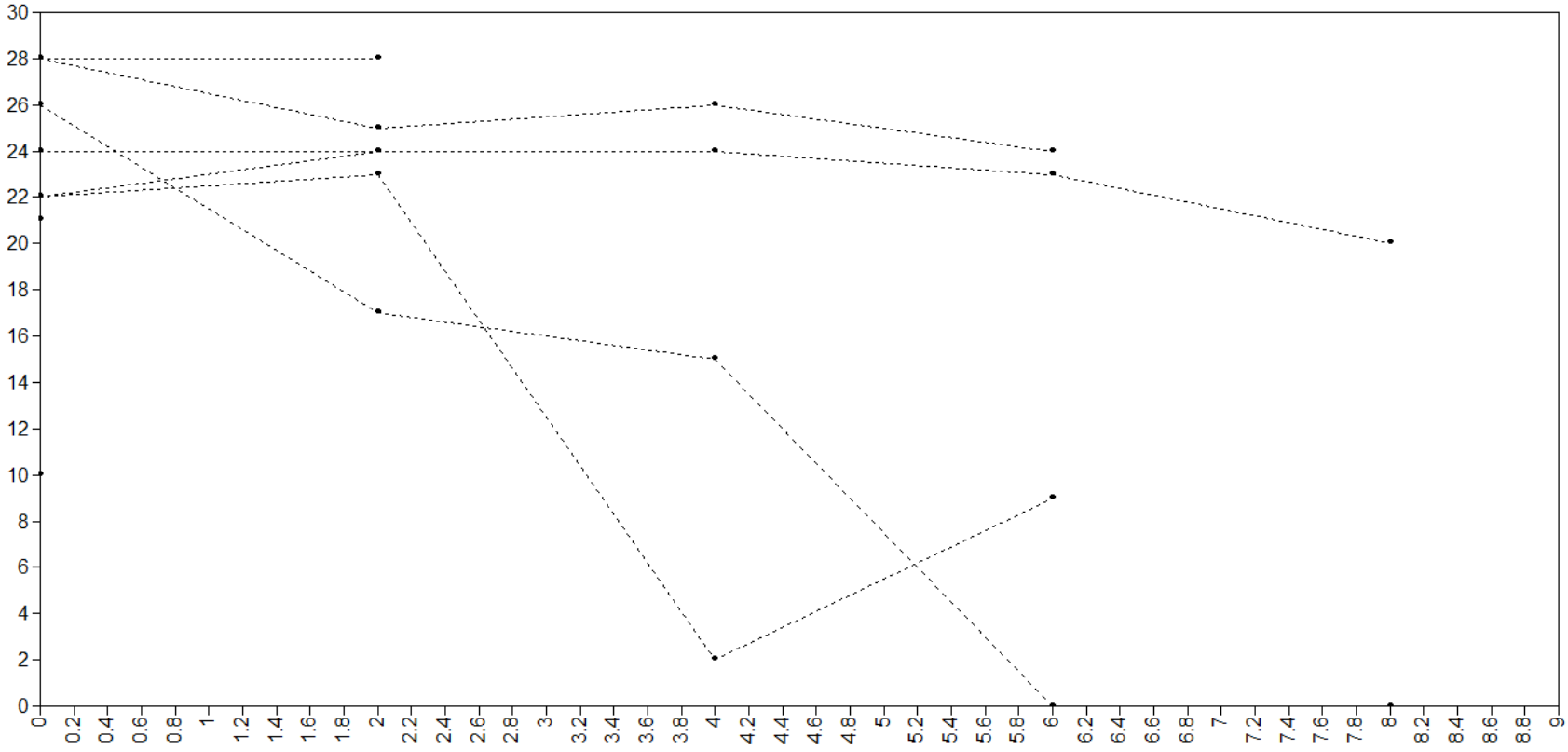


Plots data using consecutive individuals

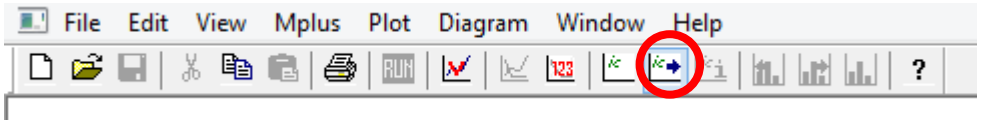
Shows 10 curves at a time

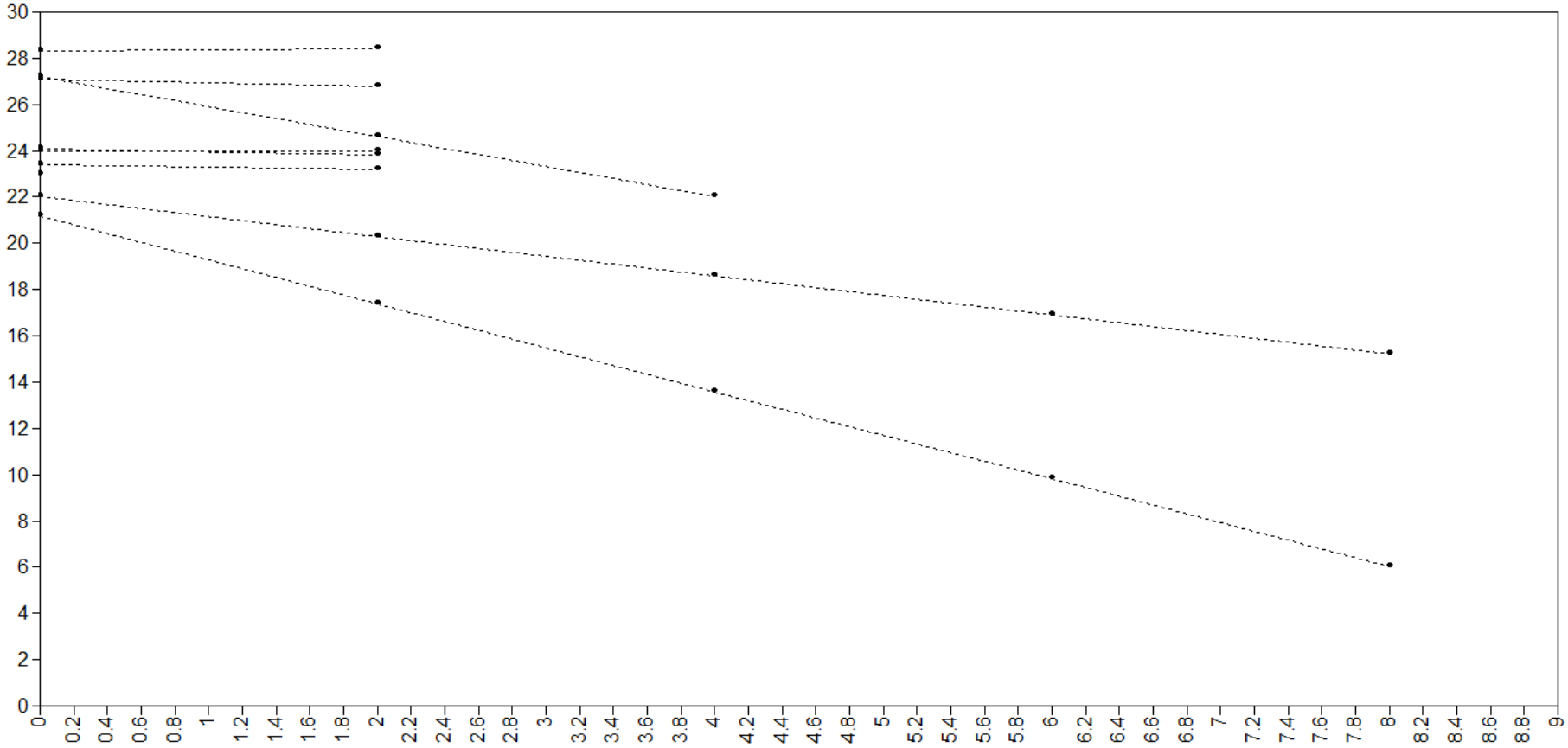
Observed data without specific curve

Individual data

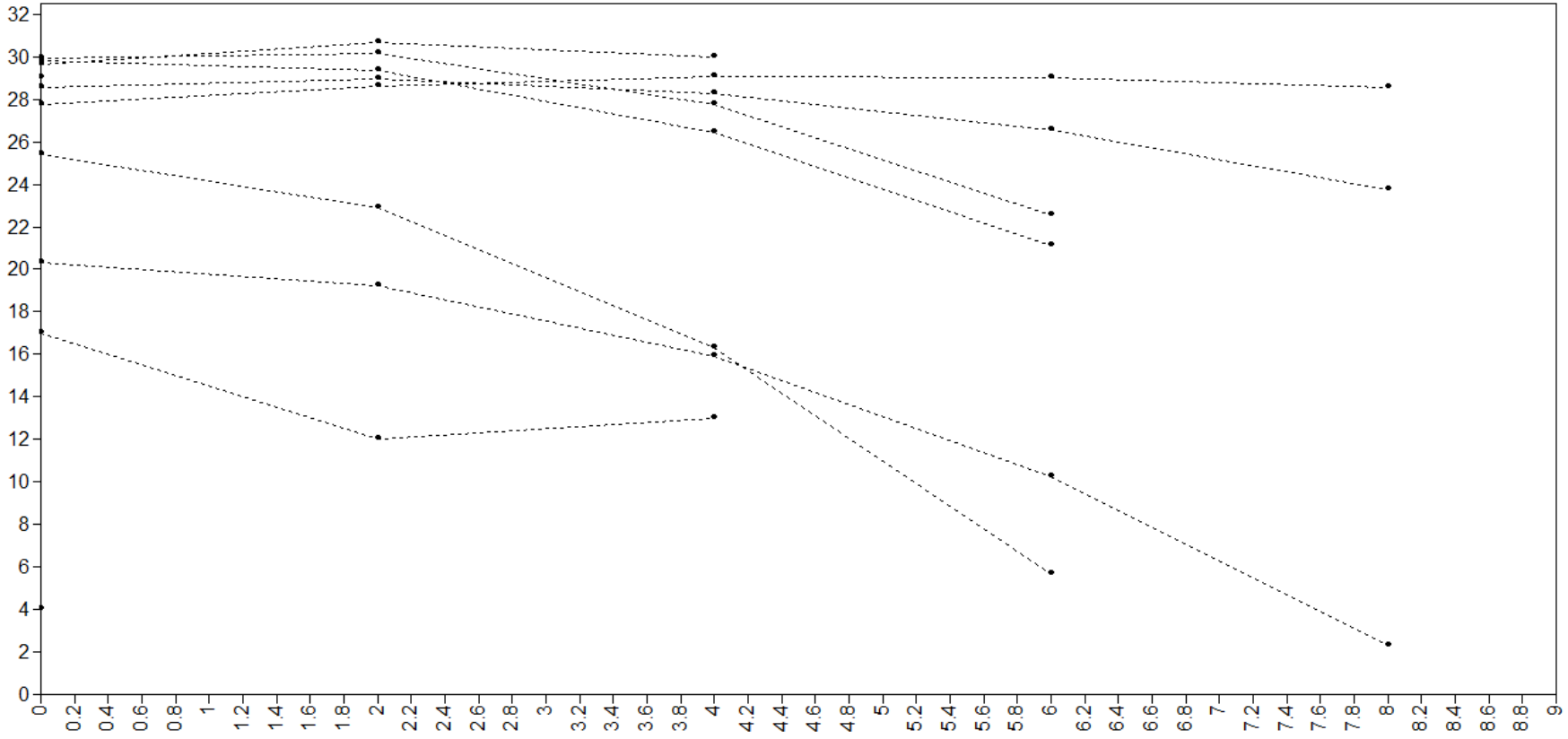


individually-fitted curves (Linear)



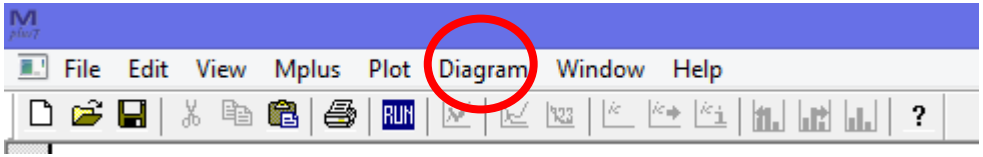


Individually-fitted curves (quadratic)

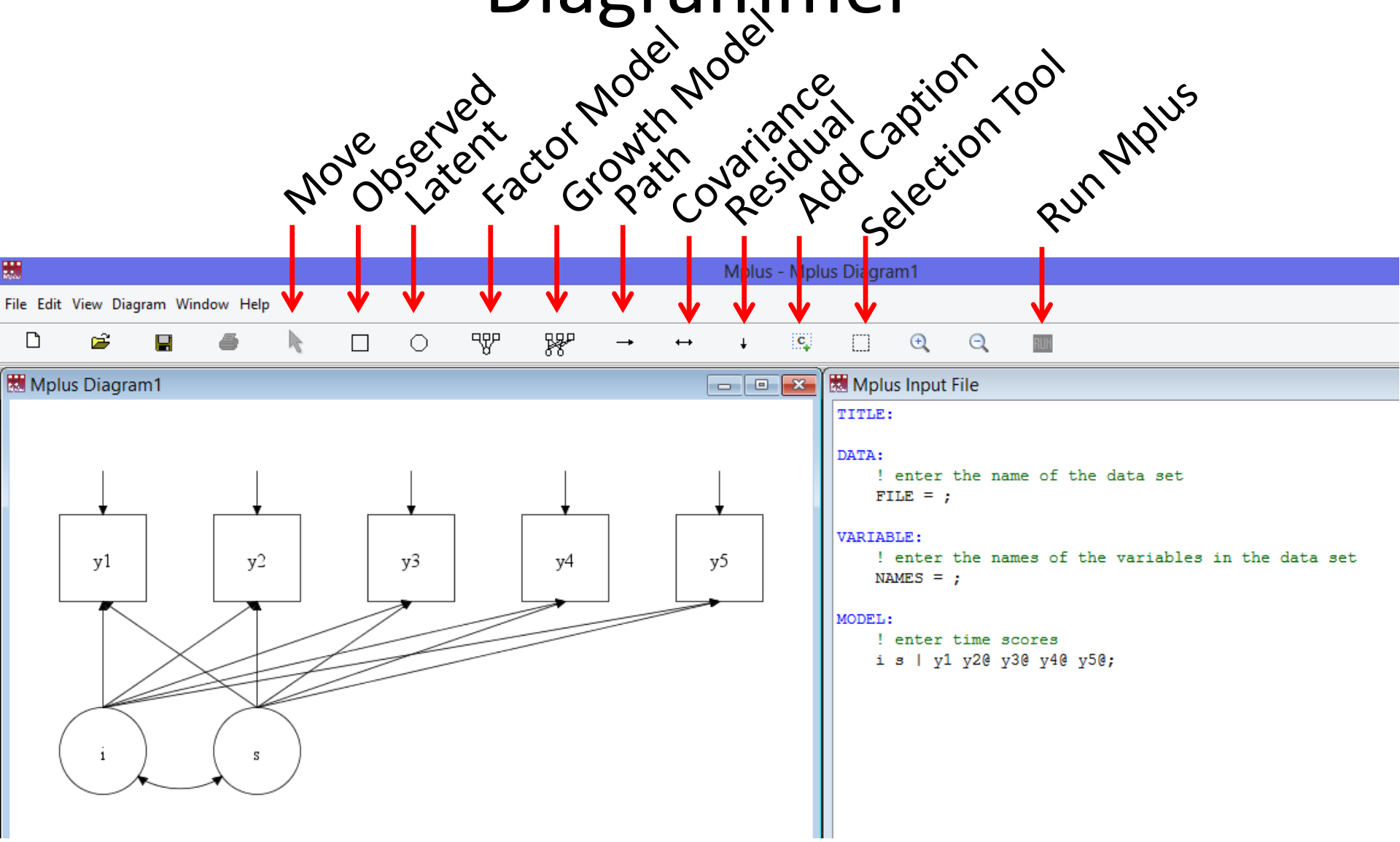


Mplus Diagrammer

- Draw a diagram
- View a diagram created from an analysis or input
- Save as PDF for publication



Diagrammer


Move
Observed
Latent
Factor Model
Growth Model
Path
Covariance
Residual
Add Caption
Selection Tool
Run Mplus
Mplus - Mplus Diagram1
File Edit View Diagram Window Help
Mplus Diagram1
y1
y2
y3
y4
y5
i
s
Mplus Input File
TITLE:
DATA:
! enter the name of the data set
FILE = ;
VARIABLE:
! enter the names of the variables in the data set
NAMES = ;
MODEL:
! enter time scores
i s | y1 y2@ y3@ y4@ y5@;

Common errors

- Input line exceeded 90 characters. Some input may be truncated.
- Variable name contains more than 8 characters
- The file specified for the FILE option cannot be found.
- Missing ;
- Forgot to include a variable or wrong order of variables on USEVAR
- A variable is misspelled
- Number of variables in data set is different than input file
- At least one variable is uncorrelated with all other variables in the model
  - E.g. List more variables on the USEVAR command than are used in the Model section

Suggested Readings

Muthén, L. K., & Muthén, B. O. (1998 –2012). *Mplus User's Guide (7th ed.)* Los Angeles, CA: Muthén & Muthén.