Mplus: A Tutorial

Abby L. Braitman, Ph.D.
Old Dominion University
November 7, 2014
NOTE: Multigroup Analysis code was updated May 3, 2016

1

About Me

- B.A. from UMD
- Briefly at NYU
- Ph.D. from ODU in 2012 (AE)
- 2-year postdoc from NIAAA
- Two great loves:
  - Alcohol research
  - Complex data modeling

2

Contacting Me

- Handouts for this workshop series (and others)
  - https://sites.google.com/site/abbybraitman/home/handouts
- abraitma@odu.edu
- MGB 132-B

3

Today

- Introduction to Mplus and basic functions
  - Intro:
    - Exporting data from SPSS
    - Code terminology
    - Reading output
  - Basics:
    - Path analyses
    - Latent variable modeling
    - Full SEM
    - Indirect effects (mediation)
    - Bootstrapping
    - Diagrammer
    - Troubleshooting

4

Today

- Intermediate functions
  - Latent growth modeling
  - Fixing and freeing paths
  - Non-continuous outcomes
  - Multilevel modeling
  - Other forms of estimation
  - Adding and relaxing equality constraints
  - LPA/LCA

5

Today

- Hands-on training
  - Sample dataset and suggested activities and models
  - Walk through an example together
  - I will give immediate hands-on training for those who are able to bring the software on their laptop
  - I will also provide ad hoc hands-on training for those who want help as they explore the software in their labs and offices for up to one week after the workshop ends

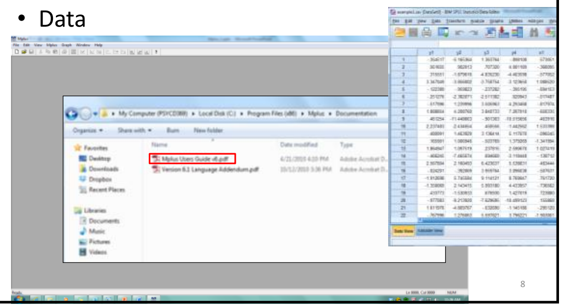
6

Why Mplus?

- Wide choice of data estimators and algorithms
  - It excels at handling categorical, nominal, binary, censored, and continuous non-normal data
- Several output options
- Beyond traditional SEM:
  - Multilevel modeling (longitudinal and cross-sectional, up to three levels of nesting)
  - Mixture modeling (latent profiles, latent classes, growth mixture)
  - Simulation analyses (Monte Carlo)
- Error messages are somewhat helpful (model is not identified versus need more iterations to reach convergence)
- Support: manual, website, Muthéns themselves
- New: Pictures!
  - Helpful for double-checking yourself, and sharing with others

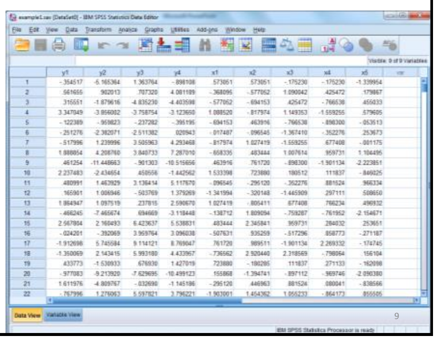
What You Need

- The editor (a big, grey expanse)
- The Users Guide (in Program Files by default)
- Data

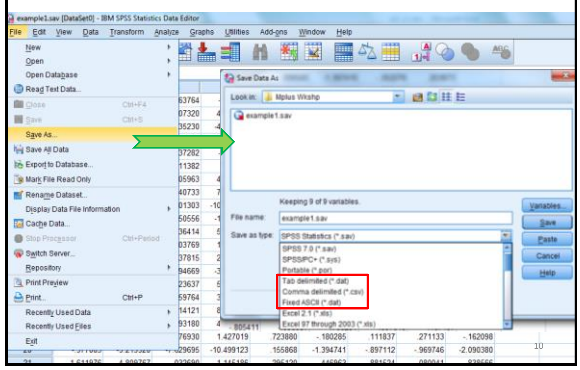


Exporting Your Data

- Must be numeric
- NAMES must be ≤ 8 characters
- y1-y4
- X1-x5

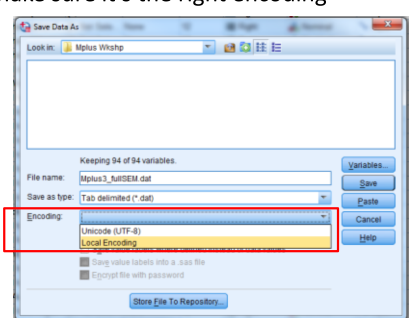


Exporting Your Data



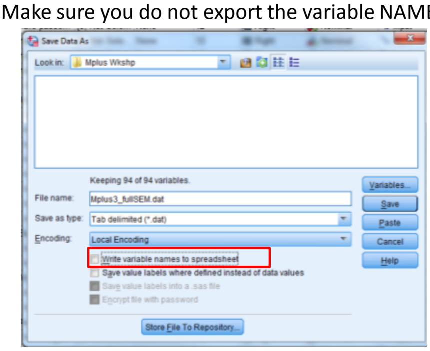
Exporting Your Data

- Make sure it's the right encoding

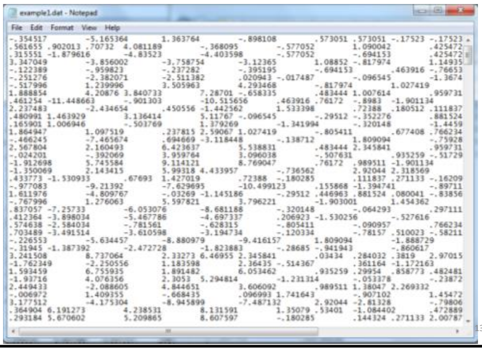


Exporting Your Data

- Make sure you do not export the variable NAMES.



Data File



Exporting Missing Data

- Missing data cannot be blank
- 5, 7, 8, [.], 32   becomes   5, 7, 8, 32

X1	X2	X3	Drinks	Age
5	7	8	.	32

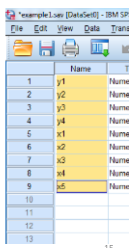
→

X1	X2	X3	Drinks	...
5	7	8	32	...

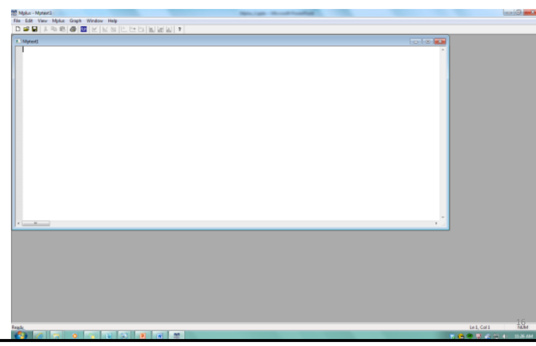
- You need some sort of indicator (that is not a plausible value)
- 5, 7, 8, 999, 32 becomes 5, 7, 8, [missing], 32
- You must tell Mplus what your indicator is
  – The language gets longer if you use different indicators for different variables, but it is possible

Exporting Your Data

- You may want to copy-paste your variable names from SPSS into Mplus when it's time to enter them
- If you accidentally omit one typing by hand, data will be mis-matched
- This is the time to shorten them if you haven't already: PROBLEMSt2 → PROBSt2

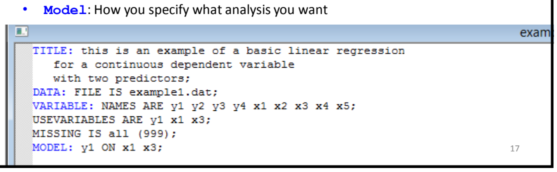


Getting Started



Getting Started

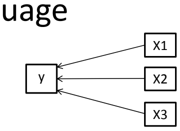
- **`Title`**: Optional, but helpful
- **`Data`**: Required
  – Exported from SPSS
- **`Variable`**: Required
  – **`NAMES ARE`** [your variable names];
  – Lists ALL variables in the dataset
- **`USEVARIABLES ARE`**: Required if you're using only some of the variables in the dataset for your model
- **`MISSING IS`** `all (999);`
  – Saying 999 is the missing data indicator, and that's true for all variables
- **`Model`**: How you specify what analysis you want



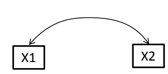
```
TITLE: this is an example of a basic linear regression
   for a continuous dependent variable
   with two predictors;
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE y1 y2 y3 y4 x1 x2 x3 x4 x5;
USEVARIABLES ARE y1 x1 x3;
MISSING IS all (999);
MODEL: y1 ON x1 x3;
```


Basic Model Language

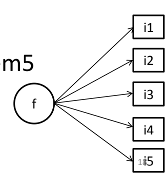
- y **ON** x1 x2 x3 (regression)



- x1 **WITH** x2 (correlation)

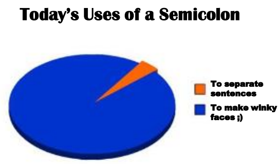


- f **BY** item1 item2 item3 item4 item5 (factors or latent variables)



Semi-Colons and Exclamation Points

- Semi-colons are how you complete a command/item in mplus. Every statement must end with it.



- Exclamation Points are how you make notes to yourself (or inactivate code).

19

Semi-Colons and Exclamation Points

Mediation_PlacePlanLimit_TYPEcomplex.inp

```
TITLE:  Mplus multilevel mediation for daily drinking
        with PBS (daily) as mediator.
        Place Context and PBS as predictors.
        Drinks as outcome.
        No mediation yet to replicate HLM findings;
DATA:   FILE is DailyL1mplus.csv;
VARIABLE: Names are SONA WeekID Home Bar Rest Party
        Other Alone Friend Fam OPlace drinks pbsplan
        pbsdo pbsall time Weekend age gendD raceD
        greekD residD marryD;
        USEVariables are Home Bar Rest Party
        Other drinks pbsplan gendD;
        CLUSTER = SONA;
        CENTERING = GRANDMEAN (pbsplan);
ANALYSIS: TYPE = COMPLEX;
!BOOT = 100;
MODEL:
        Drinks ON Home Bar Rest Party Other pbsplan;
        pbsplan ON Home Bar Rest Party Other;
        drinks ON gendD;
        pbsplan ON gendD;

!        %WITHIN%
!        s6 | drinks ON pbsdo;
!        drinks ON Home Bar Rest Party Other;
!        s1 | pbsdo ON Home;
!        s2 | pbsdo ON Bar;
!        s3 | pbsdo ON Rest;
!        s4 | pbsdo ON Party;
```

20

Double-Checking

- Make sure your data were read correctly by asking for descriptives
- Match with your descriptives from SPSS
- `Analysis: TYPE = BASIC;`

```
TITLE: this is an example of a basic linear regression
   for a continuous dependent variable
   with two predictors;
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE y1 y2 y3 y4 x1 x2 x3 x4 x5;
USEVARIABLES ARE y1 x1 x3;
MISSING IS all (999);
Analysis: TYPE = BASIC;
!MODEL: y1 ON x1 x3;
```

21

Double-Checking

example1.out

```
RESULTS FOR BASIC ANALYSIS

     ESTIMATED SAMPLE STATISTICS

          Means
             Y1            X1            X3
             ________      ________      ________
      1      0.485         0.001        -0.042

          Covariances
             Y1            X1            X3
             ________      ________      ________
 Y1          2.408
 X1          1.078         1.094
 X3          0.648         0.028         0.957

          Correlations
             Y1            X1            X3
             ________      ________      ________
 Y1          1.000
 X1          0.665         1.000
 X3          0.427         0.028         1.000
```

22

Double-Checking

Descriptive Statistics

	N	Mean	Std. Deviation
y1	500	.48484627	1.553195733
x1	500	.00128901	1.046763906
x3	500	-.04216123	.979130863
Valid N (listwise)	500		

Correlations

		y1	x1	x3
y1	Pearson Correlation	1	.665**	.427**
	Covariance	2.412	1.081	.650
	N	500	500	500
x1	Pearson Correlation	.665**	1	.028
	Covariance	1.081	1.096	.028
	N	500	500	500
x3	Pearson Correlation	.427**	.028	1
	Covariance	.650	.028	.959
	N	500	500	500

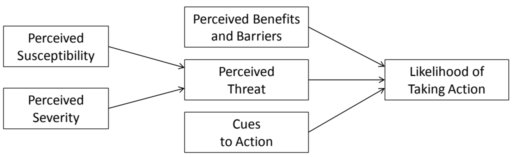
23

Path Analysis

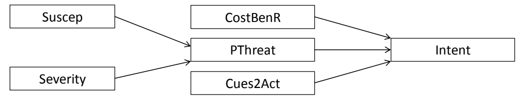
- Series of regressions: but DV's can now also be IV's!
- Great for testing models/theories
- PICTURE and CODE and OUTPUT

24

Path Analysis: Health Belief Model



Path Analysis



HealthBelief.inp

```
TITLE: this is an example of a basic path analysis
   with all continuous variables.
   Using the Health Belief Model though the data are fake;
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE dem1 intent CostBenR Pthreat dem2 Cues2Act dem3 Suscept Severity;
   USEVARIABLES ARE intent CostBenR Pthreat Cues2Act Suscept Severity;
   !MISSING IS all (999);
MODEL:
   Intent ON CostBenR Pthreat Cues2Act;
   Pthreat ON Suscept Severity;
OUTPUT: stand cint;
```


Output

- Defaults:
  - Fit statistics (not always relevant)
  - *B*, *SE*, *t*, *p*
- Can easily request additional information
  - `stand` = standardized values (e.g., βs)
  - `CINT` = confidence intervals
    - Gives 99%, 95%, 90%,

Reading Your Output

healthbelief.out

```
INPUT READING TERMINATED NORMALLY


this is an example of a basic path analysis
with all continuous variables.
Using the Health Belief Model though the data are fake;

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                         500

Number of dependent variables                                    2
Number of independent variables                                  4
Number of continuous latent variables                            0

Observed dependent variables

  Continuous
   INTENT      PTHREAT

Observed independent variables
   COSTBENR    CUES2ACT    SUSCEPT     SEVERITY


Estimator                                                       ML
Information matrix                                        OBSERVED
```


Reading Your Output

healthbelief.out

```
THE MODEL ESTIMATION TERMINATED NORMALLY


MODEL FIT INFORMATION

Number of Free Parameters                        9

Loglikelihood

          H0 Value                       -2357.703
          H1 Value                       -1568.616

Information Criteria

          Akaike (AIC)                    4733.407
          Bayesian (BIC)                  4771.338
          Sample-Size Adjusted BIC        4742.772
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                           1578.175
          Degrees of Freedom                     4
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.887
          90 Percent C.I.                    0.851  0.924
          Probability RMSEA <= .05           0.000
```


Reading Your Output

healthbelief.out

```
MODEL RESULTS
                         B          SE        t          p
                                                     Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 INTENT   ON
    COSTBENR          -0.305      0.085     -3.585      0.000
    PTHREAT            0.883      0.047     18.861      0.000
    CUES2ACT          -0.557      0.166     -3.361      0.001

 PTHREAT  ON
    SUSCEPT            3.656      0.144     25.380      0.000
    SEVERITY           2.163      0.143     15.074      0.000

 Intercepts
    INTENT            -1.515      0.092    -16.430      0.000
    PTHREAT            0.622      0.148      4.199      0.000

 Residual Variances
    INTENT             3.896      0.246     15.811      0.000
    PTHREAT           10.970      0.694     15.811      0.000
```

Reading Your Output

healthbelief.out

```
STANDARDIZED MODEL RESULTS


STDYX Standardization

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

 INTENT   ON
    COSTBENR          -0.248      0.064     -3.862      0.000
    PTHREAT            0.979      0.028     34.627      0.000
    CUES2ACT          -0.117      0.037     -3.200      0.001

 PTHREAT  ON
    SUSCEPT            0.670      0.023     29.167      0.000
    SEVERITY           0.398      0.027     14.744      0.000

 Intercepts
    INTENT            -0.298      0.020    -15.002      0.000
    PTHREAT            0.110      0.026      4.172      0.000

 Residual Variances
    INTENT             0.151      0.015     10.196      0.000
    PTHREAT            0.345      0.025     13.819      0.000


STDY Standardization
```

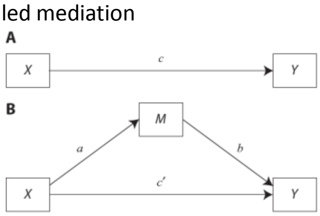

Reading Your Output

healthbelief.out

```
R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    INTENT             0.849      0.015     57.390      0.000
    PTHREAT            0.655      0.025     26.185      0.000


QUALITY OF NUMERICAL RESULTS

     Condition Number for the Information Matrix              0.477E-02
       (ratio of smallest to largest eigenvalue)


CONFIDENCE INTERVALS OF MODEL RESULTS

                  Lower .5%  Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%    Upper .5%

 INTENT   ON
    COSTBENR        -0.524      -0.471      -0.445      -0.305      -0.165      -0.138      -0.086
    PTHREAT          0.762       0.791       0.806       0.883       0.960       0.975       1.003
    CUES2ACT        -0.985      -0.883      -0.830      -0.557      -0.285      -0.232      -0.130

 PTHREAT  ON
    SUSCEPT          3.285       3.374       3.419       3.656       3.893       3.939       4.027
    SEVERITY         1.793       1.882       1.927       2.163       2.399       2.444       2.533
```

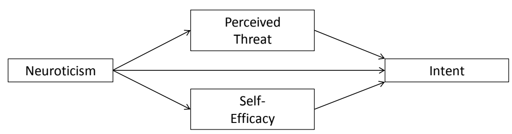

Indirect Effects

- Also called mediation



- c = total effect
- c' = direct effect
- ab = indirect effect

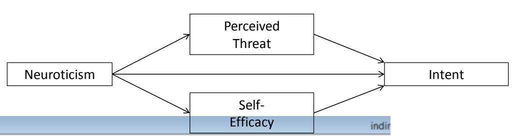
Indirect Effects




```
TITLE: this is an example of a basic path analysis
    with all continuous variables.
    Incorporating an indirect effect (data are still fake);
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE dem1 intent SelfEff Pthreat dem2 Cues2Act dem3 S
    USEVARIABLES ARE intent Pthreat Neuro SelfEff;
    !MISSING IS all (999);
MODEL:
    Intent ON Pthreat SelfEff Neuro;
    Pthreat ON Neuro;
    SelfEff ON Neuro;
MODEL INDIRECT: Intent IND Pthreat Neuro;
    Intent IND SelfEff Neuro;
OUTPUT: stand cint;
```

MODEL INDIRECT:
y IND mediator x

Indirect Effects



indir

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


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

Effects from NEURO to INTENT

  Sum of indirect      1.250      0.130      9.631      0.000

  Specific indirect

    INTENT
    PTHREAT
    NEURO              1.408      0.137     10.286      0.000

    INTENT
    SELFEFF
    NEURO             -0.158      0.042     -3.803      0.000

STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS
```


Bootstrapping

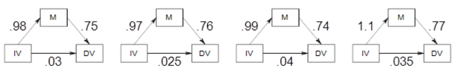
- Bootstrapping example with means:

	Original		sample 1	sample 2	sample 3		sample 1000
	31.15		31.15	31.15	31.15		31.15
	26.41		26.41	26.41	31.15		26.41
	30.82		30.82	30.82	30.82		26.41
	21.59		21.59	30.82	21.59		21.59
	26.76	⇒	26.76	26.76	26.76	…	26.76
	26.02		26.02	26.76	26.02		26.02
	28.32		28.32	28.32	28.32		26.02
	21.26		28.32	21.26	21.26		21.26
	19.50		19.50	19.50	21.26		19.50
	24.03		19.50	24.03	24.03		24.03
μ's:	25.586		26.32	26.03	26.24		24.59

- Notice that some values are repeated in the samples because they were sampled with replacement.

Bootstrapping

- Applied to mediation:



sample 1 sample 2 sample 3 sample 4

- Each parameter estimate gets a set of possible values (in this case, 1000 of them)

Bootstrapping

- Final estimate = midpoint of ordered estimates
- Significance assessed with middle 95% of ordered estimates

$$a = \begin{bmatrix} 0.968 \\ 0.969 \\ 0.971 \\ \ldots \\ 1.031 \\ \ldots \\ 1.062 \\ 1.062 \\ 1.064 \end{bmatrix}, b = \begin{bmatrix} 0.737 \\ 0.739 \\ 0.740 \\ \ldots \\ 0.762 \\ \ldots \\ 0.781 \\ 0.783 \\ 0.783 \end{bmatrix}, \therefore ab = \begin{bmatrix} 0.713 \\ 0.716 \\ 0.719 \\ \ldots \\ 0.786 \\ \ldots \\ 0.829 \\ 0.832 \\ 0.833 \end{bmatrix}$$

- Indirect effect = 0.833 (95% CI: 0.719,0.829)

Bootstrapping

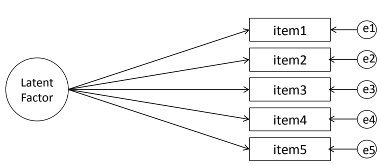
- No longer requires normally distributed variables (or coefficients)
- Create our own (normal) sampling distribution
- Assess significance using 95% CI (no 0 ~ p < .05)

```
TITLE: this is an example of a basic path analysis
   with all continuous variables.
   Incorporating an indirect effect (data are still fake);
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE dem1 intent SelfEff Pthreat dem2 Cues2Act dem3 Suscept
   USEVARIABLES ARE intent Pthreat Neuro SelfEff;
ANALYSIS: BOOTSTRAP = 5000;
MODEL:
   Intent ON Pthreat SelfEff Neuro;
   Pthreat ON Neuro;
   SelfEff ON Neuro;
MODEL INDIRECT: Intent IND Pthreat Neuro;
   Intent IND SelfEff Neuro;
OUTPUT: stand cint(bcboot);
```


Latent Variable Modeling (CFAs)

- Crux of SEM (1 minute review)
- Assumes underlying, unobserved, ***latent*** construct is driving observed items
- Different from composite scores
  - Allows for ***Measurement Error***
  - Allows for best combination ***weighting*** of items

Latent Variable Modeling (CFAs)




```
TITLE: this is an example of a basic CFA
   with all continuous variables.
   All data are fake;
DATA: FILE IS example1.dat;
VARIABLE: NAMES ARE item1 item2 item3 item4 item5 x1 x2 x3 y;
   USEVARIABLES ARE item1 item2 item3 item4 item5;
   !MISSING IS all (999);
MODEL:
   F BY item1 item2 item3 item4 item5;
!OUTPUT: stand cint;
```

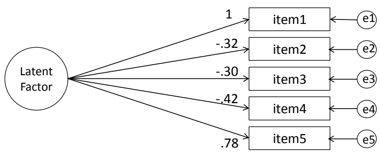

Latent Variable Modeling (CFAs)

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F BY				
ITEM1	1.000	0.000	999.000	999.000
ITEM2	-0.317	0.183	-1.731	0.083
ITEM3	-0.297	0.184	-1.613	0.107
ITEM4	-0.416	0.252	-1.654	0.098
ITEM5	0.781	0.149	5.249	0.000
Intercepts				
ITEM1	0.485	0.069	6.987	0.000
ITEM2	-1.108	0.187	-5.932	0.000
ITEM3	0.027	0.185	0.144	0.886
ITEM4	0.499	0.252	1.980	0.048
ITEM5	0.001	0.047	0.028	0.978
Variances				
F	1.375	0.287	4.788	0.000
Residual Variances				
ITEM1	1.032	0.260	3.963	0.000
ITEM2	17.294	1.097	15.764	0.000
ITEM3	16.982	1.077	15.761	0.000
ITEM4	31.518	2.000	15.757	0.000

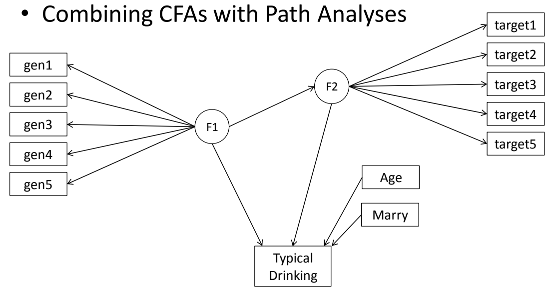
- Default is to set first item loading to 1 (to scale factor)
- Alternative is to set variance of factor to 1

Latent Variable Modeling (CFAs)



Full SEM Models

- Combining CFAs with Path Analyses



Full SEM Models

- Code (old and new)

File Edit View Mplus Graph Window Help

```
TITLE: this is an example of a full SEM
   model with all continuous variables.
   All data are fake;
DATA: FILE IS fullSEM.dat;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14
   Q60 Q61 Q62 Q63 Q64 Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2
   gen3 gen4 gen5 target1 target2 target3 target4 target5 age
   FTPT res greek gpa race hisp year athlete gender marry
   DrinkTyp DrinkHvy Probs;
   USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1 target2
   target3 target4 target5 age marry DrinkTyp;
   !MISSING IS all (999);
ANALYSIS:  ITERATIONS = 10000;
MODEL:
   F1 BY gen1 gen2 gen3 gen4 gen5;
   F2 BY target1 target2 target3 target4 target5;   } Factor Loadings
   DrinkTyp on F1 F2 age marry;   } Paths
   F2 ON F1;
OUTPUT: mod(all) stand cint;
```


Full SEM Models

```
MODEL FIT INFORMATION

Number of Free Parameters                       37

Loglikelihood

          H0 Value                       -4725.046
          H1 Value                       -4616.352

Information Criteria

          Akaike (AIC)                    9524.092
          Bayesian (BIC)                  9668.083
          Sample-Size Adjusted BIC        9550.699
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                            217.388
          Degrees of Freedom                    62
          P-Value                           0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.083
          90 Percent C.I.                    0.071  0.095
          Probability RMSEA <= .05           0.000

CFI/TLI

          CFI                                0.935
          TLI                                0.920
```


Full SEM Models

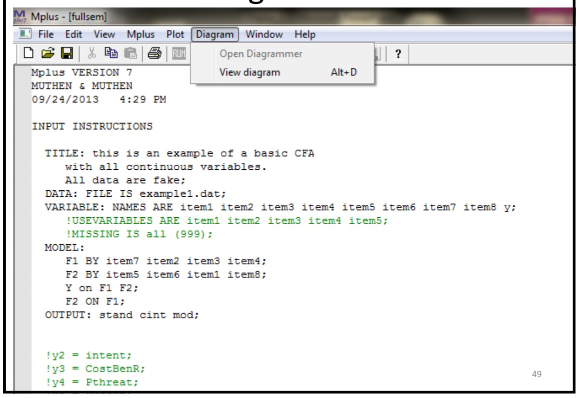
```
MODEL RESULTS

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

 F1       BY
    GEN1               1.000      0.000    999.000    999.000
    GEN2               1.311      0.145      9.021      0.000
    GEN3               0.946      0.093     10.217      0.000
    GEN4               1.532      0.136     11.244      0.000
    GEN5               1.369      0.130     10.546      0.000

 F2       BY
    TARGET1            1.000      0.000    999.000    999.000
    TARGET2            1.002      0.047     21.182      0.000
    TARGET3            1.067      0.046     23.003      0.000
    TARGET4            1.115      0.042     26.279      0.000
    TARGET5            1.111      0.042     26.550      0.000

 F2       ON
    F1                 0.942      0.126      7.455      0.000

 DRINKTYP ON
    F1                -4.652      1.133     -4.107      0.000
    F2                 1.151      0.509      2.260      0.024

 DRINKTYP ON
    AGE               -0.003      0.049     -0.065      0.948
    MARRY             -0.315      0.258     -1.221      0.222

 Intercepts
    GEN1               4.738      0.036    132.289      0.000
    GEN2               4.337      0.049     87.645      0.000
```

- Factor Loadings
- Main Paths
- Covariate Paths

Full SEM Models

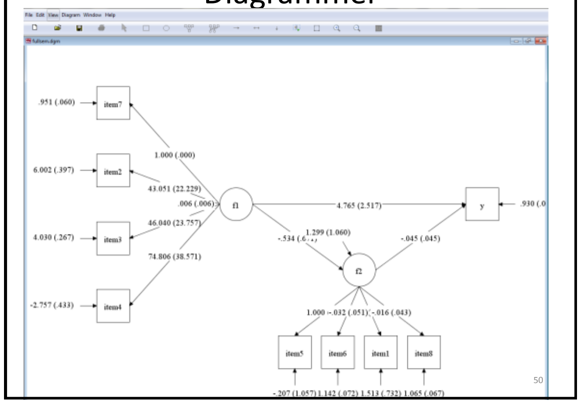
```
MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index    10.000

                              M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

ON/BY Statements

TARGET1  ON F1       /
F1       BY TARGET1         10.246    -0.241     -0.102       -0.110

ON Statements

F1       ON TARGET1         10.252    -0.256     -0.604       -0.562
F2       ON TARGET1         10.277     0.812      0.974        0.906
GEN2     ON GEN5            15.868     0.410      0.410        0.337
GEN2     ON TARGET2         13.831     0.170      0.170        0.185
GEN5     ON GEN2            15.867     0.185      0.185        0.226
TARGET1  ON GEN4            14.424    -0.122     -0.122       -0.109
TARGET1  ON TARGET3         11.819     0.180      0.180        0.204
TARGET2  ON GEN2            43.512     0.241      0.241        0.222
TARGET3  ON TARGET1         11.812     0.333      0.333        0.295
TARGET4  ON GEN4            14.800     0.135      0.135        0.109
TARGET5  ON GEN5            14.237     0.138      0.138        0.105

WITH Statements

GEN5     WITH GEN2          15.868     0.107      0.107        0.276
TARGET1  WITH F1            10.246    -0.043     -0.102       -0.249
```

Diagrammer



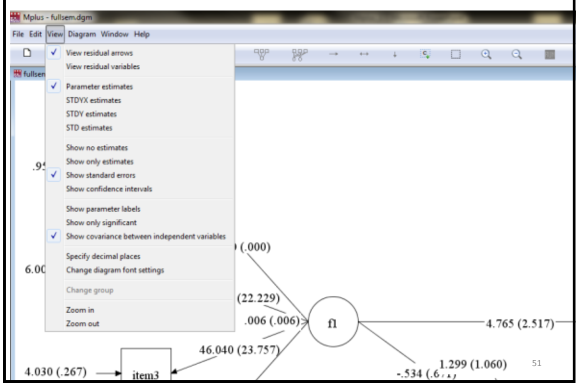
49

Diagrammer



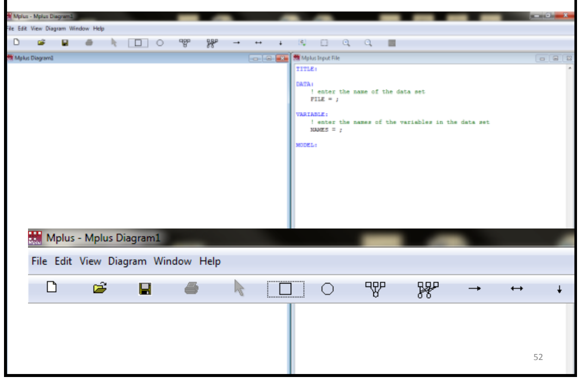
50

Diagrammer



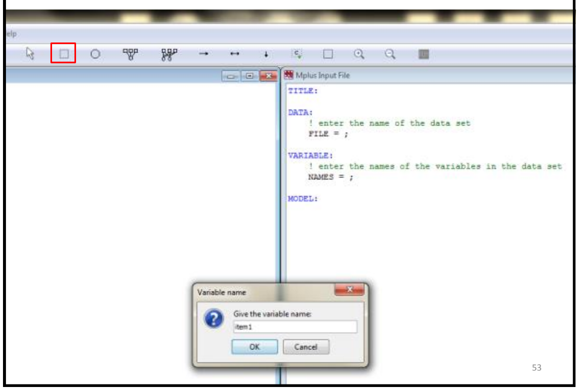
51

Diagrammer



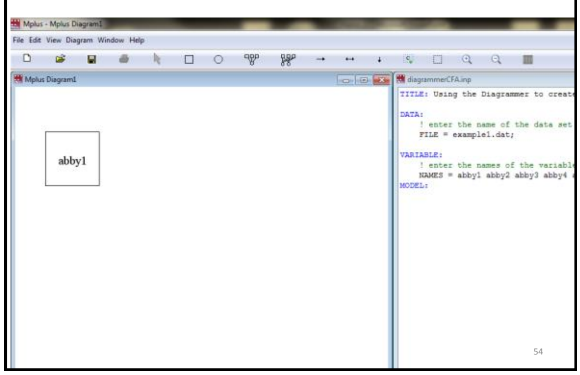
52

Diagrammer



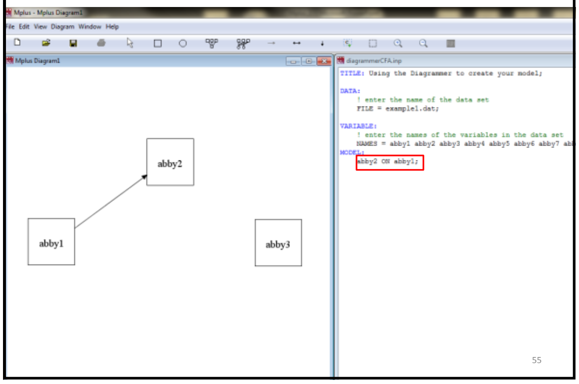
53

Diagrammer



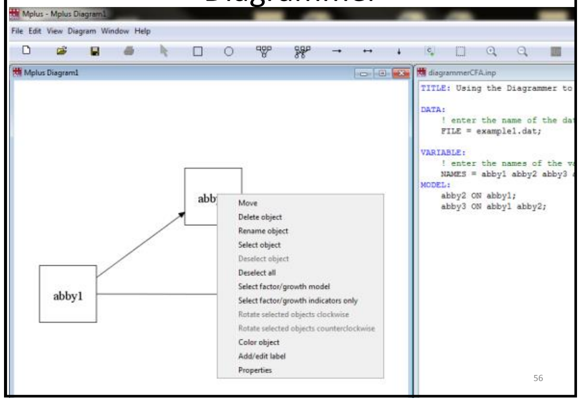
54

Diagrammer



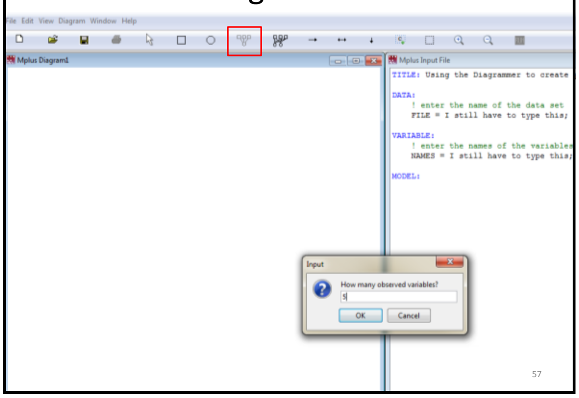
55

Diagrammer



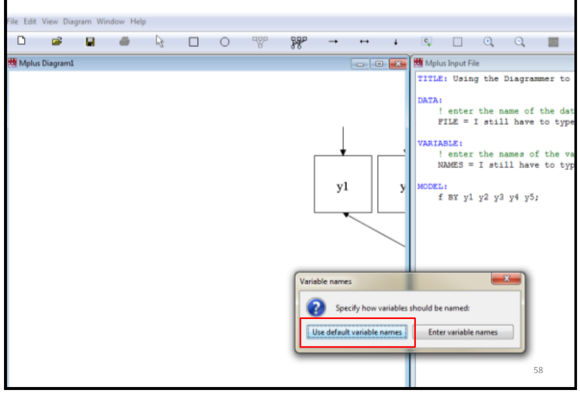
56

Diagrammer



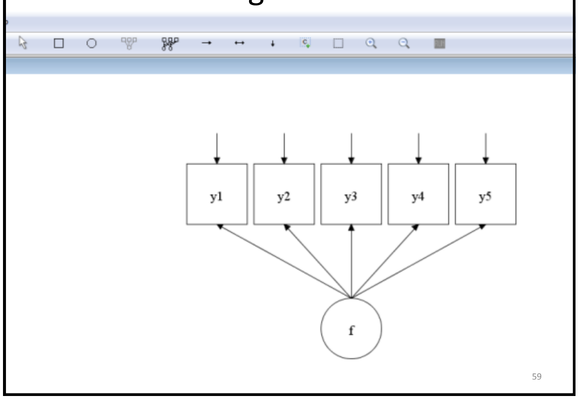
57

Diagrammer



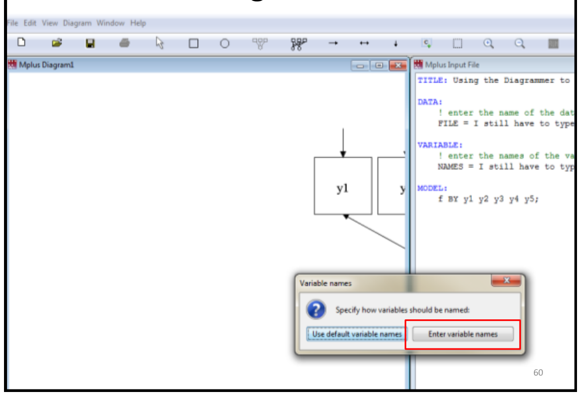
58

Diagrammer



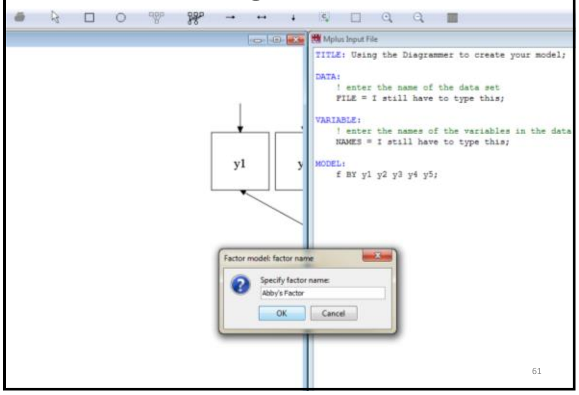
59

Diagrammer



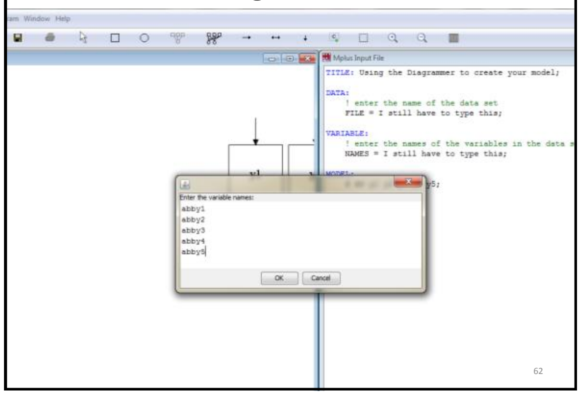
60

Diagrammer



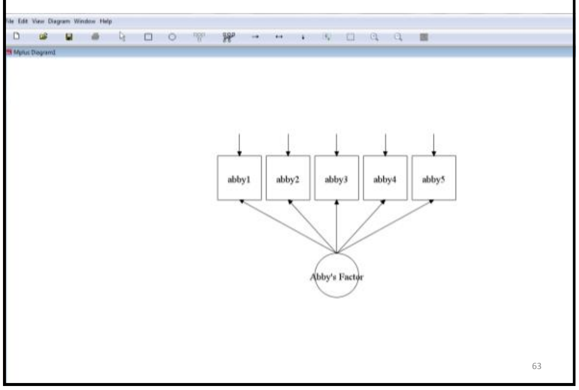
61

Diagrammer



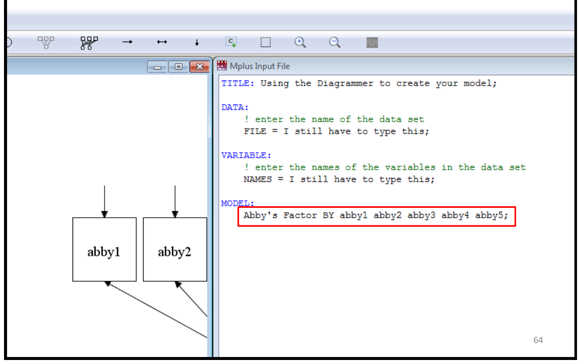
62

Diagrammer



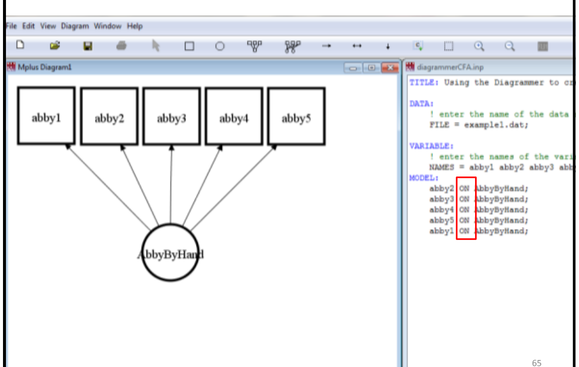
63

Diagrammer



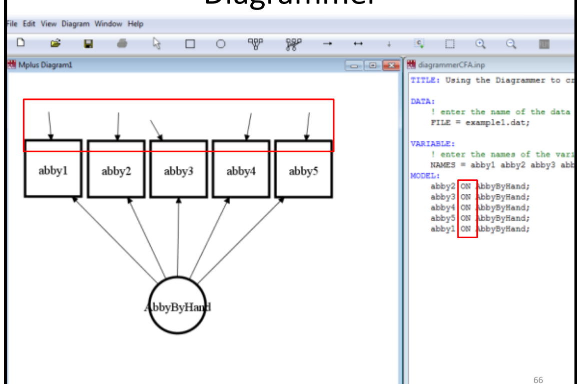
64

Diagrammer



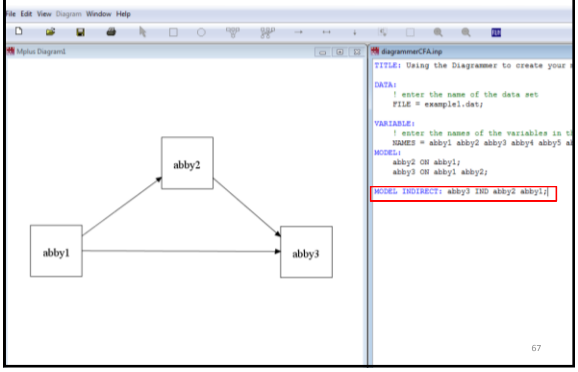
65

Diagrammer

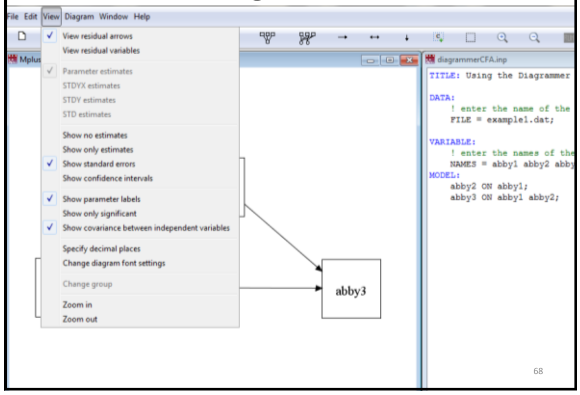


66

Diagrammer



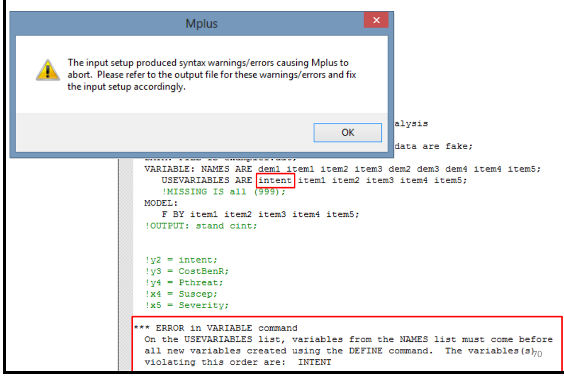
Diagrammer



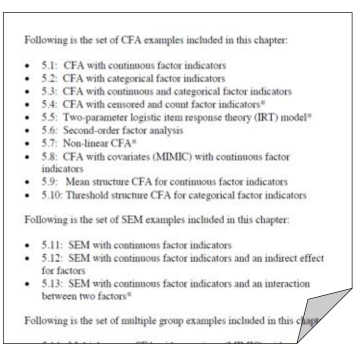
Troubleshooting

- The manual
- The website
- The software itself

The Software



Using the Manual


Following is the set of CFA examples included in this chapter:

- 5.1: CFA with continuous factor indicators
- 5.2: CFA with categorical factor indicators
- 5.3: CFA with continuous and categorical factor indicators
- 5.4: CFA with censored and count factor indicators*
- 5.5: Two-parameter logistic item response theory (IRT) model*
- 5.6: Second-order factor analysis
- 5.7: Non-linear CFA*
- 5.8: CFA with covariates (MIMIC) with continuous factor indicators
- 5.9: Mean structure CFA for continuous factor indicators
- 5.10: Threshold structure CFA for categorical factor indicators

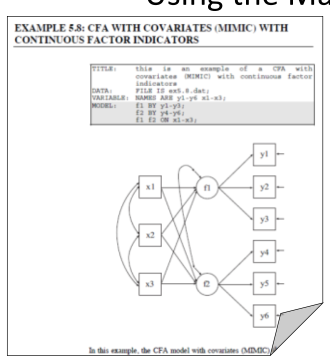
Following is the set of SEM examples included in this chapter:

- 5.11: SEM with continuous factor indicators
- 5.12: SEM with continuous factor indicators and an indirect effect for factors
- 5.13: SEM with continuous factor indicators and an interaction between two factors*

Following is the set of multiple group examples included in this chap

- Lots of examples (hundreds)
- Start with their code and alter to fit your data/model

Using the Manual


EXAMPLE 5.8: CFA WITH COVARIATES (MIMIC) WITH CONTINUOUS FACTOR INDICATORS

```
TITLE:     this is an example of a CFA with
           covariates (MIMIC) with continuous factor
           indicators
DATA:      FILE IS ex5.8.dat;
VARIABLE:  NAMES ARE y1-y6 x1-x3;
MODEL:     f1 BY y1-y3;
           f2 BY y4-y6;
           f1 f2 ON x1-x3;
```


In this example, the CFA model with covariates (MIMIC)

- Examples include code, pictures, and written explanation

Using the Manual

The first BY statement specifies that f1 is measured by y1, y2, and y3. The second BY statement specifies that f2 is measured by y4, y5, and y6. The metric of the factors is set automatically by the program by fixing the first factor loading in each BY statement to 1. This option can be overridden. The intercepts and residual variances of the factor indicators are estimated and the residuals are not correlated as the default. The residual variances of the factors are estimated as the default. The residuals of the factors are correlated as the default because residuals are correlated for latent variables that do not influence any other variable in the model except their own indicators. The ON statement describes the linear regressions of f1 and f2 on the covariates x1, x2, and x3. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. An explanation of the other commands can be found in Example 5.1.

9: MEAN STRUCTURE CFA FOR CONTINUOUS
ICATORS

- Explain defaults when relevant
  - Setting factor metric
  - Correlating factors
  - Correlating exogenous variables

Using the Website

- http://www.statmodel.com/

Available

FAQ
MPLUS DEMO VERSION
TRAINING
Short Courses
Short Course Videos and Handouts
Web Training
DOCUMENTATION
Mplus User's Guide
Mplus Diagrammer
Technical Appendices
Mplus Web Notes
User's Guide Examples
ANALYSES/RESEARCH
Mplus Examples
Papers
References
SPECIAL MPLUS TOPICS
Alignment (MG CFA)
BSEM (Bayesian SEM)
Complex Survey Data
ESEM (Exploratory SEM)
Genetics
IRT
Missing Data
Randomized Trials
HOW-TO
Using Mplus via R
Chi-Square Difference Test for MLM and MLR
Power Calculation
Monte Carlo Utility
SEARCH
Go

Latest News

- Mplus Version 7.11 is now available. Click here to see the new features. Registered users who purchased Mplus within the last year or those with a current Mplus Upgrade and Support Contract can download using our online system at no cost.
- Revised paper: Asparouhov & Muthén (2013). Multiple-group factor analysis alignment. Web note 18: Version 3. Mplus scripts are available here.
- Mplus pre-conference workshop at the European Survey Research Association (ESRA) meeting in Ljubljana, Slovenia, July 15: New Developments in Latent Variable Modeling Using Mplus (Bengt Muthén). Handouts for the workshop and related July 16 talk are available here.
- New FAQ: Growth mixture model confidence intervals for estimated trajectory means.
- Revised paper: Asparouhov & Muthén (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. Web note 15: Version 7.

Mplus Demo
The Mplus Demo download at no c demo. The demo capabilities of th only limited by t that can be used
Student Prici
Special student p student version c regular version.
Mplus Versior Examples
Click here for the and to download the Mplus User's
Mplus Web Tr
Videos and hand Mplus Short Cou viewing on the w includes web talk overview course, lecture course or
Papers Using
Click here to find date.

Using the Website

Message/Author

Monica Oxford posted on Wednesday, November 29, 2000 - 8:51 am

I am running a 3 class 13 wave quadratic growth mixture model and have two questions.

1) The model converged and terminated normally, however, the variance estimate for intercepts for the 1st and 3rd classes were negative. This seems to indicate some problems with the model, what are your suggestions (this doesn't happen in the linear model, only in the quadratic)?

2) The standard errors for some of the estimates are quite large for the quadratic 3 class model (e.g., class 2 the variance and mean of the intercept). What might be the problem? Again this is not the case with the linear model.

Thanks in advance.
Moncia Oxford

bmuthen posted on Wednesday, November 29, 2000 - 10:03 am

The fact that growth mixture modeling has more than one class tends to reduce the within-class variation and in some cases it can be set at zero. You may not get a significant worsening of fit (e.g by likelihood-ratio chi-square difference testing) if you fix the negative variance estimates at zero. If you do get a significant worsening of fit, this could indicate that the model is not appropriate for the data.

If you have class-specific parameters, standard errors could be large due to small class sizes.

The fact that your quadratic model seems to have more problems than the linear might point to the fact that you might only need a linear model once you allow several classes. A good way to vizualize your model-data fit on an individual level is mentioned in the growth mixture modeling paper number 87 as listed on this web site.

Using the Website

%c#3%
Y1 with Y2;
[Y1 Y2];

bmuthen posted on Sunday, December 12, 2004 - 7:38 pm

You have to do this by using starting values that make whatever class you want the last class. So you rerun your analysis giving some key starting values for the last class which you take from class 2 of your current solution.

Anonymous posted on Sunday, December 12, 2004 - 9:54 pm

Thanks a lot for the suggestion. I have tried it and got some strange results. When there was no starting value, the correlation between two variables in one class was positive while it became negative when starting values were given.

All other fit statistics and parameter estimates were exactly the same except the mentioned correlation.

Linda K. Muthen posted on Monday, December 13, 2004 - 6:42 am

Send the two outputs to support@statmodel.com so I can see exactly what you are doing -- the output with no starting values and the output where you use the ending values as starting values.

Individual Trouble-shooting!

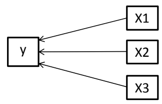
BREAK

Section 2

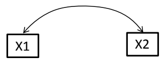
- Advanced model language
- Latent Growth Models
- Non-continuous outcomes
- Multi-group analyses
  - Fixing and freeing paths
  - Adding and relaxing equality constraints
- Latent Profile/Class Analysis
- Multilevel Modeling

Basic Model Language

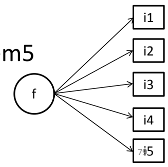
- y **ON** x1 x2 x3 (regression)



- x1 **WITH** x2 (correlation)



- f **BY** item1 item2 item3 item4 item5 (factors or latent variables)



Mathematical Operators

Symbol CODE	Definition	Example
+	Addition	`y + x;`
–	Subtraction	`y – x;`
*	Multiplication	`y * x;`
/	Division	`y / x;`
**	Exponentiation	`y**2;`

CODE	Definition	Alternate Symbol CODE
EQ	Equal	==
NE	Not Equal	/=
GE	Greater than or Equal to	>=
LE	Less than or Equal to	<=
GT	Greater Than	>
LT	Less Than	<

CODE	Definition
AND	logical and
OR	logical or
NOT	logical not

80

Advanced Model Language: DEFINE

DEFINE NewVariable = mathematical expression;

- `Depress = MEAN (dep1 dep2 dep3 dep4 dep5);`
- `DepSum = SUM (dep1 dep2 dep3 dep4 dep5);`

} Decided to create composite instead of CFA

- IF-THEN NewVariable transformation statements;
  - `IF (gender EQ 1 AND ses EQ 1) THEN group = 1;`
  - `IF (gender EQ 1 AND ses EQ 2) THEN group = 2;`
  - `IF (gender EQ 2 AND ses EQ 1) THEN group = 3;`
  - `IF (gender EQ 2 AND ses EQ 2) THEN group = 4;`
- CUT variable or list of variables (cutpoints);
  - `CUT y1 (30 40);`
  - 0 = less than or equal to 30,
  - 1 = greater than 30 and less than or equal to 40
  - 2 = greater than 40

} Continuous into ordered groups; Variable keeps same name; Cutpoint is included is lower group;

- _MISSING
  - `IF (y EQ 0) THEN newvar = _MISSING;`
  - `IF (y = _MISSING) THEN newvar = 0;`

} Changing a value to MISSING, or MISSING to a value

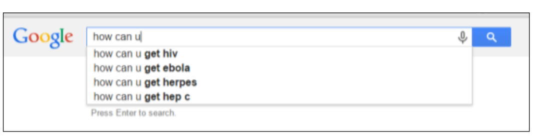
81

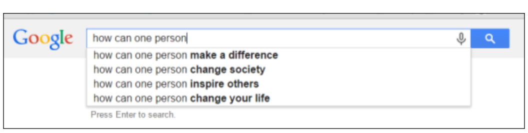
Advanced Model Language

- USEOBS or USEOBSERVATIONS
  - conditional statement to select observations
  - `USEOBS ARE gender EQ 1` } Running the model with just males
  - `USEOBS ARE x3 NE 1` } Use everyone EXCEPT group 1
  - `USEOBS ARE age GE 18` } Excluding those who are underage
- Combine with DEFINE
  - **`DEFINE:`** `IF (drinks LT 5 AND probs EQ 0) THEN group = 1;` } Identifying low to moderate drinkers (< 5 drinks, no problems)
  - `USEOBS ARE group NE 1;` } Use everyone EXCEPT them

82

Language Matters





83

Advanced Model Language

Constraints

- * frees a parameter, or denotes a specific starting value
  - example: y1*.5; - The variance of y1 will be freely estimated, starting with examining the likelihood that it is 0.5
- @ fixes a parameter at a a specific value
  - example: y1@0; - The variance of y1 is constrained or set to 0
- (*number*) constrains parameters to be ***equal***
  - example: f1 ON x1 (1);
  - f2 ON x1 (1);

} The influence of x1 predicting f1 is the same as the influence of x1 predicting f2.

84

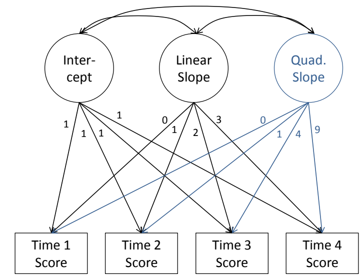
Advanced Model Language

[intercept] versus not

- list of variables without brackets refers to variances and residual variances
  - example: f1 y1-y9;
  - f1@0; - The *variance* of f1 is set to 0 Var. if exogenous; resid. var. if endogenous
- [list of variables] refers to means, intercepts, thresholds
  - example: [f1, y1-y9];
  - [f1]@0; - The *mean* of f1 is set to 0 Mean if exogenous; intcpt if endogenous

85

Latent Growth Models



86

Latent Growth Models

- Can specify with long code:

```
Int BY time1@1;
Int BY time2@1;
Int BY time2@1;
Int BY time2@1;
Slope BY time1@0;
Slope BY time2@1;
Slope BY time3@2;
Slope BY time4@3;
Int WITH Slope;
[time1 time2 time3 time4]@0;
```

"Int" and "Slope" are names I created.
- Not already in the data.
- Mplus does not require any specific name.

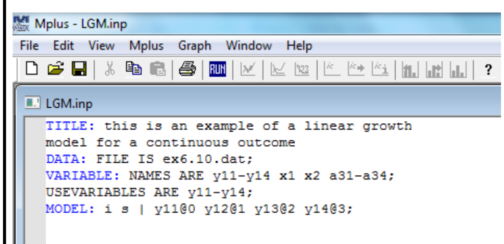
time1-time4 are variable names

- Or use Mplus' shortcut
  - `Intercept slope | time1@0 time2@1 time3@2 time4@3;`
  - Assumes intercept is 1's all around
  - Creates paths you specify for slope
  - Allows intercept and slope to correlate
  - Sets variable intercepts to 0 so that all prediction is in the mean of the latent variables (Intercept and Slope)

"Intercept" and "slope" are still labels I created. Can be whatever you want.

87

Latent Growth Models



I shortened intercept and slope to "i" and "s"; they can be whatever name you want

time1-time4 are labeled y11-y14 (studying adolescents)

88

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I |				
Y11	1.000	0.000	999.000	999.000
Y12	1.000	0.000	999.000	999.000
Y13	1.000	0.000	999.000	999.000
Y14	1.000	0.000	999.000	999.000
S |				
Y11	0.000	0.000	999.000	999.000
Y12	1.000	0.000	999.000	999.000
Y13	2.000	0.000	999.000	999.000
Y14	3.000	0.000	999.000	999.000
S WITH				
I	0.559	0.060	9.282	0.000
Means				
I	0.620	0.069	9.048	0.000
S	1.049	0.035	29.972	0.000
Intercepts				
Y11	0.000	0.000	999.000	999.000
Y12	0.000	0.000	999.000	999.000
Y13	0.000	0.000	999.000	999.000
Y14	0.000	0.000	999.000	999.000
Variances				
I	1.943	0.152	12.772	0.000
S	0.490	0.040	12.148	0.000
Residual Variances				
Y11	0.545	0.074	7.412	0.000
Y12	0.694	0.056	12.343	0.000

Loadings we specified.

All 1's for intercept.

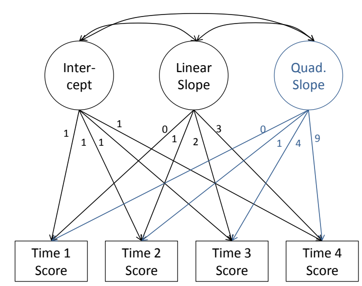
0,1,2,3 for linear growth slope.

Focus of analysis. What is initial value for construct? What is growth?

Set to zero so that prediction is all captured in the means

89

Latent Growth Models



90

Latent Growth Models

```
TITLE: this is an example of a linear growth
model for a continuous outcome
DATA: FILE IS ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
USEVARIABLES ARE y11-y14;
MODEL: i s q | y11@0 y12@1 y13@2 y14@3;
```

- Added “q” for the quadratic term
- Assigned loadings for *linear* term
- Mplus knows to square loadings for “q”

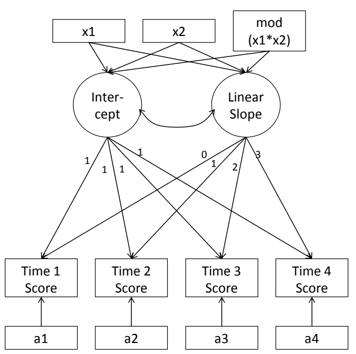
Latent Growth Models

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

 I        |
    Y11                1.000      0.000    999.000    999.000
    Y12                1.000      0.000    999.000    999.000
    Y13                1.000      0.000    999.000    999.000
    Y14                1.000      0.000    999.000    999.000

 S        |
    Y11                0.000      0.000    999.000    999.000
    Y12                1.000      0.000    999.000    999.000
    Y13                2.000      0.000    999.000    999.000
    Y14                3.000      0.000    999.000    999.000

 Q        |
    Y11                0.000      0.000    999.000    999.000
    Y12                1.000      0.000    999.000    999.000
    Y13                4.000      0.000    999.000    999.000
    Y14                9.000      0.000    999.000    999.000

 S        WITH
    I                  0.722      0.272      2.656      0.008

 Q        WITH
    I                 -0.059      0.065     -0.912      0.362
    S                 -0.122      0.060     -2.027      0.043

 Means
    I                  0.611      0.069      8.889      0.000
    S                  1.082      0.067     16.196      0.000
    Q                 -0.011      0.018     -0.597      0.551

 Intercepts
    Y11                0.000      0.000    999.000    999.000
    Y12                0.000      0.000    999.000    999.000
```

Q loadings are squared S loadings

Construct starts at 0.611
Grows 1.082 each year/week/etc.
Quadratic growth term was not sig.

Latent Growth Models



Latent Growth Models

```
TITLE: this is an example of a linear growth
model for a continuous outcome with predictors;
DATA: FILE IS ex6.10.dat;
VARIABLE: NAMES ARE y11-y14 x1 x2 a31-a34;
USEVARIABLES ARE y11-y14 x1 x2 a31-a34 mod;
DEFINE: mod = x1 * x2;
MODEL: i s | y11@0 y12@1 y13@2 y14@3;
i s ON x1 x2 mod;
y11 ON a31;
y12 ON a32;
y13 ON a33;
y14 ON a34;
```

- Not in dataset. Create in DEFINE.
- LGM language
- ON statements (path analyses)

- Combining LGM language with ON statements
- Time-***invariant*** predictors for ***i*** and ***s***
- Time-***varying*** predictors for individual ***timepoints***

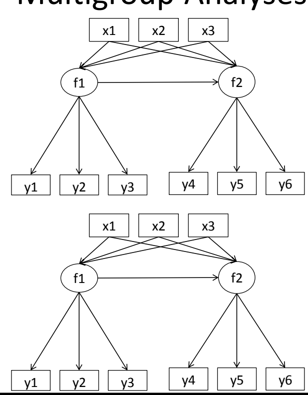

```
                                                 Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 I        |
    Y11                1.000      0.000    999.000    999.000
    Y12                1.000      0.000    999.000    999.000
    Y13                1.000      0.000    999.000    999.000
    Y14                1.000      0.000    999.000    999.000

 S        |
    Y11                0.000      0.000    999.000    999.000
    Y12                1.000      0.000    999.000    999.000
    Y13                2.000      0.000    999.000    999.000
    Y14                3.000      0.000    999.000    999.000

 I        ON
    X1                 0.569      0.054     10.475      0.000
    X2                 0.713      0.055     12.887      0.000
    MOD               -0.110      0.055     -1.990      0.047

 S        ON
    X1                 0.262      0.025     10.393      0.000
    X2                 0.474      0.026     18.436      0.000
    MOD                0.021      0.026      0.834      0.404

 Y11      ON
    A31                0.186      0.044      4.197      0.000

 Y12      ON
    A32                0.323      0.038      8.447      0.000

 Y13      ON
    A33                0.344      0.038      8.982      0.000

 Y14      ON
    A34                0.301      0.051      5.947      0.000
```

- Predictors’ influence on baseline values
- Predictors’ influence on growth slopes
- Controlling for time-specific covariates (or main predictors)

Multigroup Analyses

Males:

Females:



Multigroup Analysis Code

- Approach changed in newer versions
- Version 6 and older
  - The "MODEL" describes the overall model to be estimated for each group
  - Default is that ***ALL*** code under the "MODEL" command was constrained to equality across groups unless an exception was made
  - Exceptions were specified using "MODEL [group]" command after the overall command
- Version 7 and newer
  - The "MODEL" describes the overall model to be estimated for each group
  - Default is for ***measurement*** to be constrained, but ***structure*** to be different
    - Factor loadings are held equal across groups
    - Intercepts (for continuous variables) and thresholds (for categorical variables) are held equal across groups
    - Paths such as ON and WITH are estimated separately for each group
  - Exceptions were specified using "MODEL [group]" command after the overall command

Multigroup Analyses: Version 6 and older

```
TITLE: this is an example of a multigroup
   analysis with all continuous indicators.
   Testing invariance, too.
   All data are fake;
DATA: FILE IS multigroup.dat;
VARIABLE:    NAMES ARE y1 y2 y3 y4 y5 y6 x1 x2 x3 gender;
    GROUPING IS gender (1=male 2=female);
MODEL:
    F1 BY y1 y2 y3;
    F2 BY y4 y5 y6;
    F1 F2 ON x1 x2 x3;
    F2 ON F1;
MODEL female:
!    F1 BY y1 y2 y3;
!    F2 BY y4 y5 y6;
    F1 F2 ON x1 x2 x3;
    F2 ON F1;
OUTPUT: stand cint;
```

How you indicate you are doing a multigroup analysis: Specifying grouping variable AND group labels

Second set of model code allows estimates to be different from original model

Inactivated CFA code because construct needs to be consistently measured across groups

Multigroup Analyses: Version 7 and newer

multigroupTEST.inp

```
TITLE: testing multigroup code in version 7.
All data are fake;
DATA: FILE IS multigroup.dat;
VARIABLE: NAMES are y1 y2 y3 y4 y5 y6 x1 x2 x3 gender;
    GROUPING IS gender (1=male 2=female);
MODEL:
    F1 BY y1 y2 y3;
    F2 BY y4 y5 y6;
    F1 F2 ON x1 x2 x3;
    F2 ON F1;
!MODEL female:
!    F1 BY y1 y2 y3;
!    F2 BY y4 y5 y6;
!    F1 F2 ON x1 x2 x3;
!    F2 ON F1;
OUTPUT: stand cint;
```

How you indicate you are doing a multigroup analysis: Specifying grouping variable AND group labels

Second set of model code allows estimates to be different from original model

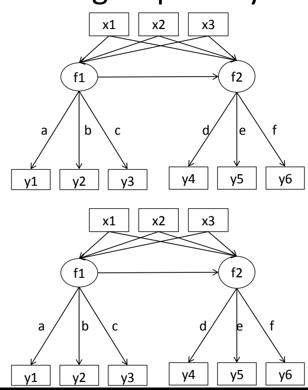
Inactivated CFA code because construct needs to be consistently measured across groups

Inactivated path code because new default for v7+ is to allow these to vary across groups

Multigroup Analyses

Males:

Females:



Multigroup Analyses

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Group MALE

 F1       BY
    Y1                 1.000      0.000    999.000    999.000
    Y2                 1.016      0.021     48.830      0.000
    Y3                 0.644      0.026     24.837      0.000

 F2       BY
    Y4                 1.000      0.000    999.000    999.000
    Y5                 1.001      0.018     55.703      0.000
    Y6                 1.007      0.018     54.562      0.000

 F2       ON
    F1                 0.285      0.050      5.744      0.000

 F1       ON
    X1                 0.515      0.027     18.963      0.000
    X2                 0.598      0.032     18.596      0.000
    X3                 0.719      0.045     15.863      0.000

 F2       ON
    X1                 0.517      0.034     15.006      0.000
    X2                 0.419      0.040     10.447      0.000
    X3                 0.218      0.052      4.204      0.000

 Intercepts
    Y1                 0.061      0.109      0.558      0.577
```

Factor loadings

Predictive paths

Multigroup Analyses

```
Group FEMALE

 F1       BY
    Y1                 1.000      0.000    999.000    999.000
    Y2                 1.016      0.021     48.830      0.000
    Y3                 0.644      0.026     24.837      0.000

 F2       BY
    Y4                 1.000      0.000    999.000    999.000
    Y5                 1.001      0.018     55.703      0.000
    Y6                 1.007      0.018     54.562      0.000

 F2       ON
    F1                 0.404      0.055      7.385      0.000

 F1       ON
    X1                 0.422      0.023     18.415      0.000
    X2                 0.572      0.027     20.841      0.000
    X3                 0.615      0.037     16.510      0.000

 F2       ON
    X1                 0.511      0.032     15.929      0.000
    X2                 0.407      0.040     10.055      0.000
    X3                 0.256      0.049      5.268      0.000

 Intercepts
    Y1                 0.061      0.109      0.558      0.577
    Y2                 0.036      0.110      0.330      0.741
```

Factor loadings (identical)

Predictive paths (unique)

Testing Measurement Invariance

```
TITLE: this is an example of a multigroup
   analysis with all continuous indicators.
   Testing invariance, too.
   All data are fake;
DATA: FILE IS multigroup.dat;
VARIABLE:   NAMES ARE y1 y2 y3 y4 y5 y6 x1 x2 x3 gender;
    GROUPING IS gender (1=male 2=female);
MODEL:
   F1 BY y1 y2 y3;
   F2 BY y4 y5 y6;
   F1 F2 ON x1 x2 x3;
   F2 ON F1;
MODEL female:
   F1 BY y1 y2 y3;
   F2 BY y4 y5 y6;
   F1 F2 ON x1 x2 x3;
   F2 ON F1;
OUTPUT: stand cint;
```

Activated CFA code for females

- Run with factor loadings free, and constrained
- Compare model fit
- Can conduct likelihood ratio test (nested models)

103

Testing Measurement Invariance

Constrained Model

```
MODEL FIT INFORMATION

Number of Free Parameters                        42

Loglikelihood

          H0 Value                        -9093.735
          H1 Value                        -8790.546

Information Criteria

          Akaike (AIC)                    18271.470
          Bayesian (BIC)                  18481.598
          Sample-Size Adjusted BIC        18348.196
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             606.379
          Degrees of Freedom                     48
          P-Value                            0.0000

Chi-Square Contributions From Each Group

          MALE                              451.523
          FEMALE                            154.856

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.145
          90 Percent C.I.                     0.135
          Probability RMSEA <= .05            0.000

CFI/TLI

          CFI                                 0.936
```

Unconstrained Model

```
MODEL FIT INFORMATION

Number of Free Parameters                        48

Loglikelihood

          H0 Value                        -8809.887
          H1 Value                        -8790.546

Information Criteria

          Akaike (AIC)                    17715.774
          Bayesian (BIC)                  17955.921
          Sample-Size Adjusted BIC        17803.462
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              38.683
          Degrees of Freedom                     42
          P-Value                            0.6174

Chi-Square Contributions From Each Group

          MALE                               20.286
          FEMALE                             18.397

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.000
          90 Percent C.I.                     0.000
          Probability RMSEA <= .05            1.000

CFI/TLI

          CFI                                 1.000
```

$\chi^2(6) = 567.696$, $p < .001$ – Significant Misfit

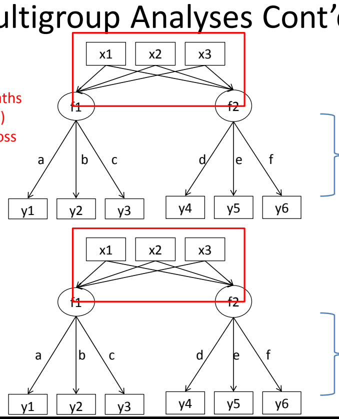
104

Multigroup Analyses Cont'd

Males:

Testing to see if paths (structural models) are consistent across groups

Females:



constrained loadings

constrained loadings

105

Multigroup Analyses Cont'd: Version 6 and older

```
TITLE: this is an example of a multigroup
   analysis with all continuous indicators.
   Testing invariance, too.
   All data are fake;
DATA: FILE IS multigroup.dat;
VARIABLE:   NAMES ARE y1 y2 y3 y4 y5 y6 x1 x2 x3 gender;
    GROUPING IS gender (1=male 2=female);
MODEL:
   F1 BY y1 y2 y3;
   F2 BY y4 y5 y6;
   F1 ON x1 (1);
   F1 ON x2 (2);
   F1 ON x3 (3);
   F2 ON x1 (4);
   F2 ON x2 (5);
   F2 ON x3 (6);
   F2 ON F1 (7);
MODEL female:
!   F1 BY y1 y2 y3;
!   F2 BY y4 y5 y6;
   F1 ON x1 (1);
   F1 ON x2 (2);
   F1 ON x3 (3);
   F2 ON x1 (4);
   F2 ON x2 (5);
   F2 ON x3 (6);
   F2 ON F1 (7);
OUTPUT: stand cint;
```

- Equality Constraints
- Making each path consistent across groups
- Paths with the same (#) are constrained to equality with one another

Not technically necessary under "MODEL female" because default is to keep everything the same unless otherwise specified

Does not hurt to include them to be sure software is doing exactly what you want.

106

Multigroup Analyses Cont'd: Version 7 and newer

```
TITLE: this is an example of a multigroup
   analysis with all continuous indicators.
   Testing invariance, too.
   All data are fake;
DATA: FILE IS multigroup.dat;
VARIABLE:   NAMES ARE y1 y2 y3 y4 y5 y6 x1 x2 x3 gender;
    GROUPING IS gender (1=male 2=female);
MODEL:
   F1 BY y1 y2 y3;
   F2 BY y4 y5 y6;
   F1 ON x1 (1);
   F1 ON x2 (2);
   F1 ON x3 (3);
   F2 ON x1 (4);
   F2 ON x2 (5);
   F2 ON x3 (6);
   F2 ON F1 (7);
MODEL female:
!   F1 BY y1 y2 y3;
!   F2 BY y4 y5 y6;
   F1 ON x1 (1);
   F1 ON x2 (2);
   F1 ON x3 (3);
   F2 ON x1 (4);
   F2 ON x2 (5);
   F2 ON x3 (6);
   F2 ON F1 (7);
OUTPUT: stand cint;
```

- Equality Constraints
- Making each path consistent across groups
- Paths with the same (#) are constrained to equality with one another

Are definitely necessary under "MODEL female" because default in V7+ is to allow ON paths to be estimated separately for each group unless otherwise specified.

This code specifies that they should be the same across groups.

107

Multigroup Analyses Cont'd

MODEL RESULTS

Group MALE

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1	BY				
	Y1	1.000	0.000	999.000	999.000
	Y2	1.017	0.021	49.122	0.000
	Y3	0.626	0.025	25.508	0.000
F2	BY				
	Y4	1.000	0.000	999.000	999.000
	Y5	0.999	0.018	55.773	0.000
	Y6	1.007	0.018	54.575	0.000
F2	ON				
	F1	0.329	0.037	8.994	0.000
F1	ON				
	X1	0.461	0.018	25.596	0.000
	X2	0.586	0.021	27.379	0.000
	X3	0.657	0.029	22.319	0.000
F2	ON				
	X1	0.522	0.024	21.978	0.000
	X2	0.421	0.029	14.744	0.000
	X3	0.249	0.035	7.042	0.000
Intercepts					
	Y1	0.195	0.082	2.387	0.017
	Y2	0.173	0.082	2.098	0.036
	Y3	0.138	0.061	2.268	0.023
	Y4	0.076	0.076	1.001	0.317

Factor loadings equal (like before)

Predictive/structural paths

108

Multigroup Analyses Cont'd

```
Group FEMALE

 F1       BY
    Y1              1.000      0.000    999.000    999.000
    Y2              1.017      0.021     49.122      0.000
    Y3              0.626      0.025     25.508      0.000

 F2       BY
    Y4              1.000      0.000    999.000    999.000
    Y5              0.999      0.018     55.773      0.000
    Y6              1.007      0.018     54.575      0.000

 F2       ON
    F1              0.329      0.037      8.994      0.000

 F1       ON
    X1              0.461      0.018     25.596      0.000
    X2              0.586      0.021     27.379      0.000
    X3              0.657      0.029     22.319      0.000

 F2       ON
    X1              0.522      0.024     21.978      0.000
    X2              0.421      0.029     14.744      0.000
    X3              0.249      0.035      7.042      0.000

 Intercepts
    Y1              0.195      0.082      2.387      0.017
    Y2              0.173      0.082      2.098      0.036
    Y3              0.138      0.061      2.268      0.023
    Y4              0.076      0.076      1.001      0.317
```

Factor loadings equal (like before)

Predictive/structural paths now ALSO equal

109

Multigroup Analyses Cont'd

Constrained Model

```
MODEL FIT INFORMATION

Number of Free Parameters                        35

Loglikelihood

          H0 Value                        -9104.524
          H1 Value                        -8790.546

Information Criteria

          Akaike (AIC)                    18279.048
          Bayesian (BIC)                  18454.156
          Sample-Size Adjusted BIC        18342.987
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             627.957
          Degrees of Freedom                     55
          P-Value                            0.0000

Chi-Square Contributions From Each Group

          MALE                              480.793
          FEMALE                            147.164

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.138
          90 Percent C.I.                     0.128
          Probability RMSEA <= .05            0.000

CFI/TLI

          CFI
```

Unconstrained Model

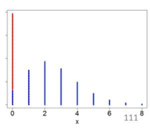
```
MODEL FIT INFORMATION

Number of Free Parameters                        42

Loglikelihood

          H0 Value                        -9093.735
          H1 Value                        -8790.546

Information Criteria

          Akaike (AIC)                    18271.470
          Bayesian (BIC)                  18481.598
          Sample-Size Adjusted BIC        18348.196
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                             606.379
          Degrees of Freedom                     48
          P-Value                            0.0000

Chi-Square Contributions From Each Group

          MALE                              451.523
          FEMALE                            154.856

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.145
          90 Percent C.I.                     0.135
          Probability RMSEA <= .05            0.000

CFI/TLI

          CFI                                 0.936
```

$\chi^2(7) = 21.578$, $p = .001$ – **Significant Misfit**

110

Noncontinuous Variables

- **CATEGORICAL ARE** (or **IS**)**:** names of binary and ordered categorical (ordinal) variables;
- **NOMINAL ARE** (or **IS**)**:** names of unordered categorical (nominal) variables;
- **COUNT ARE** (or **IS**)**:** names of count variables;
  - Poisson distribution models
  - Zero-Inflated Poisson (ZIP) models

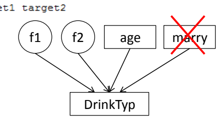


111

Noncontinuous Variables

- Still must follow rules of SEM (and regression)
- No nominal predictors
  - Need to dummy code into relevant groups (0,1)
  - EXAMPLE: Marital status = 6 groups (unordered)

```
   USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1 target2
   target3 target4 target5 age marry DrinkTyp;
   !MISSING IS all (999);
   NOMINAL ARE marry;
ANALYSIS:  ITERATIONS = 10000;
MODEL:
   F1 BY gen1 gen2 gen3 gen4 gen5;
   F2 BY target1 target2 target3 target4 target5;
   DrinkTyp on F1 F2 age marry;
   F2 ON F1;
```



fullsem_nominalx.out

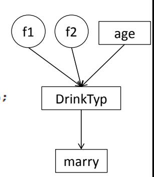
```
*** ERROR in MODEL command
  A nominal variable may not appear on the right-hand side of an ON
  statement:  MARRY
```

112

Noncontinuous Variables

- CAN run analyses with categorical outcomes
- Somewhat equivalent to logistic regressions

```
   USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1 target2
   target3 target4 target5 age marry DrinkTyp;
   !MISSING IS all (999);
   NOMINAL ARE marry;
ANALYSIS:  ITERATIONS = 10000;
MODEL:
   F1 BY gen1 gen2 gen3 gen4 gen5;
   F2 BY target1 target2 target3 target4 target5;
   DrinkTyp on F1 F2 age;
   Marry ON DrinkTyp;
   F2 ON F1;
OUTPUT: mod(all) stand cint;
```



113

Noncontinuous Variables

fullsem_nominaly.out

```
    TARGET3          1.066      0.048     22.072      0.000
    TARGET4          1.115      0.067     16.727      0.000
    TARGET5          1.111      0.067     16.639      0.000

 F2         ON
    F1               0.942      0.154      6.110      0.000

 DRINKTYP   ON
    F1              -4.809      1.361     -3.534      0.000
    F2               1.185      0.337      3.514      0.000

 DRINKTYP   ON
    AGE             -0.006      0.046     -0.124      0.901

 MARRY#1    ON
    DRINKTYP         0.353      0.060      5.894      0.000

 MARRY#2    ON
    DRINKTYP         0.336      0.024     14.258      0.000

 MARRY#3    ON
    DRINKTYP         0.243      0.035      7.032      0.000

 MARRY#4    ON
    DRINKTYP         0.275      0.064      4.266      0.000

 MARRY#5    ON
    DRINKTYP         0.309      0.025     12.404      0.000

 Intercepts
    GEN1             4.738      0.036    132.288      0.000
    GEN2             4.337      0.049     87.645      0.000
    GEN3             4.779      0.032    151.544      0.000
```

Traditional structural paths

Logistic structural paths

For *k* classes, has *k*-1 estimates: Reflect probability of being in current class versus final class, given x

Default is to use final class for comparison purposes (may want to recode prior to analysis)

114

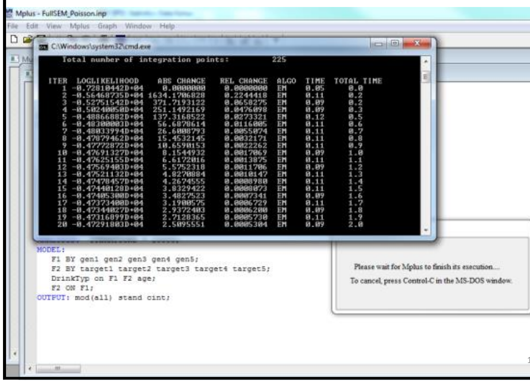
Noncontinuous Variables

- Poisson (count) variables

FullSEM_Poisson.inp

```
TITLE: this is an example of a full SEM
   model with all continuous variables.
   All data are fake;
DATA: FILE IS fullSEM.dat;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14
   Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25 Q26 Q27 Q28 Q29
   Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38 Q39 Q40 Q41 Q42 Q43 Q44
   Q45 Q46 Q47 Q48 Q49 Q50 Q51 Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59
   Q60 Q61 Q62 Q63 Q64 Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2
   gen3 gen4 gen5 target1 target2 target3 target4 target5 age
   FTPT res greek gpa race hisp year athlete gender marry
   DrinkTyp DrinkHvy Probs;
   USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1 target2
   target3 target4 target5 age DrinkTyp;
   !MISSING IS all (999);
   COUNT ARE DrinkTyp;
ANALYSIS:  ITERATIONS = 10000;
MODEL:
   F1 BY gen1 gen2 gen3 gen4 gen5;
   F2 BY target1 target2 target3 target4 target5;
   DrinkTyp on F1 F2 age;
   F2 ON F1;
OUTPUT: mod(all) stand cint;
```


Processing



Noncontinuous Variables

fullsem_poisson.out

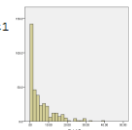
TARGET4	1.116	0.068	16.503	0.000
TARGET5	1.124	0.071	15.931	0.000
F2 ON				
F1	1.856	0.426	4.358	0.000
DRINKTYP ON				
F1	-7.010	1.651	-4.246	0.000
F2	2.628	0.559	4.701	0.000
DRINKTYP ON				
AGE	0.009	0.013	0.676	0.499
Intercepts				
GEN1	4.738	0.036	132.315	0.000
GEN2	4.337	0.049	87.647	0.000
GEN3	4.779	0.032	151.592	0.000
GEN4	4.586	0.044	105.039	0.000
GEN5	4.630	0.041	113.954	0.000
TARGET1	4.593	0.049	93.616	0.000
TARGET2	4.436	0.054	82.281	0.000
TARGET3	4.535	0.055	81.867	0.000
TARGET4	4.499	0.054	82.837	0.000
TARGET5	4.552	0.054	84.542	0.000
DRINKTYP	0.778	0.157	4.953	0.000
Variances				
F1	0.104	0.043	2.394	0.017

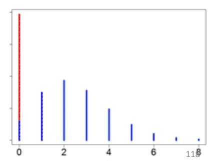
Traditional coefficients (adjusted)

Noncontinuous Variables

- Zero-Inflated Poisson (ZIP models)

```
   USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1
   target3 target4 target5 age DrinkTyp;
   !MISSING IS all (999);
   COUNT ARE DrinkTyp (i);
ANALYSIS:  ITERATIONS = 10000;
MODEL:
   F1 BY gen1 gen2 gen3 gen4 gen5;
   F2 BY target1 target2 target3 target4 target5;
   DrinkTyp on F1 F2 age;
   DrinkTyp#1 on F1 F2 age;
   F2 ON F1;
OUTPUT: mod(all) stand cint;
```





Noncontinuous Variables

fullsem_zippoisson.out

F2 ON				
F1	1.281	0.217	5.893	0.000
DRINKTYP ON				
F1	-3.048	0.647	-4.711	0.000
F2	1.531	0.307	4.988	0.000
DRINKTYP#1 ON				
F1	-0.732	1.352	-0.541	0.588
F2	1.245	0.987	1.261	0.207
DRINKTYP ON				
AGE	-0.005	0.011	-0.483	0.629
DRINKTYP#1 ON				
AGE	-0.038	0.026	-1.455	0.146
Intercepts				
GEN1	4.738	0.036	132.562	0.000
GEN2	4.338	0.049	87.719	0.000
GEN3	4.779	0.031	151.862	0.000
GEN4	4.586	0.044	105.232	0.000
GEN5	4.631	0.041	114.157	0.000
TARGET1	4.595	0.049	94.161	0.000
TARGET2	4.437	0.054	82.631	0.000

Traditional coefficients

Additional ***logit*** coefficients

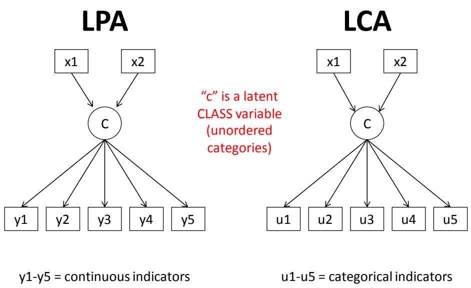
If they are a drinker, how does age/f1/f2 impact ***how much*** they drink?

How does age/f1/f2 impact the ***probability*** of being a drinker (a ***non-zero***)?

Noncontinuous Variables

- One exception to rule:
- **LATENT VARIABLES** may be nominal predictors
- Most common version of this…

Latent Profile/Class Analysis

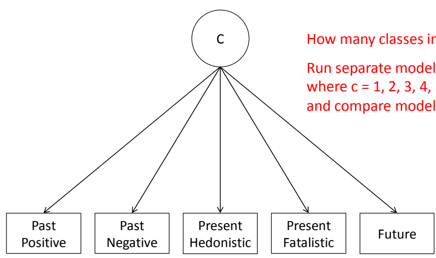


Latent Profile/Class Analysis

- Mplus calls this "mixture modeling with cross-sectional data" (chapter 7)
- Longitudinal version is often called Growth Mixture Modeling (chapter 8)
- Only covering cross-section data today (LPA/LCA), but same principles apply to longitudinal data (GMM)

Latent **Profile**/Class Analysis

- Zimbardo's Time Perspective (5 Facets)



How many classes in c?

Run separate models where c = 1, 2, 3, 4, 5, etc. and compare model fit

Latent **Profile**/Class Analysis

LPA_c3.inp

```
TITLE: Latent Profile Analysis of Zimbardo Time Perspective.
Continuous indicators.
DATA: FILE IS LPA.dat;
VARIABLE: NAMES ARE id sex relat alc1-alc6 typweek heavweek BTP ABTP
peakBAC TDDtyp TDDheav age raceD liveD greekD Qtyp Qheav
ztpi_pp ztpi_pn ztpi_ph ztpi_pf ztpi_f;
USEVARIABLES ARE ztpi_pp ztpi_pn ztpi_ph ztpi_pf ztpi_f;
CLASSES = c (3);
Missing are ALL (999);
Analysis:
TYPE = MIXTURE;
starts=20 5;
Output:
Tech11;
!savedata:
!save=cprobabilities;
!file is cprob.dat;
```

C does not exist in the data. Creating it (like any other latent factor). "(3)" indicates we believe there are three groups/profiles.

←Default type is GENERAL, so need to include code to change it to MIXTURE

←Number of initial starts and final stage optimizations. Default of 10, 2 is often not enough.

←Allows you to save the probabilities of being in each class for each participant. Helpful if you plan on predicting classes from covariates, or using class to predict outcomes. Not necessary YET.

Tech11 output includes Lo-Mendell-Rubin Adjusted LRT (compares fit for current number of classes to one fewer)

Latent **Profile**/Class Analysis

- Warning!

lpa_c3.out

```
RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:

          -1856.443  462953           7
          -1856.443  unperturbed      0
          -1856.443  107446           12
          -1863.848  573096           20
          -1863.848  650371           14


    WARNING:  WHEN ESTIMATING A MODEL WITH MORE THAN TWO CLASSES, IT MAY BE
    NECESSARY TO INCREASE THE NUMBER OF RANDOM STARTS USING THE STARTS OPTION
    TO AVOID LOCAL MAXIMA.

THE MODEL ESTIMATION TERMINATED NORMALLY


MODEL FIT INFORMATION

Number of Free Parameters                       22

Loglikelihood

         H0 Value                        -1856.443
```

Already took care of this with **STARTS 20 5;**

Should say **HELPFUL TIP: When estimating…**

Latent **Profile**/Class Analysis

- Fit (abbreviated list)

lpa_c3.out

```
MODEL FIT INFORMATION

Number of Free Parameters                       22

Loglikelihood

          H0 Value                        -1856.443
          H0 Scaling Correction Factor        1.140
            for MLR

Information Criteria

          Akaike (AIC)                     3756.886
          Bayesian (BIC)                   3849.253
          Sample-Size Adjusted BIC         3779.425
            (n* = (n + 2) / 24)


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
```

Latent **Profile**/Class Analysis

- ***Relative*** Entropy and class counts/proportions

```
lpa_c3.out
CLASSIFICATION QUALITY

     Entropy                         0.700

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

    Latent
   Classes

       1             36          0.07317
       2            338          0.68699
       3            118          0.23984
```

7.3% of sample in class 1

Latent **Profile**/Class Analysis

- Lo-Mendell-Rubin adjusted LRT

```
lpa_c3.out
TECHNICAL 11 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                  20
        Number of final stage optimizations                     5



     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

          H0 Loglikelihood Value                        -1878.192
          2 Times the Loglikelihood Difference             43.497
          Difference in the Number of Parameters                6
          Mean                                             15.586
          Standard Deviation                               16.985
          P-Value                                          0.0648

     LO-MENDELL-RUBIN ADJUSTED LRT TEST

          Value                                            42.358
          P-Value                                          0.0694
```


Latent **Profile**/Class Analysis

- Compare model fit across number of classes

Classes:	AIC	BIC	Adjusted BIC	Relative Entropy	LMR *p*	Proportion of smallest group
1	3952.805	3994.790	3963.050	--	--	--
2	3788.383	3855.559	3804.775	0.620	.0000	.341
3	3756.886	3849.253	3779.425	0.700	.0694	.091
4	3736.655	3854.213	3765.341	0.754	.1910	.012
5	3722.782	3865.530	3757.614	0.773	.1497	.013
6	3715.194	3883.134	3756.174	0.753	.7717	.012
7	3703.795	3896.925	3750.921	0.766	.7638	.011

Groups with ptn < .05

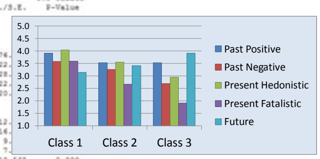
Separate Analyses:

```
CLASSES = c(1), CLASSES = c(2), CLASSES = c(3), CLASSES = c(4),
CLASSES = c(5), CLASSES = c(6), CLASSES = c(7)
```


Latent **Profile**/Class Analysis

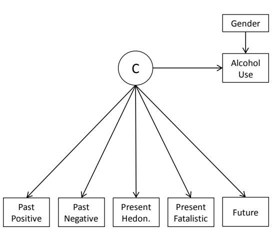
- Class Means (for final set of classes)

```
lpa_c3.out
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Latent Class 1

 Means
    ZTPI_PP            3.826      0.050      76.
    ZTPI_PN            3.493      0.153      22.
    ZTPI_PH            3.989      0.138      28.
    ZTPI_PF            3.424      0.153      22.
    ZTPI_F             3.161      0.156      20.

 Variances
    ZTPI_PP            0.191      0.015      12.
    ZTPI_PN            0.399      0.025      16.
    ZTPI_PH            0.177      0.018       9.
    ZTPI_PF            0.157      0.020       7.
    ZTPI_F             0.219      0.016     13.565      0.000

Latent Class 2

 Means
    ZTPI_PP            3.528      0.031    114.823      0.000
    ZTPI_PN            3.244      0.044     74.001      0.000
    ZTPI_PH            3.537      0.038     92.097      0.000
    ZTPI_PF            2.661      0.057     46.373      0.000
    ZTPI_F             3.412      0.044     78.050      0.000

 Variances
    ZTPI_PP            0.191      0.015     12.783      0.000
```

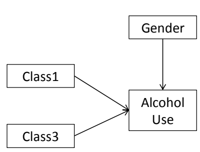


Latent **Profile**/Class Analysis

- Zimbardo's Time Perspective (5 Facets)



No nominal predictors, so…
Dummy Code!



Latent **Profile**/Class Analysis

- Model with classes predicting drinking

```
LPA_c3_qtyp.inp
TITLE: Latent Profiles predicting typical drinking.
Bootstrapping results (because drinking is non-normal).
DATA: FILE IS LPAwithclasses.dat;
VARIABLE: NAMES ARE id sex relat alc1-alc6 typweek heavweek BTP ABTP
peakBAC TDDtyp TDDheav age raceD liveD greekD Qtyp Qheav
ztpi_pp ztpi_pn ztpi_ph ztpi_pf ztpi_f class class1 class3;
USEVARIABLES ARE sex Qtyp class1 class3;
Missing are ALL (999);
Analysis:
Bootstrap = 5000;
Estimator = ml;
MODEL: Qtyp ON class1 class3 sex;
output:
sampstat stand cint(bcboot);
```

- Copy-pasted classifications from **`cprob.dat`** into my original dataset
- Exported into **`LPAwithclasses.dat`**
- **`class`** = nominal variable (1, 2, 3)
- **`class1`** and **`class3`** are dummy coded (0,1) with second class as category of reference (largest class)

Latent **Profile**/Class Analysis

- Prediction results

lpa_c3_qtyp.out

```
MODEL RESULTS

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

 QTYP     ON
    CLASS1             2.110      1.863      1.133      0.257
    CLASS3            -2.701      1.068     -2.528      0.011
    SEX                4.157      1.249      3.327      0.001

 Intercepts
    QTYP               5.244      0.604      8.685      0.000

 Residual Variances
    QTYP             106.851     19.453      5.493      0.000

STANDARDIZED MODEL RESULTS

                       StdYX       StdY        Std
                    Estimate   Estimate   Estimate

 QTYP     ON
    CLASS1             0.052      0.198      2.110
    CLASS3            -0.106     -0.254     -2.701
    SEX                0.186      0.391      4.157

 Intercepts
```

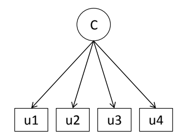
Dummy coding:
B = average number of drinks increase/decrease compared to class 2

Remember, dummy coding is one of the cases where we only want to standardize Y, because now β represents the standardized increase for one class versus another (x=0 versus x=1)

Latent Profile/**Class** Analysis

LCA.inp

```
TITLE:   this is an example of a LCA with
binary latent class indicators.
DATA:    FILE IS LCA.dat;
VARIABLE:    NAMES ARE u1-u4 x1-x10;
    USEVARIABLES = u1-u4;
    CLASSES = c (2);
    CATEGORICAL = u1-u4;
ANALYSIS:    TYPE = MIXTURE;
OUTPUT:  TECH11;
```



Remember, **CATEGORICAL** means binary or ordinal.
If you have unordered 3+ categories, you need to use **NOMINAL**.

Latent Profile/**Class** Analysis

lca.out

```
MODEL FIT INFORMATION

Number of Free Parameters                        9

Loglikelihood

          H0 Value                        -965.244
          H0 Scaling Correction Factor       1.013
            for MLR

Information Criteria

          Akaike (AIC)                    1948.488
          Bayesian (BIC)                  1986.420
          Sample-Size Adjusted BIC        1957.853
            (n* = (n + 2) / 24)
```

Model Fit

```
     Entropy                       0.904

CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

   Latent
  Classes

     1             127         0.25400
     2             373         0.74600
```

Entropy and proportions

Latent Profile/**Class** Analysis

lca.out

```
RESULTS IN PROBABILITY SCALE

Latent Class 1

 U1
    Category 1         0.113      0.037      3.025      0.002
    Category 2         0.887      0.037     23.799      0.000
 U2
    Category 1         0.151      0.038      3.934      0.000
    Category 2         0.849      0.038     22.056      0.000
 U3
    Category 1         0.911      0.031      28.
    Category 2         0.089      0.031      2.
 U4
    Category 1         0.889      0.032      28.
    Category 2         0.111      0.032      3.

Latent Class 2

 U1
    Category 1         0.890      0.018     50.016      0.000
    Category 2         0.110      0.018      6.181      0.000
 U2
    Category 1         0.887      0.018     48.873      0.000
    Category 2         0.113      0.018      6.256      0.000
 U3
    Category 1         0.101      0.018      5.472      0.000
    Category 2         0.899      0.018     48.748      0.000
 U4
    Category 1         0.126      0.020      6.267      0.000
    Category 2         0.874      0.020     43.498      0.000
```

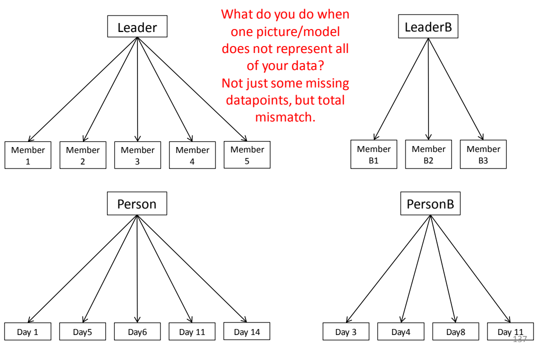
	u1	u2	u3	u4
Class 1	no	no	yes	yes
Class 2	yes	yes	no	no

Probability of membership for each indicator by class

Lo-Mendell-Rubin adjusted LRT available, but omitted for space

Multilevel Modeling

What do you do when one picture/model does not represent all of your data?
Not just some missing datapoints, but total mismatch.



Multilevel Modeling

- ABANDON SEM!
  - Latent variables/structures are not appropriate
- Different number of units within cluster, different spacing of time, etc.
- Conduct Multilevel Modeling (MLM), Hierarchical Linear Modeling (HLM), nested models, mixed models, random effects models, random coefficient models, etc.

Multilevel Modeling

- Other than LGMs (with matching timepoints), multilevel modeling is impossible in most SEM software packages
- Can use HLM (software by SSI)
  - Limited functionality beyond HLM
  - No bootstrapping
  - No path analyses where outcomes are also predictors (e.g., mediation)
- Can use SAS, MIXOR, MLWIN, VARCL, BUGS, or R, but need to learn another language

139

Multilevel Modeling

Level 1:

$$Drinks_{ti} = \pi_{0i} + \pi_{1i}(Bar_{ti}) + \pi_{2i}(Rest_{ti}) + \pi_{3i}(Party_{ti}) + \pi_{4i}(Other_{ti}) + \pi_{5i}(PBS_{ti}) + e_{ti}$$

Level 1: Drinks for person *i* at time *t* depends on: their personal intercept, plus where they drank that day (dummy coded across 4 variables), plus their PBS that day, plus random error

Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01}(Gender_i) + r_{0i}$$
$$\pi_{1i} = \beta_{10},$$
$$\pi_{2i} = \beta_{20},$$
$$\pi_{3i} = \beta_{30},$$
$$\pi_{4i} = \beta_{40},$$
$$\pi_{5i} = \beta_{50}$$

Level 2: A person's personal intercept depends on: their gender.
The effect of location does not vary by gender.
The influence of PBS does not vary by gender.

140

Multilevel Modeling in Mplus

- TYPE = TWOLEVEL RANDOM

```
%WITHIN%
drinks ON Home Bar Rest Party Other PBS;

%BETWEEN%
drinks ON gendD;
```

Level 1: Regress drinks on level-1 predictors (location and PBS).
Drinks for person *i* at time *t* depends on: where they drank that day and plus their PBS that day.
Personal intercept (π_{0i}) and random error (e_{ti}) are included as default.
Level 2: Regress drinks on level-2 predictor (gender).
Personal intercept is influenced by gender.

- TYPE = COMPLEX

```
Drinks ON Home Bar Rest Party Other pbsdo;
Drinks ON gendD;
```

Same interpretation as above.

141

Multilevel Modeling in Mplus

- Code relevant to both:
  `TYPE=COMPLEX` & **`TYPE=TWOLEVEL RANDOM`**
  - **`CLUSTER =`** *name of grouping variable*;
  - **`CENTERING IS`** `GRAND`**`MEAN`** (*variable names*);
  - `GROUP`**`MEAN`** (*variable names*);
- **`TYPE = TWOLEVEL RANDOM`** code only
  - **`WITHIN ARE`** *names of level-1 observed variables*;
  - **`BETWEEN ARE`** *names of level-2 observed variables*;

142

Multilevel Modeling: **TWOLEVEL RANDOM**

HLM_noboot.inp

```
TITLE:   Mplus multilevel mediation for daily drinking
         with PBS (daily) as mediator.
         Place Context and PBS as predictors.
         Drinks as outcome.
         No mediation yet to replicate HLM findings;
DATA:    FILE is DailyL1mplus.csv;
VARIABLE: Names are SONA WeekID Home Bar Rest Party
         Other Alone Friend Fam OPlace drinks pbsplan
         pbsdo pbsall time Weekend age gendD raceD
         greekD residD marryD;
         USEVariables are Home Bar Rest Party
         Other drinks pbsdo gendD;
         WITHIN = Home Bar Rest Party Other;
         BETWEEN = gendD;
         CLUSTER = SONA;
         CENTERING = GRANDMEAN (pbsdo);
ANALYSIS: TYPE = TWOLEVEL RANDOM;
MODEL:
         %WITHIN%
         drinks ON Home Bar Rest Party Other PBSdo;
         %BETWEEN%
         drinks ON gendD;
!Output:  stand;
```

Note "drinks" is not under **`WITHIN`** or **`BETWEEN`**. Outcome does not need to be specified by level.

143

Multilevel Modeling: **TWOLEVEL RANDOM**

hlm_noboot.out

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Within Level				
DRINKS ON				
HOME	0.926	0.359	2.583	0.010
BAR	1.800	0.458	3.928	0.000
REST	0.018	0.396	0.047	0.963
PARTY	3.279	0.368	8.908	0.000
OTHER	2.038	0.559	3.645	0.000
PBSDO	-0.136	0.092	-1.480	0.139
Residual Variances				
DRINKS	9.047	1.041	8.693	0.000
Between Level				
DRINKS ON				
GENDD	1.823	0.447	4.081	0.000
Intercepts				
DRINKS	1.637	0.424	3.863	0.000
Residual Variances				
DRINKS	5.601	0.983	5.697	0.000

144

Multilevel Modeling: COMPLEX

HLM_TYPEcomplex.inp

```
TITLE:   Mplus multilevel mediation for daily drinking
         with PBS (daily) as mediator.
         Place Context and PBS as predictors.
         Drinks as outcome.
         No mediation yet to replicate HLM findings;
DATA:    FILE is DailyL1mplus.csv;
VARIABLE: Names are SONA WeekID Home Bar Rest Party
         Other Alone Friend Fam OPlace drinks pbsplan
         pbsdo pbsall time Weekend age gendD raceD
         greekD residD marryD;
         USEVariables are Home Bar Rest Party
         Other drinks pbsdo gendD;
         CLUSTER = SONA;
         CENTERING = GRANDMEAN (pbsdo);
ANALYSIS: TYPE = COMPLEX;
MODEL:
         Drinks ON Home Bar Rest Party Other pbsdo;
         Drinks ON gendD;

!Output:  stand;
```

No **WITHIN** or **BETWEEN** anywhere.

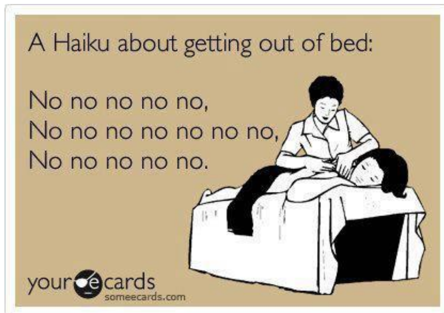
Multilevel Modeling: COMPLEX

hlm_typecomplex.out

```
MODEL RESULTS

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

 DRINKS   ON
    HOME               1.427      0.496      2.880      0.004
    BAR                2.422      0.609      3.977      0.000
    REST              -0.537      0.481     -1.116      0.265
    PARTY              3.654      0.457      7.994      0.000
    OTHER              2.436      0.615      3.959      0.000
    PBSDO             -0.184      0.089     -2.073      0.038
    GENDD              1.745      0.449      3.887      0.000

 Intercepts
    DRINKS             1.505      0.533      2.823      0.005

 Residual Variances
    DRINKS            14.300      1.321     10.829      0.000


QUALITY OF NUMERICAL RESULTS

     Condition Number for the Information Matrix              0.170E-01
       (ratio of smallest to largest eigenvalue)


     Beginning Time:  17:57:40
        Ending Time:  17:57:40
```


Remember… Language Matters

BREAK

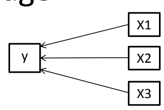


This Section

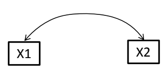
- Analyze examples together
  - I provide SPSS data
  - Together: export datafile
  - Write necessary language for Mplus to read data
  - Write model language for desired analyses
- Full SEM with continuous outcomes
- Latent Growth Models

Basic Model Language

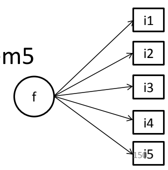
- y **ON** x1 x2 x3 (regression)



- x1 **WITH** x2 (correlation)



- f **BY** item1 item2 item3 item4 item5 (factors or latent variables)



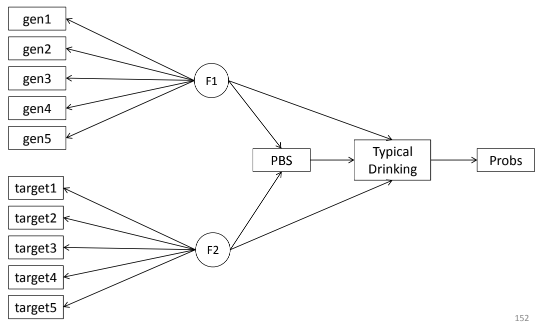
Data

- Mplus3_fullSEM.sav
- No missing data
- Convert to Mplus-compatible file
  - Save as
  - .dat (tab delimited), .cvs (comma delimited), .dat (fixed ASCII)
  - Don't "Write variable names"

151

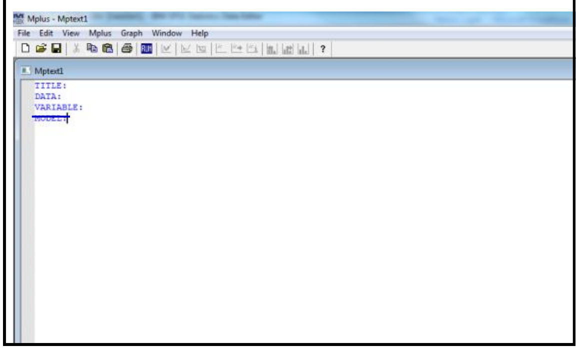
Main Model

- Combining CFAs with Path Analyses



152

Fill In The Blanks

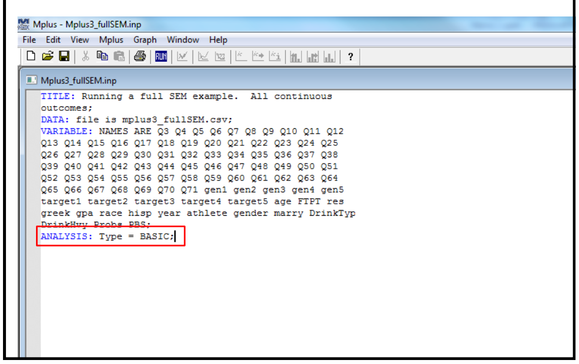


Variable List

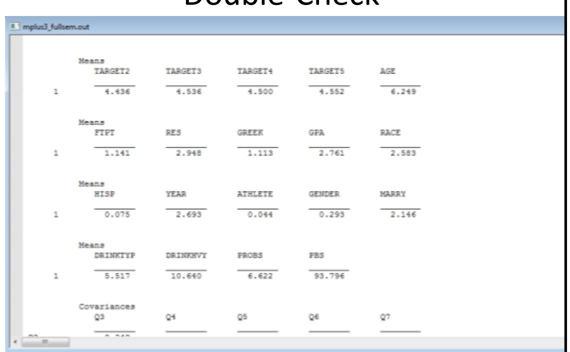
NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;

154

Double-Check

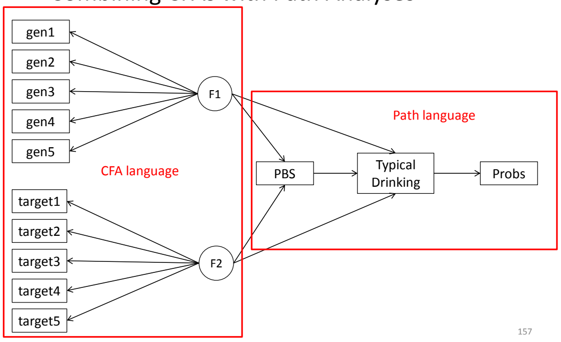


Double-Check



Main Model

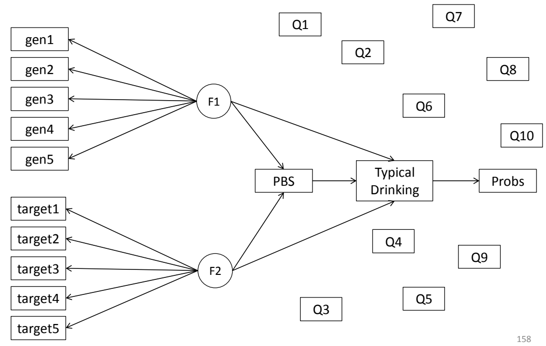
- Combining CFAs with Path Analyses



157

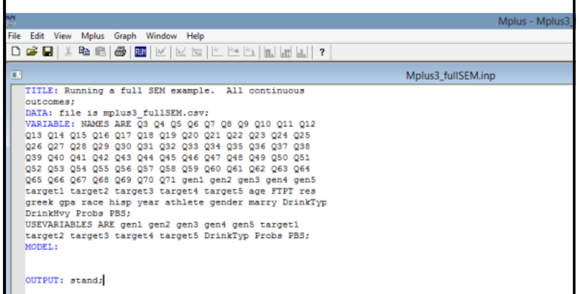
Main Model: Don't End Up With…

- USEVARIABLES are…



158

Model: ???



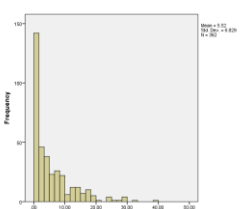

```
TITLE: Running a full SEM example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1
target2 target3 target4 target5 DrinkTyp Probs PBS;
MODEL:


OUTPUT: stand;
```

159

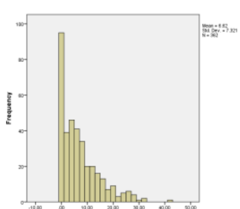
Run It!

160

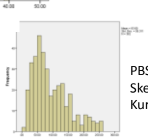
But Wait!



DrinkTyp



Probs

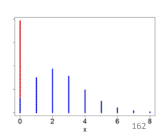


PBS:
Skewness < 1
Kurtosis < 1

161

Noncontinuous Variables

- **`CATEGORICAL ARE`** (or **`IS`**)**:** names of binary and ordered categorical (ordinal) variables;
- **`NOMINAL ARE`** (or **`IS`**)**:** names of unordered categorical (nominal) variables;
- **`COUNT ARE`** (or **`IS`**)**:** names of count variables;
  - Poisson distribution models
  - Zero-Inflated Poisson (ZIP) models



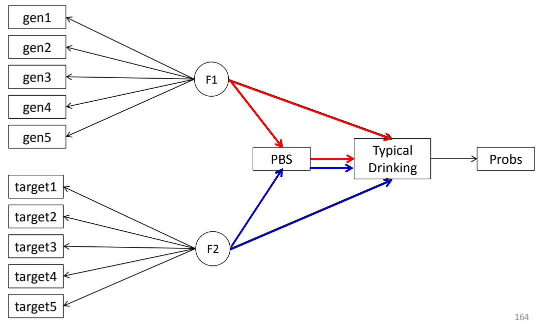
162

Run It!

163

But Wait!

- INDIRECT EFFECT: …



164

But Wait!

- We cannot assume that indirect effects (the combined ab paths) are normally distributed
- What do we do?

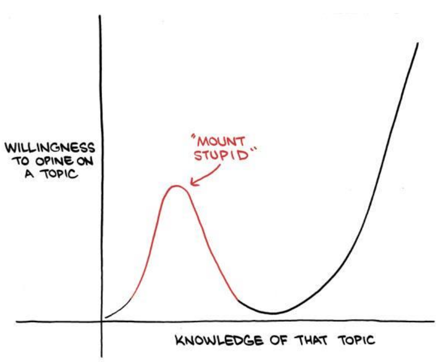
- **Bootstrap!!**
  - **Can delete Poisson code (bootstrap also corrects for non-normality)**

165

Run It!

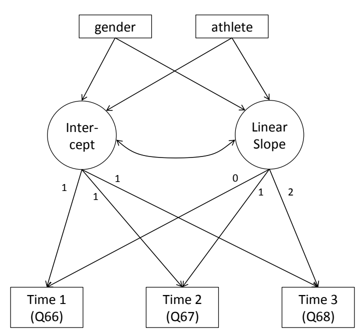
166

On Your Own Time



167

Latent Growth Model



168

Model: ???

```
Mplus3_LatentGrowth.i
TITLE: Running latent growth model example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE Q66 Q67 Q68 athlete gender;
MODEL:


OUTPUT: stand;
```

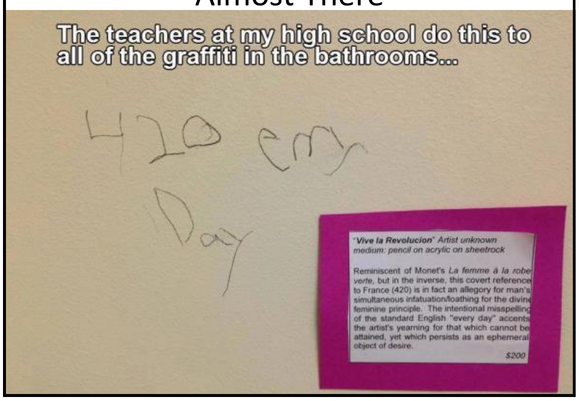
169

Answers

- Coming up!

170

Almost There



ANSWERS! Main Model

```
Mplus - Mplus3_fullSEM.inp
TITLE: Running a full SEM example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1
target2 target3 target4 target5 DrinkTyp Probs PBS;
MODEL:
f1 BY gen1 gen2 gen3 gen4 gen5;
f2 BY target1 target2 target3 target4 target5;
PBS ON f1 f2;
DrinkTyp ON PBS f1 f2;
Probs ON DrinkTyp;
OUTPUT: stand;
```

Factor loadings (BY)

Paths (ON)

172

Results

mplus3_fullsem.out

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
F1 BY				
GEN1	1.000	0.000	999.000	999.000
GEN2	1.313	0.145	9.028	0.000
GEN3	0.946	0.093	10.214	0.000
GEN4	1.533	0.136	11.247	0.000
GEN5	1.369	0.130	10.549	0.000
F2 BY				
TARGET1	1.000	0.000	999.000	999.000
TARGET2	1.002	0.047	21.201	0.000
TARGET3	1.065	0.046	22.969	0.000
TARGET4	1.115	0.042	26.278	0.000
TARGET5	1.111	0.042	26.588	0.000
PBS ON				
F1	6.496	9.181	0.708	0.479
F2	11.501	4.384	2.624	0.009
DRINKTYP ON				
F1	-4.817	1.130	-4.261	0.000
F2	1.189	0.515	2.309	0.021
DRINKTYP ON				
PBS	0.000	0.006	-0.042	0.967
PROBS ON				
DRINKTYP	0.628	0.046	13.760	0.000
F2 WITH				
F1	0.169	0.026	6.484	0.000
Intercepts				
GEN1	4.738	0.036	132.289	0.000

Factor Loadings

Paths

173

ANSWERS! Zero-Inflated Poisson

```
Mplus3_fullSEM_count.inp
TITLE: Running a full SEM example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1
target2 target3 target4 target5 DrinkTyp Probs PBS;
COUNT ARE DrinkTyp(i) Probs(i);
MODEL:
f1 BY gen1 gen2 gen3 gen4 gen5;
f2 BY target1 target2 target3 target4 target5;
PBS ON f1 f2;
DrinkTyp ON PBS f1 f2;
Probs ON DrinkTyp;
DrinkTyp#1 ON PBS f1 f2;
Probs#1 ON DrinkTyp;
OUTPUT: stand;
```

Remember "(i)" means zero-inflated, and is associated with the DV#1 code.

Can omit for regular Poisson distributions.

174

Results

mplus3_fullsem_count.out

```
    GEN4              1.434      0.192      7.468      0.000
    GEN5              1.285      0.210      6.118      0.000

 F2       BY
    TARGET1           1.000      0.000    999.000    999.000
    TARGET2           1.003      0.062     16.091      0.000
    TARGET3           1.063      0.048     22.120      0.000
    TARGET4           1.113      0.066     16.965      0.000
    TARGET5           1.124      0.069     16.251      0.000

 PBS        ON
    F1                7.488     11.510      0.651      0.515
    F2               11.182      4.394      2.545      0.011

 DRINKTYP   ON
    F1               -3.057      0.653     -4.683      0.000
    F2                1.508      0.312      4.840      0.000

 DRINKTYP#1 ON
    F1               -0.457      1.369     -0.334      0.739
    F2                0.970      0.982      0.987      0.324

 DRINKTYP   ON
    PBS               0.000      0.001     -0.116      0.908

 PROBS      ON
    DRINKTYP          0.043      0.005      8.564      0.000

 DRINKTYP#1 ON
    PBS              -0.001      0.003     -0.272      0.785

 PROBS#1    ON
    DRINKTYP         -0.497      0.105     -4.712      0.000

 F2       WITH
    F1                0.195      0.045      4.336      0.000
```

Paths from original model

New paths (#1) identifying impact on likelihood of drinking at all (anything other than 0).

175

ANSWERS! Bootstrapping and Indirect Effects

Mplus3_fullSEM_bootIND.inp

```
TITLE: Running a full SEM example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE gen1 gen2 gen3 gen4 gen5 target1
target2 target3 target4 target5 DrinkTyp Probs PBS;
ANALYSIS: boot = 5000;
MODEL:
f1 BY gen1 gen2 gen3 gen4 gen5;
f2 BY target1 target2 target3 target4 target5;
PBS ON f1 f2;
DrinkTyp ON PBS f1 f2;
Probs ON DrinkTyp;
MODEL INDIRECT: DrinkTyp PBS f1;
MODEL INDIRECT: DrinkTyp PBS f1;
OUTPUT: stand cint(bcboot);
```

Bootstrap samples of n = 5,000

Assessing two different indirect effects (y IND m x)

176

Results

mplus3_fullsem_bootind.out

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


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

Effects from F1 to DRINKTYP

  Sum of indirect     -0.003      0.107     -0.031      0.975

  Specific indirect

    DRINKTYP
    PBS
    F1                -0.002      0.053     -0.031      0.975

    DRINKTYP
    PBS
    F1                -0.002      0.053     -0.031      0.975


STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


STDYX Standardization
```

Combined impact of both indirect effects

Estimates of each individual indirect effect

177

Results

mplus3_fullsem_bootind.out

```
CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS


                 Lower .5%  Lower 2.5%   Lower 5%   Estimate   Upper 5%  Upper 2.5%   Upper .5%

Effects from F1 to DRINKTYP

  Sum of indirect  -0.441     -0.273     -0.198     -0.003      0.136      0.190      0.358

  Specific indirect

    DRINKTYP
    PBS
    F1             -0.221     -0.136     -0.099     -0.002      0.068      0.095      0.179

    DRINKTYP
    PBS
    F1             -0.221     -0.136     -0.099     -0.002      0.068      0.095      0.179

CONFIDENCE INTERVALS OF STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT,
AND DIRECT EFFECTS


STDYX Standardization

                 Lower .5%  Lower 2.5%   Lower 5%   Estimate   Upper 5%  Upper 2.5%   Upper .5%

Effects from F1 to DRINKTYP
```

- 95% CI's use 2.5% upper and lower boundaries
- Zero is in all three intervals, so no indirect effects are significant

178

Latent Growth Models

- Remember: Mplus shortcut
  - `i s | time1@0 time2@1 time3@2;`
  - Assumes intercept is 1's all around
  - Creates paths you specify for slope
  - Allows intercept and slope to correlate
  - Sets variable intercepts to 0 so that all prediction is in the mean of the latent variables (Intercept and Slope)

"i" and "s" are still labels I created for the latent variables. Can be whatever you want.

179

Answers! LGM

Mplus3_LatentGrowth.inp

```
TITLE: Running latent growth model example.  All continuous
outcomes;
DATA: file is mplus3_fullSEM.csv;
VARIABLE: NAMES ARE Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12
Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20 Q21 Q22 Q23 Q24 Q25
Q26 Q27 Q28 Q29 Q30 Q31 Q32 Q33 Q34 Q35 Q36 Q37 Q38
Q39 Q40 Q41 Q42 Q43 Q44 Q45 Q46 Q47 Q48 Q49 Q50 Q51
Q52 Q53 Q54 Q55 Q56 Q57 Q58 Q59 Q60 Q61 Q62 Q63 Q64
Q65 Q66 Q67 Q68 Q69 Q70 Q71 gen1 gen2 gen3 gen4 gen5
target1 target2 target3 target4 target5 age FTPT res
greek gpa race hisp year athlete gender marry DrinkTyp
DrinkHvy Probs PBS;
USEVARIABLES ARE Q66 Q67 Q68 athlete gender;
MODEL:
i s | Q66@0 Q67@1 Q68@2;
i s ON athlete gender;
OUTPUT: stand;
```

First row = LGM
Second row = prediction paths

180

Results

mplus3_latentgrowth.out

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
I |				
Q66	1.000	0.000	999.000	999.000
Q67	1.000	0.000	999.000	999.000
Q68	1.000	0.000	999.000	999.000
S |				
Q66	0.000	0.000	999.000	999.000
Q67	1.000	0.000	999.000	999.000
Q68	2.000	0.000	999.000	999.000
I ON				
ATHLETE	-0.246	0.840	-0.293	0.769
GENDER	-0.165	0.380	-0.433	0.665
S ON				
ATHLETE	-0.486	0.508	-0.956	0.339
GENDER	-0.410	0.230	-1.782	0.075
S WITH				
I	0.938	0.556	1.687	0.092
Intercepts				
Q66	0.000	0.000	999.000	999.000
Q67	0.000	0.000	999.000	999.000
Q68	0.000	0.000	999.000	999.000
I	3.315	0.206	16.061	0.000
S	0.835	0.125	6.683	0.000
Residual Variances				
Q66	6.003	1.040	5.771	0.000
Q67	5.537	0.636	8.711	0.000
Q68	1.615	1.287	1.256	0.209
I	5.876	1.034	5.680	0.000
S	[illegible]	[illegible]	[illegible]	[illegible]

Loadings

Influence of gender and athletic status on baseline levels (i) and linear growth over time (s)

Baseline levels for female non-athletes (i) and their growth over time (s)

181

Thank You!!

World's Most Accurate Pie Chart

182