

9555A Structural Equation Modeling Introduction and Overview

Introductions -- and why am I here teaching this course?

Course Outline

- Content

- Lecturing style

- Readings

- Evaluation

- Expectations

- What you will get out of this course

Projects

- Overview of SEM

- Mplus and other software

Where Does SEM Fit Within Mathematical Modeling?

- Axiomatic Models
- Algebraic Models
- Computational Models
 - Algorithmic
 - Connectionist
 - Bayesian



The algebraic model – like the linear regression model

$$y = ax + b$$

This model describes how input stimuli and model parameters produce output

Logic of SEM

- A model reveals something (limited) but meaningful about reality
- e.g., The mean is a model that describes the reality of a collection of numbers
- The linear regression equation is a model of the relation between two variables
- SEM “translates” the variance-covariance data matrix into a more parsimonious representation of the associations between the variables
 - The variance-covariance matrix

	Covariances				
	STRESS1B	STRESS2B	STRESS3B	STRESS4B	HARM1B
STRESS1B	2.949				
STRESS2B	2.241	2.860			
STRESS3B	2.304	2.434	3.099		
STRESS4B	2.202	2.191	2.311	2.900	
HARM1B	0.624	0.577	0.625	0.508	2.542
HARM2B	0.566	0.645	0.631	0.589	1.554
HARM3B	0.638	0.721	0.730	0.682	1.559
HARM4B	0.458	0.571	0.639	0.584	1.354
DEP1B	2.450	2.767	3.218	2.901	2.086
DEP2B	2.273	2.627	3.052	2.738	1.787
DEP3B	2.415	2.645	3.304	3.085	1.770
DEP4B	2.425	2.733	3.248	3.230	1.249

	Covariances				
	HARM2B	HARM3B	HARM4B	DEP1B	DEP2B
HARM2B	2.818				
HARM3B	1.799	3.100			
HARM4B	1.635	1.785	2.805		
DEP1B	2.244	1.816	2.293	36.575	
DEP2B	2.279	1.588	2.257	32.445	33.133
DEP3B	2.228	2.136	2.399	33.827	33.264
DEP4B	1.885	1.419	2.196	32.774	31.523

	Covariances	
	DEP3B	DEP4B
DEP3B	38.344	
DEP4B	34.109	35.753

Logic of SEM

- Note $\text{COV}_{xy} = r_{xy} sd_x sd_y$

	Covariances				
	STRESS1B	STRESS2B	STRESS3B	STRESS4B	HARM1B
STRESS1B	2.949				
STRESS2B	2.241	2.860			
STRESS3B	2.304	2.434	3.099		
STRESS4B	2.202	2.191	2.311	2.900	
HARM1B	0.624	0.577	0.625	0.508	2.542
HARM2B	0.566	0.645	0.631	0.589	1.554
HARM3B	0.638	0.721	0.730	0.682	1.559
HARM4B	0.458	0.571	0.639	0.584	1.354
DEP1B	2.450	2.767	3.218	2.901	2.086
DEP2B	2.273	2.627	3.052	2.738	1.787
DEP3B	2.415	2.645	3.304	3.085	1.770
DEP4B	2.425	2.733	3.248	3.230	1.249

	Correlations				
	STRESS1B	STRESS2B	STRESS3B	STRESS4B	HARM1B
STRESS1B	1.000				
STRESS2B	0.772	1.000			
STRESS3B	0.762	0.818	1.000		
STRESS4B	0.753	0.761	0.771	1.000	
HARM1B	0.228	0.214	0.223	0.187	1.000
HARM2B	0.196	0.227	0.214	0.206	0.581
HARM3B	0.211	0.242	0.235	0.227	0.555
HARM4B	0.159	0.202	0.217	0.205	0.507
DEP1B	0.236	0.271	0.302	0.282	0.216
DEP2B	0.230	0.270	0.301	0.279	0.195
DEP3B	0.227	0.253	0.303	0.293	0.179
DEP4B	0.236	0.270	0.309	0.317	0.131

Logic of SEM

- Can we say something meaningful about all the elements in the matrix

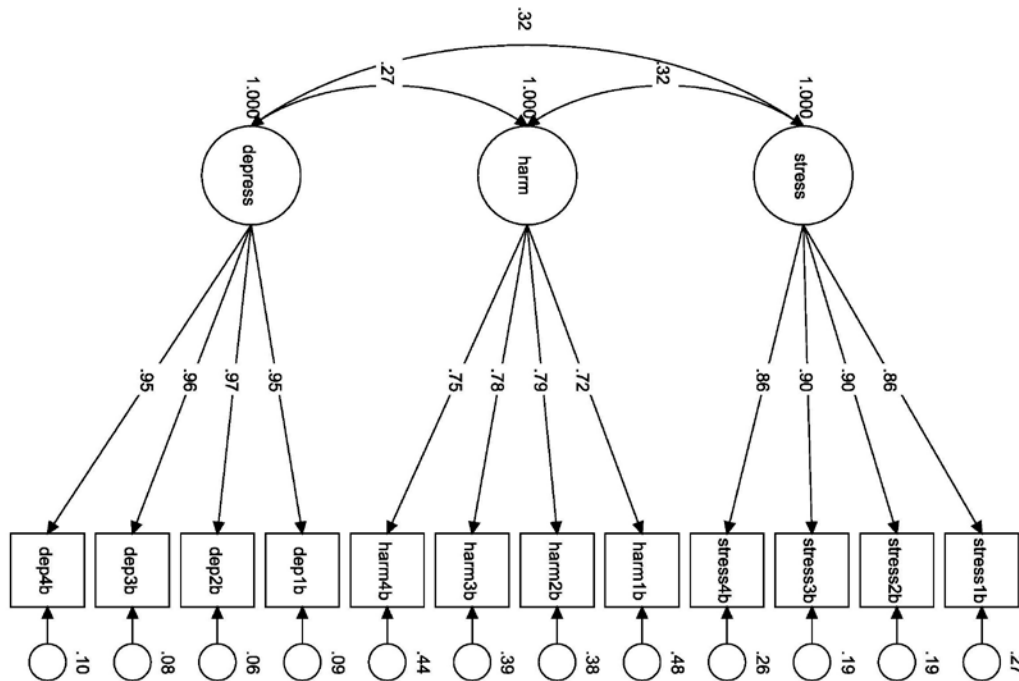
- There are a lot of elements:

$$\frac{v(v+1)}{2}$$

- With 12 variables, the variance-covariance matrix consists of 78 elements:
 - 12 variances + 66 covariances
 - We start with 78 degrees of freedom (dfs) (think of the analogy of dollars)
 - We could create a replica of this matrix but it would cost us 78 dfs
 - Let create a simplified but close replica of the matrix with fewer than 78 parts in the model (i.e., model parameters)
 - Each part (model parameter) costs 1df
 - The more degrees of freedom left over the more parsimonious the model
 - BUT there is a balance between parsimony and goodness of model fit to the original data matrix

Logic of SEM

- Let's use the following model and see how closely we can reproduce the original variance-covariance matrix of 78 elements
- There are 27 parts (parameters) in this model (we will learn how to count them)
- $78 - 27 = 51$ dfs



Logic of SEM

- How good is this model?
- The answer lies in how well we have reproduced the original variance-covariance matrix
- In other words: goodness of fit
- The RAM model in SEM is a visual representation of the algebraic equations used to reproduce the original variance-covariance matrix
- Instead of RAM, SEM can also be represented in a set of
 - Matrices (as in LISREL, the original formulation)
 - Regression equations (EQS)

Logic of SEM

- Here is our reproduced variance-covariance matrix:

	Model Estimated Covariances/Correlations/Residual Correlations				
	STRESS1B	STRESS2B	STRESS3B	STRESS4B	HARM1B
STRESS1B	2.949				
STRESS2B	2.236	2.859			
STRESS3B	2.335	2.413	3.100		
STRESS4B	2.150	2.222	2.320	2.900	
HARM1B	0.539	0.557	0.582	0.536	2.561
HARM2B	0.623	0.644	0.672	0.619	1.529
HARM3B	0.646	0.667	0.697	0.642	1.585
HARM4B	0.588	0.608	0.635	0.584	1.444
DEP1B	2.750	2.842	2.967	2.732	1.792
DEP2B	2.670	2.760	2.881	2.653	1.740
DEP3B	2.835	2.930	3.058	2.817	1.847
DEP4B	2.711	2.801	2.924	2.693	1.766

	Model Estimated Covariances/Correlations/Residual Correlations				
	HARM2B	HARM3B	HARM4B	DEP1B	DEP2B
HARM2B	2.831				
HARM3B	1.831	3.100			
HARM4B	1.667	1.729	2.812		
DEP1B	2.069	2.146	1.954	36.575	
DEP2B	2.009	2.084	1.898	32.234	33.134
DEP3B	2.133	2.212	2.015	34.221	33.231
DEP4B	2.040	2.115	1.926	32.720	31.774

	Model Estimated Covariances/Correlations/Residual Correlations	
	DEP3B	DEP4B
DEP3B	38.327	
DEP4B	33.732	35.752

Logic of SEM

- To see goodness of fit, we can look at deviations from the original variance-covariance matrix (also called the observed matrix) to the reproduced matrix (also called the model matrix)
- Original – reproduced = residual
- Here is the residual matrix:

	Residuals for Covariances/Correlations/Residual Correlations				
	STRESS1B	STRESS2B	STRESS3B	STRESS4B	HARM1B
STRESS1B	0.000				
STRESS2B	0.004	0.000			
STRESS3B	-0.031	0.021	0.000		
STRESS4B	0.051	-0.031	-0.009	0.000	
HARM1B	0.085	0.020	0.043	-0.028	-0.020
HARM2B	-0.057	0.001	-0.041	-0.029	0.025
HARM3B	-0.008	0.053	0.033	0.040	-0.026
HARM4B	-0.130	-0.037	0.004	-0.001	-0.090
DEP1B	-0.300	-0.075	0.251	0.169	0.295
DEP2B	-0.398	-0.133	0.171	0.084	0.047
DEP3B	-0.420	-0.285	0.245	0.268	-0.077
DEP4B	-0.285	-0.068	0.323	0.536	-0.517

	Residuals for Covariances/Correlations/Residual Correlations				
	HARM2B	HARM3B	HARM4B	DEP1B	DEP2B
HARM2B	-0.013				
HARM3B	-0.031	0.001			
HARM4B	-0.032	0.056	-0.007		
DEP1B	0.175	-0.329	0.339	0.000	
DEP2B	0.269	-0.496	0.359	0.211	-0.001
DEP3B	0.094	-0.076	0.384	-0.393	0.033
DEP4B	-0.155	-0.696	0.269	0.055	-0.250

	Residuals for Covariances/Correlations/Residual Correlations	
	DEP3B	DEP4B
DEP3B	0.017	
DEP4B	0.377	0.001

Logic of SEM

- Various indices of goodness of fit
- Some of these take into account parsimony
- Tests of significance of the parameters
- A powerful feature of SEM is the ability to compare models
- We should take this feature more seriously

Features of SEM

- SEM as combination of factor analysis and multiple linear regression
- Measurement model and structural model
- Confirmatory factor analysis vs. SEM
- Path analysis vs. SEM
- The Measurement Model
 - The common factor model (may have seen this in EFA)
 - Latent variables
 - Observed variables (e.g., the indicator variables)
 - The residuals (also called errors, uniqueness)
 - Difference between CFA and EFA
- The Structural Model
 - Exogenous and endogenous variables
 - Residuals
 - Different types of parameters

SEM and Your Future as a Researcher

- Focus on concepts and latent variables
- Construct validity and the nomological network
- Focus on processes
- How to develop a model or a theory
 - *Theory construction and model building skills* (Jaccard & Jacoby, 2010)
 - My application of their 26 heuristics to Intrinsic Motivation:
<http://publish.uwo.ca/~ptrembla/heuristics-revised.pdf>
- a few examples:
 - analyzing your own experiences
 - case studies
 - participant observation
 - Your grandmother isn't always right

Mplus Website

FRIDAY
JULY 05, 2013

MPLUS
Mplus at a Glance
General Description
Mplus Programs
Pricing
Version History
System Requirements
Platforms
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
Complex Survey Data
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

Mplus Website Updates

HOME ORDER CONTACT US LOGIN MPLUS DISCUSSION

[Last updated:](#) July 03, 2013



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.
- New FAQ: [Growth mixture model confidence intervals](#) for estimated trajectory means.
- New [paper](#): Asparouhov & Muthén (2013). Multiple group factor analysis alignment. Web note 18: Version 2.
- Revised [paper](#): Asparouhov & Muthén (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. Web note 15: Version 7.
- Handouts and video for Mplus-related talks at the May 2013 UConn M3 conference 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).
- New [talk](#): Muthén (2013). Overview of recent Mplus developments. Conference call presentation to the Prevention Science Methodology Group, January 29.
- New [paper](#): Muthén & Asparouhov (2013). Item response modeling in Mplus: A multi-dimensional, multi-level, and multi-timpoint example.

Mplus [Papers](#)
Using Special Mplus Features

Mplus Demo Version

The Mplus Demo version is available for download at no cost. Click [here](#) to download the demo. The demo version contains all of the capabilities of the regular version of Mplus and is only limited by the number of observed variables that can be used in an analysis.

Student Pricing for Mplus Version 7.11

Special student pricing is available for Mplus. The student version of the program is identical to the regular version. Click [here](#) for more information.

Mplus Version 7 User's Guide and Examples

Click [here](#) for the Mplus Version 7 User's Guide and to download the input, output, and data for the Mplus User's Guide examples.

Mplus Web Training and Handouts

Videos and handouts for the 9 topics of the [Mplus Short Courses](#) are now available for viewing on the web. Other [Mplus web training](#) includes web talks, a seminar series, a one-day overview course, a two-day course, and a 20-lecture course on Mplus analyses.

Papers Using Special Mplus Features

Click [here](#) to find papers ordered by topic and by date.

Mplus Web Notes

Mplus Web Notes is a vehicle for presenting information on Mplus analysis techniques. Click [here](#) to go to Mplus Web Notes.

Mplus Complex Survey Data Project

Mplus Basic Syntax

Ten basic commands (with sub-commands)

Most common:

TITLE:

DATA:

VARIABLE:

MODEL:

Others:

DEFINE:

ANALYSIS:

OUTPUT:

SAVEDATA:

PLOT:

MONTECARLO:

Free Format Data File Specification in Mplus

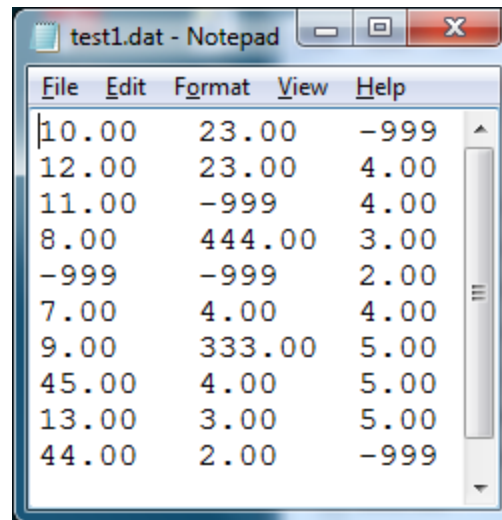
DATA:

file is test1.dat;

VARIABLE:

names are v1 v2 v3;

missing all (-999);



The screenshot shows a Notepad window with the title 'test1.dat - Notepad'. The window contains a data file with 11 rows and 3 columns of values. The values are as follows:

v1	v2	v3
10.00	23.00	-999
12.00	23.00	4.00
11.00	-999	4.00
8.00	444.00	3.00
-999	-999	2.00
7.00	4.00	4.00
9.00	333.00	5.00
45.00	4.00	5.00
13.00	3.00	5.00
44.00	2.00	-999

Fixed Format Data File Specification in Mplus

data: File is vignette.dat;

Format is 21F8.2;

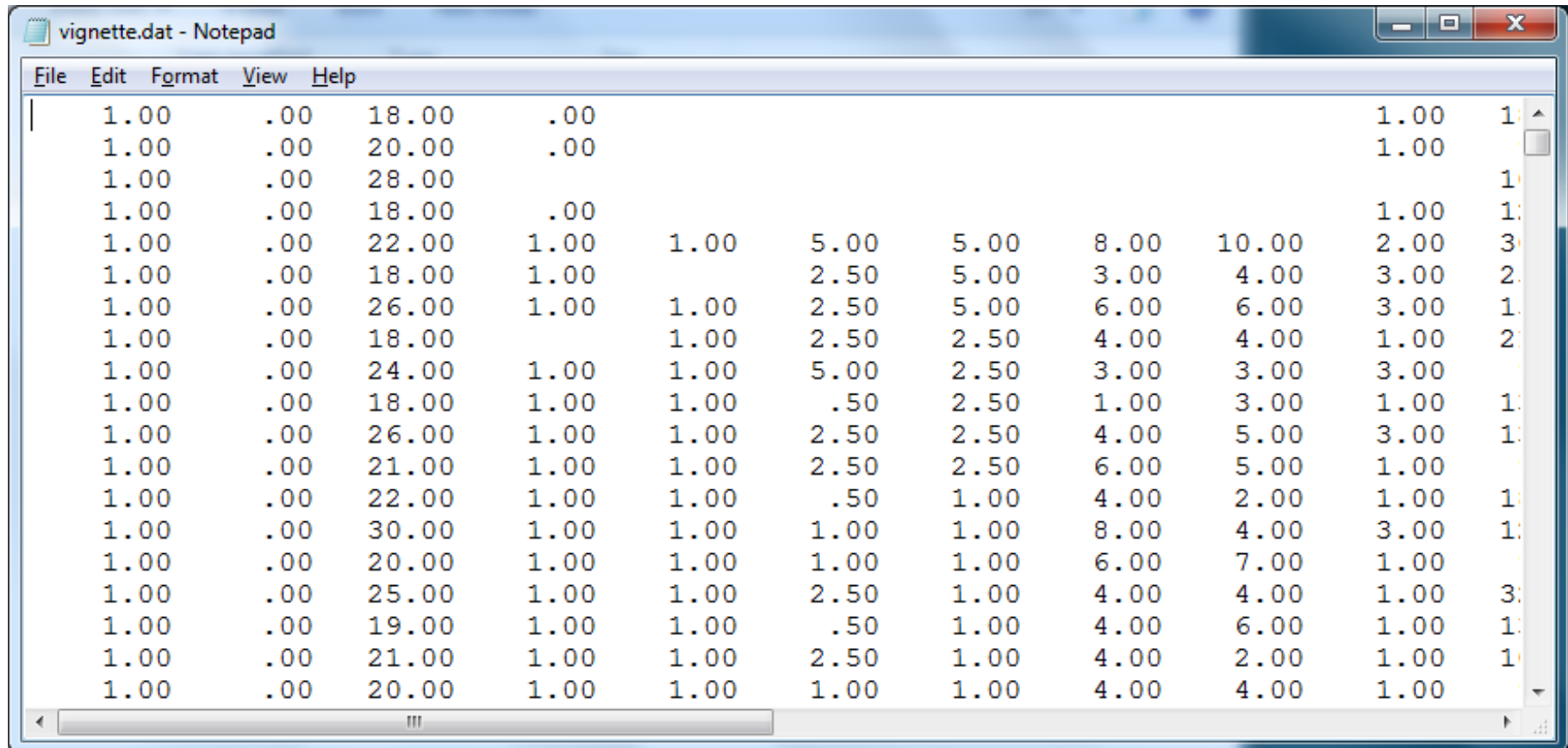
variable: names are clust gender age

everdr intox daysl2 daysse avgdr12 avgdrse smoke

bp_PA bp_V bp_A bp_H bp_tot AE_TR AE_LC AE_CB AE_RA

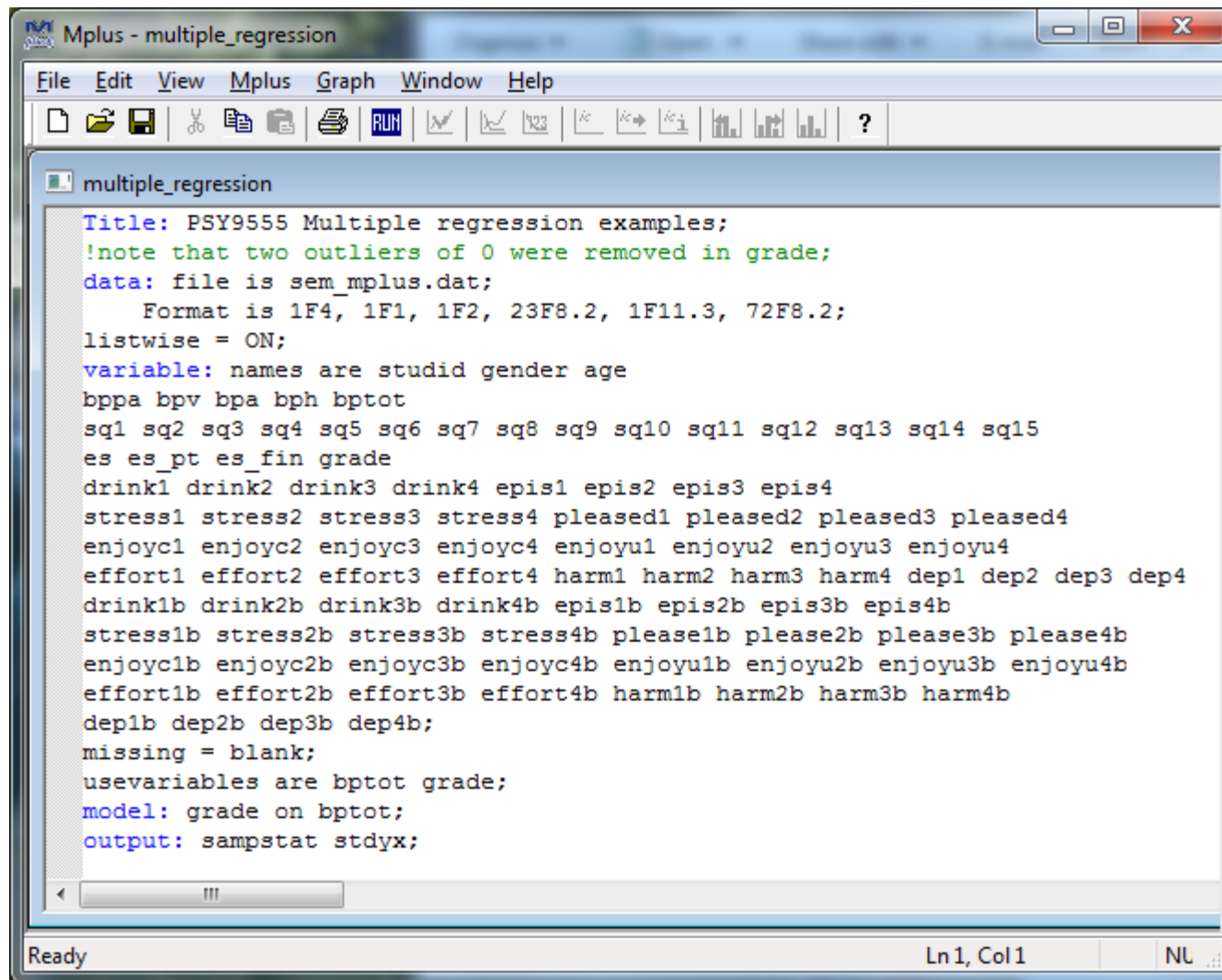
uni_num program2;

missing = blank;



1.00	.00	18.00	.00							1.00	1
1.00	.00	20.00	.00							1.00	
1.00	.00	28.00									1
1.00	.00	18.00	.00							1.00	1
1.00	.00	22.00	1.00	1.00	5.00	5.00	8.00	10.00	2.00	3.00	3
1.00	.00	18.00	1.00		2.50	5.00	3.00	4.00	3.00	2.00	2
1.00	.00	26.00	1.00	1.00	2.50	5.00	6.00	6.00	3.00	1.00	1
1.00	.00	18.00		1.00	2.50	2.50	4.00	4.00	1.00	2.00	2
1.00	.00	24.00	1.00	1.00	5.00	2.50	3.00	3.00	3.00		
1.00	.00	18.00	1.00	1.00	.50	2.50	1.00	3.00	1.00	1.00	1
1.00	.00	26.00	1.00	1.00	2.50	2.50	4.00	5.00	3.00	1.00	1
1.00	.00	21.00	1.00	1.00	2.50	2.50	6.00	5.00	1.00		
1.00	.00	22.00	1.00	1.00	.50	1.00	4.00	2.00	1.00	1.00	1
1.00	.00	30.00	1.00	1.00	1.00	1.00	8.00	4.00	3.00	1.00	1
1.00	.00	20.00	1.00	1.00	1.00	1.00	6.00	7.00	1.00		
1.00	.00	25.00	1.00	1.00	2.50	1.00	4.00	4.00	1.00	3.00	3
1.00	.00	19.00	1.00	1.00	.50	1.00	4.00	6.00	1.00	1.00	1
1.00	.00	21.00	1.00	1.00	2.50	1.00	4.00	2.00	1.00	1.00	1
1.00	.00	20.00	1.00	1.00	1.00	1.00	4.00	4.00	1.00		

Mplus: Regression – One Predictor



The screenshot shows the Mplus software window titled "Mplus - multiple_regression". The window has a menu bar with "File", "Edit", "View", "Mplus", "Graph", "Window", and "Help". Below the menu bar is a toolbar with various icons for file operations, editing, and running the model. The main text area contains the following Mplus input code:

```
multiple_regression

Title: PSY9555 Multiple regression examples;
!note that two outliers of 0 were removed in grade;
data: file is sem_mplus.dat;
      Format is 1F4, 1F1, 1F2, 23F8.2, 1F11.3, 72F8.2;
listwise = ON;
variable: names are studid gender age
bppa bpv bpa bph bptot
sq1 sq2 sq3 sq4 sq5 sq6 sq7 sq8 sq9 sq10 sq11 sq12 sq13 sq14 sq15
es es_pt es_fin grade
drink1 drink2 drink3 drink4 epis1 epis2 epis3 epis4
stress1 stress2 stress3 stress4 pleased1 pleased2 pleased3 pleased4
enjoyc1 enjoyc2 enjoyc3 enjoyc4 enjoyu1 enjoyu2 enjoyu3 enjoyu4
effort1 effort2 effort3 effort4 harm1 harm2 harm3 harm4 dep1 dep2 dep3 dep4
drink1b drink2b drink3b drink4b epis1b epis2b epis3b epis4b
stress1b stress2b stress3b stress4b please1b please2b please3b please4b
enjoyc1b enjoyc2b enjoyc3b enjoyc4b enjoyu1b enjoyu2b enjoyu3b enjoyu4b
effort1b effort2b effort3b effort4b harm1b harm2b harm3b harm4b
dep1b dep2b dep3b dep4b;
missing = blank;
usevariables are bptot grade;
model: grade on bptot;
output: sampstat stdyx;
```

At the bottom of the window, the status bar shows "Ready", "Ln 1, Col 1", and "NL".

Mplus: Regression – One Predictor

SAMPLE STATISTICS

MODEL RESULTS

[Insert](#) | [Tables](#) | [Images](#) | [Links](#) | [Files](#) | [Reco](#)

SAMPLE STATISTICS			Estimate				Two-Tailed P-Value
			S.E.				Est./S.E.
Means			GRADE ON				
GRADE			BPTOT				
1			Intercepts				
74.698			GRADE				
59.829			Residual Variances				
Covariances			GRADE				
GRADE			BPTOT				
94.201			STANDARDIZED MODEL RESULTS				
-23.893			STDYX Standardization				
304.857			Estimate				
Correlations			S.E.				Two-Tailed
GRADE			Est./S.E.				P-Value
1.000			GRADE ON				
-0.141			BPTOT				
1.000			Intercepts				
0.007			GRADE				
0.000			Residual Variances				
0.015			GRADE				
66.535			0.980				

Mplus: Regression – One Predictor MLR (Maximum Likelihood Robust)

```
usevariables are bptot grade;
analysis:
estimator = mlr;
model: grade on bptot;
output: sampstat stdyx;
```

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON BPTOT	-0.078	0.032	-2.476	0.013
Intercepts GRADE	79.387	1.865	42.556	0.000
Residual Variances GRADE	92.330	9.062	10.189	0.000

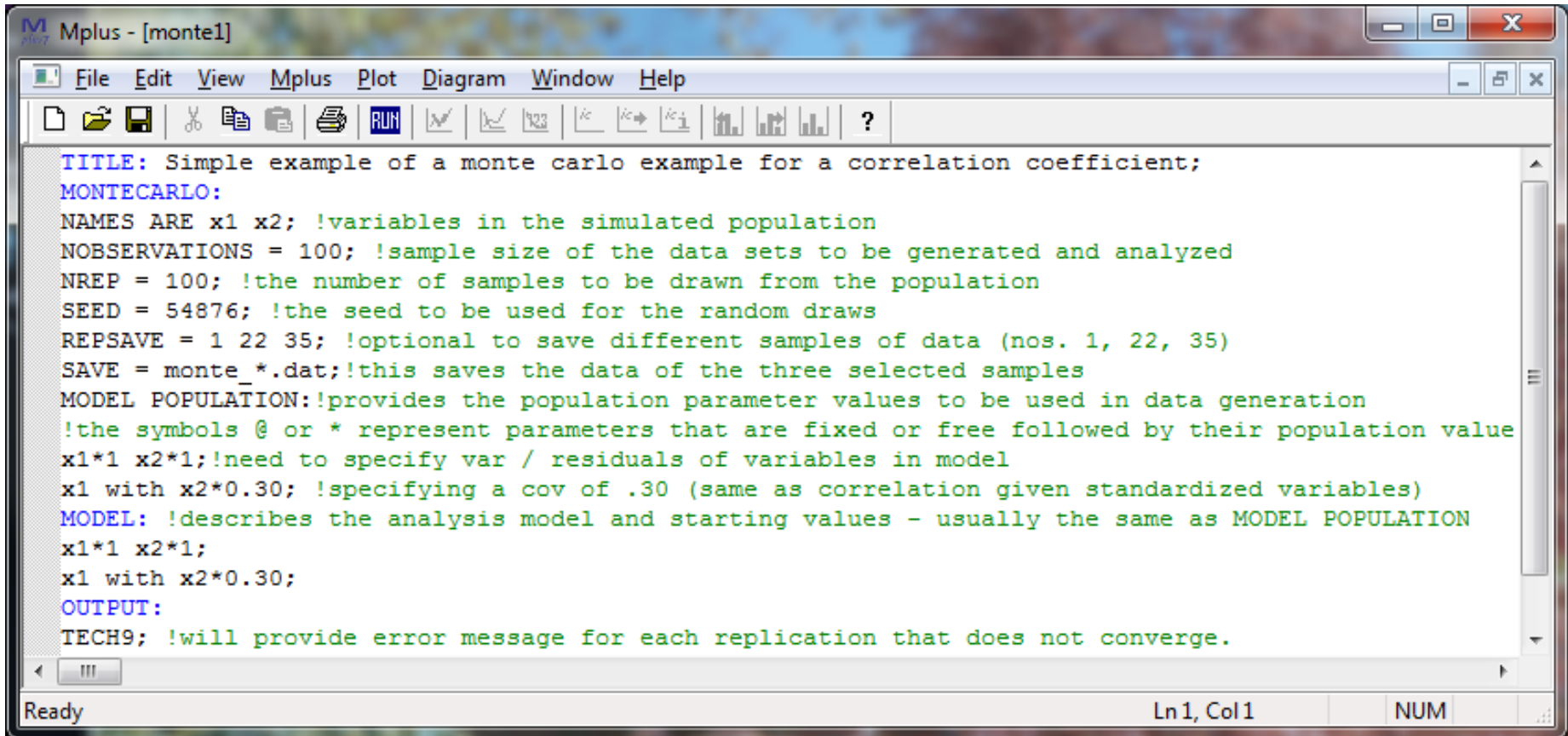
STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON BPTOT	-0.141	0.056	-2.522	0.012
Intercepts GRADE	8.179	0.421	19.444	0.000
Residual Variances GRADE	0.980	0.016	62.179	0.000

MONTECARLO in Mplus and Power

Example : A Correlation between two variables



The screenshot shows the Mplus software window titled "Mplus - [monte1]". The menu bar includes File, Edit, View, Mplus, Plot, Diagram, Window, and Help. The toolbar contains icons for file operations, a RUN button, and various plot and diagram tools. The main text area contains the following code:

```
TITLE: Simple example of a monte carlo example for a correlation coefficient;
MONTECARLO:
NAMES ARE x1 x2; !variables in the simulated population
NOBSERVATIONS = 100; !sample size of the data sets to be generated and analyzed
NREP = 100; !the number of samples to be drawn from the population
SEED = 54876; !the seed to be used for the random draws
REPSAVE = 1 22 35; !optional to save different samples of data (nos. 1, 22, 35)
SAVE = monte_*.dat; !this saves the data of the three selected samples
MODEL POPULATION: !provides the population parameter values to be used in data generation
!the symbols @ or * represent parameters that are fixed or free followed by their population value
x1*1 x2*1; !need to specify var / residuals of variables in model
x1 with x2*0.30; !specifying a cov of .30 (same as correlation given standardized variables)
MODEL: !describes the analysis model and starting values - usually the same as MODEL POPULATION
x1*1 x2*1;
x1 with x2*0.30;
OUTPUT:
TECH9; !will provide error message for each replication that does not converge.
```

The status bar at the bottom shows "Ready", "Ln1, Col1", and "NUM".

AMOS in SPSS

The screenshot shows the IBM SPSS Statistics Data Editor window for a file named 'Phase1_Final.sav'. The 'Analyze' menu is open, displaying a list of statistical analysis options. The data table in the background contains 19 rows and 6 columns of data.

Analyze Menu Options:

- Reports
- Descriptive Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Simulation...
- Quality Control
- ROC Curve...
- IBM SPSS Amos...

Data Table:

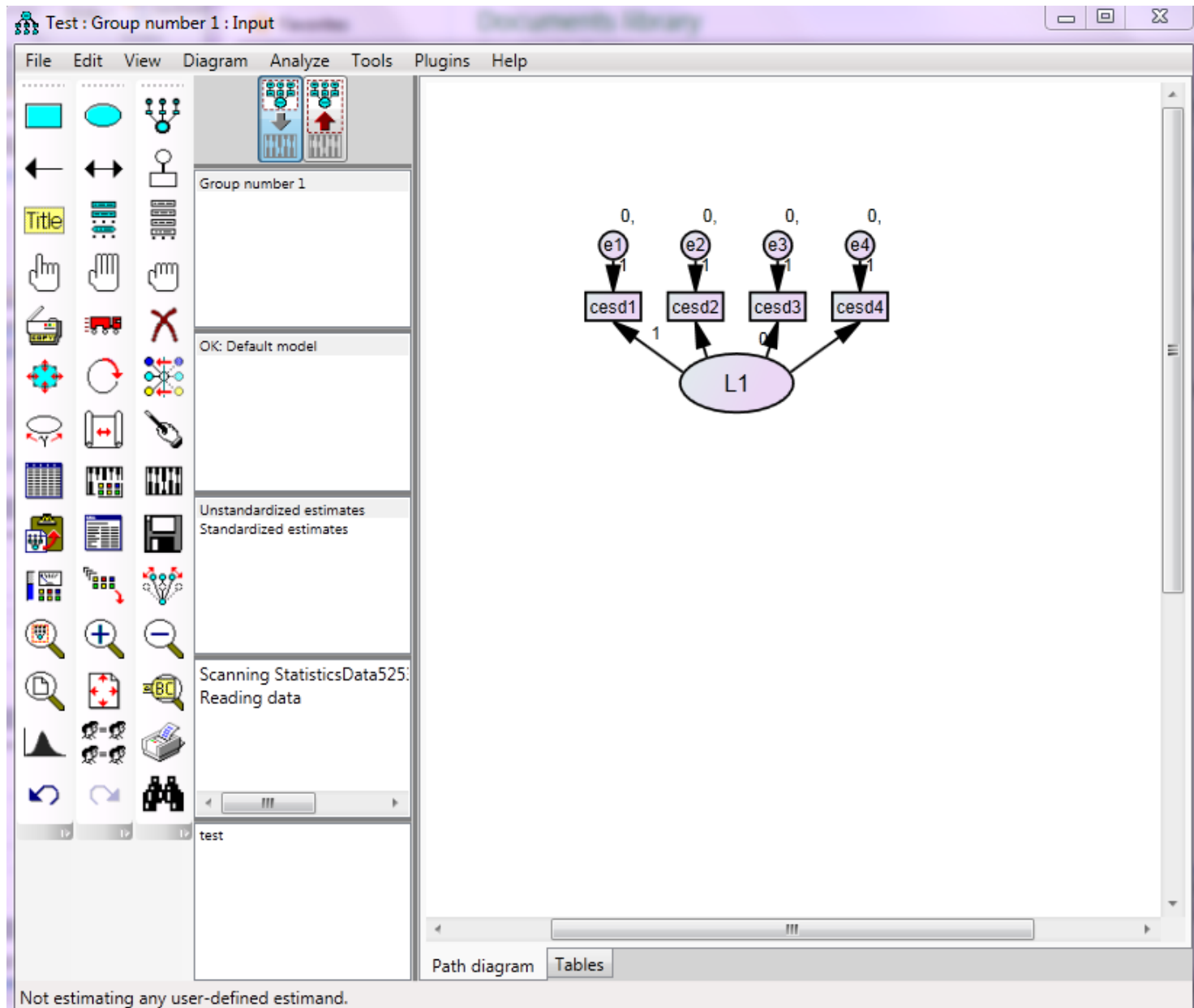
		cesd10	cesd14	cesd15	cesd16	cesd17	cesd18	cesd19
1	1	2						
2	1	2						
3	1	2						
4	1	.						
5	2	4						
6	1	1						
7	1	1						
8	1	2						
9	.	.						
10	4	4						
11	2	3						
12	1	2						
13	.	.						
14	1	2						
15	1	2						
16	1	1						
17	1	3						
18	4	3						
19	1	1						

Visible: 482 of 482 Variables

IBM SPSS Amos...

IBM SPSS Statistics Processor is ready

AMOS



lavaan (*la*tent *va*riable *an*alysis) in R

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. Journal of Statistical Software, 48, 2, <http://www.jstatsoft.org/v48/i02>

<http://lavaan.org/>

If you are new to R and want to start with the help of a menu driven environment (GUI; Graphic User Interface), you can download the Rcommander (John Fox). See <http://socserv.mcmaster.ca/jfox/Misc/Rcmdr/>