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 $\mathrm{COV}_{xy} = \mathrm{r}_{xy}\, sd_x sd_y$

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

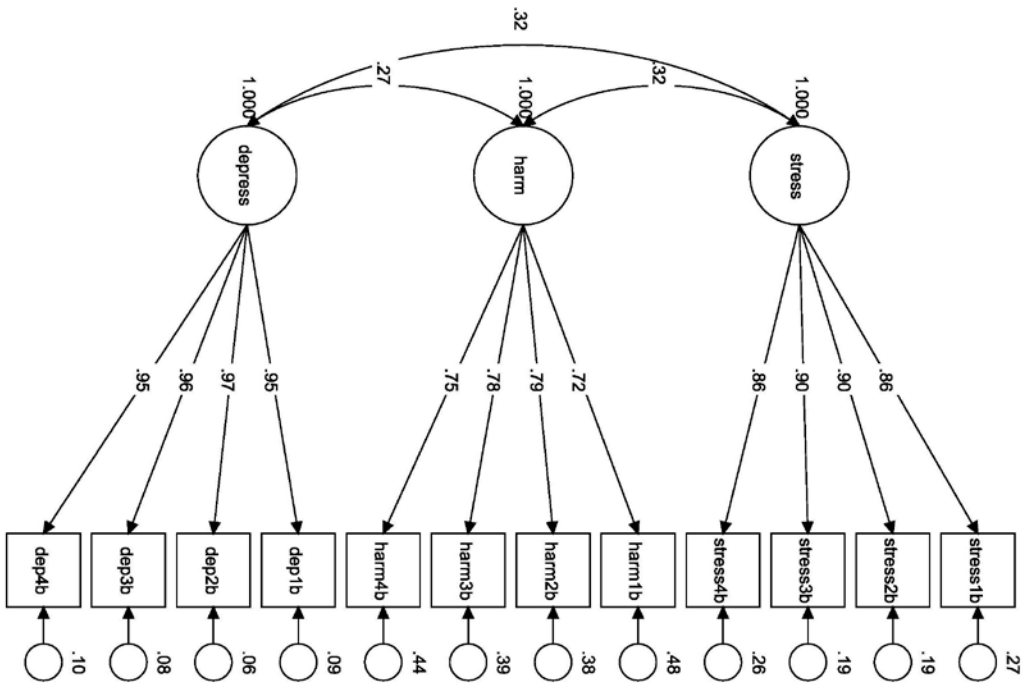
- Can we say something meaningful about all the elements in the matrix
  - There are a lot of elements:

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

  - 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

Mplus Basic Syntax

Ten basic commands (with sub-commands)
Most common:

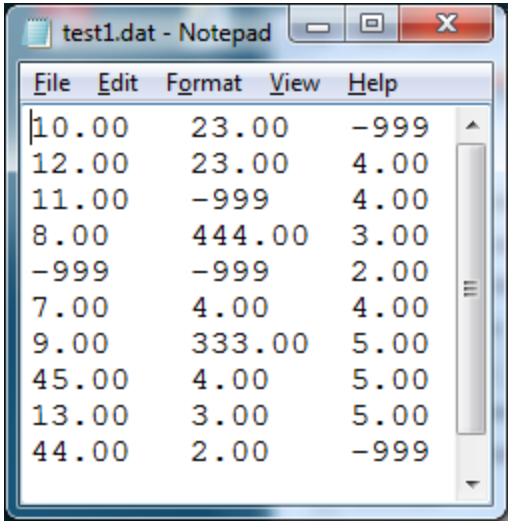
- TITLE:
- DATA:
- VARIABLE:
- MODEL:

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);
```



Fixed Format Data File Specification in Mplus

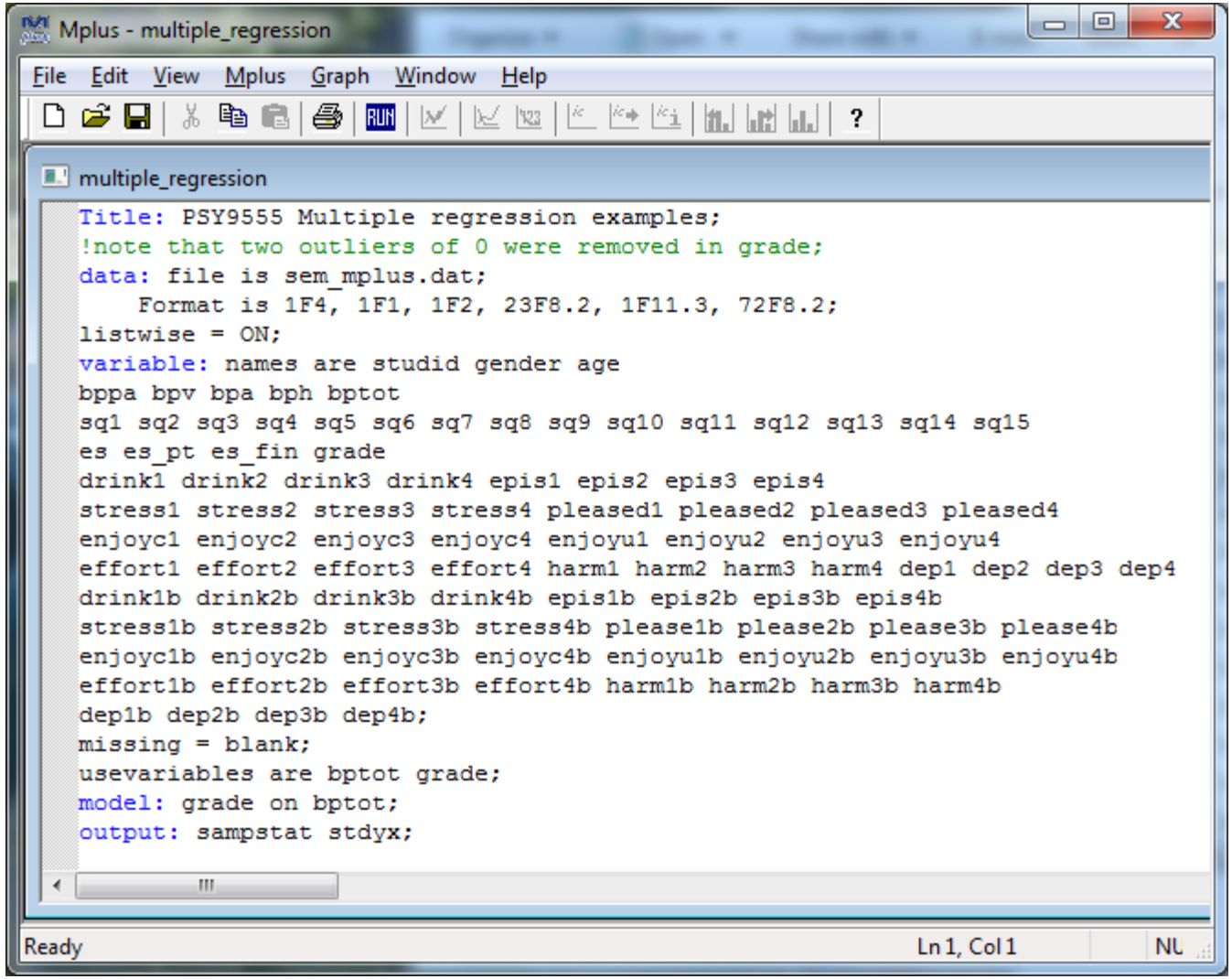
```
data: File is vignette.dat;
    Format is 21F8.2;
variable: names are clust gender age
everdr intox days12 daysse avgdr12 avgdrse smoke
bp_PA bp_V bp_A bp_H bp_tot AE_TR AE_LC AE_CB AE_RA
uni_num program2;
missing = blank;
```

vignette.dat - Notepad

File Edit Format View Help

```
    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
    1.00      .00    18.00     1.00              2.50     5.00     3.00     4.00    3.00    2
    1.00      .00    26.00     1.00     1.00     2.50     5.00     6.00     6.00    3.00    1
    1.00      .00    18.00               1.00     2.50     2.50     4.00     4.00    1.00    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
    1.00      .00    26.00     1.00     1.00     2.50     2.50     4.00     5.00    3.00    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
    1.00      .00    30.00     1.00     1.00     1.00     1.00     8.00     4.00    3.00    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
    1.00      .00    19.00     1.00     1.00      .50     1.00     4.00     6.00    1.00    1
    1.00      .00    21.00     1.00     1.00     2.50     1.00     4.00     2.00    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



Mplus: Regression – One Predictor

SAMPLE STATISTICS

SAMPLE STATISTICS

Means

	GRADE	BPTOT
1	74.698	59.829

Covariances

	GRADE	BPTOT
GRADE	94.201	
BPTOT	-23.893	304.857

Correlations

	GRADE	BPTOT
GRADE	1.000	
BPTOT	-0.141	1.000

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON				
BPTOT	-0.078	0.029	-2.672	0.008
Intercepts				
GRADE	79.387	1.828	43.425	0.000
Residual Variances				
GRADE	92.330	6.960	13.266	0.000

STANDARDIZED MODEL RESULTS

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON				
BPTOT	-0.141	0.052	-2.699	0.007
Intercepts				
GRADE	8.179	0.329	24.831	0.000
Residual Variances				
GRADE	0.980	0.015	66.535	0.000

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

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

MODEL RESULTS

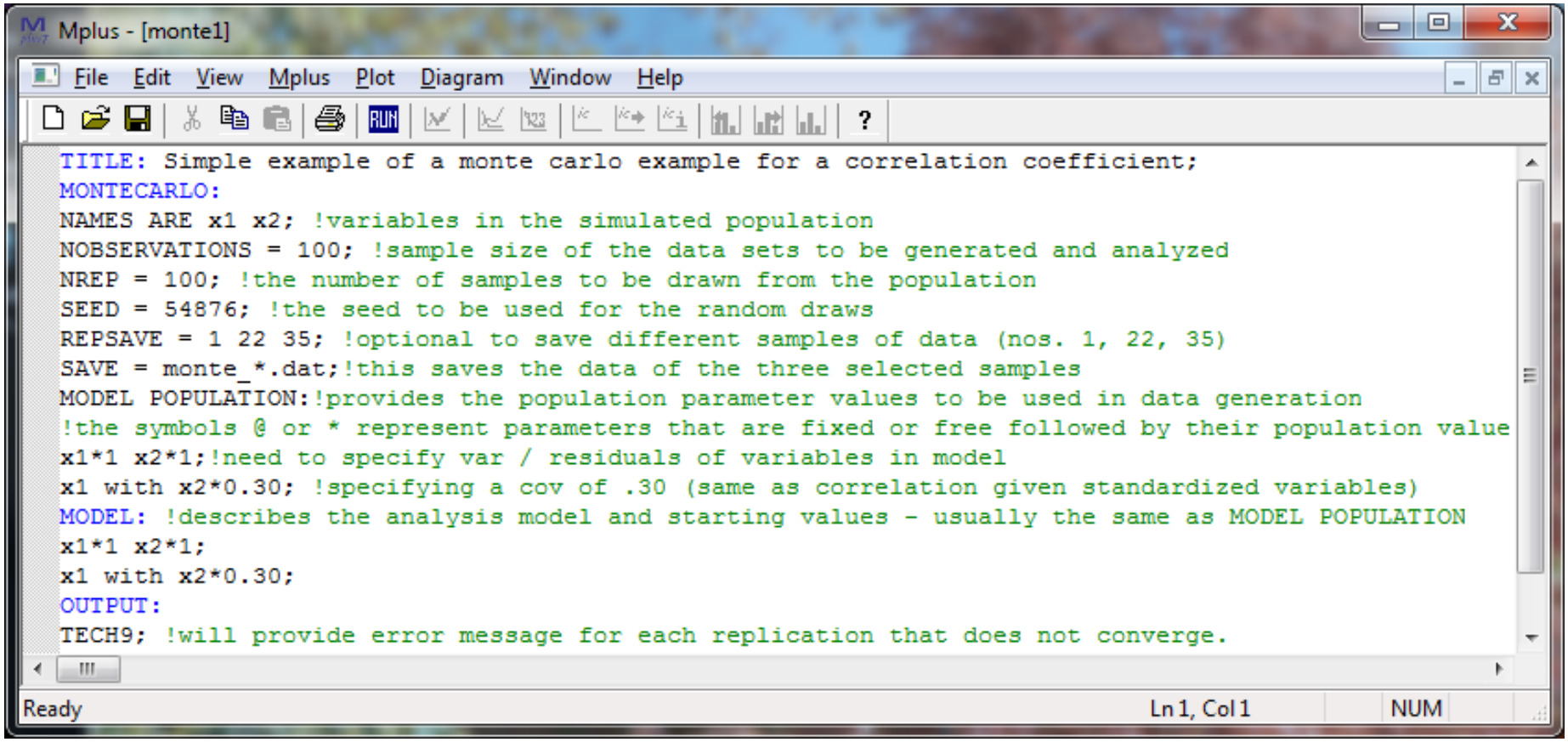
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON				
BPTOT	-0.078	0.032	-2.476	0.013
Intercepts				
GRADE	79.387	1.865	42.556	0.000
Residual Variances				
GRADE	92.330	9.062	10.189	0.000

STANDARDIZED MODEL RESULTS

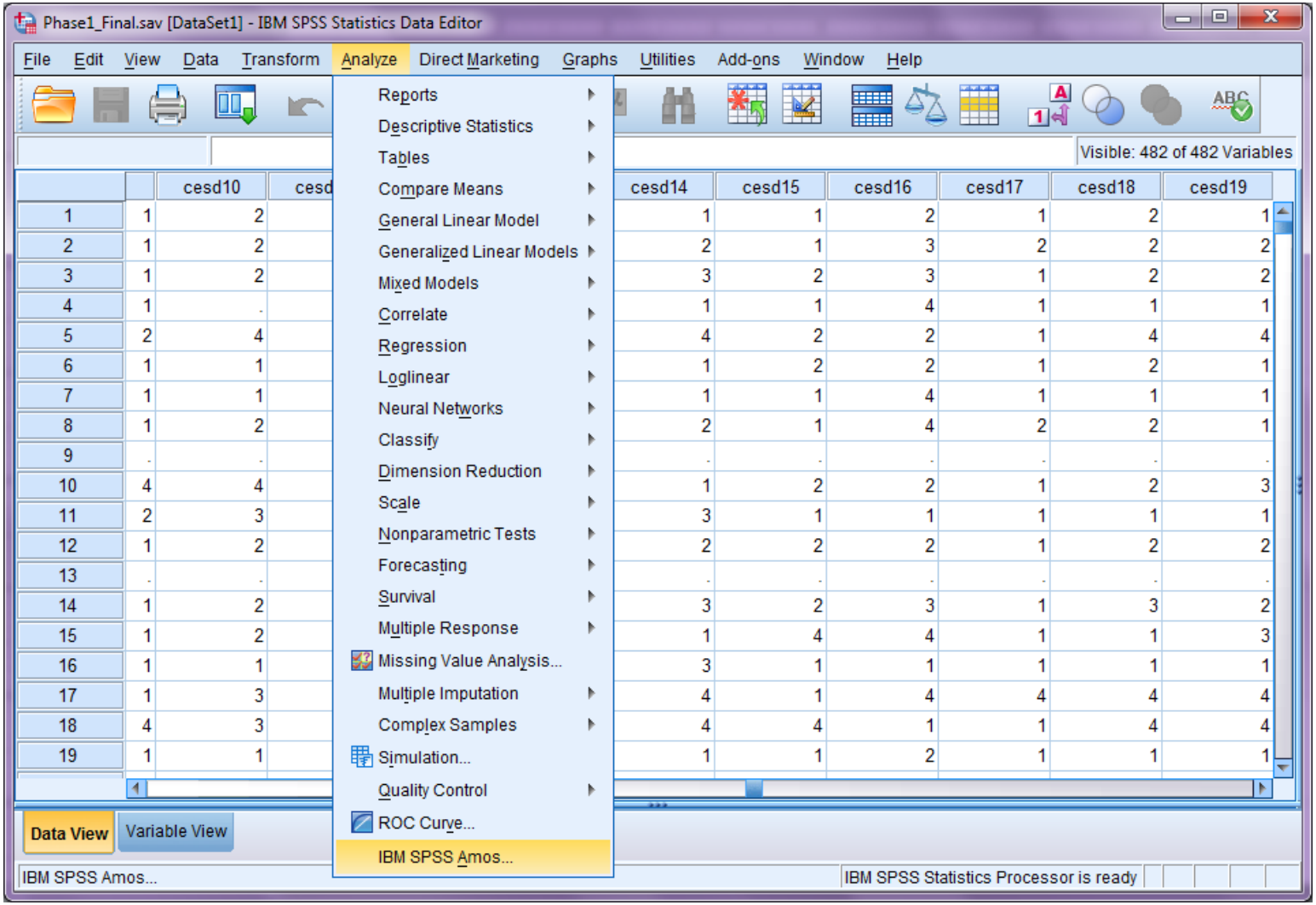
STDYX Standardization

	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

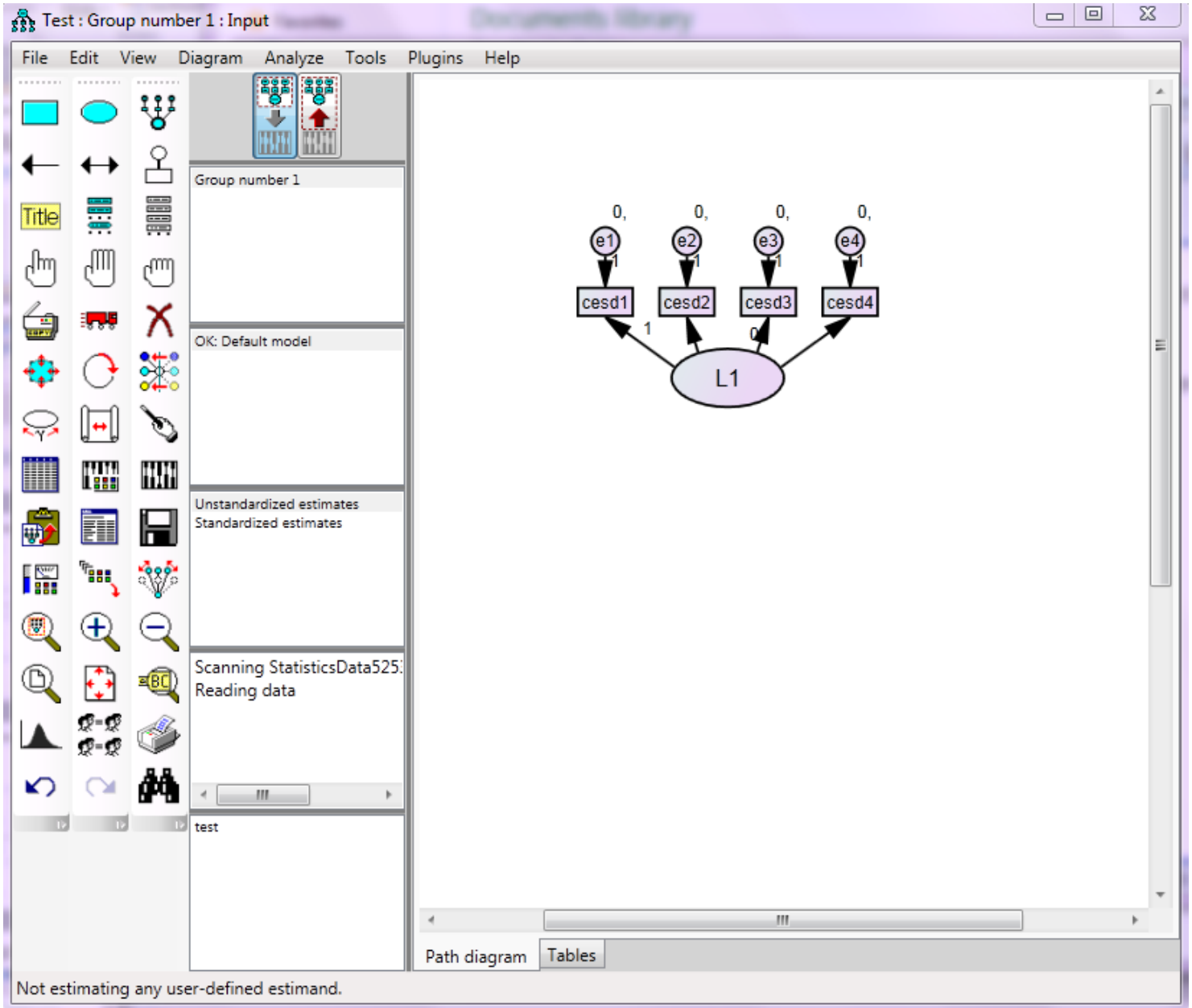
MONTECARLO in Mplus and Power
Example : A Correlation between two variables



AMOS in SPSS


Phase1_Final.sav [DataSet1] - IBM SPSS Statistics Data Editor
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
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...
Visible: 482 of 482 Variables
cesd10 cesd14 cesd15 cesd16 cesd17 cesd18 cesd19
Data View
Variable View
IBM SPSS Amos...
IBM SPSS Statistics Processor is ready

AMOS


Test : Group number 1 : Input
File Edit View Diagram Analyze Tools Plugins Help
Group number 1
OK: Default model
Unstandardized estimates
Standardized estimates
Scanning StatisticsData525
Reading data
test
0, 0, 0, 0,
e1 e2 e3 e4
1 1 1 1
cesd1 cesd2 cesd3 cesd4
1 0
L1
Path diagram Tables
Not estimating any user-defined estimand.

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/