

PSY 9555A – SEM: Sep 18 – Building Blocks (Multiple Regression & Factor Analysis)

Fundamental Concepts

$$cov_{xy} = r_{xy}SD_xSD_y$$

$$r_{TxTy} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}$$

$$\beta_1 = \frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} \quad \beta_2 = \frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2}$$

$$R_{y_{12}}^2 = \beta_1^2 + \beta_2^2 + 2r_{12}\beta_1\beta_2$$

$$R_{y_{12}}^2 = \beta_1r_{y1} + \beta_2r_{y2}$$

Fundamental Concepts

- Assumptions in multiple regression
 - see <http://publish.uwo.ca/~ptrembla/resources/Assumptions-Multiple-Regression.pdf>
- Suppression
- Partial and part (semi-partial correlations)
- Spuriousness and specification error
- Different types of correlation coefficients
- Censored variables

Fundamental Concepts

- Statistical tests
- Critical ratio
- Power, effect size, sample size
- Standard error of measurement estimates in SEM
- Bootstrapping

Data preparation (Chapter 3 – Kline)

- Raw data vs. summary (correlations/covariance matrix)
- Positive definiteness
- Missing data (ask yourself why data are missing)
 - Alternative procedures (see next slide)
- Normality and dealing with non-normality

Missing Data

- use multiple imputation or maximum likelihood
- not perfect but much better than conventional procedures
- Mplus
- See also http://www.ats.ucla.edu/stat/mplus/faq/fiml_counts.htm

Baraldi, A. N. & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5-37.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576.

Missing Data: Mplus

In the second use, the AUXILIARY option is used in conjunction with TYPE=GENERAL with continuous dependent variables and maximum likelihood estimation to identify a set of variables that will be used as missing data correlates in addition to the analysis variables (Collins, Schafer, & Kam, 2001; Graham, 2003; Asparouhov & Muthén, 2008b; Enders, 2010). This use is not available with MODINDICES, BOOTSTRAP, and models with a set of exploratory factor analysis (EFA) factors in the MODEL command. The setting M in parentheses is placed behind the variables on the AUXILIARY statement that will be used as missing data correlates. Following is an example of how to specify the M setting:

AUXILIARY = z1-z4 (M);

where z1, z2, z3, and z4 are variables that will be used as missing data correlates in addition to the analysis variables.

From Mplus version 7 p. 552.

Dichotomous (Binary), Categorical-Ordinal, and Count Outcomes

Continuous

- continuous ordinal
- interval
- ratio

Discrete

- dichotomous, binary
- nominal (can have more than two unordered categories)
- ordinal

In Mplus after “variable: names are...” include
categorical are

- For dichotomous variables
 - 1 and 2 will be automatically recoded as 0 and 1
- For ordinal variables
 - no more than 10 categories
 - integer values only
 - categories automatically recoded as 0, 1, 2, ...

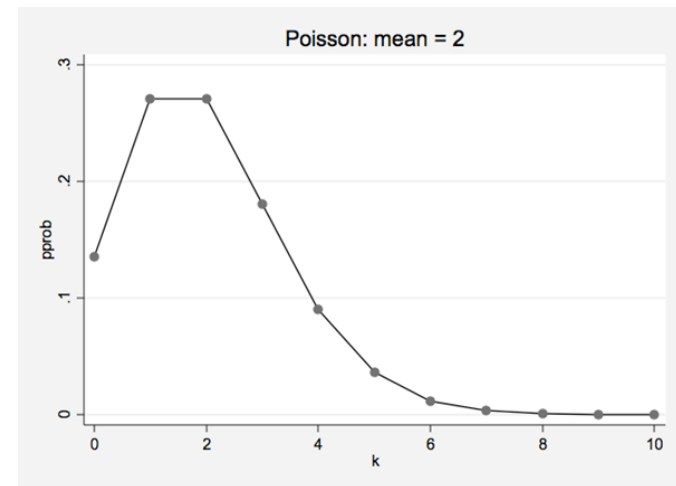
nominal are

- categories automatically recoded as 0, 1, 2, ...
- last category is reference

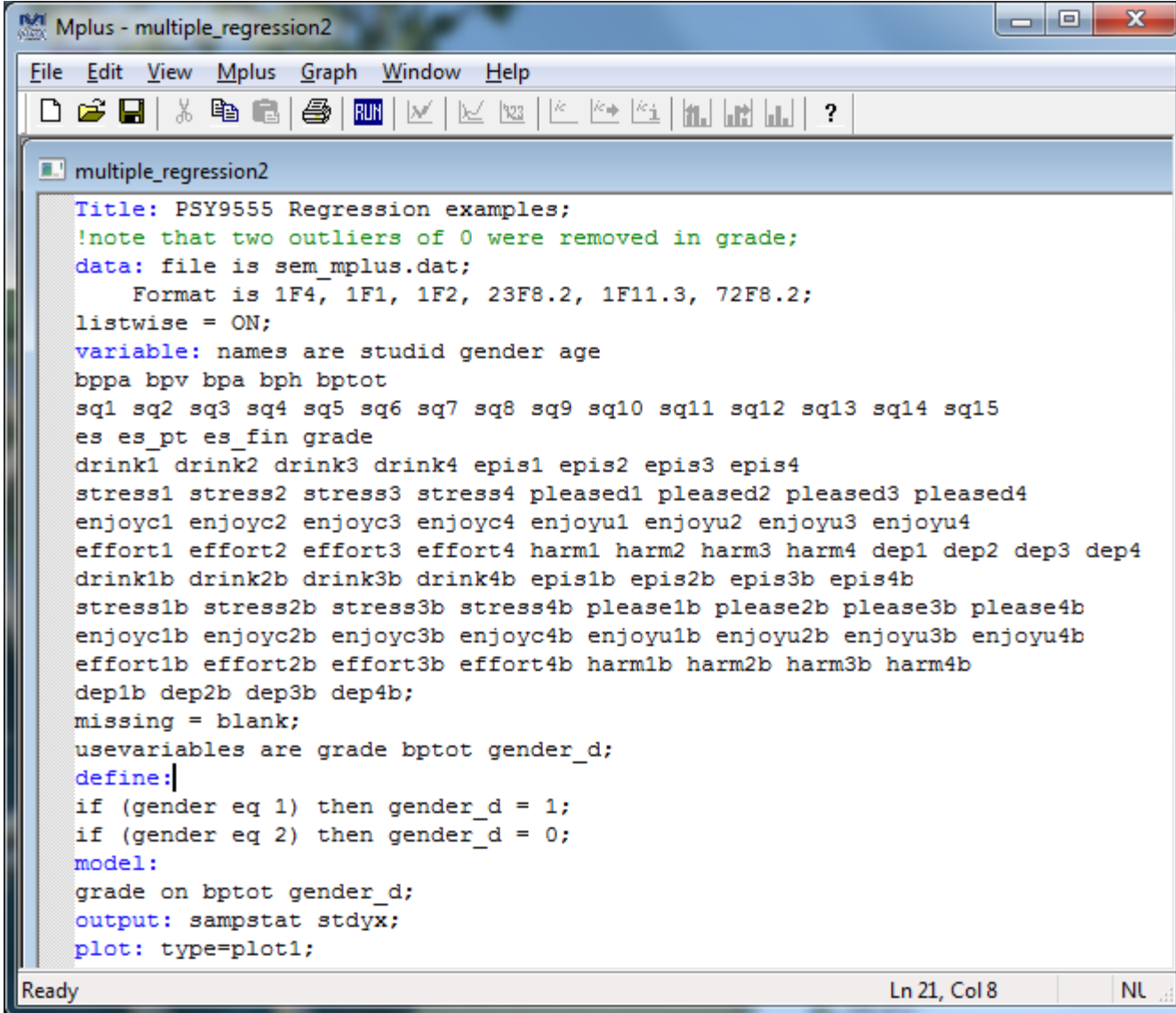
Dichotomous (Binary), Categorical-Ordinal, and Count Outcomes

Count Variables

- e.g., number of accidents at a particular highway
- Number of aggressive acts in the previous month
- Poisson distribution
 - To model count data that varies randomly over time
 - Often used when probability is small
 - Discrete values (positive integers)



Mplus: Regression – Two Predictors



The screenshot shows the Mplus software window titled "multiple_regression2". The interface includes a menu bar (File, Edit, View, Mplus, Graph, Window, Help) and a toolbar with icons for file operations, running, and graphing. The main text area contains the following Mplus input code:

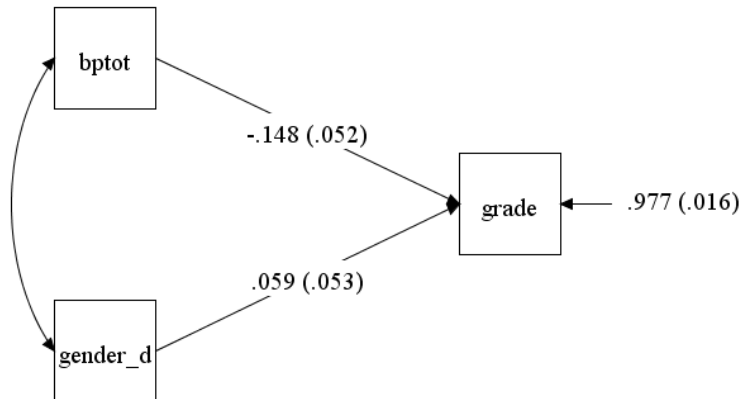
```
multiple_regression2

Title: PSY9555 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 grade bptot gender_d;
define:
if (gender eq 1) then gender_d = 1;
if (gender eq 2) then gender_d = 0;
model:
grade on bptot gender_d;
output: sampstat stdyx;
plot: type=plot1;
```

The status bar at the bottom indicates "Ready", "Ln 21, Col 8", and "NL".

Mplus: Multiple Regression – Two Predictors

Means			
	GRADE	BPTOT	GENDER_D
1	74.698	59.829	0.355
Covariances			
	GRADE	BPTOT	GENDER_D
GRADE	94.201		
BPTOT	-23.893	304.857	
GENDER_D	0.193	0.978	0.229
Correlations			
	GRADE	BPTOT	GENDER_D
GRADE	1.000		
BPTOT	-0.141	1.000	
GENDER_D	0.041	0.117	1.000



MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON				
BPTOT	-0.082	0.029	-2.788	0.005
GENDER_D	1.192	1.076	1.108	0.268

Intercepts				
GRADE	79.193	1.833	43.195	0.000

Residual Variances				
GRADE	92.007	6.935	13.267	0.000

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE ON				
BPTOT	-0.148	0.052	-2.818	0.005
GENDER_D	0.059	0.053	1.110	0.267

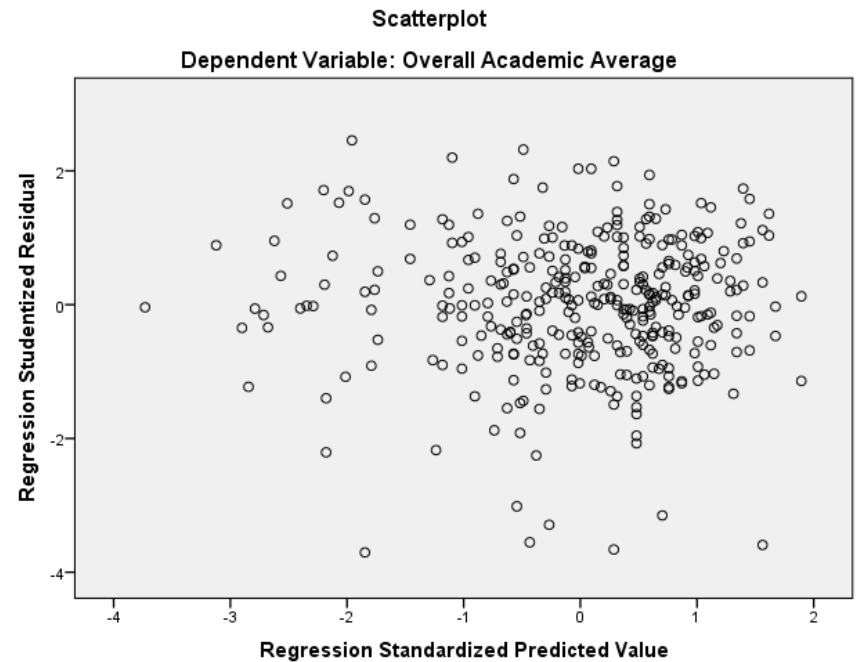
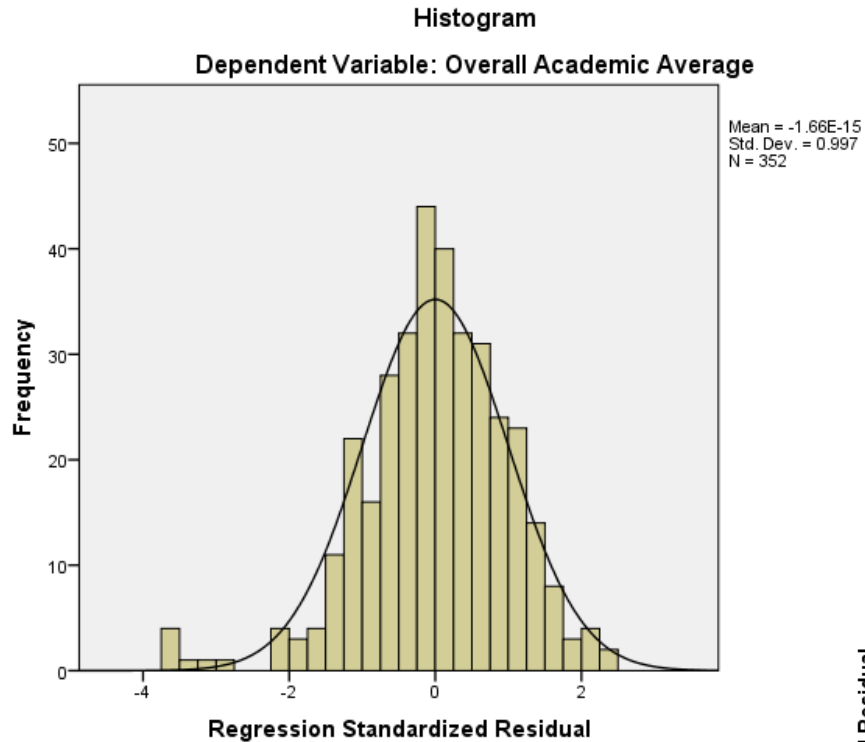
Intercepts				
GRADE	8.159	0.331	24.682	0.000

Residual Variances				
GRADE	0.977	0.016	61.478	0.000

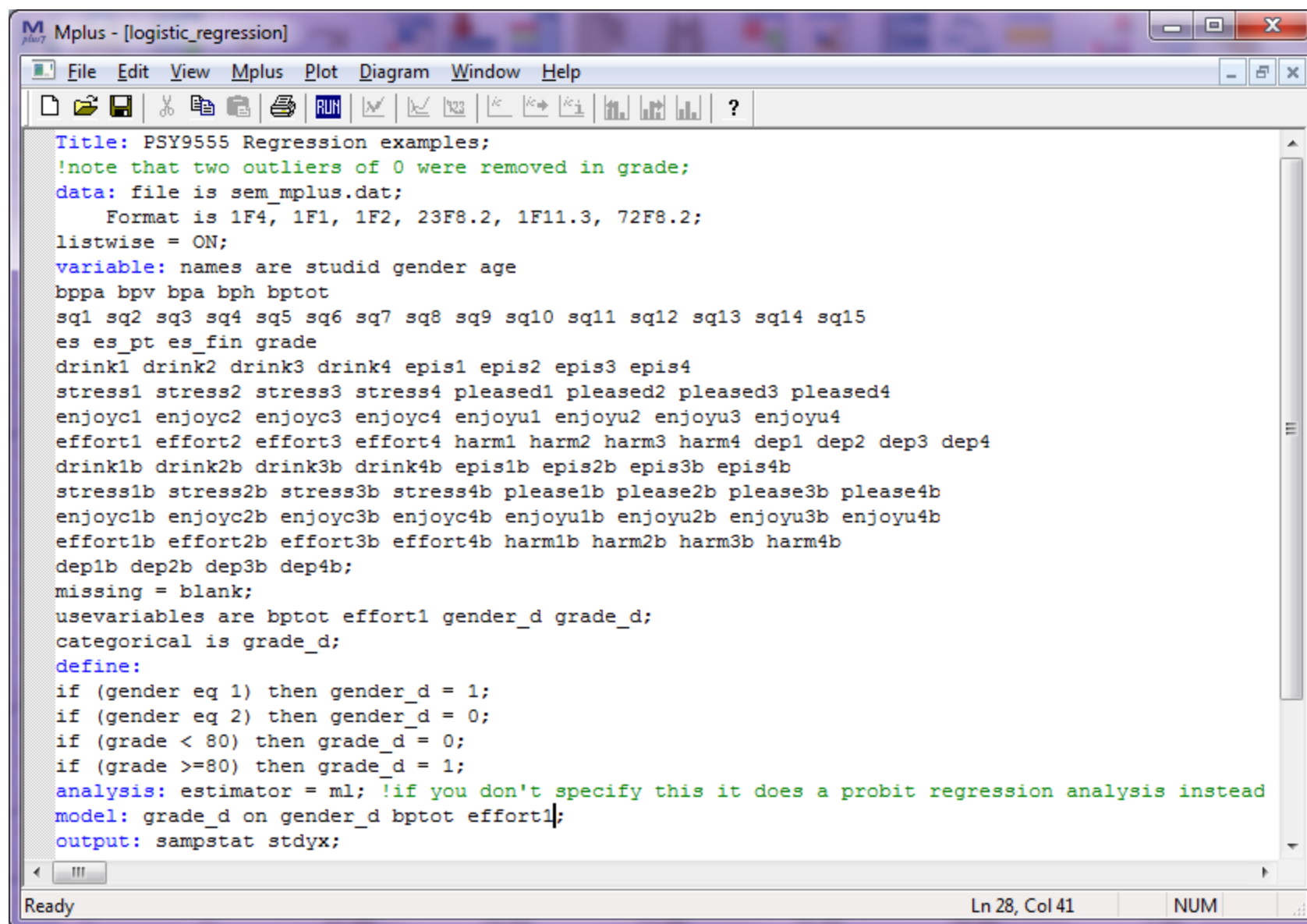
R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE	0.023	0.016	1.466	0.143

Multiple Regression – Two Predictors



Mplus: Logistic Regression



The screenshot shows the Mplus software window titled "Mplus - [logistic_regression]". The window has a menu bar (File, Edit, View, Mplus, Plot, Diagram, Window, Help) and a toolbar with icons for file operations, running, and plotting. The main text area contains the following model specification code:

```
Title: PSY9555 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 effort1 gender_d grade_d;  
categorical is grade_d;  
define:  
if (gender eq 1) then gender_d = 1;  
if (gender eq 2) then gender_d = 0;  
if (grade < 80) then grade_d = 0;  
if (grade >=80) then grade_d = 1;  
analysis: estimator = ml; !if you don't specify this it does a probit regression analysis instead  
model: grade_d on gender_d bptot effort1;  
output: sampstat stdyx;
```

The status bar at the bottom indicates "Ready", "Ln 28, Col 41", and "NUM".

Mplus: Logistic Regression

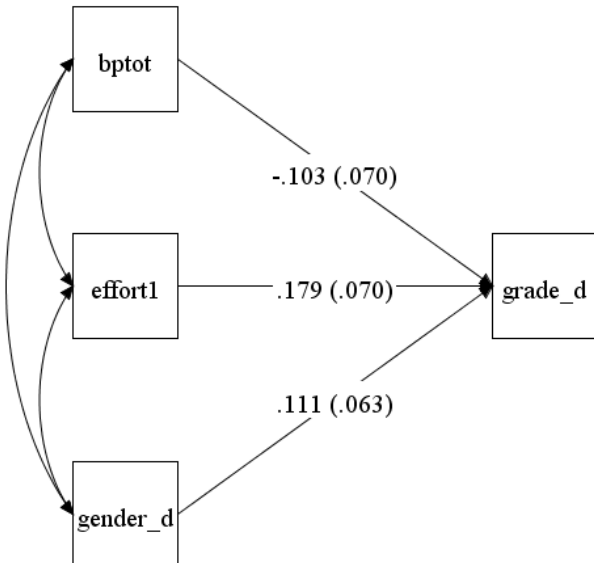
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

GRADE_D		
Category 1	0.705	248.000
Category 2	0.295	104.000

SAMPLE STATISTICS

SAMPLE STATISTICS

	Means		
	BPTOT	EFFORT1	GENDER_D
1	59.829	6.671	0.355
	Covariances		
	BPTOT	EFFORT1	GENDER_D
BPTOT	305.726		
EFFORT1	-8.854	2.841	
GENDER_D	0.981	-0.018	0.230



MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE_D ON				
GENDER_D	0.435	0.248	1.751	0.080
BPTOT	-0.011	0.008	-1.453	0.146
EFFORT1	0.199	0.080	2.480	0.013

Thresholds				
GRADE_D\$1	1.735	0.801	2.166	0.030

LOGISTIC REGRESSION ODDS RATIO RESULTS

GRADE_D ON	
GENDER_D	1.545
BPTOT	0.989
EFFORT1	1.221

STANDARDIZED MODEL RESULTS

STDYX Standardization

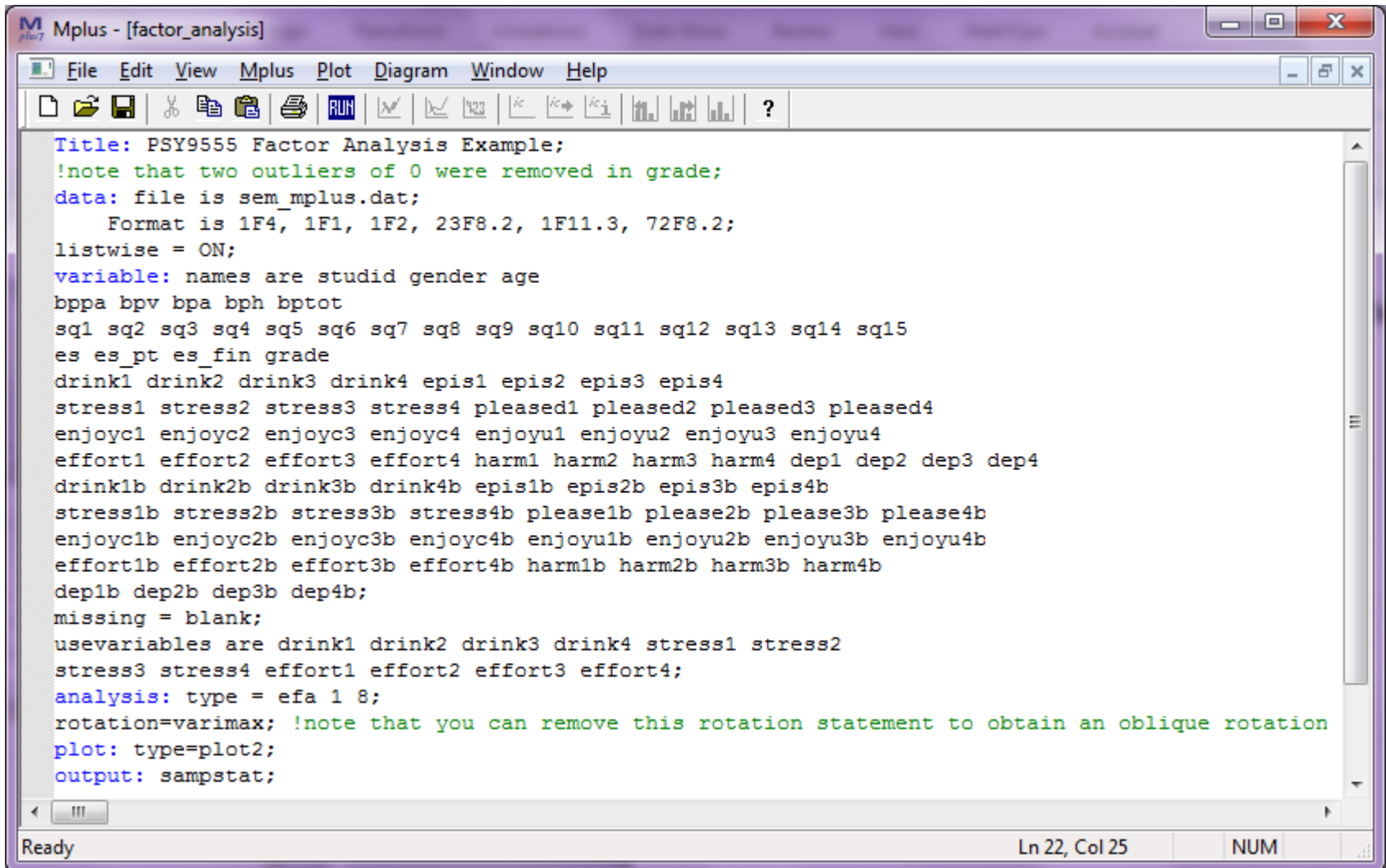
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE_D ON				
GENDER_D	0.111	0.063	1.771	0.077
BPTOT	-0.103	0.070	-1.467	0.142
EFFORT1	0.179	0.070	2.546	0.011

Thresholds				
GRADE_D\$1	0.926	0.423	2.191	0.028

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRADE_D	0.063	0.033	1.887	0.059

Mplus: Exploratory Factor Analysis



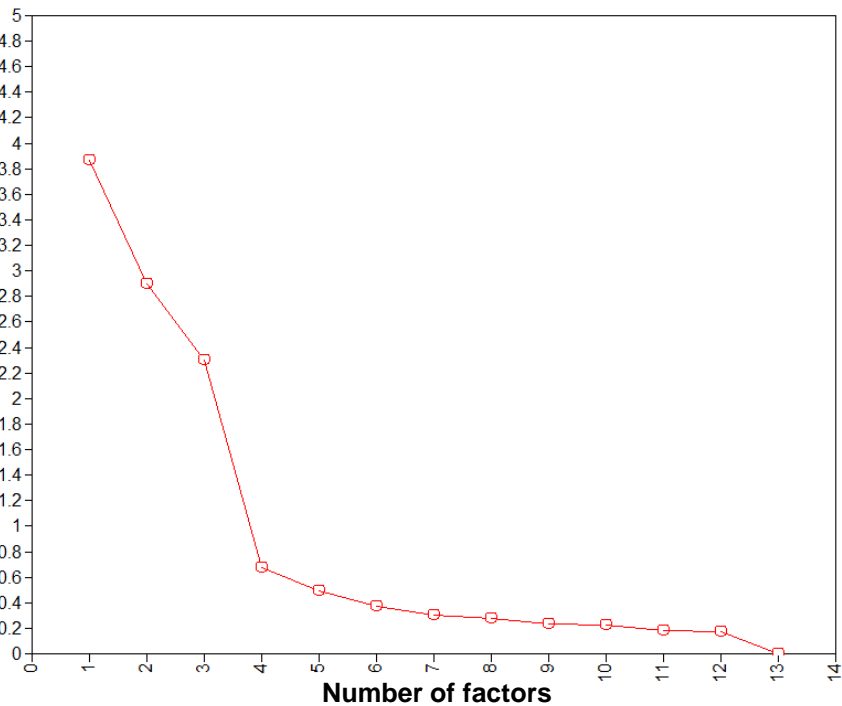
The screenshot shows the Mplus software window titled "Mplus - [factor_analysis]". The window has a menu bar with "File", "Edit", "View", "Mplus", "Plot", "Diagram", "Window", and "Help". Below the menu bar is a toolbar with various icons for file operations, editing, and analysis. The main text area contains the following script:

```
Title: PSY9555 Factor Analysis Example;  
!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 drink1 drink2 drink3 drink4 stress1 stress2  
              stress3 stress4 effort1 effort2 effort3 effort4;  
analysis: type = efa 1 8;  
rotation=varimax; !note that you can remove this rotation statement to obtain an oblique rotation  
plot: type=plot2;  
output: sampstat;
```

The status bar at the bottom shows "Ready", "Ln 22, Col 25", and "NUM".

Mplus: Exploratory Factor Analysis

Eigenvalues



Scree Plot

PROBLEM OCCURRED IN EXPLORATORY FACTOR ANALYSIS WITH 7 FACTOR(S).

SUMMARY OF MODEL FIT INFORMATION

Model	Chi-Square	Degrees of Freedom	P-Value
1-factor	1713.208	54	0.0000
2-factor	841.598	43	0.0000
3-factor	214.815	33	0.0000
4-factor	106.778	24	0.0000
5-factor	42.427	16	0.0003
6-factor	18.150	9	0.0335
7-factor	N/A		

Models Compared	Chi-Square	Degrees of Freedom	P-Value
1-factor against 2-factor	871.610	11	0.0000
2-factor against 3-factor	626.783	10	0.0000
3-factor against 4-factor	108.037	9	0.0000
4-factor against 5-factor	64.352	8	0.0000
5-factor against 6-factor	24.276	7	0.0010

Mplus: Exploratory Factor Analysis

VARIMAX ROTATED LOADINGS

	1	2	3
DRINK1	0.847	0.047	0.062
DRINK2	0.845	0.018	0.045
DRINK3	0.907	0.069	0.024
DRINK4	0.872	0.038	0.086
STRESS1	0.048	-0.748	0.003
STRESS2	-0.042	-0.788	-0.079
STRESS3	-0.123	-0.858	-0.127
STRESS4	-0.065	-0.817	-0.167
EFFORT1	-0.074	-0.057	-0.696
EFFORT2	-0.020	-0.088	-0.706
EFFORT3	-0.072	-0.064	-0.837
EFFORT4	-0.028	-0.120	-0.815

GEOMIN ROTATED LOADINGS (* significant at 5% level)

	1	2	3
DRINK1	0.849*	-0.004	-0.009
DRINK2	0.849*	0.023	0.006
DRINK3	0.910*	-0.028	0.036
DRINK4	0.874*	0.009	-0.032
STRESS1	0.088	0.769*	-0.089
STRESS2	0.004	0.797*	-0.015
STRESS3	-0.072	0.860*	0.021
STRESS4	-0.013	0.816*	0.069
EFFORT1	-0.031	-0.015	0.700*
EFFORT2	0.026	0.017	0.711*
EFFORT3	-0.020	-0.022	0.844*
EFFORT4	0.026	0.039	0.818*

GEOMIN FACTOR CORRELATIONS (* significant at 5% level)

	1	2	3
1	1.000		
2	-0.115	1.000	
3	-0.130*	0.222*	1.000

FACTOR STRUCTURE

	1	2	3
DRINK1	0.851	-0.104	-0.120
DRINK2	0.846	-0.073	-0.100
DRINK3	0.909	-0.124	-0.088
DRINK4	0.877	-0.099	-0.144
STRESS1	0.011	0.739	0.070
STRESS2	-0.085	0.793	0.161
STRESS3	-0.173	0.873	0.221
STRESS4	-0.116	0.833	0.252
EFFORT1	-0.120	0.144	0.701
EFFORT2	-0.068	0.172	0.711
EFFORT3	-0.128	0.167	0.842
EFFORT4	-0.085	0.218	0.823