An introduction to latent class analysis using Mplus

Dr. Orla McBride

orlamcbride@rcsi.ie

18th November 2011

University of Ulster, Magee

Workshop Overview

10-11:15am – Background to latent class analysis

11:15-11:30 – Coffee Break

11:30-12:30 – Estimating LC models in Mplus: Guidelines and Examples

12:30-1:30 – Lunch

1:30-2:30 – Practical

2:30-2:45 – Coffee Break

2:45-4:00 – More flexible LC models

Some useful references

- McCutcheon, A. C. (1987). *Latent class analysis*. Beverly Hills, CA: Sage.
  - (‘*Anyone with a good practical knowledge of algebra should have little difficulty reading this monograph*’!)

Some useful references

- Muthén & Muthén (2009). Categorical Latent Variable Modeling Using Mplus: Cross-Sectional Data (http://www.statmodel.com/download/Topic%205.pdf)

- Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates.

- Uebersax, J. (2009). *A Practical Guide to Conditional Dependence in Latent Class Models*. Retrieved from http://www.john-uebersax.com/stat/condep.htm

- Collins, L.M. & Lanza, S.T. (2010). *Latent class and latent transition analysis for the social, behavioral, and health sciences*. New York: Wiley.

Some useful references (applied)

- Hagenaars, J.A. & McCutcheon, A.L. (2002). Applied Latent Class Analysis. Cambridge University Press.

- Breslau, N., Reboussin, B.A. et al. (2005). The Structure of Posttraumatic Stress Disorder: latent class analysis in 2 community samples. *Archives of General Psychiatry*, 62, 1343-1351.

- Reboussin, B.A., Ip, E.H., & Wolfson, M. (2008). Locally dependent latent class models with covariates: an application to under-age drinking in the USA. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 877-897.

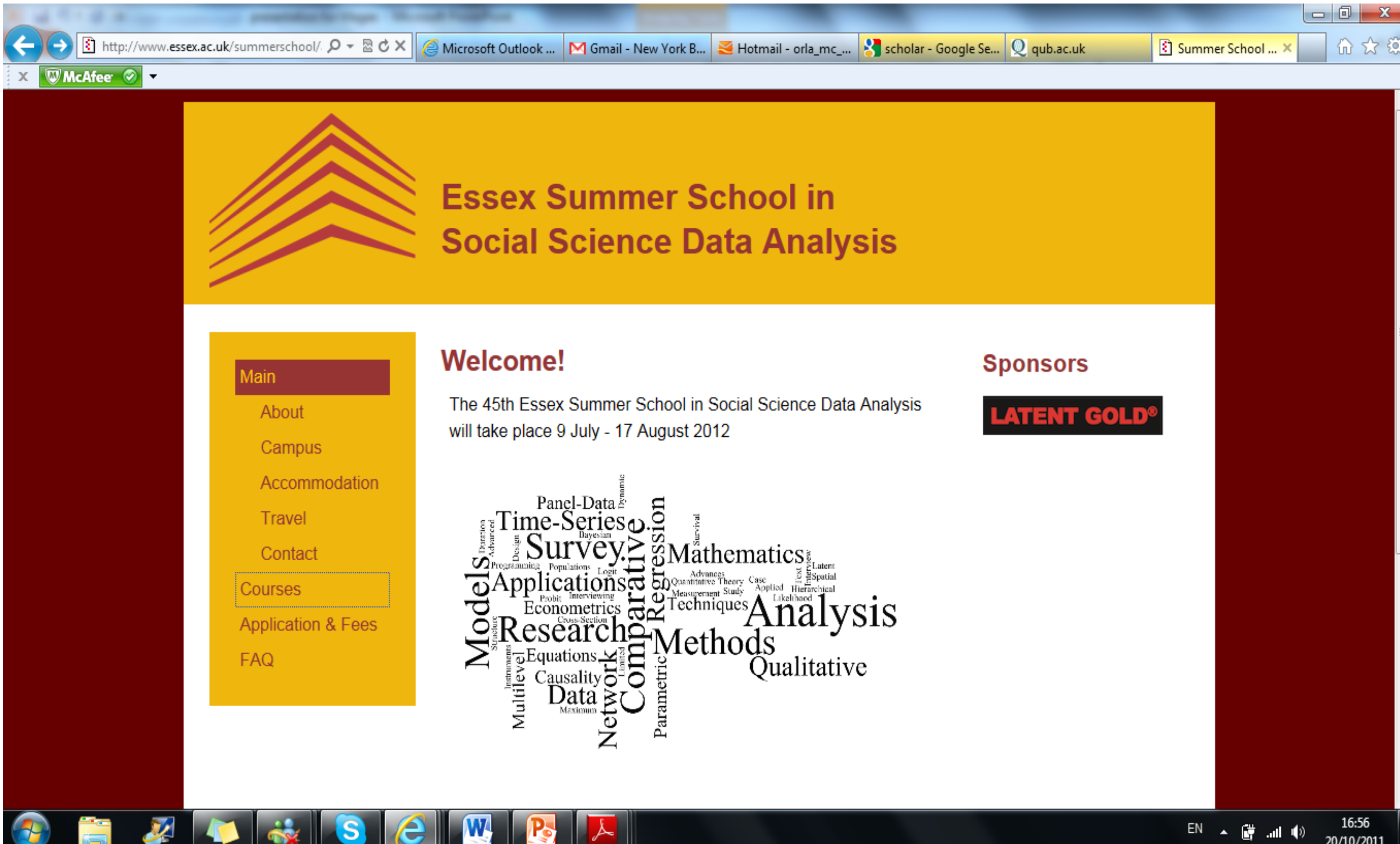
- Weich, S. McBride, O., Hussey, D. et al. (2011). Latent class analysis of co-morbidity in the Adult Psychiatric Morbidity Survey in England 2007: implications for DSM-5 and ICD-11. *Psychological Medicine*, 41, 2201-2212.

Some useful online resources

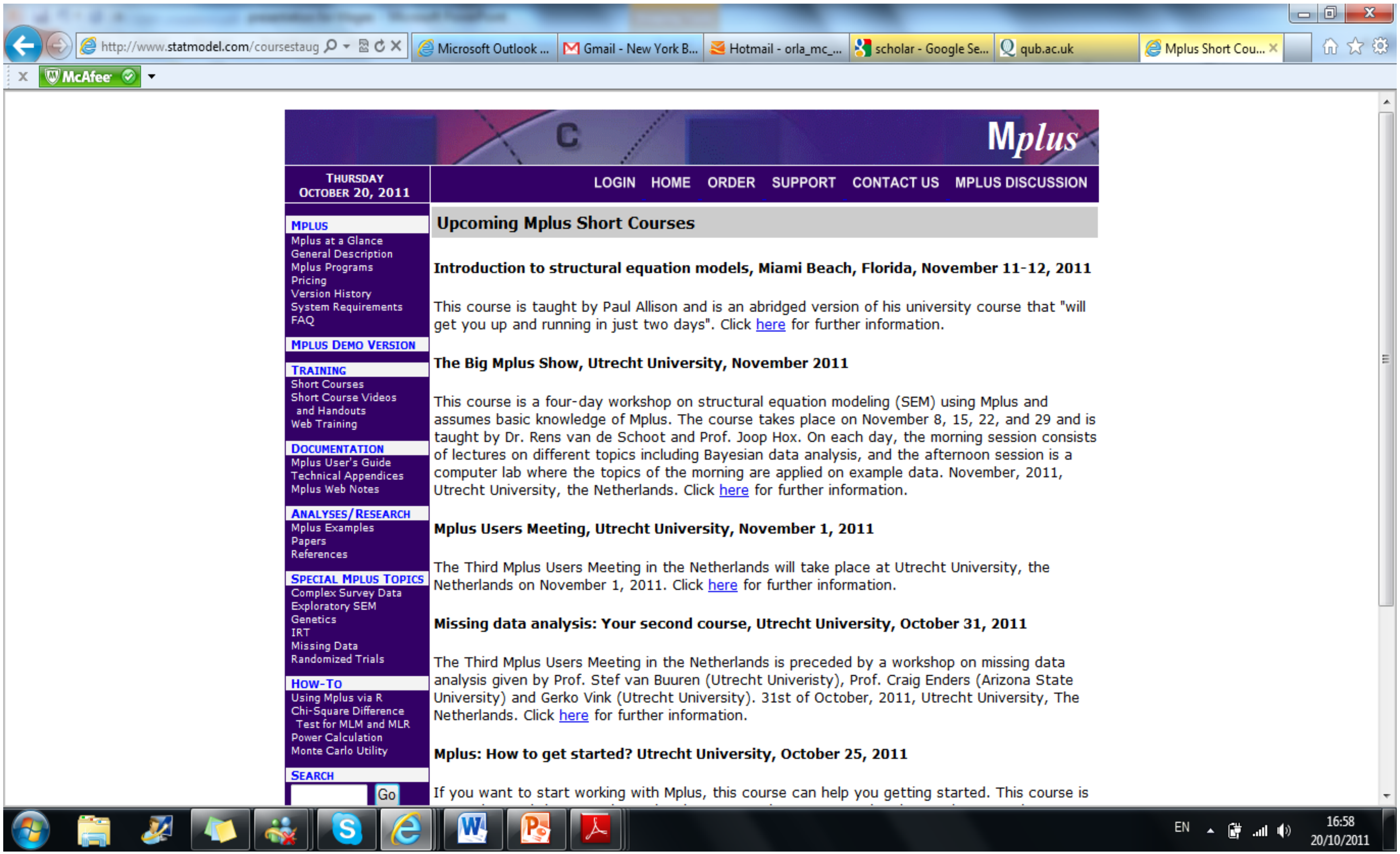
- Bengt Muthén’s UCLA homepage (lots of papers available to download for free): http://pages.gseis.ucla.edu/faculty/muthen/full_paper_list.htm

- Statistical Computing Seminars from UCLA Academic Technology Services: http://www.ats.ucla.edu/stat/mplus/seminars/lca/default.htm

Some useful training courses (beyond this workshop)



Some useful training courses (beyond this workshop)



Mplus

THURSDAY OCTOBER 20, 2011

LOGIN HOME ORDER SUPPORT CONTACT US MPLUS DISCUSSION

MPLUS
Mplus at a Glance
General Description
Mplus Programs
Pricing
Version History
System Requirements
FAQ

MPLUS DEMO VERSION

TRAINING
Short Courses
Short Course Videos and Handouts
Web Training

DOCUMENTATION
Mplus User's Guide
Technical Appendices
Mplus Web Notes

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
Go

Upcoming Mplus Short Courses

Introduction to structural equation models, Miami Beach, Florida, November 11-12, 2011

This course is taught by Paul Allison and is an abridged version of his university course that "will get you up and running in just two days". Click here for further information.

The Big Mplus Show, Utrecht University, November 2011

This course is a four-day workshop on structural equation modeling (SEM) using Mplus and assumes basic knowledge of Mplus. The course takes place on November 8, 15, 22, and 29 and is taught by Dr. Rens van de Schoot and Prof. Joop Hox. On each day, the morning session consists of lectures on different topics including Bayesian data analysis, and the afternoon session is a computer lab where the topics of the morning are applied on example data. November, 2011, Utrecht University, the Netherlands. Click here for further information.

Mplus Users Meeting, Utrecht University, November 1, 2011

The Third Mplus Users Meeting in the Netherlands will take place at Utrecht University, the Netherlands on November 1, 2011. Click here for further information.

Missing data analysis: Your second course, Utrecht University, October 31, 2011

The Third Mplus Users Meeting in the Netherlands is preceded by a workshop on missing data analysis given by Prof. Stef van Buuren (Utrecht Univeristy), Prof. Craig Enders (Arizona State University) and Gerko Vink (Utrecht University). 31st of October, 2011, Utrecht University, The Netherlands. Click here for further information.

Mplus: How to get started? Utrecht University, October 25, 2011

If you want to start working with Mplus, this course can help you getting started. This course is

Introduction

The logic of latent variables

Religious commitment

Happiness and Wellbeing

So, how do you measure wellbeing and happiness?

Can you really measure wellbeing and happiness? The Office for National Statistics has published plans to measure how content we are. But how would they work?

How do you measure happiness and wellbeing? The ONS thinks it knows. Photograph: Christopher Thomond/Guardian

How happy are you right now? Content? Satisfied? Anxious? Crucially, who else feels the way you do?

Alcohol use disorder

Basic orientation

- Belief: observed indicators are caused by an unobserved, or *latent*, variable of interest
- Covariation among the observed indicators is expected
- Study the patterns of interrelationships among the observed indicators to understand and characterise the underlying latent variable

Early work on latent variables

- Used factor analysis – continuous latent variables (generally continuous observed indicators)
- Factor analysis reduces many observed variables to a few latent factors
- Latent class analysis (LCA) is a method for studying categorically scored variables that is comparable to factor analysis

What is LCA?

- Allows researchers to empirically identify discrete latent variables from two or more discrete observed variables (Green, 1951)

- Goal - To group individuals into categories, each one of which contains individuals who are similar to each other and different from individuals in other categories (Muthén & Muthén, 2000).

Example (Adapted from McCutcheon, 1987)

- The Irish Contraception and Crisis Pregnancy survey (ICCP) is a national survey conducted in the Republic of Ireland in 2003 (replicated in 2010) – www.crisispregnancy.ie

- Survey of knowledge, attitudes, and behaviour in relation to contraception, sexual health, and pregnancy

Cross-tabulation of attitudes towards contraception - **observed**

	Variable B: "*Carrying condoms while not in a relationship, gives the message that you are 'easy' or looking for sex*"		Totals
Variable A: "*Find it difficult to talk to a sexual partner about contraception*"	+ (Agree)	- (Disagree)	
+ (Agree)	95 (A)	55 (B)	150 (A+B)
- (Disagree)	70 (C)	80 (D)	150 (C+D)
	165 (A+C)	135 (B+D)	300 (A+B+C+D=N)

Cross-tabulation of attitudes towards contraception - **expected**

	Variable B: "*Carrying condoms while not in a relationship, gives the message that you are 'easy' or looking for sex*"		Totals
Variable A: "*Find it difficult to talk to a sexual partner about contraception*"	+ (Agree)	- (Disagree)	
+ (Agree)	(165x150)/300	(135x150)/300	150
- (Disagree)	(165x150)/300	(135x150)/300	150
	165	135	300

Calculating χ² – contraception attitudes example

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(95 - 82 \cdot 5)^2}{82 \cdot 5} + \frac{(55 - 67 \cdot 5)^2}{67.5} + \frac{(70 - 82 \cdot 5)^2}{82 \cdot 5} + \frac{(80 - 67.5)^2}{67 \cdot 5}$$

$$\chi^2 = 1.89 + 2 \cdot 32 + 1 \cdot 89 + 2 \cdot 32$$

$$\chi^2 = 8 \cdot 42 (p < 0.01)$$

Upper critical values of chi-square distribution with ν degrees of freedom

ν	0.10	0.05	0.025	0.01	0.001
	Probability of exceeding the critical value				
1	2.706	3.841	5.024	$\boxed{6.635}$	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315
21	29.615	32.671	35.479	38.932	46.797
22	30.813	33.924	36.781	40.289	48.268

What does this result mean?

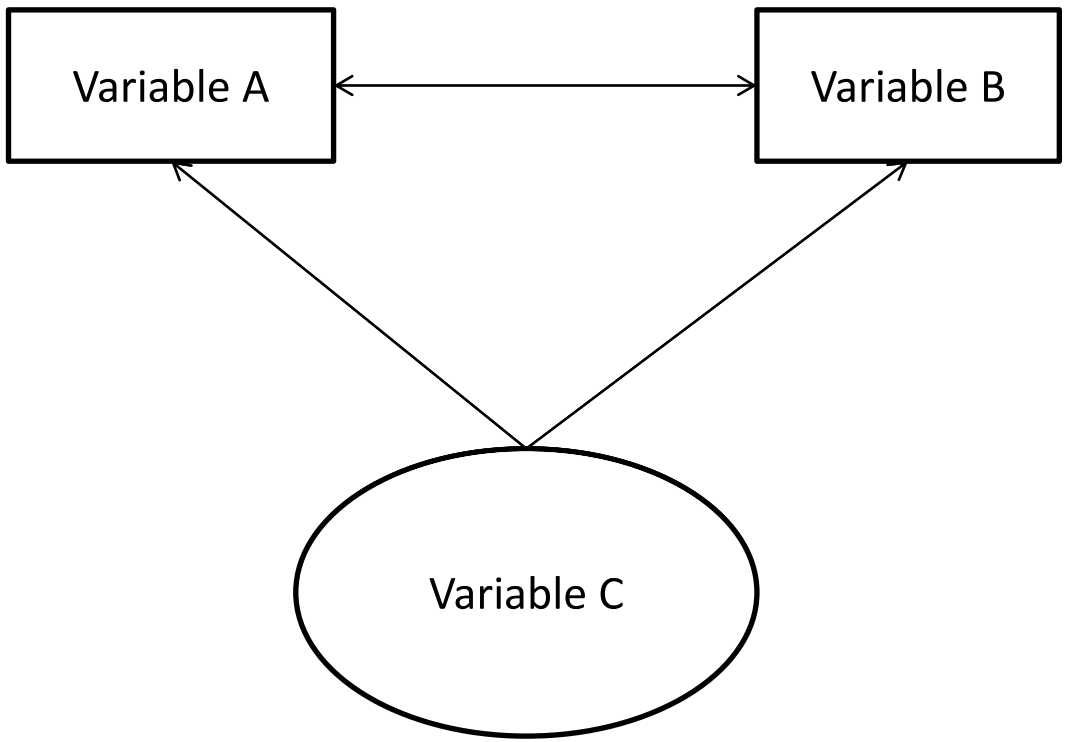
- These attitudes towards contraception are related (i.e. <u>**not**</u> independent)

Interpretations of relationships between two or more variables

- Causality

- Symmetrical relationships
  - Alternative indicators of same concept
  - Parts of a common ‘system’
  - Fortuitous

Interpretations of relationships between two or more variables



Local independence as a concept

High Conservatism – 'Conservative'

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	80 (A)	20 (B)	100 (A+B)
- (Disagree)	40 (C)	10 (D)	50 (C+D)
	120 (A+C)	30 (B+D)	150 (N)

Low Conservatism- 'Non-Conservative'

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	15 (A)	35 (B)	50 (A+B)
- (Disagree)	30 (C)	70(D)	100 (C+D)
	45 (A+C)	105 (B+D)	150 (N)

At each level of Variable C ('Conservatism'), variables A and B are independent of one another ($\chi^2 = 0.0$)

Local independence

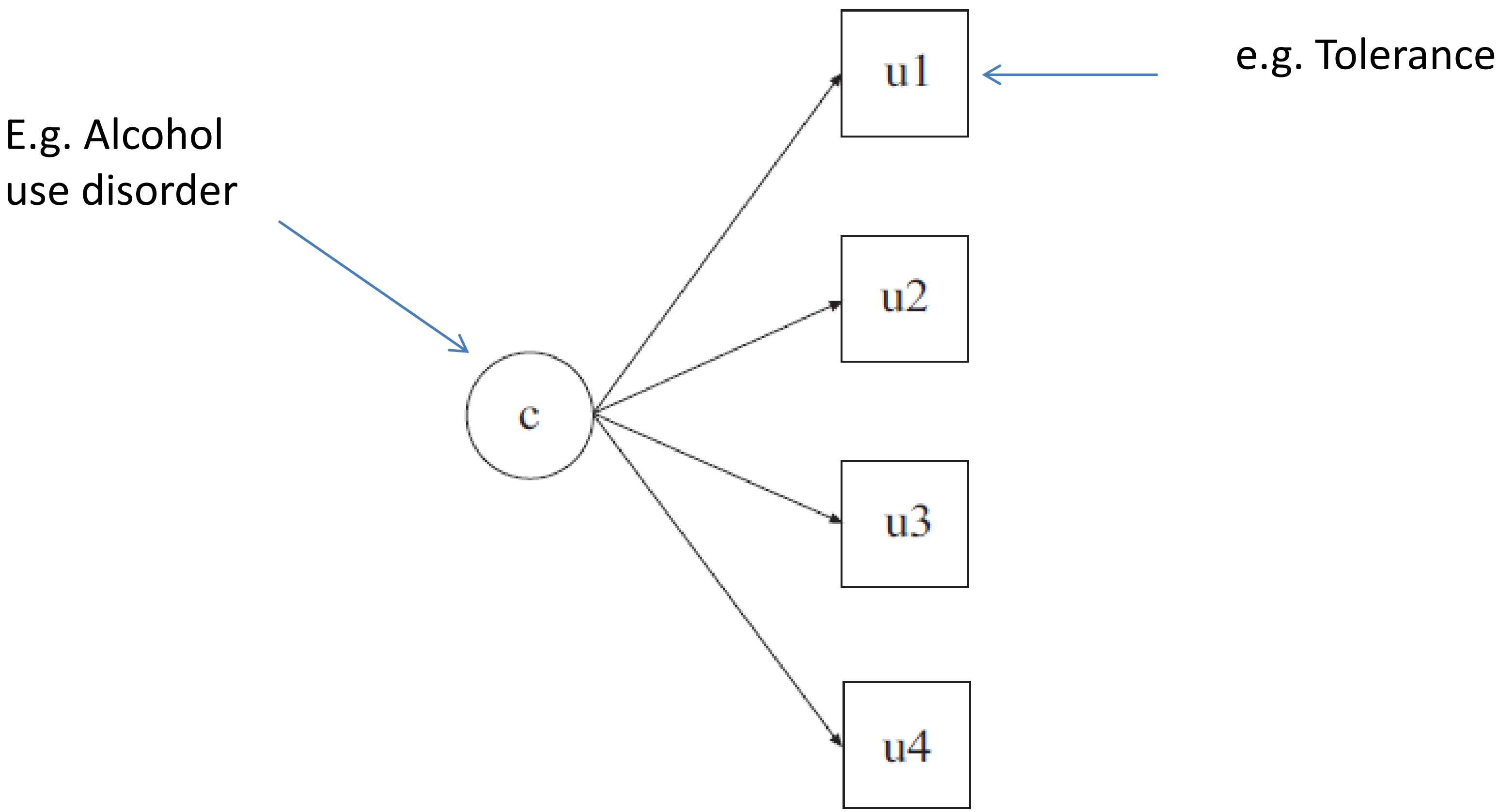
- Local independence – when the relationship observed among a set of variables are found to be zero within the categories of some other variable (Lazarsfeld & Henry, 1968).

- Key concept in LCA, but a strict one! (We'll return to this issue later).

The formal latent class (LC) model

- In the example, we measured a variable (C) that explains the symmetrical relationship between A and B
- In reality, however, we are not so fortunate as to have measured that variable!!
- **Object of LC model**: to define a latent variable – specifically, a set of classes – within which the manifest variables are locally independent

The Latent Class Model



(Source: Muthén & Muthén, 1998-2010)

Model parameters

There are **<u>two</u>** fundamental quantities in LCA (Goodman, 1974)

1. Latent class probabilities

2. Conditional probabilities for each class

Latent class probabilities (π_t^X)

- Describe distribution of classes of the latent variable (X) within which the observed measures are (locally) independent of one another.
  - Number of classes (T)
  - Relative sizes of the classes

Conditional probabilities

- Comparable to factor loadings in factor analysis
- Represent the probability of an individual in a given class of the latent variable being at a particular level of the observed variables

Conditional probability

- Recall our contraception/Conservatism example: Variables A and B were independent of one another **conditional** on Variable C

- Let’s work out how to calculate the conditional probabilities by hand (even though Mplus will do this for you!)

Example

High Conservatism – 'Conservative'

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	80 (A)	20 (B)	100 (A+B)
- (Disagree)	40 (C)	10 (D)	50 (C+D)
	120 (A+C)	30 (B+D)	150 (N)

Low Conservatism- 'Non-Conservative'

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	15 (A)	35 (B)	50 (A+B)
- (Disagree)	30 (C)	70(D)	100 (C+D)
	45 (A+C)	105 (B+D)	150 (N)

Conditional probability

- What is the probability that a **conservative** respondent agrees that:
  - ‘Find it difficult to talk to sexual partner about contraception’ (Variable A)

Can be expressed using this equation:

$$\widehat{P}\,{}_{11}^{\bar{A}C}$$

where $\widehat{P}$ = Sample probability

$\bar{A}C$ = Respondent is at level 1 of Variable A (i.e. Agree) and level 1 of Variable C (i.e. Conservative)

Conditional probability

High Conservatism – 'Conservative'

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	80 (A)	20 (B)	100 (A+B)
- (Disagree)	40 (C)	10 (D)	50 (C+D)
	120 (A+C)	30 (B+D)	150 (N)

Calculate conditional probability

$$\widehat{P}\,{}^{\bar{A}C}_{11} = 100/150 = 0.667$$

The probability that a **<u>conservative</u>** respondent agrees that it is "Difficult to talk to sexual partner about contraception" is 0.667

Conditional probability

- What is the probability that a **conservative** respondent agrees that:
  - "Carrying condoms gives the message you are 'easy'" (Variable B)

Can be expressed using this equation:

$$\widehat{P}\,{}^{\bar{B}C}_{11}$$

where $\widehat{P}$ = Probabilities are sample estimates

$\bar{B}C$ = Respondent is at level 1 of Variable B (i.e. Agree) and level 1 of Variable C (i.e. Conservative)

Conditional probability

High Conservatism – ‘Conservative’

	Variable B ‘Carrying condoms= easy’		Totals
Variable A ‘Difficult to talk to partner’	+ (Agree)	- (Disagree)	
+ (Agree)	80 (A)	20 (B)	100 (A+B)
- (Disagree)	40 (C)	10 (D)	50 (C+D)
	120 (A+C)	30 (B+D)	150 (N)

Calculate conditional probability

$$\widehat{P}\,{}^{\bar{B}C}_{11} = 120/150 = 0.800$$

The probability that a **<u>conservative</u>** respondent agrees that “Carrying condoms gives the message you are ‘easy’” is 0.800

Conditional probability

- What is the probability that a **non-conservative** respondent agrees that:
  - ‘Find it difficult to talk to sexual partner about contraception’ (Variable A)

Can be expressed using this equation:

$$\widehat{P}\,{}^{\bar{A}C}_{12}$$

where $\widehat{P}$ = Probabilities are sample estimates

$\bar{A}C$ = Respondent is at level 1 of Variable A (i.e. Agree) and level 2 of Variable C (i.e. Non-Conservative)

Conditional probability

Non-Conservative

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	15 (A)	35 (B)	50 (A+B)
- (Disagree)	30 (C)	70(D)	100 (C+D)
	45 (A+C)	105 (B+D)	150 (N)

Calculate conditional probability

$$\widehat{P}\,{}^{\bar{A}C}_{12} = 50/150 = 0.333$$

The probability that a **non-conservative** respondent agrees that it is "Difficult to talk to sexual partner about contraception" is 0.333

Conditional probability

- What is the probability that a **non-conservative** respondent agrees that:
  - "'Carrying condoms gives the message you are 'easy'"

Can be expressed using this equation:

$$\widehat{P}\,{}^{\bar{B}C}_{12}$$

where $\widehat{P}$ = Probabilities are sample estimates

$\bar{A}C$ = Respondent is at level 1 of Variable B (i.e. Agree) and level 2 of Variable C (i.e. Non-Conservative)

Conditional probability

Non-Conservative

	Variable B 'Carrying condoms= easy'		Totals
Variable A 'Difficult to talk to partner'	+ (Agree)	- (Disagree)	
+ (Agree)	15 (A)	35 (B)	50 (A+B)
- (Disagree)	30 (C)	70(D)	100 (C+D)
	45 (A+C)	105 (B+D)	150 (N)

Calculate conditional probability

$$\widehat{P}\,{}^{\bar{A}C}_{12} = 45/150 = 0.300$$

The probability that a **<u>non-conservative</u>** respondent agrees that it is "Carrying condoms gives the message you are 'easy'" is 0.300

Conditional probabilities

- Number of distinct conditional probabilities for each observed variable = Number of levels measured for that variable
  - Eg. If an observed variable has only 2 levels (finds it difficult to talk to a partner; does not find it difficult to talk to a partner), there will be 2 associated probabilities:

$$\pi_{1t}^{\bar{A}X} \quad \pi_{2t}^{\bar{A}X}$$

Conditional probabilities

- Help to characterise the nature of the ‘types’ defined by each latent class
- Indicate whether observations in a given class are likely or unlikely to have characteristics of each of the observed variables
- Within each of the latent classes (T), the conditional probabilities for each of the observed variables sum to 1

The formal latent class (LC) model

If A and B are (observed) manifest variables (indexed by *i* and *j*)

- Eg: If A_i is respondent's religious identification with, 1=Protestant, 2=Catholic, 3=Jewish, 4=Other, 5=None; (i.e. i=5), then A_2 represents the Catholics

If X is the latent variable ('Variable C')

If T is the number of latent classes (levels)

If π is the probability (when 'Variable C' is latent)

The formal latent class (LC) model

Then the formal LC model can be expressed as:

$$\pi_{ijt}^{ABX} = \pi_{it}^{\bar{A}X} \times \pi_{jt}^{\bar{B}X} \times \pi_{t}^{X}$$

(the last equation in this workshop!)


COFFEE

Estimating LC models in Mplus

Guidelines and Examples

Example: Alcohol Experiences

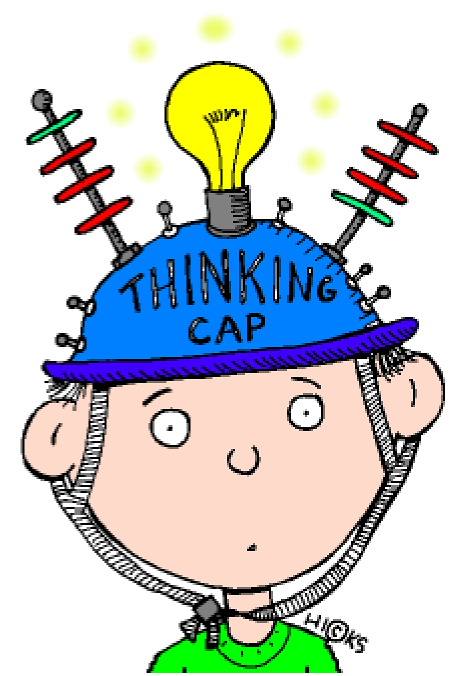
- 2 billion alcohol users worldwide; 76.3 million are diagnosable with alcohol use disorders (WHO, 2004)
- AUD are associated with a variety of medical, social, and legal consequences
- $185 billion – cost to the US economy (Li et al., 2004)
- 2001-2002 NESARC (USA): Large survey focusing on alcohol use and associated disorders

Alcohol Experiences

Code	In the last 12 months, did you…
S2BQ1B1	Find usual # of drinks had less effect than before
S2BQ1B3	Drink equivalent of a 1/5 bottle of liquor in one day
S2BQ1B6	Try unsuccessfully to stop/cut down on drinking more than once
S2BQ1B8	Have period when kept drinking longer than intended
S2BQ1B9B	Shake when effects of alcohol were wearing off
S2BQ1B9C	Feel anxious or nervous when effects of alcohol were wearing off
S2BQ1B9D	Have nausea when effects of alcohol were wearing off
S2BQ1B9F	Sweat/heart beat fast when effects of alcohol were wearing off
S2BQ1B10	Drink or use medicine/drugs (other than aspirin) to get over bad aftereffects of drinking
S2BQ1B13	Spent a lot of time being sick/getting over bad effects of drinking
S2BQ1B15	Give up or cut down pleasurable activities to drink
S2BQ1B16	Continue to drink even though depressed/uninterested/suspicious of others
S2BQ1B19	Have period when drinking interfered with taking care of home or family
S2BQ1B20	Have job/school troubles because of drinking
S2BQ1B24	Get in situations that increased chances of getting hurt while drinking
S2BQ1B25	Continue to drink despite causing trouble with family or friends
S2BQ1B26	Get into physical fights when or right after drinking
S2BQ1B27	Get arrested or have other legal problems because of drinking

Things to think about…

- Before running any analysis in Mplus, there are several things to think about:
  - Sample size
  - Response patterns/Sparseness
  - Model identification
  - Theory



Sample size

- For EFA or CFA:
  - *'The bigger the better!'*
  - 10-20 cases for each variable
  - Sample size of $\geq$ 100 is required

- Latent class models are case sensitive (i.e. they require relatively large sample sizes but general rules of thumb don't really exist)

Sparseness

- Sparseness: Many *sampling* zeros in dataset
- Difference between sampling and structural zeros
- Sparseness leads to difficulties in model evaluation
- Importance of response patterns

Model identification

- Identifiability: the degree to which there is sufficient information in the sample observations to estimate the parameters in a proposed model (McCutcheon, 1987).

- Mplus will tell you if your model is not identified!

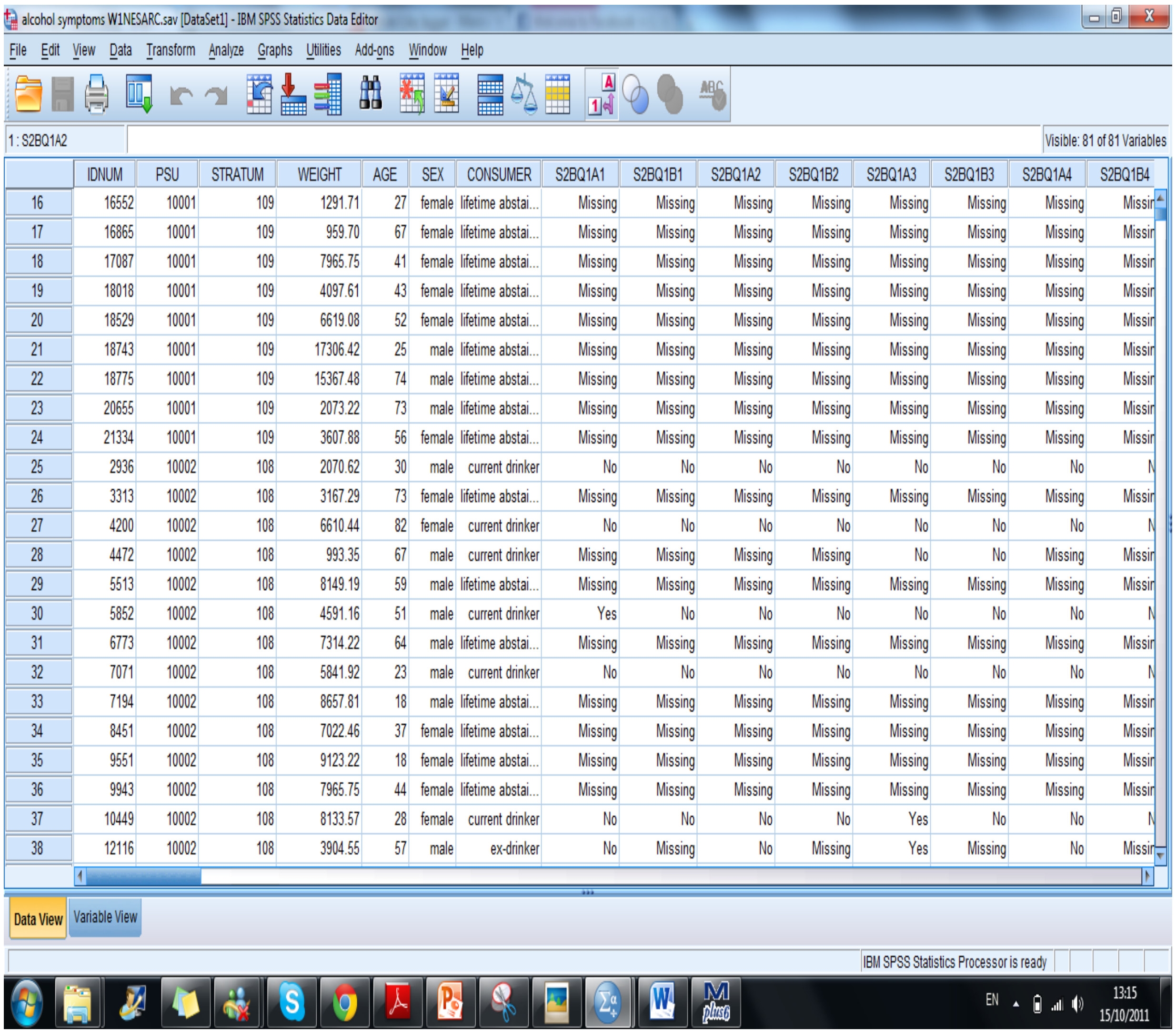
Theory

- DSM-IV: 2 mutually exclusive alcohol use disorders (abuse, dependence)
- Alcohol dependence syndrome (ADS; Edwards & Gross, 1976)
- DSM-5: addiction severity (‘alcohol use disorder’)

Now we’re ready to…

- Learn how to prepare an SPSS data file for analysis in Mplus
- Explore the different functions of the main commands and subcommands used in Mplus input syntax
- Understand the specific sections in the Mplus output file

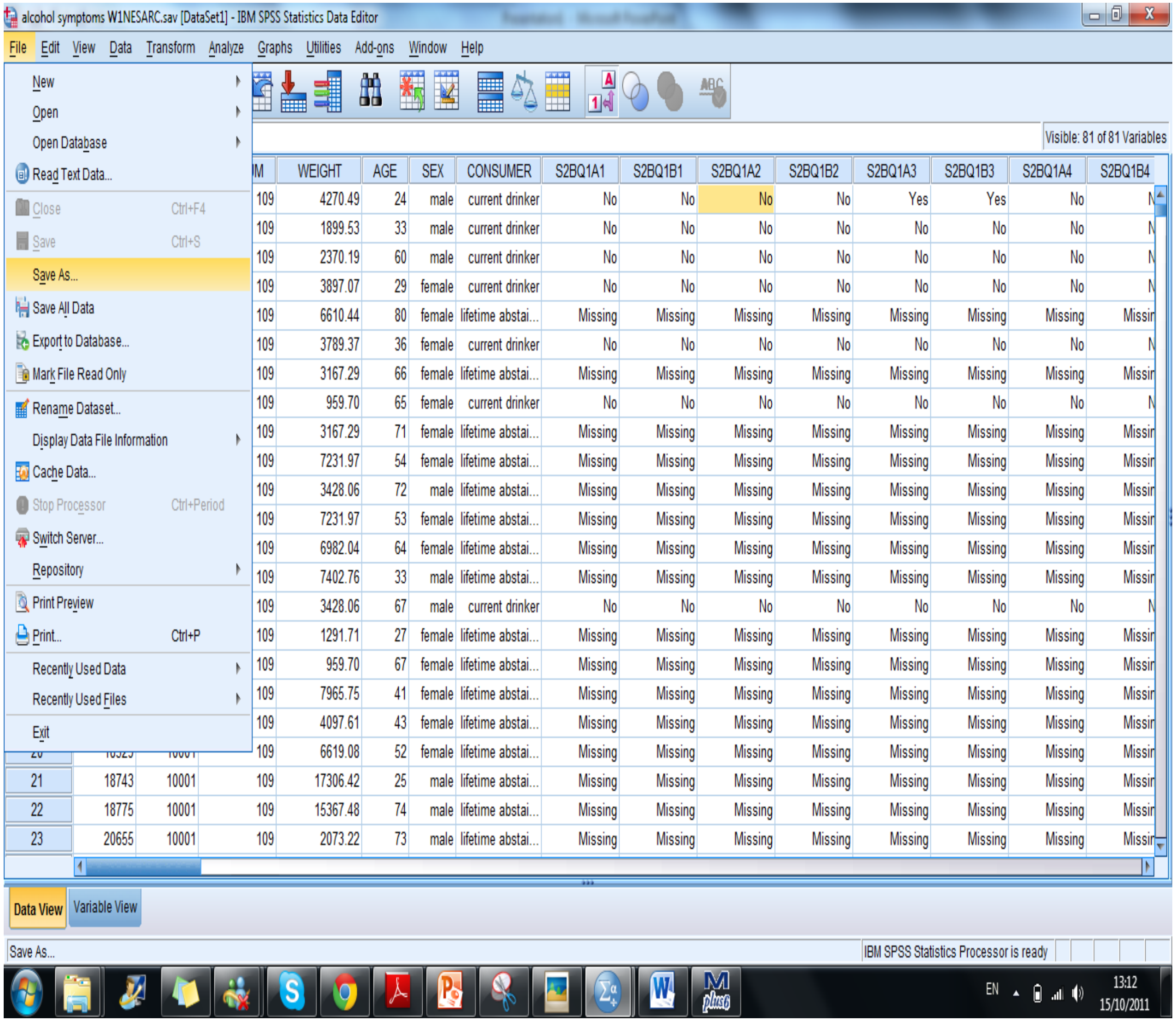
1. Prepare data file in SPSS



	IDNUM	PSU	STRATUM	WEIGHT	AGE	SEX	CONSUMER	S2BQ1A1	S2BQ1B1	S2BQ1A2	S2BQ1B2	S2BQ1A3	S2BQ1B3	S2BQ1A4	S2BQ1B4
16	16552	10001	109	1291.71	27	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
17	16865	10001	109	959.70	67	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
18	17087	10001	109	7965.75	41	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
19	18018	10001	109	4097.61	43	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
20	18529	10001	109	6619.08	52	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
21	18743	10001	109	17306.42	25	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
22	18775	10001	109	15367.48	74	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
23	20655	10001	109	2073.22	73	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
24	21334	10001	109	3607.88	56	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
25	2936	10002	108	2070.62	30	male	current drinker	No	No	No	No	No	No	No	N
26	3313	10002	108	3167.29	73	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
27	4200	10002	108	6610.44	82	female	current drinker	No	No	No	No	No	No	No	N
28	4472	10002	108	993.35	67	male	current drinker	Missing	Missing	Missing	Missing	No	No	Missing	Missir
29	5513	10002	108	8149.19	59	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
30	5852	10002	108	4591.16	51	male	current drinker	Yes	No	No	No	No	No	No	N
31	6773	10002	108	7314.22	64	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
32	7071	10002	108	5841.92	23	male	current drinker	No	No	No	No	No	No	No	N
33	7194	10002	108	8657.81	18	male	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
34	8451	10002	108	7022.46	37	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
35	9551	10002	108	9123.22	18	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
36	9943	10002	108	7965.75	44	female	lifetime abstai...	Missing	Missing	Missing	Missing	Missing	Missing	Missing	Missir
37	10449	10002	108	8133.57	28	female	current drinker	No	No	No	No	Yes	No	No	N
38	12116	10002	108	3904.55	57	male	ex-drinker	No	Missing	No	Missing	Yes	Missing	No	Missir

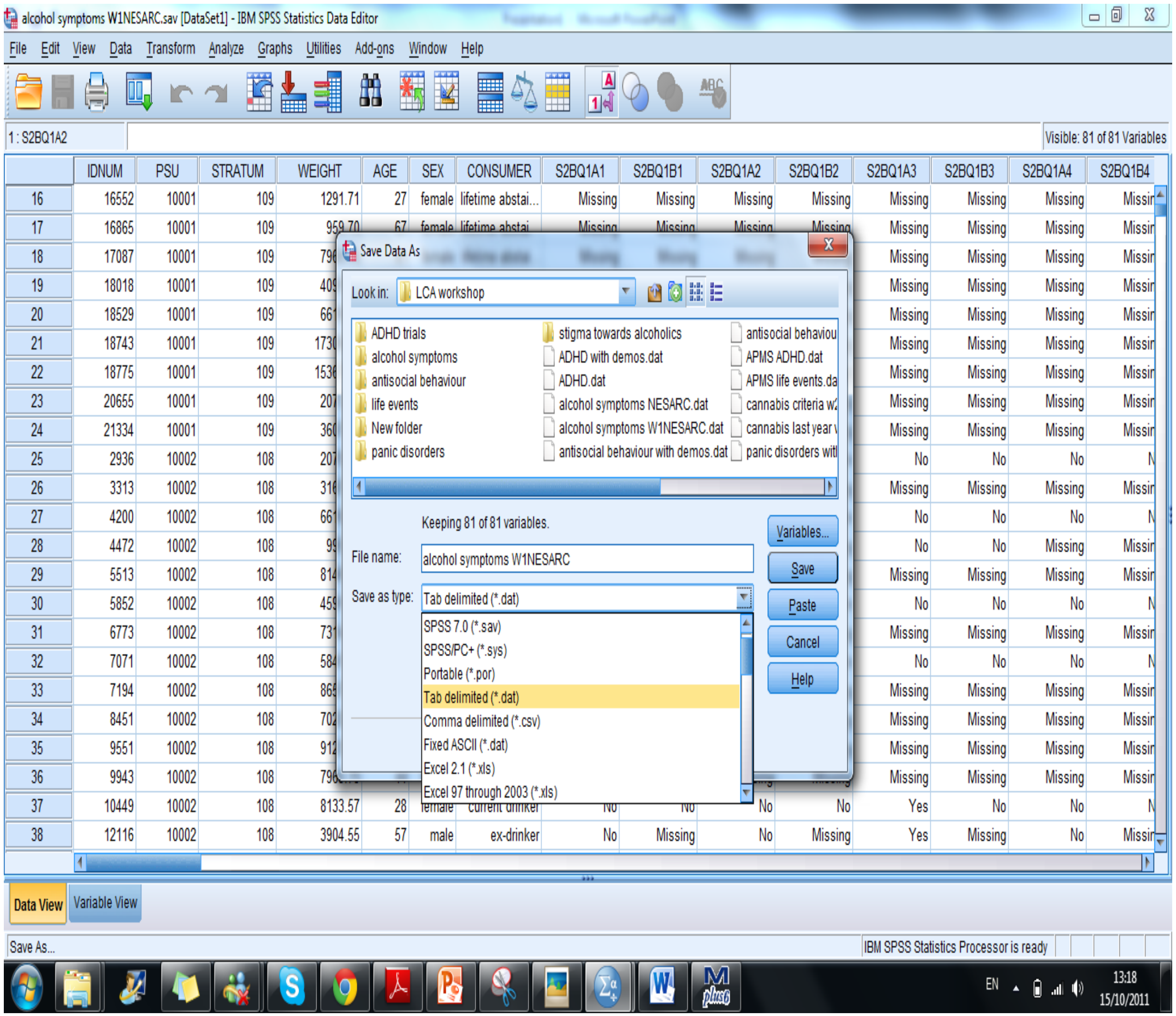
- Most researchers prepare their data file in SPSS first, prior to using Mplus
- You can re-structure data in Mplus – can be complicated!
- Go with what you know best ☺

2. Saving the data as a tab delimited file (.DAT file)



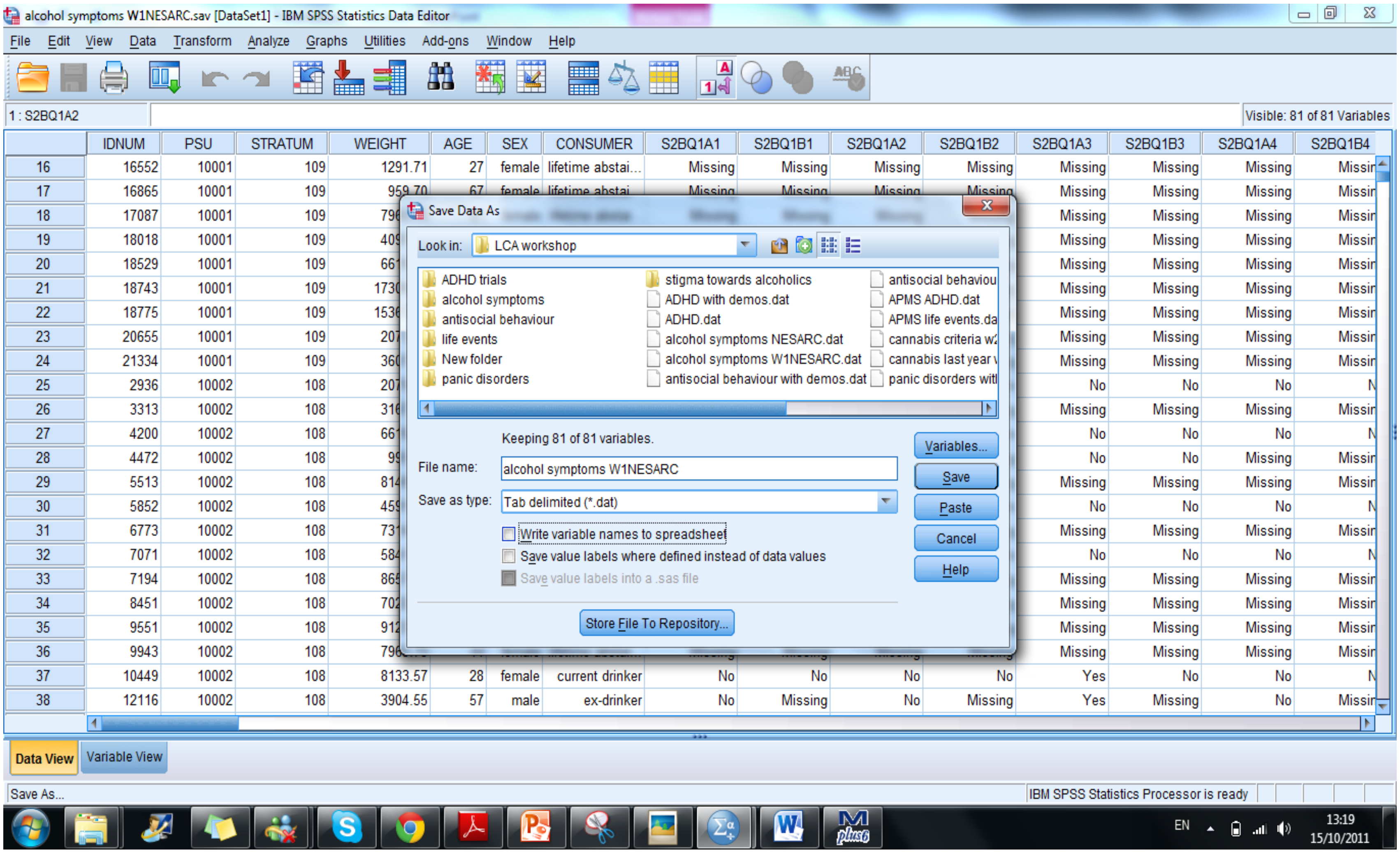
- Click on the ‘File’ tab in the Data View window
- Click on the ‘Save As’ sub-menu
- This will open the dialogue box

3. Give the .DAT file a (short) name

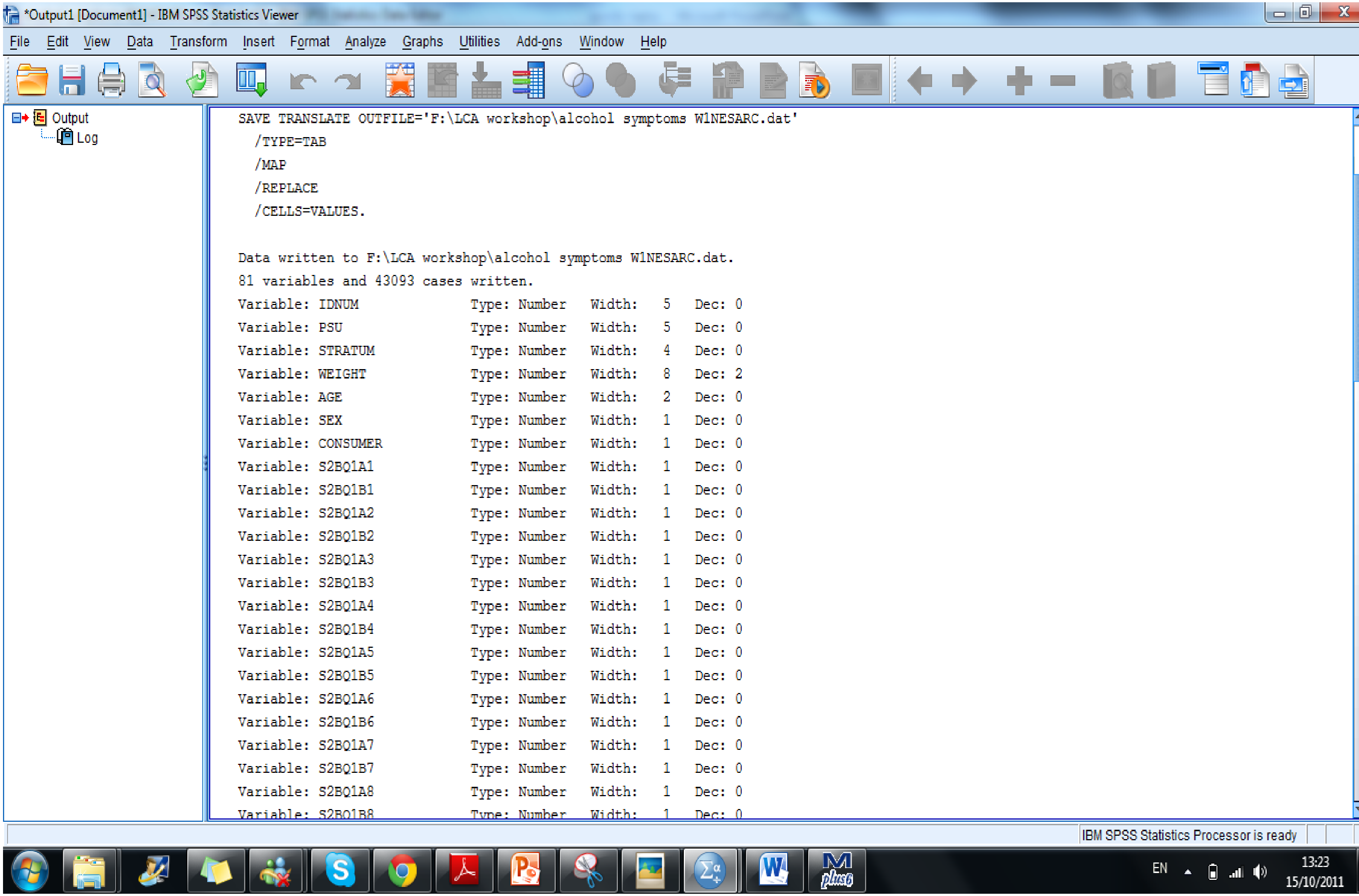


- Under the 'Save As Type' command, change the file type from .SAV (SPSS data file) to a .DAT (Tab delimited) file

4. Make sure to uncheck the 'Write variable names to spreadsheet box'



5. Check SPSS output to ensure file conversion is correct



6. Check the format of the .DAT file using Wordpad

(Mplus will be sure to tell you if there are any errors!)

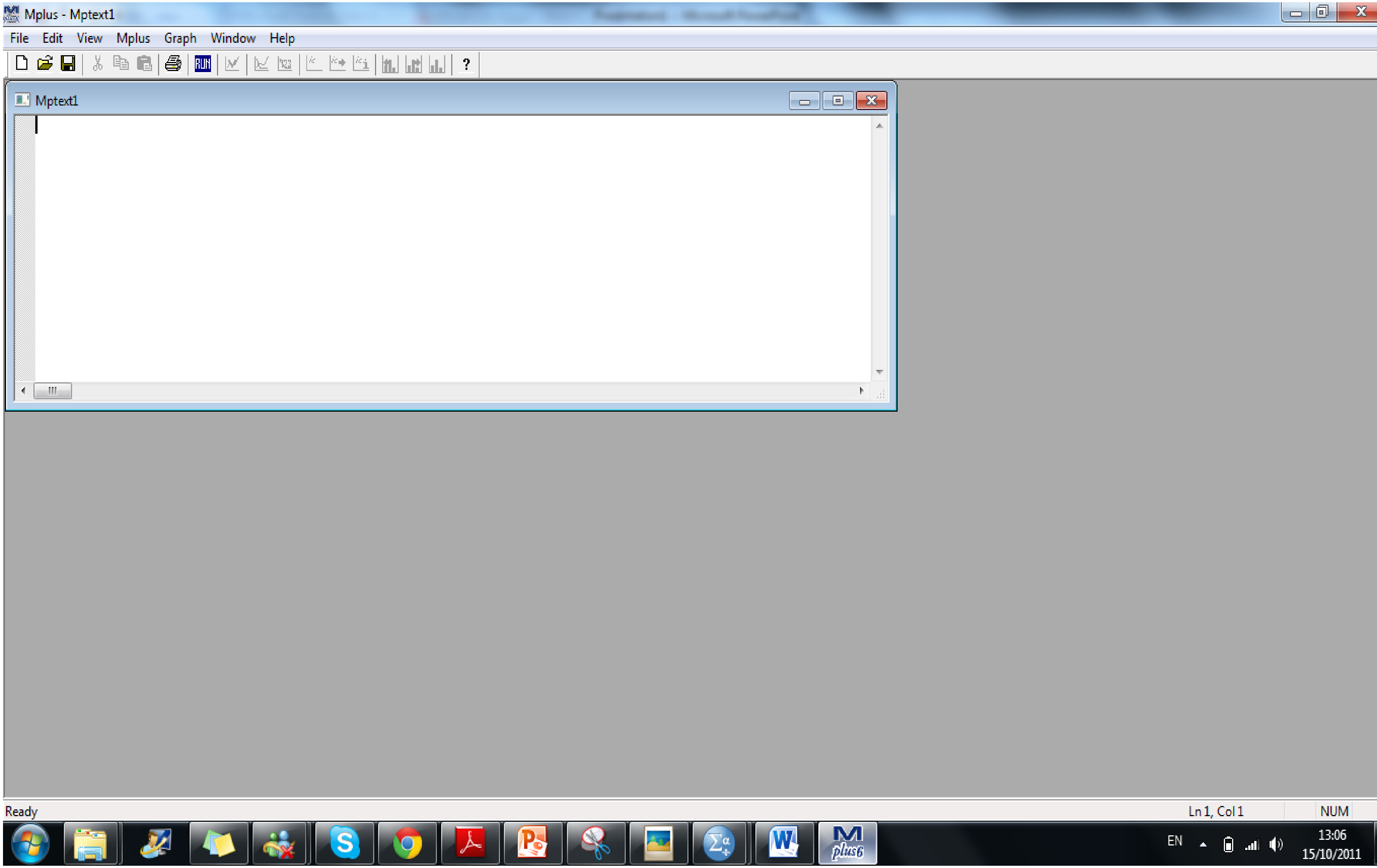
alcohol symptoms NESARC - Notepad

File Edit Format View Help

3618	10001	109	4270.4888476	24	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0
4066	10001	109	1899.529264	33	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4401	10001	109	2370.1947983	60	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4966	10001	109	3897.0659111	29	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6843	10001	109	6610.4425204	80	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
10528	10001	109	3789.3731736	36	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1
10646	10001	109	3167.2861306	66	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
11222	10001	109	959.70159185	65	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12263	10001	109	3167.2861306	71	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
12463	10001	109	7231.9748457	54	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
13307	10001	109	3428.0580713	72	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
13928	10001	109	7231.9748457	53	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
14717	10001	109	6982.0386479	64	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
15320	10001	109	7402.7605365	33	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
16457	10001	109	3428.0580713	67	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
16552	10001	109	1291.7140343	27	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
16865	10001	109	959.70159185	67	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
17087	10001	109	7965.7483787	41	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
18018	10001	109	4097.6073802	43	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
18529	10001	109	6619.0826265	52	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
18743	10001	109	17306.422802	25	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
18775	10001	109	15367.477506	74	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
20655	10001	109	2073.2228431	73	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
21334	10001	109	3607.880222	56	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
2936	10002	108	2070.6193795	30	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3313	10002	108	3167.2861306	73	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
4200	10002	108	6610.4425204	82	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1
4472	10002	108	993.35104061	67	1	1	-9	-9	-9	-9	-9	-9	0	0	0	-9	-9	-9	1
5513	10002	108	8149.190025	59	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
5852	10002	108	4591.1575816	51	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
6773	10002	108	7314.2170633	64	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
7071	10002	108	5841.9154731	23	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7194	10002	108	8657.8143906	18	1	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
8451	10002	108	7022.4605547	37	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
9551	10002	108	9123.2225829	18	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
9943	10002	108	7965.7483787	44	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
10449	10002	108	8133.5658168	28	2	1	0	0	0	0	0	0	1	0	1	0	0	0	0
12116	10002	108	3904.5519967	57	1	2	0	-9	0	0	-9	0	1	-9	1	0	-9	0	1
12969	10002	108	9577.6526417	27	1	2	0	-9	0	0	-9	0	0	-9	0	0	-9	0	0
13258	10002	108	7022.4605547	37	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
13341	10002	108	2746.484805	31	1	2	0	-9	0	0	-9	0	0	-9	0	0	-9	0	0
13555	10002	108	3816.6589123	44	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
13576	10002	108	2037.2961406	28	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
14182	10002	108	875.36425293	28	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
14294	10002	108	17264.891432	43	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
15210	10002	108	7863.6654797	35	1	2	0	-9	0	0	-9	0	0	-9	0	0	-9	0	1
15778	10002	108	3428.0580713	79	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
16652	10002	108	1869.2523709	40	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
16818	10002	108	7994.4573187	41	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1
18076	10002	108	5312.1676054	65	2	3	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
18473	10002	108	7196.1910537	33	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1

EN 17:50 19/10/2011

7. Open Mplus


Mplus - Mptext1
File Edit View Mplus Graph Window Help
Mptext1
Ready
Ln 1, Col 1
NUM
EN
13:06
15/10/2011

Some general notes for writing Mplus syntax

- Any text in Mplus input that is preceded by ! will turn to **GREEN** text and will not be read by Mplus as an input command/subcommand.

- This feature to useful for writing notes for yourself about the analysis you are conducting etc.

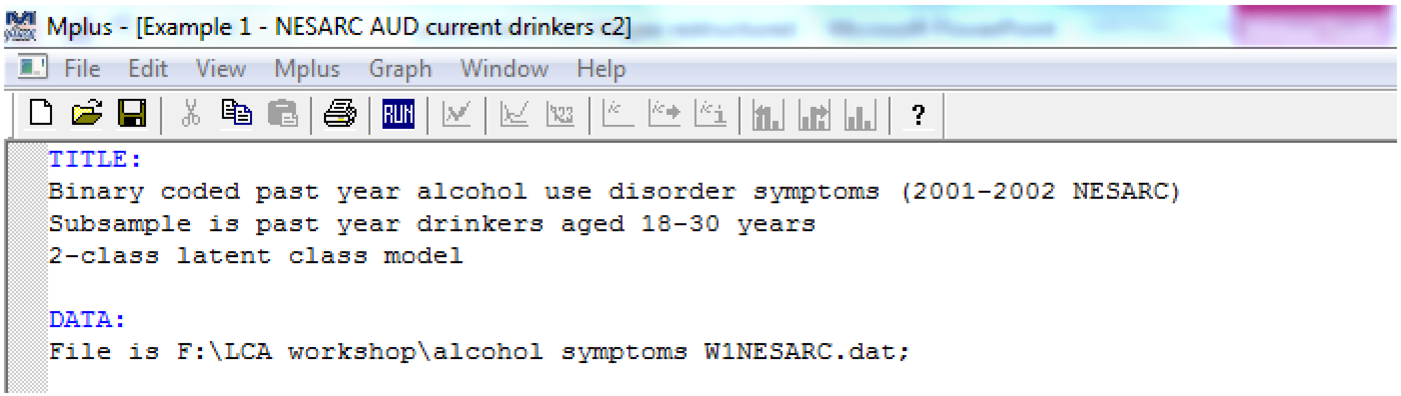
- This feature can also be used to exclude certain pieces of syntax you may not want Mplus to include (examples will be shown during this workshop).

- Every single line of text that you do not want to include as an input instruction must be preceded by an !

The Mplus Input Commands

- Mplus has several command headings that, when typed into an input file and followed by a colon (:), will turn **BLUE** automatically.
- The main command headings are: **TITLE**, **DATA**, **VARIABLES**, **ANALYSIS**, **MODEL**, **SAVEDATA**, **OUTPUT**, and **PLOT.**
- Mplus recognises these commands as key elements of the input to run the programme.
- Generally, all of these command headings are used when running a latent class model.
- Many of the command headings have subcommands, which are used to tell Mplus exactly how to run the model you are specifying.

TITLE and DATA



TITLE: Optional

DATA: Provides information about the data that is to be analysed

VARIABLE

```
VARIABLE:
NAMES are IDNUM PSU STRATUM WEIGHT AGE SEX CONSUMER
S2BQ1A1 S2BQ1B1 S2BQ1A2 S2BQ1B2 S2BQ1A3 S2BQ1B3 S2BQ1A4 S2BQ1B4
S2BQ1A5 S2BQ1B5 S2BQ1A6 S2BQ1B6 S2BQ1A7 S2BQ1B7 S2BQ1A8 S2BQ1B8
S2BQ1A9A S2BQ1B9A S2BQ1A9B S2BQ1B9B S2BQ1A9C S2BQ1B9C
S2BQ1A9D S2BQ1B9D S2BQ1A9E S2BQ1B9E S2BQ1A9F S2BQ1B9F
S2BQ1A9G S2BQ1B9G S2BQ1A9H S2BQ1B9H S2BQ1A9I S2BQ1B9I
S2BQ1A10 S2BQ1B10 S2BQ1A11 S2BQ1B11 S2BQ1A12 S2BQ1B12
S2BQ1A13 S2BQ1B13 S2BQ1A14 S2BQ1B14 S2BQ1A15 S2BQ1B15
S2BQ1A16 S2BQ1B16 S2BQ1A17 S2BQ1B17 S2BQ1A18 S2BQ1B18
S2BQ1A19 S2BQ1B19 S2BQ1A20 S2BQ1B20 S2BQ1A21 S2BQ1B21
S2BQ1A22 S2BQ1B22 S2BQ1A23 S2BQ1B23 S2BQ1A24 S2BQ1B24
S2BQ1A25 S2BQ1B25 S2BQ1A26 S2BQ1B26 S2BQ1A27 S2BQ1B27
S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;

USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
S2BQ1B24 S2BQ1B25 S2BQ1B26 S2BQ1B27;

CATEGORICAL are S2BQ1B1-S2BQ1B27;

IDVARIABLE IS IDNUM;
WEIGHT IS WEIGHT;
CLUSTER IS PSU;
STRATIFICATION IS STRATUM;

CLASSES = c(2);

MISSING are all (-9);

SUBPOPULATION is (CONSUMER EQ 1) AND (AGE LE 30);
```

VARIABLE: provides information about the variables in the data file to be analysed

ANALYSIS and MODEL

```
ANALYSIS:
type = complex mixture;
starts = 50 5;

MODEL:
%Overall%
```

ANALYSIS: describes the technical details of the analysis including the type of analysis and the statistical estimator

MODEL: describes the model to be estimated.

SAVEDATA, OUTPUT, and PLOT

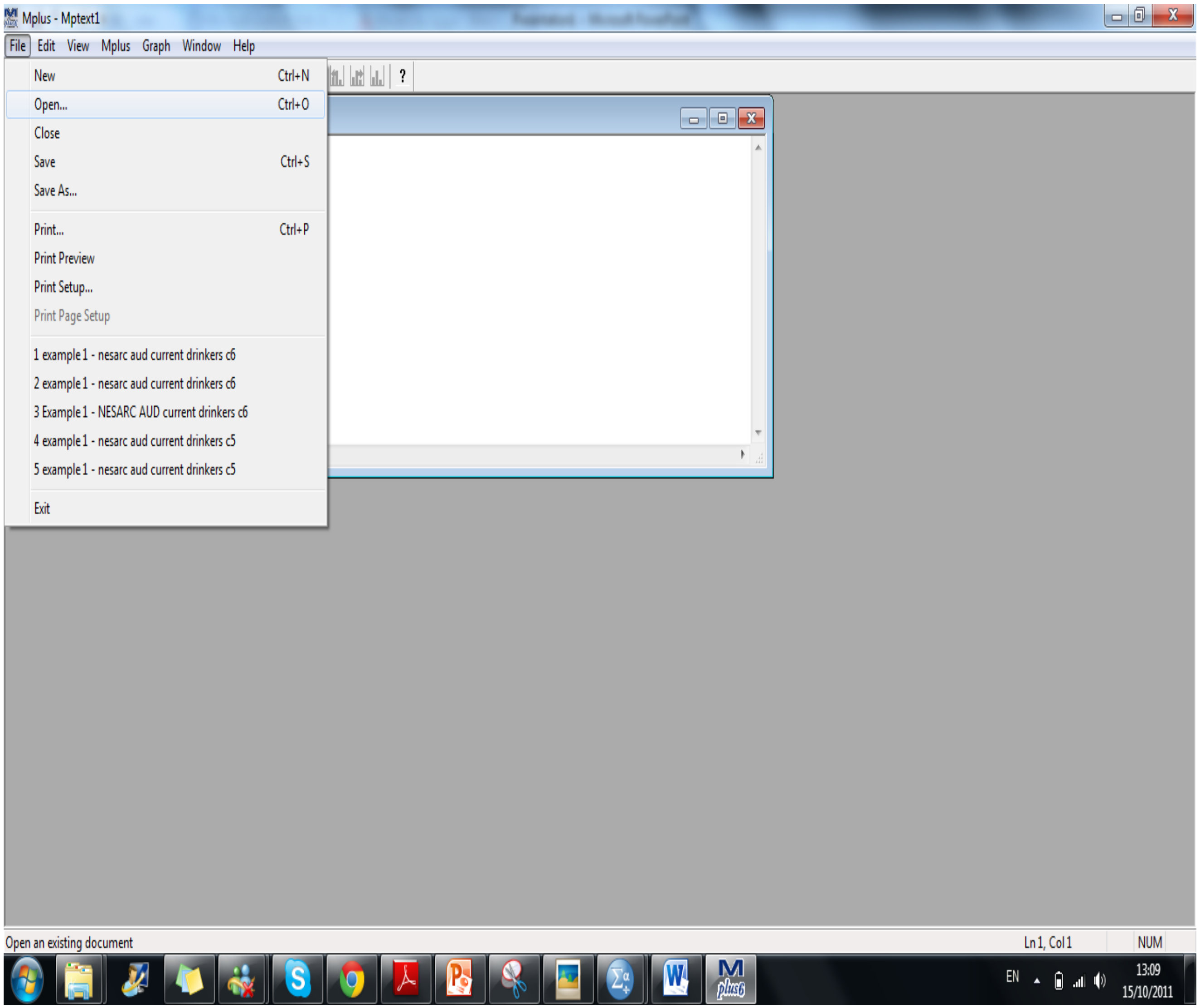
```
SAVEDATA:
File is AUD c2.dat;
Save is cprob;

OUTPUT:
tech10 tech11;

PLOT:
type = plot3;
series = S2BQ1B1-S2BQ1B27(*);
```

SAVEDATA: Used to save the analysis data, auxiliary variables, and a variety of analysis results

OUTPUT: Optional

PLOT: Optional

8. Opening an Mplus Input File

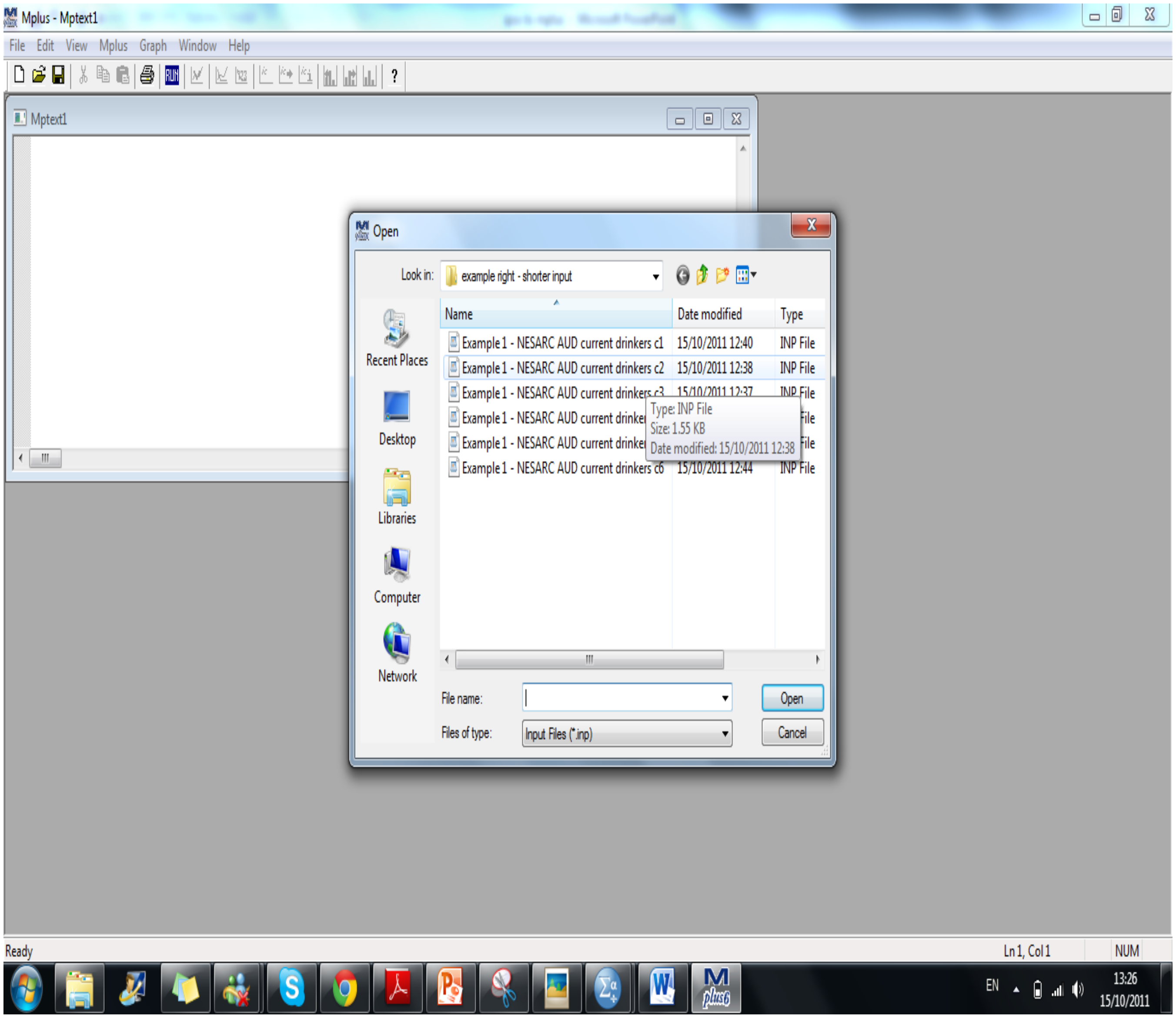


- Under the 'File' tab:

a) Click 'New' if you want to create a new syntax file

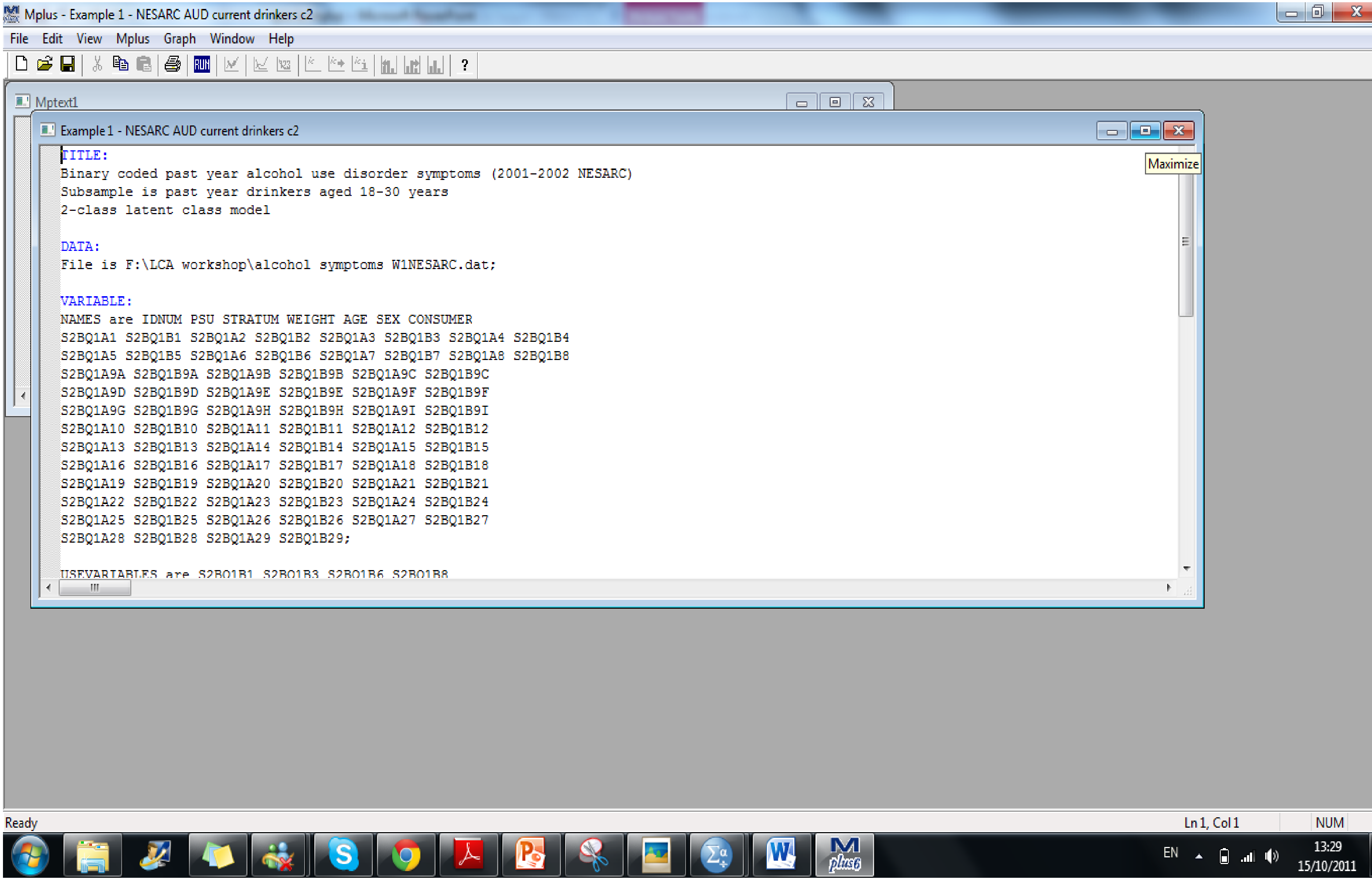
b) Click 'Open' if you want to open a saved syntax file

9. Opening an existing Mplus Input File



- Mplus will open an Input file as the default
- If you want to view an output or graphic file, change the 'Files of type' option

10. Minimized view of Mplus Input file



11. Maximized view of Mplus Input file (part 1)

Mplus - [Example 1 - NESARC AUD current drinkers c2]

File Edit View Mplus Graph Window Help

```
TITLE:
Binary coded past year [illegible] use disorder symptoms (2001-2002 NESARC)
Subsample is past year drinkers aged 18-30 years
2-class latent class model

DATA:
File is F:\LCA workshop\alcohol symptoms W1NESARC.dat;

VARIABLE:
NAMES are IDNUM PSU STRATUM WEIGHT AGE SEX CONSUMER
S2BQ1A1 S2BQ1B1 S2BQ1A2 S2BQ1B2 S2BQ1A3 S2BQ1B3 S2BQ1A4 S2BQ1B4
S2BQ1A5 S2BQ1B5 S2BQ1A6 S2BQ1B6 S2BQ1A7 S2BQ1B7 S2BQ1A8 S2BQ1B8
S2BQ1A9A S2BQ1B9A S2BQ1A9B S2BQ1B9B S2BQ1A9C S2BQ1B9C
S2BQ1A9D S2BQ1B9D S2BQ1A9E S2BQ1B9E S2BQ1A9F S2BQ1B9F
S2BQ1A9G S2BQ1B9G S2BQ1A9H S2BQ1B9H S2BQ1A9I S2BQ1B9I
S2BQ1A10 S2BQ1B10 S2BQ1A11 S2BQ1B11 S2BQ1A12 S2BQ1B12
S2BQ1A13 S2BQ1B13 S2BQ1A14 S2BQ1B14 S2BQ1A15 S2BQ1B15
S2BQ1A16 S2BQ1B16 S2BQ1A17 S2BQ1B17 S2BQ1A18 S2BQ1B18
S2BQ1A19 S2BQ1B19 S2BQ1A20 S2BQ1B20 S2BQ1A21 S2BQ1B21
S2BQ1A22 S2BQ1B22 S2BQ1A23 S2BQ1B23 S2BQ1A24 S2BQ1B24
S2BQ1A25 S2BQ1B25 S2BQ1A26 S2BQ1B26 S2BQ1A27 S2BQ1B27
S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;

USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
S2BQ1B24 S2BQ1B25 S2BQ1B26 S2BQ1B27;

CATEGORICAL are S2BQ1B1-S2BQ1B27;

IDVARIABLE IS IDNUM;
WEIGHT IS WEIGHT;
CLUSTER IS PSU;
STRATIFICATION IS STRATUM;

CLASSES = c(2);

MISSING are all (-9);
```

Run Mplus

Run Mplus Ln 1, Col 1 NUM

EN 16:24 15/10/2011

12. Maximized view of Mplus Input file (part 2)

Mplus - [Example 1 - NESARC AUD current drinkers c2]

File Edit View Mplus Graph Window Help

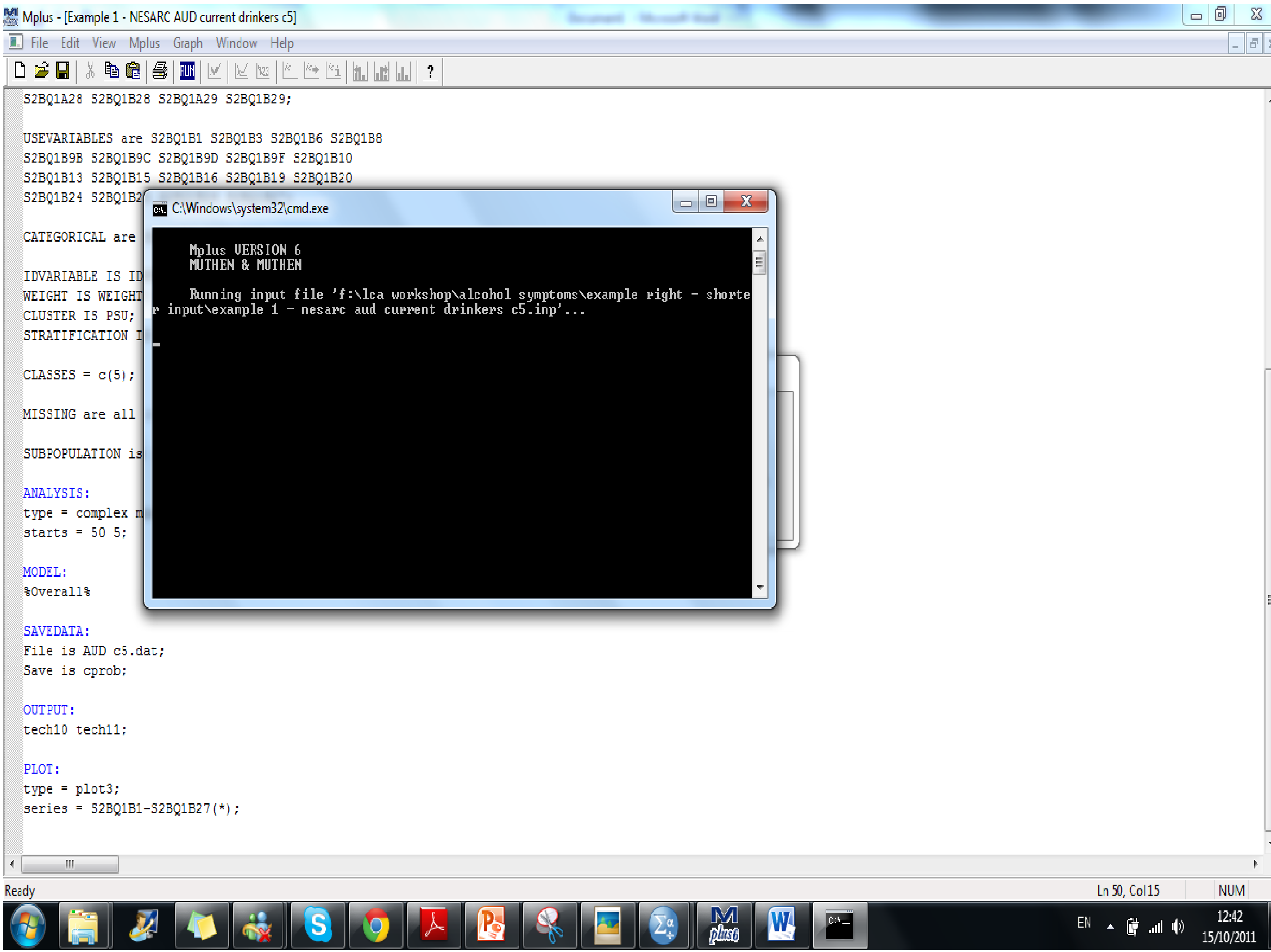
```
S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;

USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
S2BQ1B24 S2BQ1B25 S2BQ1B26 S2BQ1B27;

CATEGORICAL are S2BQ1B1-S2BQ1B27;

IDVARIABLE IS IDNUM;
WEIGHT IS WEIGHT;
CLUSTER IS PSU;
STRATIFICATION IS STRATUM;

CLASSES = c(2);

MISSING are all (-9);

SUBPOPULATION is (CONSUMER EQ 1) AND (AGE LE 30);

ANALYSIS:
type = complex mixture;
starts = 50 5;

MODEL:
%Overall%

SAVEDATA:
File is AUD c2.dat;
Save is cprob;

OUTPUT:
tech10 tech11;

PLOT:
type = plot3;
series = S2BQ1B1-S2BQ1B27(*);
```

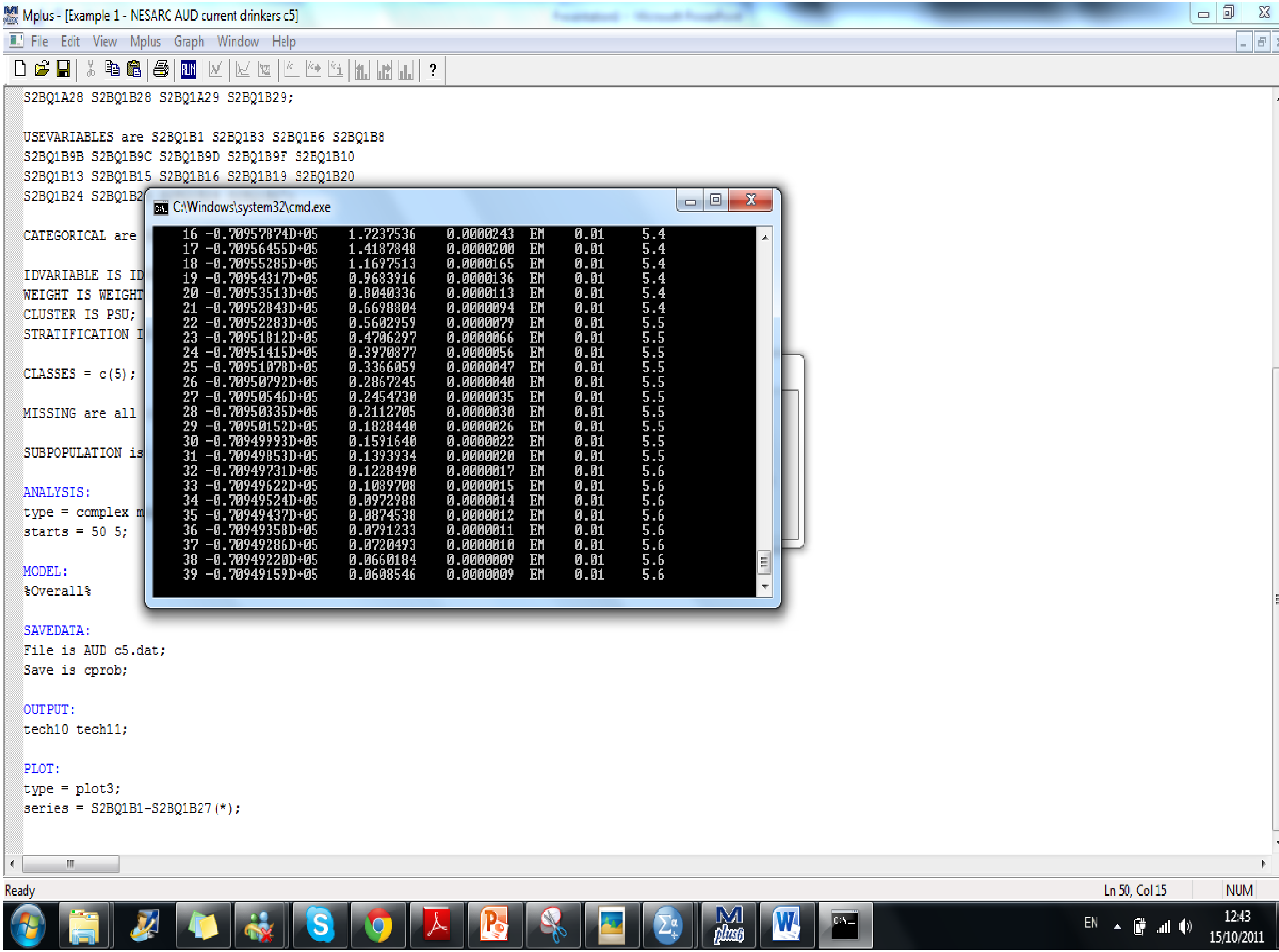
Ready Ln 1, Col 1 NUM

EN 16:33 15/10/2011

Estimating latent class models

- A one-class model - the ‘complete independence’ test
- If a one-class model is a good fit: the observed variables are <u>not</u> interrelated, and therefore no latent variable is needed since there is no relationship between the manifest variables that requires explanation (RARELY happens!)
- Generally 1-6 class models are estimated (maybe more!)


Mplus - [Example 1 - NESARC AUD current drinkers c5]
File Edit View Mplus Graph Window Help
S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;
USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
S2BQ1B24 S2BQ1B2
CATEGORICAL are
IDVARIABLE IS ID
WEIGHT IS WEIGHT
CLUSTER IS PSU;
STRATIFICATION I
CLASSES = c(5);
MISSING are all
SUBPOPULATION is
ANALYSIS:
type = complex m
starts = 50 5;
MODEL:
%Overall%
SAVEDATA:
File is AUD c5.dat;
Save is cprob;
OUTPUT:
tech10 tech11;
PLOT:
type = plot3;
series = S2BQ1B1-S2BQ1B27(*);
C:\Windows\system32\cmd.exe
Mplus VERSION 6
MUTHEN & MUTHEN
Running input file 'f:\lca workshop\alcohol symptoms\example right - shorter input\example 1 - nesarc aud current drinkers c5.inp'...
Ready
Ln 50, Col 15
NUM
EN
12:42
15/10/2011


Mplus - [Example 1 - NESARC AUD current drinkers c5]
File Edit View Mplus Graph Window Help
S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;
USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
S2BQ1B24 S2BQ1B2
CATEGORICAL are
IDVARIABLE IS ID
WEIGHT IS WEIGHT
CLUSTER IS PSU;
STRATIFICATION I
CLASSES = c(5);
MISSING are all
SUBPOPULATION is
ANALYSIS:
type = complex m
starts = 50 5;
MODEL:
%Overall%
SAVEDATA:
File is AUD c5.dat;
Save is cprob;
OUTPUT:
tech10 tech11;
PLOT:
type = plot3;
series = S2BQ1B1-S2BQ1B27(*);
C:\Windows\system32\cmd.exe
16 -0.70957874D+05 1.7237536 0.0000243 EM 0.01 5.4
17 -0.70956455D+05 1.4187848 0.0000200 EM 0.01 5.4
18 -0.70955285D+05 1.1697513 0.0000165 EM 0.01 5.4
19 -0.70954317D+05 0.9683916 0.0000136 EM 0.01 5.4
20 -0.70953513D+05 0.8040336 0.0000113 EM 0.01 5.4
21 -0.70952843D+05 0.6698804 0.0000094 EM 0.01 5.4
22 -0.70952283D+05 0.5602959 0.0000079 EM 0.01 5.5
23 -0.70951812D+05 0.4706297 0.0000066 EM 0.01 5.5
24 -0.70951415D+05 0.3970877 0.0000056 EM 0.01 5.5
25 -0.70951078D+05 0.3366059 0.0000047 EM 0.01 5.5
26 -0.70950792D+05 0.2867245 0.0000040 EM 0.01 5.5
27 -0.70950546D+05 0.2454730 0.0000035 EM 0.01 5.5
28 -0.70950335D+05 0.2112705 0.0000030 EM 0.01 5.5
29 -0.70950152D+05 0.1828440 0.0000026 EM 0.01 5.5
30 -0.70949993D+05 0.1591640 0.0000022 EM 0.01 5.5
31 -0.70949853D+05 0.1393934 0.0000020 EM 0.01 5.5
32 -0.70949731D+05 0.1228490 0.0000017 EM 0.01 5.6
33 -0.70949622D+05 0.1089708 0.0000015 EM 0.01 5.6
34 -0.70949524D+05 0.0972988 0.0000014 EM 0.01 5.6
35 -0.70949437D+05 0.0874538 0.0000012 EM 0.01 5.6
36 -0.70949358D+05 0.0791233 0.0000011 EM 0.01 5.6
37 -0.70949286D+05 0.0720493 0.0000010 EM 0.01 5.6
38 -0.70949220D+05 0.0660184 0.0000009 EM 0.01 5.6
39 -0.70949159D+05 0.0608546 0.0000009 EM 0.01 5.6
Ready
Ln 50, Col 15
NUM
EN
12:43
15/10/2011

Class Enumeration

- Can be a tricky business!!
- Parsimonious model
- Information ('goodness of fit') statistics can be used to evaluate model fit
- Theoretical meaning
- Stability of model

```
Mplus VERSION 6
MUTHEN & MUTHEN
10/15/2011  12:38 PM

INPUT INSTRUCTIONS

  TITLE:
  Binary coded past year alcohol use disorder symptoms (2001-2002 NESARC)
  Subsample is past year drinkers aged 18-30 years
  2-class latent class model

  DATA:
  File is F:\LCA workshop\alcohol symptoms W1NESARC.dat;

  VARIABLE:
  NAMES are IDNUM PSU STRATUM WEIGHT AGE SEX CONSUMER
  S2BQ1A1 S2BQ1B1 S2BQ1A2 S2BQ1B2 S2BQ1A3 S2BQ1B3 S2BQ1A4 S2BQ1B4
  S2BQ1A5 S2BQ1B5 S2BQ1A6 S2BQ1B6 S2BQ1A7 S2BQ1B7 S2BQ1A8 S2BQ1B8
  S2BQ1A9A S2BQ1B9A S2BQ1A9B S2BQ1B9B S2BQ1A9C S2BQ1B9C
  S2BQ1A9D S2BQ1B9D S2BQ1A9E S2BQ1B9E S2BQ1A9F S2BQ1B9F
  S2BQ1A9G S2BQ1B9G S2BQ1A9H S2BQ1B9H S2BQ1A9I S2BQ1B9I
  S2BQ1A10 S2BQ1B10 S2BQ1A11 S2BQ1B11 S2BQ1A12 S2BQ1B12
  S2BQ1A13 S2BQ1B13 S2BQ1A14 S2BQ1B14 S2BQ1A15 S2BQ1B15
  S2BQ1A16 S2BQ1B16 S2BQ1A17 S2BQ1B17 S2BQ1A18 S2BQ1B18
  S2BQ1A19 S2BQ1B19 S2BQ1A20 S2BQ1B20 S2BQ1A21 S2BQ1B21
  S2BQ1A22 S2BQ1B22 S2BQ1A23 S2BQ1B23 S2BQ1A24 S2BQ1B24
  S2BQ1A25 S2BQ1B25 S2BQ1A26 S2BQ1B26 S2BQ1A27 S2BQ1B27
  S2BQ1A28 S2BQ1B28 S2BQ1A29 S2BQ1B29;

  USEVARIABLES are S2BQ1B1 S2BQ1B3 S2BQ1B6 S2BQ1B8
  S2BQ1B9B S2BQ1B9C S2BQ1B9D S2BQ1B9F S2BQ1B10
  S2BQ1B13 S2BQ1B15 S2BQ1B16 S2BQ1B19 S2BQ1B20
  S2BQ1B24 S2BQ1B25 S2BQ1B26 S2BQ1B27;

  CATEGORICAL are S2BQ1B1-S2BQ1B27;

  IDVARIABLE IS IDNUM;
  WEIGHT IS WEIGHT;
  CLUSTER IS PSU;
```

Output begins with a copy of the input

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
  IDVARIABLE IS IDNUM;
  WEIGHT IS WEIGHT;
  CLUSTER IS PSU;
  STRATIFICATION IS STRATUM;

  CLASSES = c(2);

  MISSING are all (-9);

  SUBPOPULATION is (CONSUMER EQ 1) AND (AGE LE 30);

  ANALYSIS:
  type = complex mixture;
  starts = 50 5;

  MODEL:
  %Overall%

  SAVEDATA:
  File is AUD c2.dat;
  Save is cprob;

  OUTPUT:
  tech10 tech11;

  PLOT:
  type = plot3;
  series = S2BQ1B1-S2BQ1B27(*);



*** WARNING
  Data set contains cases with missing on all variables.
  These cases were not included in the analysis.
  Number of cases with missing on all variables:  16240
   1 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS
```

Mplus often issues warnings –
Some are more serious than others!

This warning alerts us that there are cases with missing data on all of the variables in the analysis (we expected this!)

Ready   Ln 4, Col 1   NUM

EN 16:43 15/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Binary coded past year alcohol use disorder symptoms (2001-2002 NESARC)
Subsample is past year drinkers aged 18-30 years
2-class latent class model

SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                        6746

Number of dependent variables                                   18
Number of independent variables                                  0
Number of continuous latent variables                            0
Number of categorical latent variables                           1

Observed dependent variables

  Binary and ordered categorical (ordinal)
   S2BQ1B1     S2BQ1B3     S2BQ1B6     S2BQ1B8     S2BQ1B9B    S2BQ1B9C
   S2BQ1B9D    S2BQ1B9F    S2BQ1B10    S2BQ1B13    S2BQ1B15    S2BQ1B16
   S2BQ1B19    S2BQ1B20    S2BQ1B24    S2BQ1B25    S2BQ1B26    S2BQ1B27

Categorical latent variables
   C

Variables with special functions

  Stratification        STRATUM
  Cluster variable      PSU
  Weight variable       WEIGHT
  ID variable           IDNUM

Estimator                                                      MLR
Information matrix                                        OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations                                 100
  Convergence criterion                                  0.100D-05
Optimization Specifications for the EM Algorithm
  Maximum number of iterations                                 500
```

Repeat of TITLE for analysis

This tells us we have only 1 group and 6746 cases were included in the analysis (*Check this is what you intended*)

18 dependent variables included in the analysis

1 categorical latent variable

These are the survey design variables

Ready Ln 4, Col 1 NUM

EN 16:44 15/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Estimator                                                  MLR
Information matrix                                    OBSERVED
Optimization Specifications for the Quasi-Newton Algorithm for
Continuous Outcomes
  Maximum number of iterations                             100
  Convergence criterion                              0.100D-05
Optimization Specifications for the EM Algorithm
  Maximum number of iterations                             500
  Convergence criteria
    Loglikelihood change                             0.100D-06
    Relative loglikelihood change                    0.100D-06
    Derivative                                       0.100D-05
Optimization Specifications for the M step of the EM Algorithm for
Categorical Latent variables
  Number of M step iterations                                1
  M step convergence criterion                       0.100D-05
  Basis for M step termination                       ITERATION
Optimization Specifications for the M step of the EM Algorithm for
Censored, Binary or Ordered Categorical (Ordinal), Unordered
Categorical (Nominal) and Count Outcomes
  Number of M step iterations                                1
  M step convergence criterion                       0.100D-05
  Basis for M step termination                       ITERATION
  Maximum value for logit thresholds                        15
  Minimum value for logit thresholds                       -15
  Minimum expected cell size for chi-square          0.100D-01
Maximum number of iterations for H1                       2000
Convergence criterion for H1                         0.100D-03
Optimization algorithm                                     EMA
Random Starts Specifications
  Number of initial stage random starts                     50
  Number of final stage optimizations                        5
  Number of initial stage iterations                        10
  Initial stage convergence criterion                0.100D+01
  Random starts scale                                0.500D+01
  Random seed for generating random starts                   0
Link                                                     LOGIT
```

This section of the output outlines the procedure Mplus used to run the analysis

Ready   Ln 4, Col 1   NUM

EN 16:47 15/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Input data file(s)
  F:\LCA workshop\alcohol symptoms W1NESARC.dat
Input data format  FREE



UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

    S2BQ1B1
      Category 1    0.893    23887.930
      Category 2    0.107     2852.054
    S2BQ1B3
      Category 1    0.934    25041.562
      Category 2    0.066     1759.300
    S2BQ1B6
      Category 1    0.975    26154.740
      Category 2    0.025      670.967
    S2BQ1B8
      Category 1    0.853    22847.521
      Category 2    0.147     3931.052
    S2BQ1B9B
      Category 1    0.972    26037.262
      Category 2    0.028      756.480
    S2BQ1B9C
      Category 1    0.973    26047.424
      Category 2    0.027      732.579
    S2BQ1B9D
      Category 1    0.788    21103.145
      Category 2    0.212     5686.684
    S2BQ1B9F
      Category 1    0.944    25243.566
      Category 2    0.056     1502.920
    S2BQ1B10
      Category 1    0.955    25587.010
      Category 2    0.045     1201.210
    S2BQ1B13
      Category 1    0.982    26338.828
      Category 2    0.018      472.070
```

This tells you what data file was analysed and the format of that file

This section tells you the proportion and number of cases falling into each category of the observed variables

Ready   Ln 33, Col 36   NUM

EN   16:40 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
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:

          -73295.292  568859           49
          -73295.292  573096           20
          -73295.292  963053           43
          -73295.292  153942           31
          -73295.292  195873           6



     WARNING:  THE VARIANCE CONTRIBUTION FROM A STRATUM WITH A SINGLE CLUSTER
     (PSU) IS BASED ON THE DIFFERENCE BETWEEN THE SINGLE CLUSTER VALUE AND
     THE OVERALL CLUSTER MEAN.


THE MODEL ESTIMATION TERMINATED NORMALLY



TESTS OF MODEL FIT

Loglikelihood

          H0 Value                        -18413.214
          H0 Scaling Correction Factor         1.569
            for MLR

Information Criteria

          Number of Free Parameters               37
          Akaike (AIC)                     36900.428
          Bayesian (BIC)                   37152.646
          Sample-Size Adjusted BIC         37035.069
            (n* = (n + 2) / 24)
```

Check to ensure that the best loglikelihood value was replicated

Ready Ln 229, Col 1 NUM

EN 16:44 27/10/2011

Global maximum

- When estimating LC models using ML techniques, there are several solutions around which a model can converge (local maxima)
- Only one solution is best (global maximum)
- If model converges around a specific local maximum, instead of global maximum, the best fitting solution can be missed
- Different sets of random starting values important!

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
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:

          -73295.292  568859           49
          -73295.292  573096           20
          -73295.292  963053           43
          -73295.292  153942           31
          -73295.292  195873           6



     WARNING:  THE VARIANCE CONTRIBUTION FROM A STRATUM WITH A SINGLE CLUSTER
     (PSU) IS BASED ON THE DIFFERENCE BETWEEN THE SINGLE CLUSTER VALUE AND
     THE OVERALL CLUSTER MEAN.

THE MODEL ESTIMATION TERMINATED NORMALLY


TESTS OF MODEL FIT

Loglikelihood

          H0 Value                      -18413.214
          H0 Scaling Correction Factor       1.569
            for MLR

Information Criteria

          Number of Free Parameters             37
          Akaike (AIC)                   36900.428
          Bayesian (BIC)                 37152.646
          Sample-Size Adjusted BIC       37035.069
            (n* = (n + 2) / 24)
```

Check to ensure that the best loglikelihood value was replicated

This just lets you know that the Mplus was able to estimate the model you specified

This section outlines the fit statistics for the model

Useful for comparing across models

Information criteria – generally reported

Ready   Ln 229, Col 1   NUM

EN 16:44 27/10/2011

Loglikelihood

- The best solution is the solution with the largest loglikelihood

- If the best (highest) loglikelihood value is not replicated in at least two final stage solutions and preferably more, it is possible that a local solution has been reached

Goodness of fit indices

- Akaike Information Criterion (AIC; Akaike, 1974)
- Bayesian Information Criterion (BIC; Schwartz, 1978)
- Sample-size adjusted BIC (SSABIC; Sclove, 1987)
- LC model with the <u>smallest</u> values on these three statistics is considered to be the best fitting model

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes**

          Pearson Chi-Square

          Value                           132555.049
          Degrees of Freedom                  261967
          P-Value                             1.0000

          Likelihood Ratio Chi-Square

          Value                           131810.145
          Degrees of Freedom                  261967
          P-Value                             1.0000

** Of the 5058700 cells in the latent class indicator table, 139
   were deleted in the calculation of chi-square due to extreme values.

Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model

          Pearson Chi-Square

          Value                          2404332.630
          Degrees of Freedom                 4796428
          P-Value                             1.0000

          Likelihood Ratio Chi-Square

          Value                          2404305.714
          Degrees of Freedom                 4796428
          P-Value                             1.0000
```

Generally reported for nested models

This notice is generally common when analysing many observed variables

Chi-square test

- Mplus will output values for:
  - Likelihood Ratio Test (L^2)

- L^2 can be useful particularly for more hypothesis-driven models (i.e. confirmatory LCA)

- Once the number of classes (T) in the latent variable (X) has been established, hypotheses about the values of the conditional probabilities and latent class probabilities can be tested.

Mplus - [example 1 - nesarc aud current drinkers c2]

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

    Latent
   Classes

      1       1140.99118          0.16914
      2       5605.00882          0.83086
```

Start of main output/results for latent class model

```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

      1       1140.99103          0.16914
      2       5605.00897          0.83086
```

Outlines the number of cases in each class (should agree)

```
CLASSIFICATION QUALITY

     Entropy                        0.882
```

Degree of classification accuracy

```
CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

    Latent
   Classes

      1             1031          0.15276
      2             5715          0.84724
```

Extent to which these counts and proportions concur with final class counts depends on classification accuracy

Ready   Ln 229, Col 1   NUM

EN 16:48 27/10/2011

Entropy

- Entropy statistic - ranges from 0 to 1

- Standardised summary measure of the classification accuracy of placing participants into classes based on their model-based posterior probabilities.

- Higher entropy values reflect better classification of individuals (Ramaswany et al., 1993).

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)

           1        2

    1   0.937    0.063
    2   0.031    0.969


MODEL RESULTS

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

Latent Class 1

 Thresholds
    S2BQ1B1$1          0.257      0.097      2.640      0.008
    S2BQ1B3$1          0.659      0.103      6.383      0.000
    S2BQ1B6$1          2.037      0.136     14.939      0.000
    S2BQ1B8$1         -0.487      0.110     -4.439      0.000
    S2BQ1B9B$1         1.697      0.129     13.164      0.000
    S2BQ1B9C$1         1.813      0.127     14.321      0.000
    S2BQ1B9D$1        -0.878      0.098     -8.936      0.000
    S2BQ1B9F$1         0.999      0.097     10.244      0.000
    S2BQ1B10$1         1.335      0.123     10.888      0.000
    S2BQ1B13$1         2.177      0.147     14.837      0.000
    S2BQ1B15$1         2.852      0.186     15.344      0.000
    S2BQ1B16$1         1.884      0.139     13.557      0.000
    S2BQ1B19$1         2.516      0.167     15.079      0.000
    S2BQ1B20$1         2.546      0.193     13.218      0.000
    S2BQ1B24$1         0.735      0.113      6.490      0.000
    S2BQ1B25$1         1.930      0.140     13.795      0.000
    S2BQ1B26$1         1.298      0.107     12.093      0.000
    S2BQ1B27$1         2.202      0.128     17.213      0.000
```

Summary of how well model is classifying people into groups

Detailed model results

Thresholds ($) – conditional model estimates based on class membership – can be difficult to interpret!

Ready Ln 353, Col 1 NUM

EN 16:49 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
Latent Class 2

 Thresholds
    S2BQ1B1$1          3.192      0.114     27.948      0.000
    S2BQ1B3$1          4.645      0.226     20.554      0.000
    S2BQ1B6$1          5.014      0.218     22.965      0.000
    S2BQ1B8$1          2.934      0.113     26.071      0.000
    S2BQ1B9B$1         6.024      0.503     11.971      0.000
    S2BQ1B9C$1         5.431      0.316     17.177      0.000
    S2BQ1B9D$1         2.074      0.088     23.643      0.000
    S2BQ1B9F$1         4.345      0.191     22.789      0.000
    S2BQ1B10$1         4.451      0.164     27.205      0.000
    S2BQ1B13$1         7.674      0.772      9.939      0.000
    S2BQ1B15$1         8.339      0.821     10.158      0.000
    S2BQ1B16$1         7.102      0.810      8.766      0.000
    S2BQ1B19$1        15.000      0.000    999.000    999.000
    S2BQ1B20$1        10.105      3.177      3.181      0.001
    S2BQ1B24$1         4.036      0.153     26.459      0.000
    S2BQ1B25$1         6.406      0.421     15.212      0.000
    S2BQ1B26$1         5.036      0.263     19.166      0.000
    S2BQ1B27$1         5.082      0.261     19.503      0.000

Categorical Latent Variables

 Means
    C#1               -1.592      0.074    -21.644      0.000


RESULTS IN PROBABILITY SCALE

Latent Class 1

 S2BQ1B1
    Category 1         0.564      0.024     23.548      0.000
    Category 2         0.436      0.024     18.210      0.000
 S2BQ1B3
    Category 1         0.659      0.023     28.407      0.000
    Category 2         0.341      0.023     14.698      0.000
 S2BQ1B6
```

Conditional probability that people within LC1 endorse item *'Find usual # of drinks had less effect than before'*

Ready   Ln 352, Col 1   NUM

EN 16:55 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

```
S2BQ1B6
   Category 1        0.885      0.014     63.567      0.000
   Category 2        0.115      0.014      8.291      0.000
S2BQ1B8
   Category 1        0.381      0.026     14.706      0.000
   Category 2        0.619      0.026     23.939      0.000
S2BQ1B9B
   Category 1        0.845      0.017     50.093      0.000
   Category 2        0.155      0.017      9.177      0.000
S2BQ1B9C
   Category 1        0.860      0.015     56.310      0.000
   Category 2        0.140      0.015      9.187      0.000
S2BQ1B9D
   Category 1        0.294      0.020     14.406      0.000
   Category 2        0.706      0.020     34.664      0.000
S2BQ1B9F
   Category 1        0.731      0.019     38.106      0.000
   Category 2        0.269      0.019     14.037      0.000
S2BQ1B10
   Category 1        0.792      0.020     39.148      0.000
   Category 2        0.208      0.020     10.306      0.000
S2BQ1B13
   Category 1        0.898      0.013     66.932      0.000
   Category 2        0.102      0.013      7.588      0.000
S2BQ1B15
   Category 1        0.945      0.010     98.550      0.000
   Category 2        0.055      0.010      5.692      0.000
S2BQ1B16
   Category 1        0.868      0.016     54.541      0.000
   Category 2        0.132      0.016      8.290      0.000
S2BQ1B19
   Category 1        0.925      0.012     80.200      0.000
   Category 2        0.075      0.012      6.476      0.000
S2BQ1B20
   Category 1        0.927      0.013     71.432      0.000
   Category 2        0.073      0.013      5.597      0.000
S2BQ1B24
   Category 1        0.676      0.025     27.240      0.000
   Category 2        0.324      0.025     13.059      0.000
```

Conditional probability that people in LC1 endorse ‘*Give up or cut down pleasurable activities to drink*’

Ready   Ln 352, Col 1   NUM

EN   16:57 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

```
 S2BQ1B25
    Category 1         0.873      0.015     56.389      0.000
    Category 2         0.127      0.015      8.185      0.000
 S2BQ1B26
    Category 1         0.786      0.018     43.434      0.000
    Category 2         0.214      0.018     11.857      0.000
 S2BQ1B27
    Category 1         0.900      0.011     78.496      0.000
    Category 2         0.100      0.011      8.683      0.000

Latent Class 2

 S2BQ1B1
    Category 1         0.961      0.004    221.778      0.000
    Category 2         0.039      0.004      9.117      0.000
 S2BQ1B3
    Category 1         0.990      0.002    464.817      0.000
    Category 2         0.010      0.002      4.468      0.000
 S2BQ1B6
    Category 1         0.993      0.001    693.944      0.000
    Category 2         0.007      0.001      4.611      0.000
 S2BQ1B8
    Category 1         0.950      0.005    175.976      0.000
    Category 2         0.050      0.005      9.358      0.000
 S2BQ1B9B
    Category 1         0.998      0.001    822.999      0.000
    Category 2         0.002      0.001      1.992      0.046
 S2BQ1B9C
    Category 1         0.996      0.001    725.354      0.000
    Category 2         0.004      0.001      3.177      0.001
 S2BQ1B9D
    Category 1         0.888      0.009    102.115      0.000
    Category 2         0.112      0.009     12.830      0.000
 S2BQ1B9F
    Category 1         0.987      0.002    409.458      0.000
    Category 2         0.013      0.002      5.313      0.000
 S2BQ1B10
    Category 1         0.988      0.002    530.029      0.000
    Category 2         0.012      0.002      6.183      0.000
```

Conditional probability that people in LC2 endorse *‘Feel anxious or nervous when effects of alcohol were wearing off’*

Ready   Ln 352, Col 1   NUM

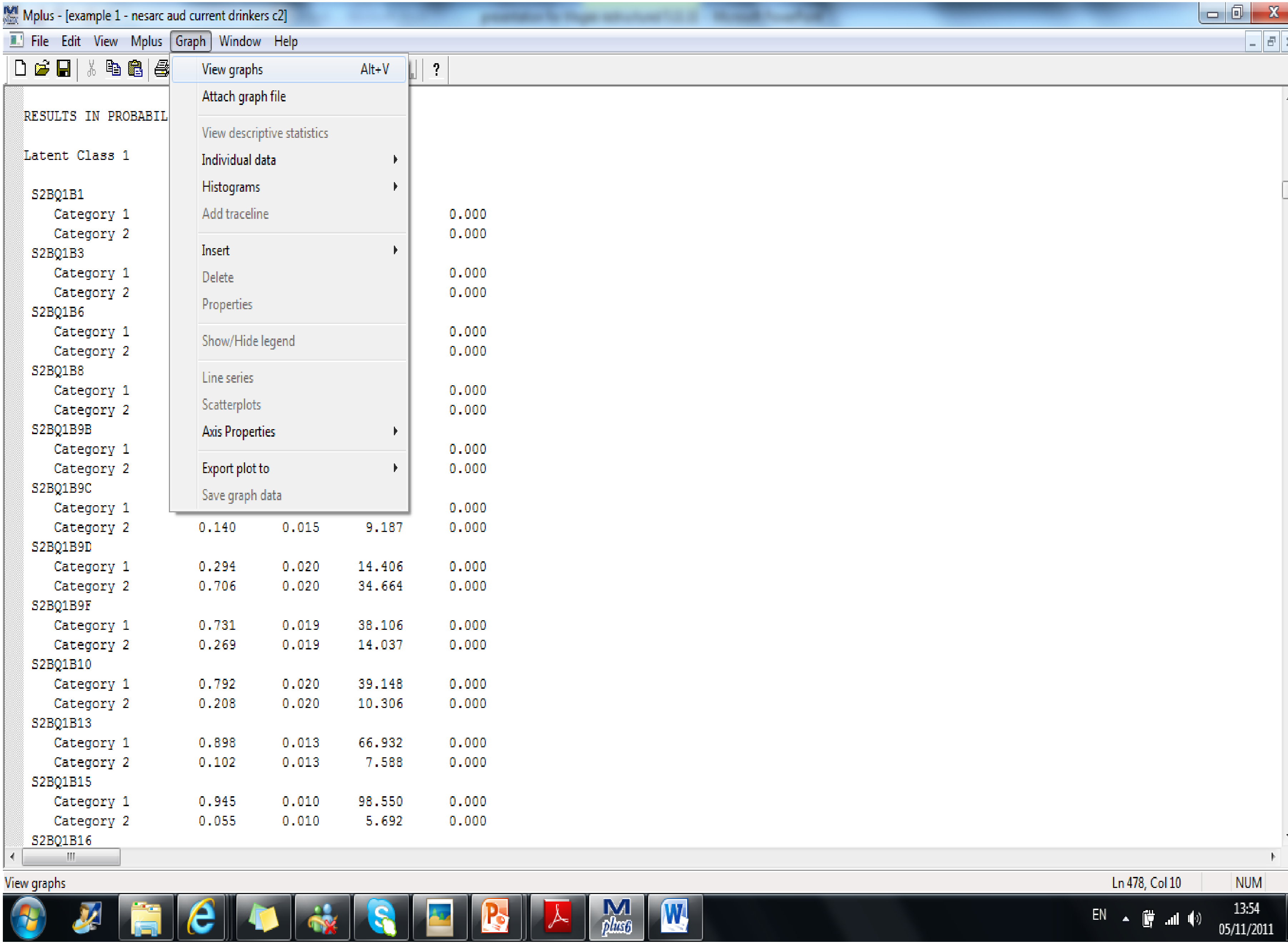
Mplus - [example 1 - nesarc aud current drinkers c2]

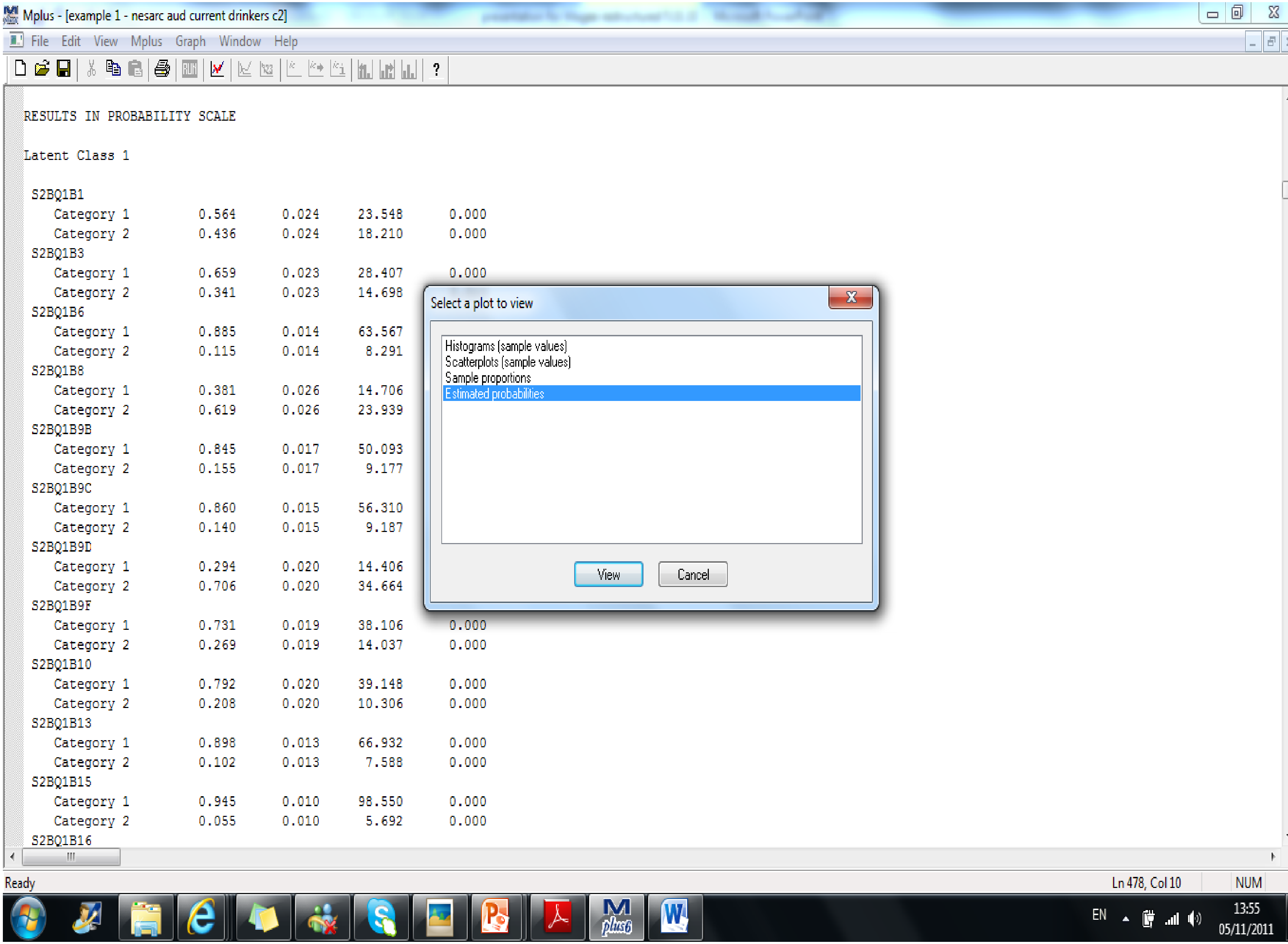
```
 S2BQ1B13
    Category 1         1.000      0.000   2788.331      0.000
    Category 2         0.000      0.000      1.296      0.195
 S2BQ1B15
    Category 1         1.000      0.000   5096.276      0.000
    Category 2         0.000      0.000      1.219      0.223
 S2BQ1B16
    Category 1         0.999      0.001   1499.859      0.000
    Category 2         0.001      0.001      1.235      0.217
 S2BQ1B19
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 S2BQ1B20
    Category 1         1.000      0.000   7700.640      0.000
    Category 2         0.000      0.000      0.315      0.753
 S2BQ1B24
    Category 1         0.983      0.003    377.544      0.000
    Category 2         0.017      0.003      6.672      0.000
 S2BQ1B25
    Category 1         0.998      0.001   1439.851      0.000
    Category 2         0.002      0.001      2.379      0.017
 S2BQ1B26
    Category 1         0.994      0.002    589.235      0.000
    Category 2         0.006      0.002      3.831      0.000
 S2BQ1B27
    Category 1         0.994      0.002    622.220      0.000
    Category 2         0.006      0.002      3.861      0.000


LATENT CLASS ODDS RATIO RESULTS

Latent Class 1 Compared to Latent Class 2

 S2BQ1B1
    Category > 1      18.812      2.294      8.200      0.000
 S2BQ1B3
    Category > 1      53.829     12.238      4.399      0.000
 S2BQ1B6
    Category > 1      19.630      4.947      3.968      0.000
```

Conditional probability of zero for people in LC 2 to endorse- '***Have period when drinking interfered with taking care of home or family***' OR '***Have job/school troubles because of drinking***'

Probability of experiencing alcohol-related problems in the last 12 months across latent classes 1 and 2 (2-class model)

	In the last 12 months, did you…	Class 1	Class 2
S2BQ1B1	Find usual # of drinks had less effect than before	**0.436**	0.039
S2BQ1B3	Drink equivalent of a 1/5 bottle of liquor in one day	**0.341**	0.010
S2BQ1B6	Try unsuccessfully to stop/cut down on drinking more than once	0.115	0.007
S2BQ1B8	Have period when kept drinking longer than intended	**0.619**	0.050
S2BQ1B9B	Shake when effects of alcohol were wearing off	0.155	0.002
S2BQ1B9C	Feel anxious or nervous when effects of alcohol were wearing off	0.140	0.004
S2BQ1B9D	Have nausea when effects of alcohol were wearing off	**0.706**	**0.112**
S2BQ1B9F	Sweat/heart beat fast when effects of alcohol were wearing off	0.269	0.013
S2BQ1B10	Drink or use medicine/drugs to get over bad aftereffects of drinking	0.208	0.012
S2BQ1B13	Spent a lot of time being sick/getting over bad effects of drinking	0.102	0.000
S2BQ1B15	Give up or cut down pleasurable activities to drink	0.055	0.000
S2BQ1B16	Continue to drink though depressed/uninterested/suspicious of others	0.132	0.001
S2BQ1B19	Have period when drinking interfered with taking care of home or family	0.075	0.000
S2BQ1B20	Have job/school troubles because of drinking	0.073	0.000
S2BQ1B24	Get in situations that increased chances of getting hurt while drinking	**0.324**	0.017
S2BQ1B25	Continue to drink despite causing trouble with family or friends	0.127	0.002
S2BQ1B26	Get into physical fights when or right after drinking	0.214	0.006
S2BQ1B27	Get arrested or have other legal problems because of drinking	0.100	0.006


Mplus - [example 1 - nesarc aud current drinkers c2]
File Edit View Mplus Graph Window Help
View graphs Alt+V
Attach graph file
View descriptive statistics
Individual data
Histograms
Add traceline
Insert
Delete
Properties
Show/Hide legend
Line series
Scatterplots
Axis Properties
Export plot to
Save graph data
RESULTS IN PROBABIL
Latent Class 1
S2BQ1B1
Category 1 0.000
Category 2 0.000
S2BQ1B3
Category 1 0.000
Category 2 0.000
S2BQ1B6
Category 1 0.000
Category 2 0.000
S2BQ1B8
Category 1 0.000
Category 2 0.000
S2BQ1B9B
Category 1 0.000
Category 2 0.000
S2BQ1B9C
Category 1 0.000
Category 2 0.140 0.015 9.187 0.000
S2BQ1B9D
Category 1 0.294 0.020 14.406 0.000
Category 2 0.706 0.020 34.664 0.000
S2BQ1B9F
Category 1 0.731 0.019 38.106 0.000
Category 2 0.269 0.019 14.037 0.000
S2BQ1B10
Category 1 0.792 0.020 39.148 0.000
Category 2 0.208 0.020 10.306 0.000
S2BQ1B13
Category 1 0.898 0.013 66.932 0.000
Category 2 0.102 0.013 7.588 0.000
S2BQ1B15
Category 1 0.945 0.010 98.550 0.000
Category 2 0.055 0.010 5.692 0.000
S2BQ1B16
View graphs
Ln 478, Col 10
NUM
EN
13:54
05/11/2011


Mplus - [example 1 - nesarc aud current drinkers c2]
File Edit View Mplus Graph Window Help
RESULTS IN PROBABILITY SCALE
Latent Class 1
S2BQ1B1
Category 1 0.564 0.024 23.548 0.000
Category 2 0.436 0.024 18.210 0.000
S2BQ1B3
Category 1 0.659 0.023 28.407 0.000
Category 2 0.341 0.023 14.698
S2BQ1B6
Category 1 0.885 0.014 63.567
Category 2 0.115 0.014 8.291
S2BQ1B8
Category 1 0.381 0.026 14.706
Category 2 0.619 0.026 23.939
S2BQ1B9B
Category 1 0.845 0.017 50.093
Category 2 0.155 0.017 9.177
S2BQ1B9C
Category 1 0.860 0.015 56.310
Category 2 0.140 0.015 9.187
S2BQ1B9D
Category 1 0.294 0.020 14.406
Category 2 0.706 0.020 34.664
S2BQ1B9F
Category 1 0.731 0.019 38.106 0.000
Category 2 0.269 0.019 14.037 0.000
S2BQ1B10
Category 1 0.792 0.020 39.148 0.000
Category 2 0.208 0.020 10.306 0.000
S2BQ1B13
Category 1 0.898 0.013 66.932 0.000
Category 2 0.102 0.013 7.588 0.000
S2BQ1B15
Category 1 0.945 0.010 98.550 0.000
Category 2 0.055 0.010 5.692 0.000
S2BQ1B16
Select a plot to view
Histograms (sample values)
Scatterplots (sample values)
Sample proportions
Estimated probabilities
View
Cancel
Ready
Ln 478, Col 10
NUM
EN
13:55
05/11/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
RESULTS IN PROBABILITY SCALE

Latent Class 1

 S2BQ1B1
    Category 1         0.564      0.024     23.548      0.000
    Category 2         0.436      0.024     18.210      0.000
 S2BQ1B3
    Category 1         0.659      0.023     28.407      0.000
    Category 2         0.341      0.023     14.698      0.000
 S2BQ1B6
    Category 1         0.885      0.014     63.567      0.000
    Category 2         0.115      0.014      8.291      0.000
 S2BQ1B8
    Category 1         0.381      0.026     14.706      0.000
    Category 2         0.619      0.026     23.939      0.000
 S2BQ1B9B
    Category 1         0.845      0.017     50.093      0.000
    Category 2         0.155      0.017      9.177      0.000
 S2BQ1B9C
    Category 1         0.860      0.015     56.310      0.000
    Category 2         0.140      0.015      9.187      0.000
 S2BQ1B9D
    Category 1         0.294      0.020     14.406      0.000
    Category 2         0.706      0.020     34.664      0.000
 S2BQ1B9F
    Category 1         0.731      0.019     38.106      0.000
    Category 2         0.269      0.019     14.037      0.000
 S2BQ1B10
    Category 1         0.792      0.020     39.148      0.000
    Category 2         0.208      0.020     10.306      0.000
 S2BQ1B13
    Category 1         0.898      0.013     66.932      0.000
    Category 2         0.102      0.013      7.588      0.000
 S2BQ1B15
    Category 1         0.945      0.010     98.550      0.000
    Category 2         0.055      0.010      5.692      0.000
 S2BQ1B16
```

Probabilities

Specify probabilities to plot:

Probability of one category: 2

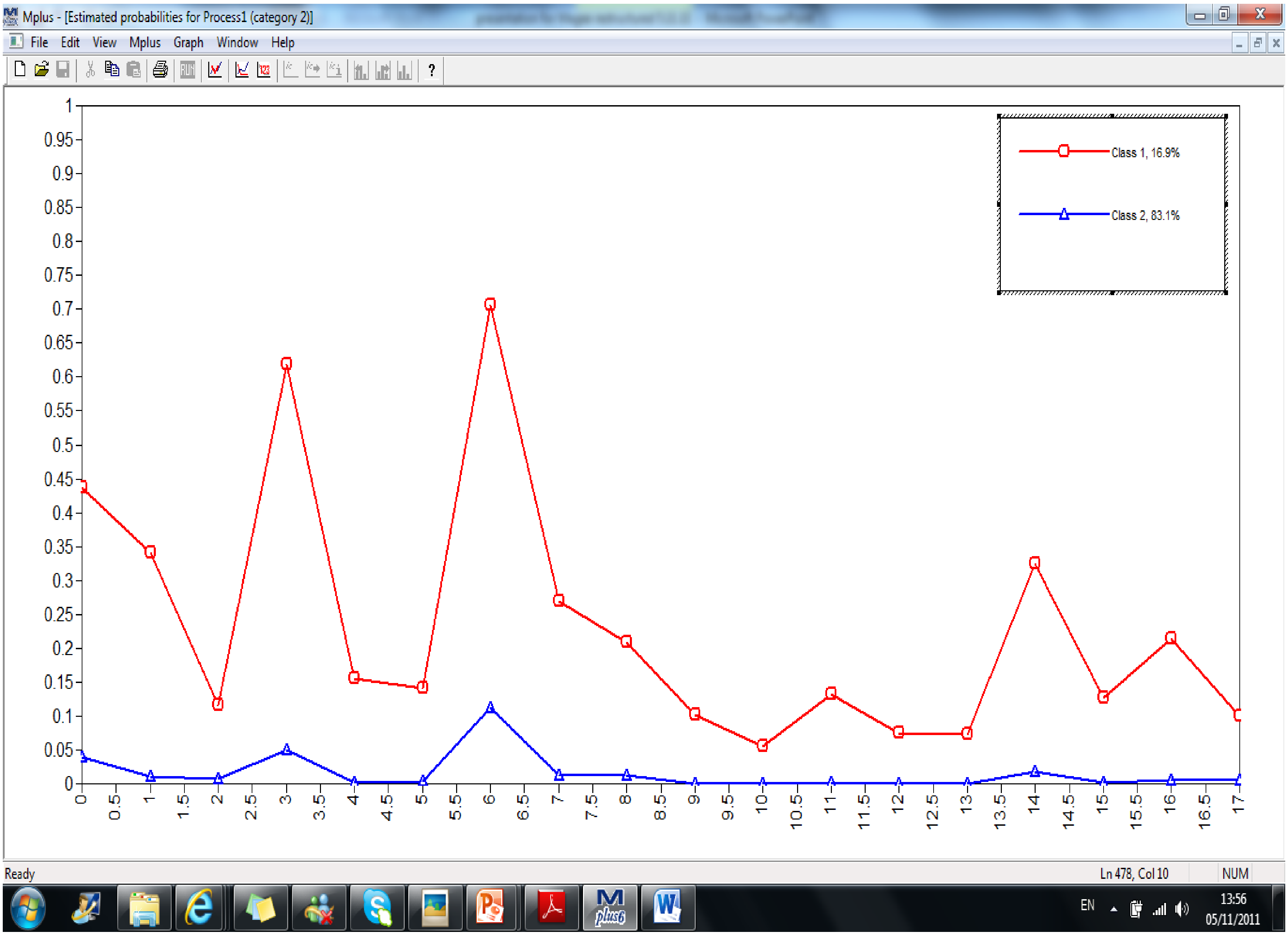
Sum of probabilities:

From category 1 to category 1

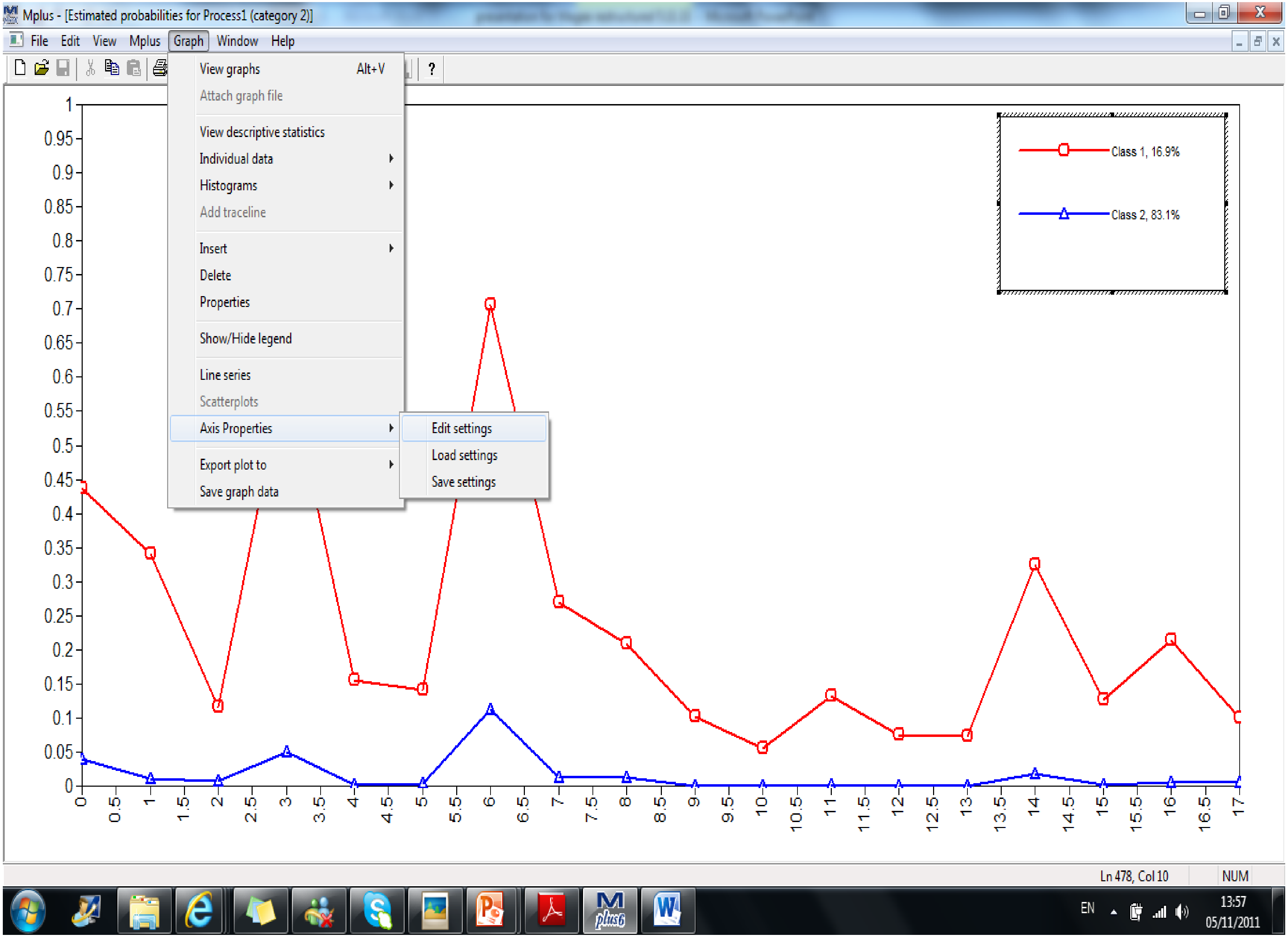
Cancel OK

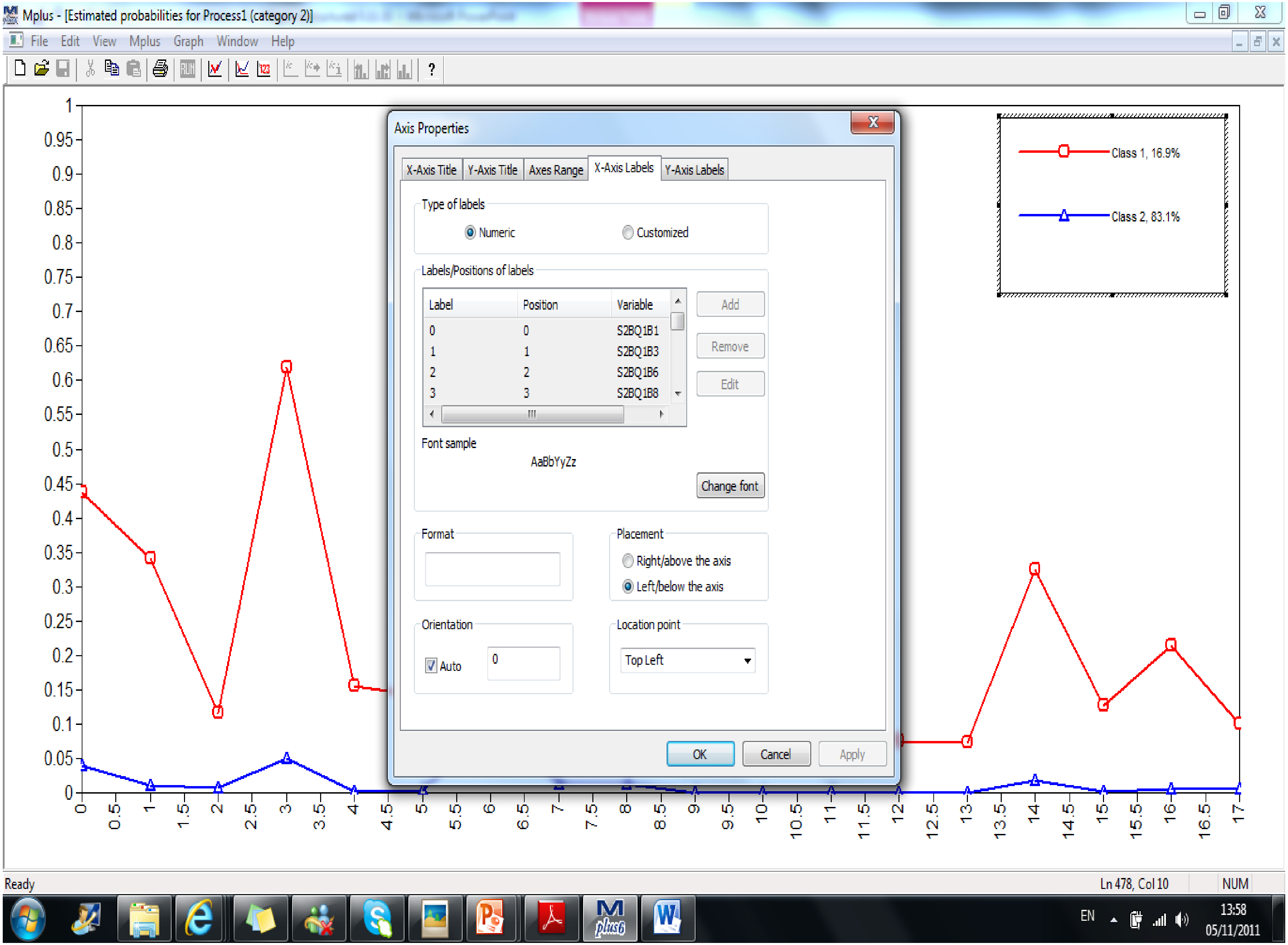
Ready Ln 478, Col 10 NUM

EN 13:55 05/11/2011

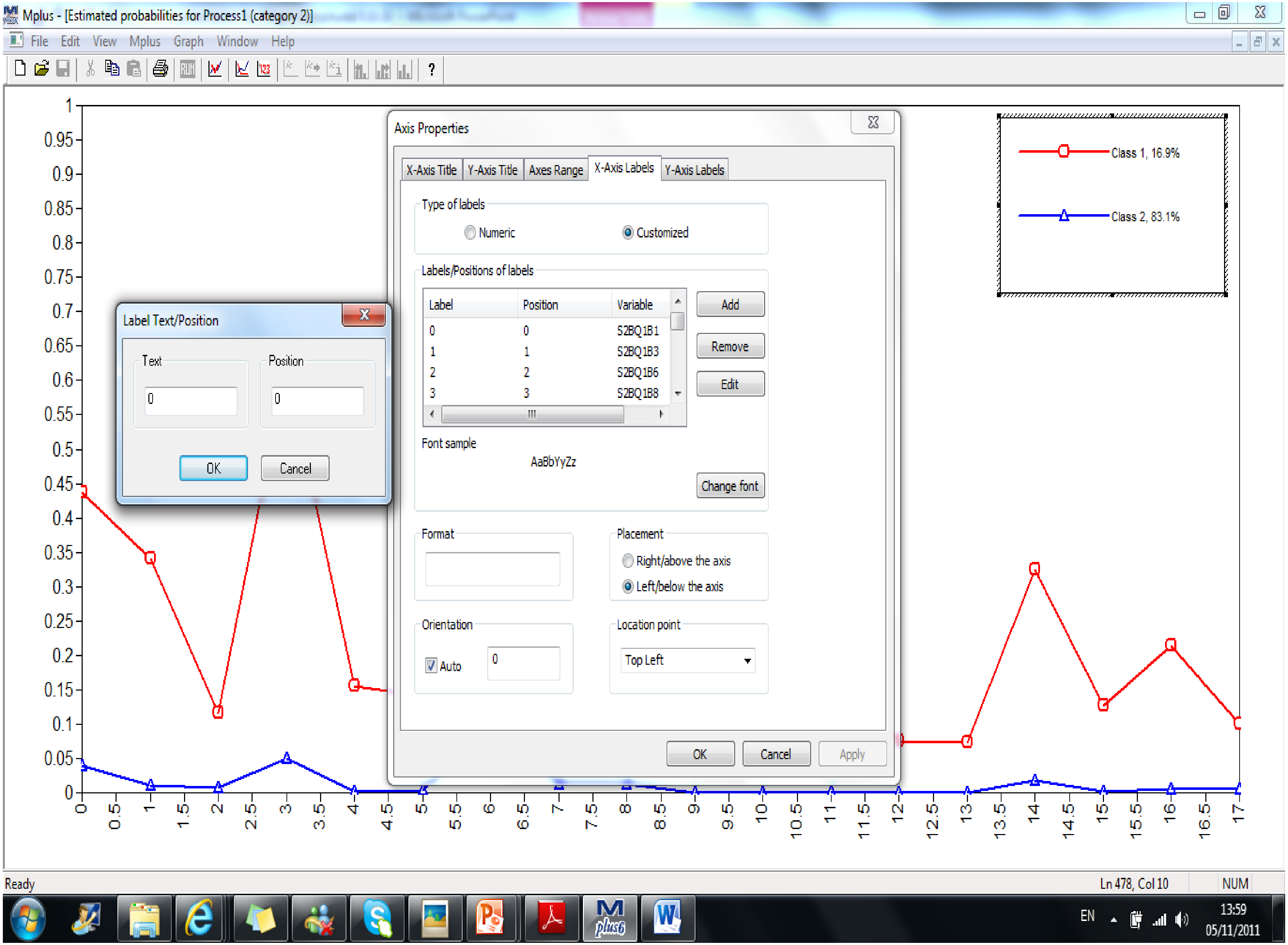


Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
Class 1, 16.9%
Class 2, 83.1%
Ready
Ln 478, Col 10
NUM
EN
13:56
05/11/2011

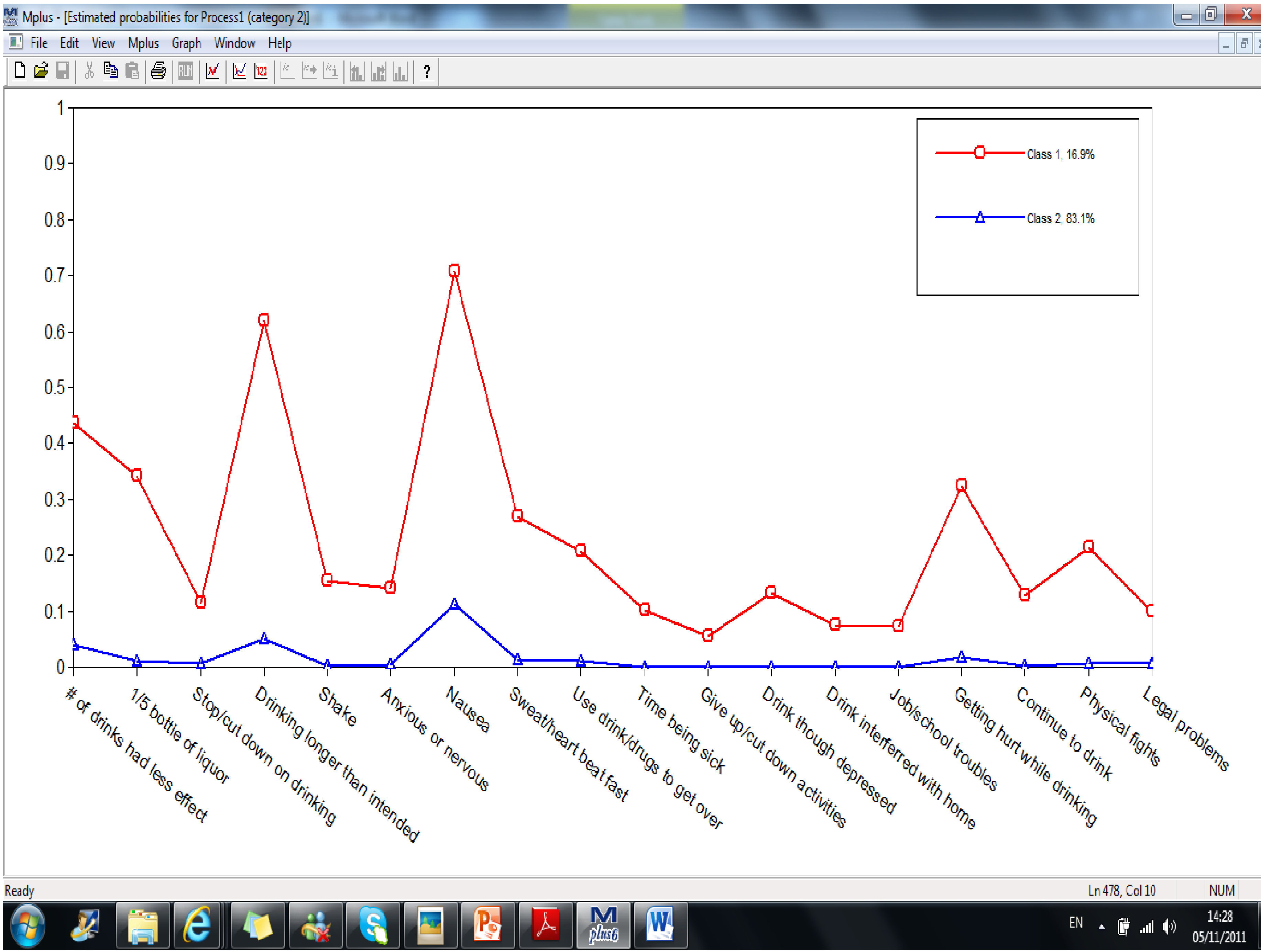

Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
View graphs Alt+V
Attach graph file
View descriptive statistics
Individual data
Histograms
Add traceline
Insert
Delete
Properties
Show/Hide legend
Line series
Scatterplots
Axis Properties
Export plot to
Save graph data
Edit settings
Load settings
Save settings
Class 1, 16.9%
Class 2, 83.1%
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17
Ln 478, Col 10
NUM
EN
13:57
05/11/2011



Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
Axis Properties
X-Axis Title
Y-Axis Title
Axes Range
X-Axis Labels
Y-Axis Labels
Type of labels
Numeric
Customized
Labels/Positions of labels
Label Position Variable
0 0 S2BQ1B1
1 1 S2BQ1B3
2 2 S2BQ1B6
3 3 S2BQ1B8
Add
Remove
Edit
Font sample
AaBbYyZz
Change font
Format
Placement
Right/above the axis
Left/below the axis
Orientation
Auto
0
Location point
Top Left
OK
Cancel
Apply
Class 1, 16.9%
Class 2, 83.1%
Ready
Ln 478, Col 10
NUM
EN
13:58
05/11/2011



Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
Axis Properties
X-Axis Title
Y-Axis Title
Axes Range
X-Axis Labels
Y-Axis Labels
Type of labels
Numeric
Customized
Labels/Positions of labels
Label Position Variable
0 0 S2BQ1B1
1 1 S2BQ1B3
2 2 S2BQ1B6
3 3 S2BQ1B8
Add
Remove
Edit
Font sample
AaBbYyZz
Change font
Format
Placement
Right/above the axis
Left/below the axis
Orientation
Auto
0
Location point
Top Left
OK
Cancel
Apply
Label Text/Position
Text
Position
0
0
OK
Cancel
Class 1, 16.9%
Class 2, 83.1%
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10 10.5 11 11.5 12 12.5 13 13.5 14 14.5 15 15.5 16 16.5 17
Ready
Ln 478, Col 10
NUM
EN
13:59
05/11/2011



Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Class 1, 16.9%
Class 2, 83.1%
# of drinks had less effect
1/5 bottle of liquor
Stop/cut down on drinking
Drinking longer than intended
Shake
Anxious or nervous
Nausea
Sweat/heart beat fast
Use drink/drugs to get over
Time being sick
Give up/cut down activities
Drink though depressed
Drink interferred with home
Job/school troubles
Getting hurt while drinking
Continue to drink
Physical fights
Legal problems
Ready
Ln 478, Col 10
NUM
EN
14:28
05/11/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
 S2BQ1B13
    Category 1         1.000      0.000   2788.331      0.000
    Category 2         0.000      0.000      1.296      0.195
 S2BQ1B15
    Category 1         1.000      0.000   5096.276      0.000
    Category 2         0.000      0.000      1.219      0.223
 S2BQ1B16
    Category 1         0.999      0.001   1499.859      0.000
    Category 2         0.001      0.001      1.235      0.217
 S2BQ1B19
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 S2BQ1B20
    Category 1         1.000      0.000   7700.640      0.000
    Category 2         0.000      0.000      0.315      0.753
 S2BQ1B24
    Category 1         0.983      0.003    377.544      0.000
    Category 2         0.017      0.003      6.672      0.000
 S2BQ1B25
    Category 1         0.998      0.001   1439.851      0.000
    Category 2         0.002      0.001      2.379      0.017
 S2BQ1B26
    Category 1         0.994      0.002    589.235      0.000
    Category 2         0.006      0.002      3.831      0.000
 S2BQ1B27
    Category 1         0.994      0.002    622.220      0.000
    Category 2         0.006      0.002      3.861      0.000


LATENT CLASS ODDS RATIO RESULTS

Latent Class 1 Compared to Latent Class 2

 S2BQ1B1
    Category > 1      18.812      2.294      8.200      0.000
 S2BQ1B3
    Category > 1      53.829     12.238      4.399      0.000
 S2BQ1B6
    Category > 1      19.630      4.947      3.968      0.000
```

(Output truncated)
Odds ratios expressing the likelihood of members in LC2 experiencing a specific alcohol experience, compared to members of LC1

Ready   Ln 352, Col 1   NUM

EN 16:59 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

```
TECHNICAL 10 OUTPUT


     MODEL FIT INFORMATION FOR THE LATENT CLASS INDICATOR MODEL PART


     RESPONSE PATTERNS

     No.  Pattern               No.  Pattern
       1  000000000000000000      2  100000000000000000
       3  010000000000000000      4  110000000000000000
       5  001000000000000000      6  101000000000000000
       7  011000000000000000      8  111000000000000000
       9  000100000000000000     10  100100000000000000
      11  010100000000000000     12  110100000000000000
      13  001100000000000000     14  101100000000000000
      15  011100000000000000     16  111100000000000000
      17  000010000000000000     18  100010000000000000
      19  010010000000000000     20  001010000000000000
      21  101010000000000000     22  000110000000000000
      23  100110000000000000     24  010110000000000000
      25  110110000000000000     26  001110000000000000
      27  111110000000000000     28  000001000000000000
      29  100001000000000000     30  010001000000000000
      31  001001000000000000     32  000101000000000000
      33  100101000000000000     34  010101000000000000
      35  110101000000000000     36  001101000000000000
      37  111101000000000000     38  000011000000000000
      39  100011000000000000     40  110011000000000000
      41  001011000000000000     42  000111000000000000
      43  001111000000000000     44  000000100000000000
      45  100000100000000000     46  010000100000000000
      47  110000100000000000     48  001000100000000000
      49  101000100000000000     50  011000100000000000
      51  000100100000000000     52  100100100000000000
      53  010100100000000000     54  110100100000000000
      55  001100100000000000     56  101100100000000000
      57  111100100000000000     58  000010100000000000
      59  100010100000000000     60  010010100000000000
```

Tech10 (output truncated here) – helpful to explore # of response patterns

Ready   Ln 545, Col 1   NUM

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

RESPONSE PATTERN FREQUENCIES AND CHI-SQUARE CONTRIBUTIONS

Tech10 (output – truncated)

Response Pattern	Frequency Observed	Frequency Estimated	Standardized Residual (z-score)	Chi-square Contribution Pearson	Chi-square Contribution Loglikelihood	Deleted
1	4289.95	4136.76	2.37	5.67	200.81	
2	136.84	173.41	-1.72	7.70	-67.24	
3	28.13	42.09	-1.32	4.63	-23.31	
4	10.66	3.47	2.36	14.90	23.67	
5	20.90	28.05	-0.83	1.82	-12.82	
6	3.91	1.59	1.12	3.39	6.94	
7	1.24	0.57	0.54	0.78	1.89	
8	1.66	0.25	1.72	7.93	6.23	
9	165.28	227.20	-2.55	16.87	-104.25	
10	21.64	14.81	1.09	3.15	16.10	
11	2.99	5.98	-0.75	1.43	-3.30	
12	5.04	3.08	0.68	1.25	4.86	
13	3.44	2.43	0.39	0.41	2.28	
14	1.55	0.81	0.50	0.66	1.94	
15	0.36	0.52	-0.13	0.05	-0.27	
16	2.18	0.39	1.75	8.23	7.46	
17	4.70	10.84	-1.14	3.48	-7.96	
18	3.75	1.06	1.60	6.82	9.38	
19	0.47	0.53	-0.05	0.01	-0.13	
20	1.68	0.18	2.20	12.91	7.56	
21	0.00	0.09	-0.18	0.09	0.00	
22	1.73	1.90	-0.08	0.01	-0.37	
23	0.00	1.08	-0.64	1.08	0.00	
24	0.00	0.71	-0.52	0.71	0.00	
25	0.60	0.55	0.04	0.00	0.09	
26	0.25	0.18	0.10	0.03	0.15	
27	0.00	0.07	-0.16	0.07	0.00	
28	15.89	18.85	-0.42	0.46	-5.76	
29	0.00	1.32	-0.70	1.32	0.00	
30	0.00	0.56	-0.46	0.56	0.00	
31	0.13	0.22	-0.11	0.03	-0.14	
32	5.93	2.18	1.55	6.45	11.73	
33	0.48	0.98	-0.31	0.25	-0.70	
34	0.00	0.64	-0.49	0.64	0.00	

Ready Ln 545, Col 1 NUM

EN 17:07 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
THE TOTAL PEARSON CHI-SQUARE CONTRIBUTION FROM EMPTY CELLS IS       606.61


UNIVARIATE MODEL FIT INFORMATION

                          Estimated Probabilities
                                             Standardized
Variable             H1            H0          Residual
                                               (z-score)
S2BQ1B1
  Category 1        0.894         0.893          0.012
  Category 2        0.106         0.107         -0.012
  Univariate Pearson Chi-Square                  0.000
  Univariate Log-Likelihood Chi-Square           0.000
S2BQ1B3
  Category 1        0.934         0.934         -0.004
  Category 2        0.066         0.066          0.004
  Univariate Pearson Chi-Square                  0.000
  Univariate Log-Likelihood Chi-Square           0.000
S2BQ1B6
  Category 1        0.975         0.975         -0.001
  Category 2        0.025         0.025          0.001
  Univariate Pearson Chi-Square                  0.000
  Univariate Log-Likelihood Chi-Square           0.000
S2BQ1B8
  Category 1        0.853         0.853          0.005
  Category 2        0.147         0.147         -0.005
  Univariate Pearson Chi-Square                  0.000
  Univariate Log-Likelihood Chi-Square           0.000
S2BQ1B9B
  Category 1        0.972         0.972         -0.003
  Category 2        0.028         0.028          0.003
  Univariate Pearson Chi-Square                  0.000
  Univariate Log-Likelihood Chi-Square           0.000
S2BQ1B9C
  Category 1        0.972         0.973         -0.049
  Category 2        0.028         0.027          0.049
  Univariate Pearson Chi-Square                  0.006
```

(Output truncated)

Ready   Ln 545, Col 1   NUM

EN   17:09 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
BIVARIATE MODEL FIT INFORMATION

                                    Estimated Probabilities
                                                       Standardized
Variable       Variable              H1           H0         Residual
                                                           (z-score)
S2BQ1B1        S2BQ1B3
  Category 1     Category 1        0.858        0.853         0.670
  Category 1     Category 2        0.035        0.040        -1.190
  Category 2     Category 1        0.076        0.081        -0.872
  Category 2     Category 2        0.030        0.025         1.488
  Bivariate Pearson Chi-Square                               11.468
  Bivariate Log-Likelihood Chi-Square                        11.330
S2BQ1B1        S2BQ1B6
  Category 1     Category 1        0.878        0.877         0.130
  Category 1     Category 2        0.015        0.016        -0.307
  Category 2     Category 1        0.097        0.098        -0.144
  Category 2     Category 2        0.010        0.009         0.419
  Bivariate Pearson Chi-Square                                0.768
  Bivariate Log-Likelihood Chi-Square                         0.759
S2BQ1B1        S2BQ1B8
  Category 1     Category 1        0.798        0.794         0.536
  Category 1     Category 2        0.095        0.099        -0.712
  Category 2     Category 1        0.055        0.059        -0.910
  Category 2     Category 2        0.052        0.047         0.994
  Bivariate Pearson Chi-Square                                5.990
  Bivariate Log-Likelihood Chi-Square                         5.988
S2BQ1B1        S2BQ1B9B
  Category 1     Category 1        0.876        0.877        -0.178
  Category 1     Category 2        0.018        0.017         0.486
  Category 2     Category 1        0.096        0.095         0.198
  Category 2     Category 2        0.010        0.011        -0.579
  Bivariate Pearson Chi-Square                                1.615
  Bivariate Log-Likelihood Chi-Square                         1.634
S2BQ1B1        S2BQ1B9C
  Category 1     Category 1        0.876        0.877        -0.052
  Category 1     Category 2        0.017        0.017         0.163
  Category 2     Category 1        0.096        0.096         0.032
  Category 2     Category 2        0.010        0.010        -0.128
```

(Output truncated)

Ready   Ln 545, Col 1   NUM

EN 17:10 27/10/2011

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
TECHNICAL 11 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                  50
        Number of final stage optimizations                     5



     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

          H0 Loglikelihood Value                       -88648.163
          2 Times the Loglikelihood Difference         140469.898
          Difference in the Number of Parameters               19
          Mean                                            768.362
          Standard Deviation                             2265.525
          P-Value                                          0.0000

     LO-MENDELL-RUBIN ADJUSTED LRT TEST

          Value                                        139748.669
          P-Value                                          0.0000


PLOT INFORMATION

The following plots are available:

  Histograms (sample values)
  Scatterplots (sample values)
  Sample proportions
  Estimated probabilities

SAVEDATA INFORMATION

  Order and format of variables

    S2BQ1B1        F10.3
    S2BQ1B3        F10.3
    S2BQ1B6        F10.3
```

Usually reported

If you asked Mplus to save data from the analysis, this information will be presented here

Ready Ln 2542, Col 75 NUM

EN 17:11 27/10/2011

Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR-LRT)

- LMR-LRT (Lo et al., 2001)

- Used to compare models with different number of classes

- A non-significant value suggests that the model with one fewer class is a better explanation of the data

Mplus - [example 1 - nesarc aud current drinkers c2]

File Edit View Mplus Graph Window Help

```
    S2BQ1B15       F10.3
    S2BQ1B16       F10.3
    S2BQ1B19       F10.3
    S2BQ1B20       F10.3
    S2BQ1B24       F10.3
    S2BQ1B25       F10.3
    S2BQ1B26       F10.3
    S2BQ1B27       F10.3
    WEIGHT         F10.3
    IDNUM          I6
    CPROB1         F10.3
    CPROB2         F10.3
    C              F10.3
    STRATUM        I6
    PSU            I6

  Save file
    AUD c2.dat

  Save file format
    19F10.3 I6 3F10.3 2I6

  Save file record length    5000


     Beginning Time:  12:38:44
        Ending Time:  12:39:16
       Elapsed Time:  00:00:32

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066

Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com

Copyright (c) 1998-2010 Muthen & Muthen
```

Ready   Ln 3682, Col 40   NUM

EN   17:12   27/10/2011

Statistical fit indices do not always provide a clear-cut answer!!

Deciding On The Number Of Classes For The ASB Items

Number of Classes	1	2	3	4	5	6
Loglikelihood	-48,168.475	-42,625.653	-41,713.142	-41,007.498	-40,808.314	-40,604.231
# par.	17	35	53	71	89	107
BIC	96,488	85,563	83,898	82,647	82,409	82,161
ABIC		85,452	83,730	82,421	82,126	81,821
AIC	96,370	85,321	83,532	82,157	81,795	81,422
2*LogL k – 1 vs. k #par. diff. = 18		11,085.644	1,825.022	1,411.288	398.368	408.166
TECH14 LRT p-value for k-1		.0000	.0000	.0000	.0000	.0000
TECH11 LRT p-value for k-1	NA	.0000	.0000	.0000	.0000	.0019
Entropy	NA	.838	.743	.742	.741	.723

Source: Muthén & Muthén, 2009

Substantive/Theoretical meaning

- No ‘guidelines’ for how to determine if model fits in with existing theory

- Need to examine the resulting class to see if they fit in with previous research or your own hypotheses

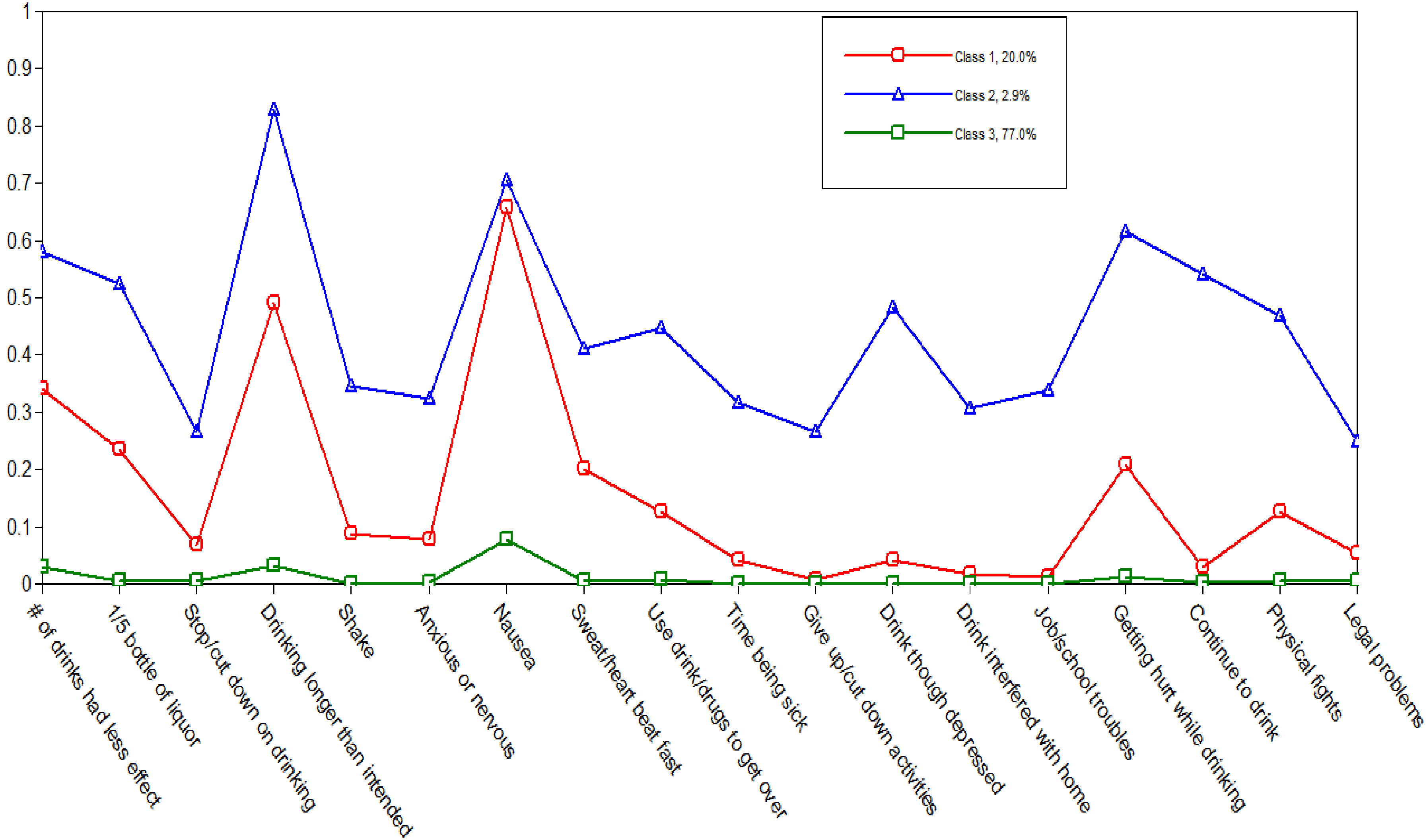
Alcohol experiences: Results

Table 1 Results from latent class analysis of 18 alcohol experiences -
2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions (n=6746)

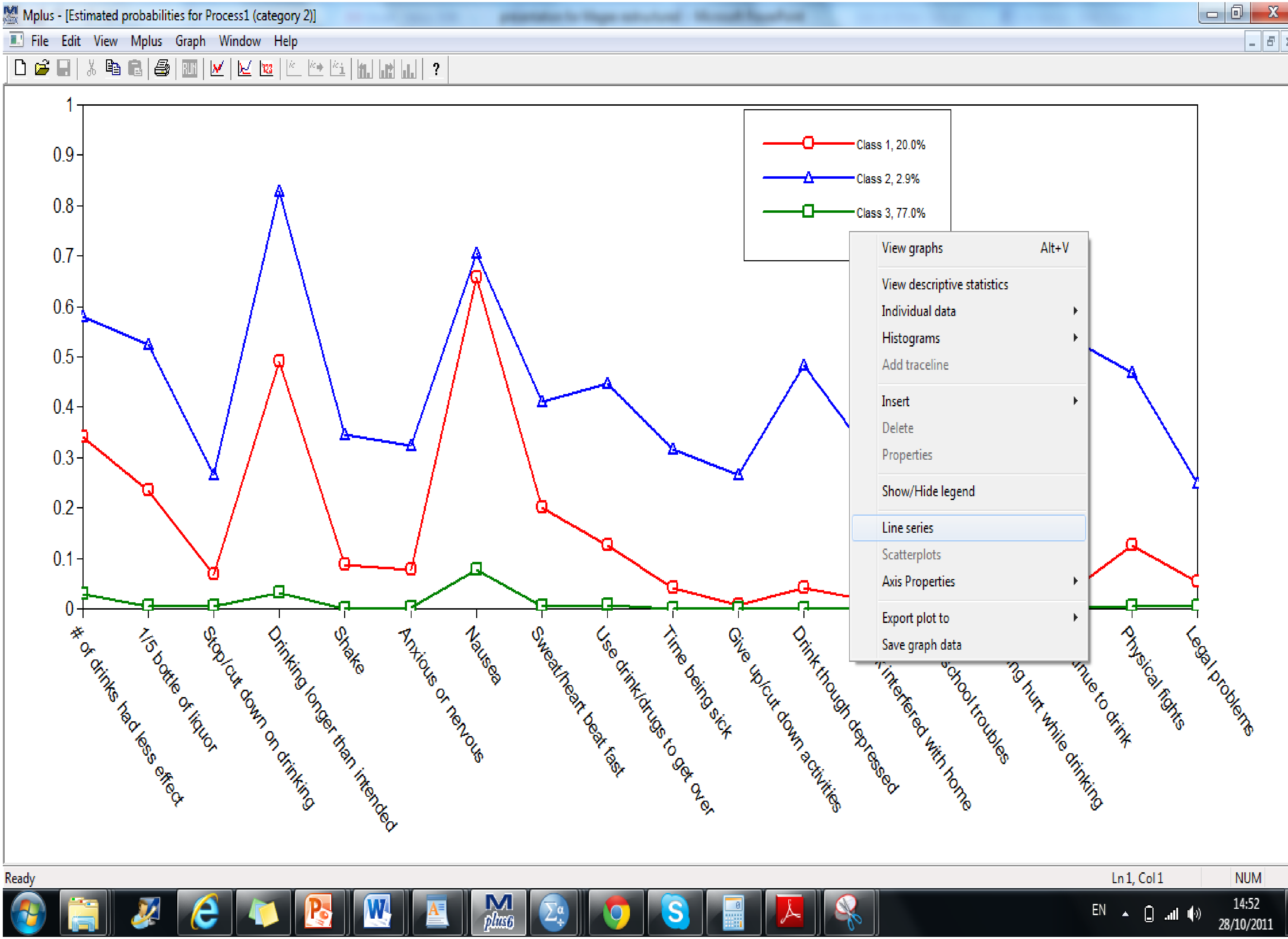
# of LC	Loglikelihood	Best H0 replicated	# para-meters	AIC	BIC	SSABIC	LMR-LRT (*p*)	Entropy
1	-22270.156	Yes	18	44576.313	44699.013	44641.814	NA	NA
2	-18413.214	Yes	37	36900.428	37152.646	37035.069	139748.669 ($p < 0.001$)	0.882
3	-17911.373	Yes	56	35934.746	36299.481	36138.527	110199.111 ($p < 0.001$)	0.865
4	-17823.187	Yes	75	35796.375	36307.627	36069.296	106399.850 ($p = 0.2294$)	0.828
5	-17553.530	Yes	94	35695.060	36335.830	36037.121	105839.995 ($p = 0.0166$)	0.811
6	-17753.530	Yes	113	35626.493	36396.781	36037.694	105394.308 ($p < 0.0001$)	0.802

Probability of experiencing alcohol-related problems in the last 12 months across latent classes 1, 2 and 3 (3-class model)

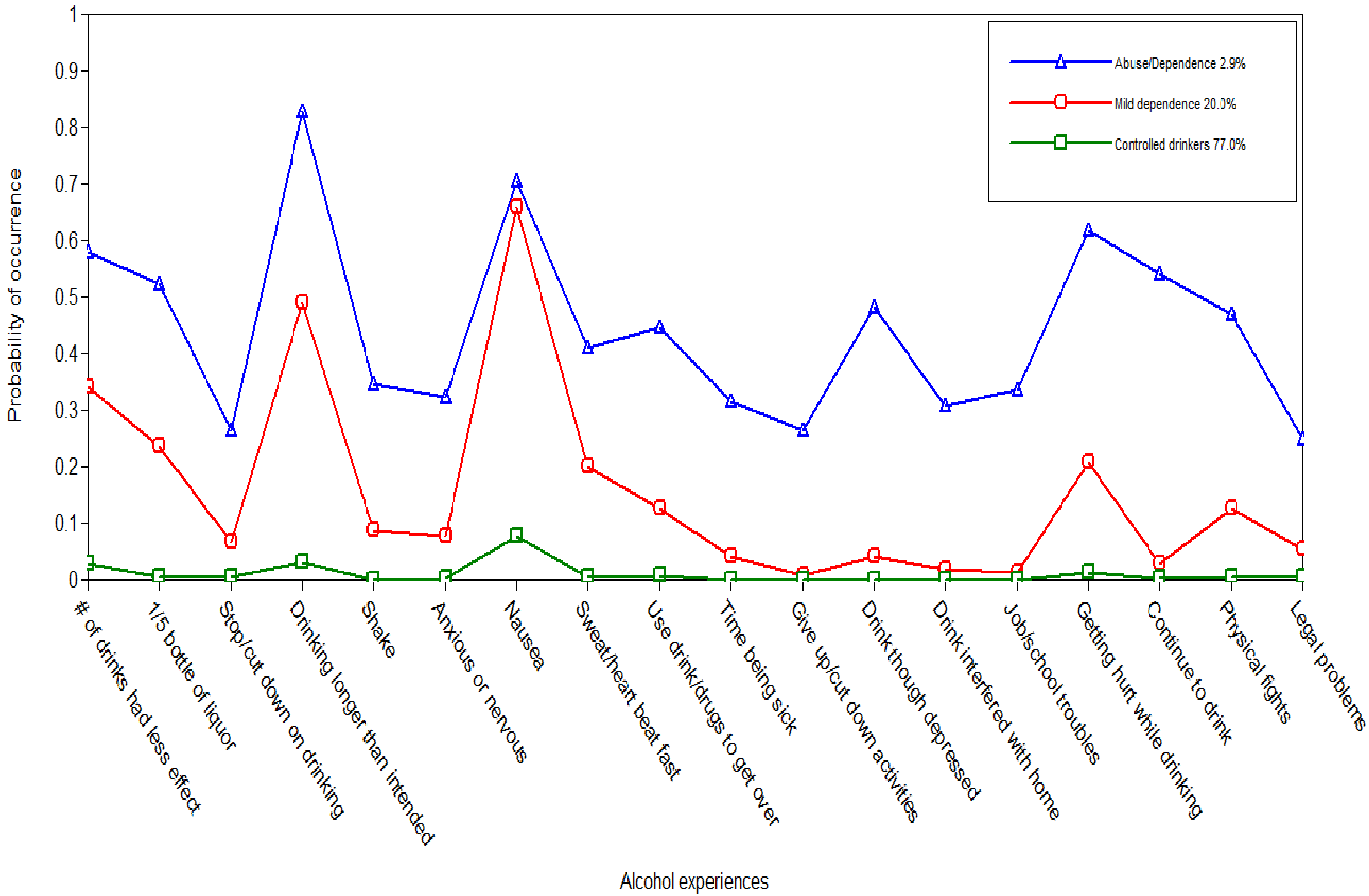
	In the last 12 months, did you…	Class 1	Class 2	Class 3
S2BQ1B1	Find usual # of drinks had less effect than before	0.341	**0.580**	0.028
S2BQ1B3	Drink equivalent of a 1/5 bottle of liquor in one day	0.235	**0.523**	0.004
S2BQ1B6	Try unsuccessfully to stop/cut down on drinking more than once	0.067	0.265	0.005
S2BQ1B8	Have period when kept drinking longer than intended	**0.490**	**0.829**	0.031
S2BQ1B9B	Shake when effects of alcohol were wearing off	0.088	0.346	0.000
S2BQ1B9C	Feel anxious or nervous when effects of alcohol were wearing off	0.078	0.342	0.003
S2BQ1B9D	Have nausea when effects of alcohol were wearing off	**0.658**	**0.706**	0.078
S2BQ1B9F	Sweat/heart beat fast when effects of alcohol were wearing off	0.201	0.410	0.005
S2BQ1B10	Drink or use medicine/drugs to get over bad aftereffects of drinking	0.126	0.447	0.008
S2BQ1B13	Spent a lot of time being sick/getting over bad effects of drinking	0.041	0.316	0.000
S2BQ1B15	Give up or cut down pleasurable activities to drink	0.007	0.265	0.000
S2BQ1B16	Continue to drink though depressed/suspicious of others	0.042	0.483	0.001
S2BQ1B19	Have period when drinking interfered with taking care of home or family	0.018	0.307	0.000
S2BQ1B20	Have job/school troubles because of drinking	0.012	0.337	0.000
S2BQ1B24	Get in situations that increased chances of getting hurt while drinking	0.207	**0.617**	0.013
S2BQ1B25	Continue to drink despite causing trouble with family or friends	0.028	**0.542**	0.002
S2BQ1B26	Get into physical fights when or right after drinking	0.125	0.469	0.004
S2BQ1B27	Get arrested or have other legal problems because of drinking	0.054	0.249	0.005



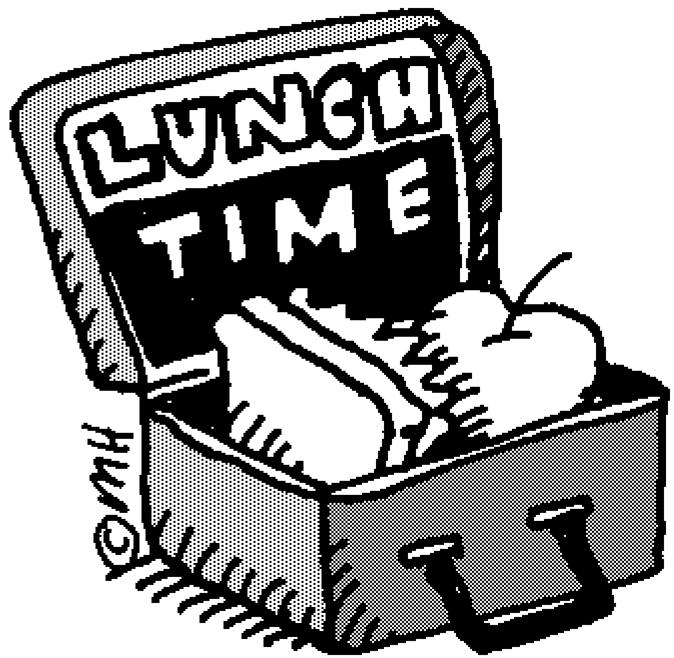
Class 1, 20.0%
Class 2, 2.9%
Class 3, 77.0%
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
# of drinks had less effect
1/5 bottle of liquor
Stop/cut down on drinking
Drinking longer than intended
Shake
Anxious or nervous
Nausea
Sweat/heart beat fast
Use drink/drugs to get over
Time being sick
Give up/cut down activities
Drink though depressed
Drink interfered with home
Job/school troubles
Getting hurt while drinking
Continue to drink
Physical fights
Legal problems



Mplus - [Estimated probabilities for Process1 (category 2)]
File Edit View Mplus Graph Window Help
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Class 1, 20.0%
Class 2, 2.9%
Class 3, 77.0%
View graphs Alt+V
View descriptive statistics
Individual data
Histograms
Add traceline
Insert
Delete
Properties
Show/Hide legend
Line series
Scatterplots
Axis Properties
Export plot to
Save graph data
# of drinks had less effect
1/5 bottle of liquor
Stop/cut down on drinking
Drinking longer than intended
Shake
Anxious or nervous
Nausea
Sweat/heart beat fast
Use drink/drugs to get over
Time being sick
Give up/cut down activities
Drink though depressed
interfered with home
school troubles
Physical fights
Legal problems
Ready
Ln 1, Col 1
NUM
EN
14:52
28/10/2011



Abuse/Dependence 2.9%
Mild dependence 20.0%
Controlled drinkers 77.0%
Probability of occurrence
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
# of drinks had less effect
1/5 bottle of liquor
Stop/cut down on drinking
Drinking longer than intended
Shake
Anxious or nervous
Nausea
Sweat/heart beat fast
Use drink/drugs to get over
Time being sick
Give up/cut down activities
Drink though depressed
Drink interfered with home
Job/school troubles
Getting hurt while drinking
Continue to drink
Physical fights
Legal problems
Alcohol experiences


LUNCH
TIME
©MH

Now it's over to you…

Practical

Stressful life events

- Stress and its impact on general health

- 2007 Adult Psychiatry Morbidity Survey in England (McManus et al., 2009)

SPSS file name is: life events (practical).sav

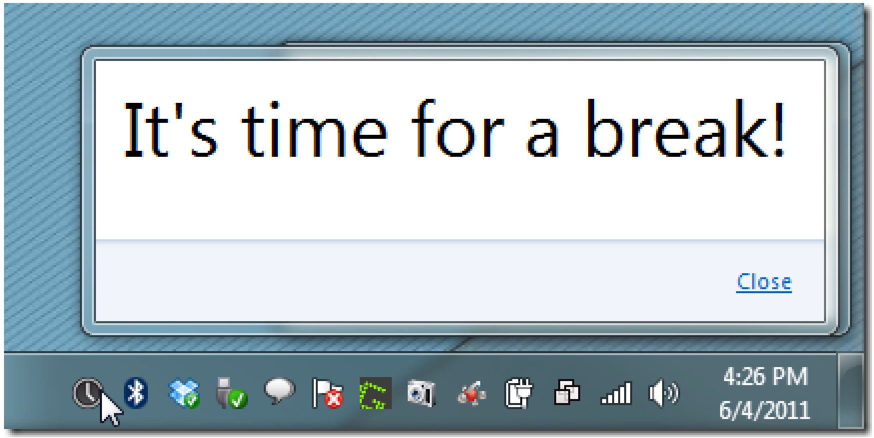
Variable names	**Details**
idnum	Serial number of individual
weight	Weight to use with phase one data
cluster	PSU - use to control for survey design (1)
strata	STRATA - use to control for survey design (2)
sex	Sex of selected respondent
marital	De facto marital status of selected respondent
injury	Serious injury to you or close family member
separate	Separation due to marital difficulties/divorce etc.
sacked	Being made redundant or sacked from your job
police	Problem with police involving court appearance
bullied	Bullying
abused	Sexual abuse

Task (40 minutes)

- Save SPSS file as a .DAT file
- Create an Mplus input file for a latent class model
- Run a series of latent class models, varying the number of latent class from 1-4
- Fill in the details of the goodness of fit indices in the table in the handout
- Compare and contrast resulting models using the goodness of fit indices to decide on the best model
- Create a graph and a table to represent the probability scale for your chosen model

So, how did you get on?

REVIEW OF PRACTICAL SESSION


It's time for a break!
Close
4:26 PM
6/4/2011

EXTENDING THE BASIC LC MODEL

(And some common error messages Mplus likes to issue!)

So far we have learned how to:

- Estimate LC models in an exploratory way – the '*unrestricted*' model (McCutcheon, 1987)

- Using a single population/group

- Local independence assumption upheld

(Life isn't always this simple!)

We might need to extend the basic LC model when…

- We need to relax the local independence assumption
- We have *a priori* hypotheses about the structure of a latent variable (i.e. testing a 'restricted' LC model)
- We have more than one group (e.g. multi-country data)

Relaxing the local independence assumption

Latent Class Analysis and Psychiatry Research

To the Editor:

The statistical procedure, latent class analysis (LCA), has been increasingly applied to problems of psychiatric typology. In the last two years, nearly two dozen major articles making use of LCA have appeared in the psychiatry literature.[1,2] These articles have been of a very high quality, but certain methodological issues have received insufficient attention. Some major concerns are as follows:

- **Conditional dependence**

  An essential assumption of LCA is that of "conditional independence." This requires that all observed variables (e.g., symptoms) be statistically independent (roughly, uncorrelated) within each latent class. Many studies have analyzed symptoms that would appear to violate this assumption a priori. For example, "increased appetite" and "decreased appetite" [2] can scarcely be independent for any group of patients. Such dependent items exert a distorting influence on results. Generally, they promote emergence of extra, spurious latent classes as the estimation algorithm tries to reconcile conditional independence assumptions with the data.

  The problem can be lessened by eliminating clearly dependent items from analysis, or by combining them to form a single item.[1] Simple graphical methods can be used to verify conditional independence.[3] Extensions of LCA exist that accommodate dependent items.[4-5] [for more information, see A Practical Guide to Local Dependence in Latent Class Models.]

- **Local maxima**

  LCA is subject to the problem of "local maxima," where the computer program, trying to find best-fitting values for quantities such as the population base rates of the latent classes, instead converges on values that are not best-fitting; the phenomenon is more common when the number of latent classes exceeds two or three. There is no reason to think such nonoptimal values will be even approximately the same as the true optimal values. Extra computation, such as beginning estimation several times with different initial parameter values, is needed for reasonable assurance that the best solution is found. Some LCA programs do this automatically. The recent articles have mostly not addressed this issue, raising concerns about how well reported results reflect best-fitting solutions.

http://www.john-uebersax.com/stat/lcletter.htm

Relaxing the local independence assumption

- Fewer classes might actually be better?

- Fit may be improved by introducing local dependencies between indicators that have high residuals – as an alternative to creating another class (Magidson & Vermunt, 2000).

- More advanced LC model!

Relaxing the local independence assumption

J R Stat Soc Ser A Stat Soc. 2008 October ; 171(4): 877–897. doi:10.1111/j.1467-985X.2008.00544.x.

Locally dependent latent class models with covariates: an application to under-age drinking in the USA

Beth A. Reboussin, **Edward H. Ip**, and **Mark Wolfson**
Wake Forest University School of Medicine, Winston-Salem, USA

Since the two-class model exhibited a significant amount of local dependences based on our residual diagnostic statistics that are shown in Fig. 1 and the three-class model provided a substantively meaningful classification of under-age drinkers, we explored whether we could improve the fit of the three-class model by relaxing local independence assumptions and thereby avoid introduction of a possibly spurious fourth class. We began by fitting a series of local dependence models for pairs of items identified through residual diagnostics in Fig. 1 as exhibiting some evidence of local dependence under the three-class model. First, we fit the least restrictive model, the unstructured model (5), that allowed the local dependence to be different for each pair of items and to vary by latent class. Because some of the local dependences that were identified through residual diagnostics were not statistically significant under this model, we used a backward elimination procedure to simplify the model. The item pair with the least significant local dependence was removed first and the model was refitted. Each subsequent step removed the least significant local dependence in the model until all the remaining local dependences had p-values that were smaller than 0.05. The final model resulted in statistically significant local dependences between

- **a.** binge drinking and becoming drunk (γ_{23m}),
- **b.** becoming drunk and experiencing headaches (γ_{37m}), and
- **c.** being unable to remember what happened while drinking and social problems (γ_{68m}).

Relaxing the local independence assumption

ORIGINAL ARTICLE

The Structure of Posttraumatic Stress Disorder

Latent Class Analysis in 2 Community Samples

Naomi Breslau, PhD; Beth A. Reboussin, PhD; James C. Anthony, PhD; Carla L. Storr, ScD

Arch Gen Psychiatry. 2005;62:1343-1351

Request Tech10 in Mplus

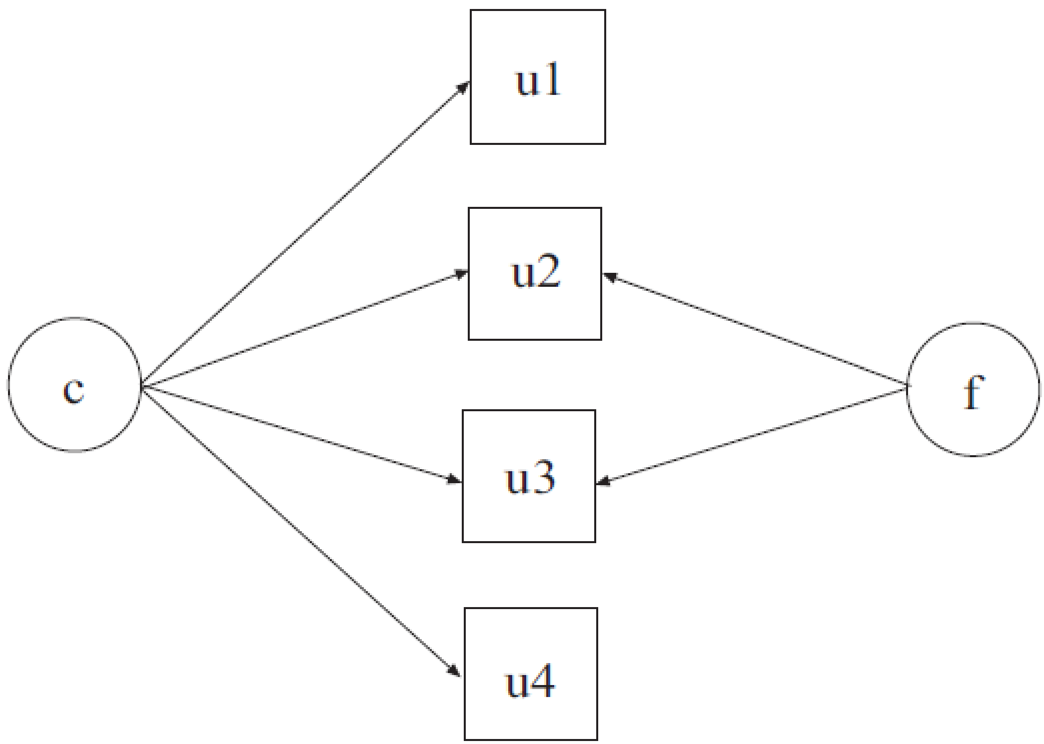
weight of evidence, the better the fit is. Finally, we examined the bivariate residuals between pairs of indicators. In general, bivariate residuals larger than 3.84 identify correlations between the associated variable pairs that have not been adequately explained by the model at $\alpha = .05$.[34]

As an alternative to adding a class to improve model fit, advanced LCA allows for residual interdependence of pairs of indicators by introducing local dependencies (direct effect) in a pair of indicators that have high residuals.[31] We used this technique in this study. Alternative techniques, outlined in Magidson and Vermunt[32] (ie, deleting one in the pair of indicators, combining them into a single "and/or" item, and adding a latent variable), yielded similar results, in terms of the number and size of latent classes and response probabilities, but less adequate model fit (not displayed).

Some researchers use 1.96 as cut-off

Relaxing the local independence assumption

EXAMPLE 7.16: LCA WITH PARTIAL CONDITIONAL INDEPENDENCE

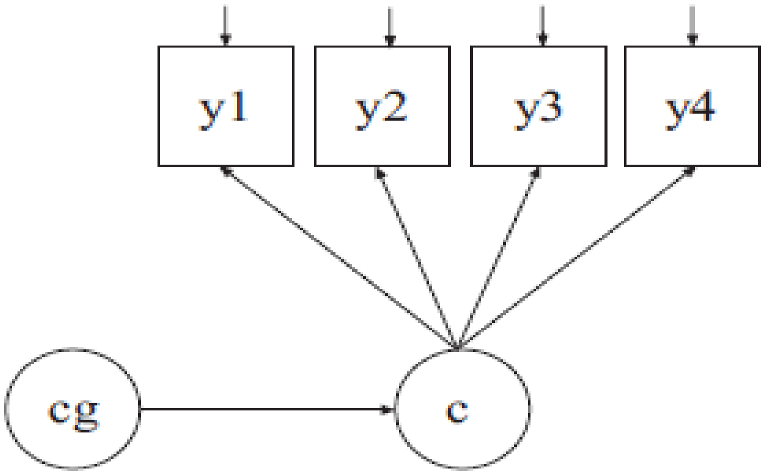


(Source: Muthén & Muthén, 1998-2010)

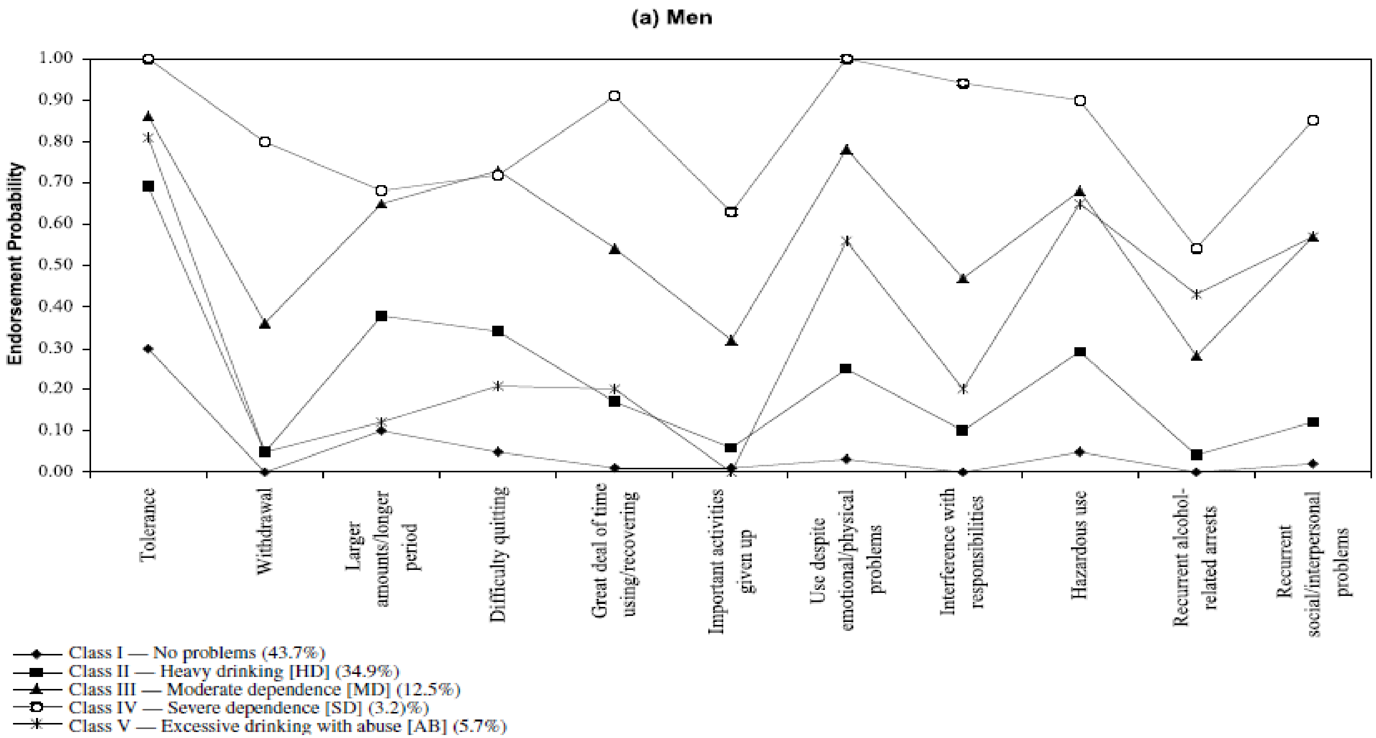
Hypothesis-drive LC models

- Test whether latent structure of a construct varies across subgroups using multiple-group LCA
  - Lynskey et al. (2005) - AUD
  - Chung & Breslau (2006) - PTSD

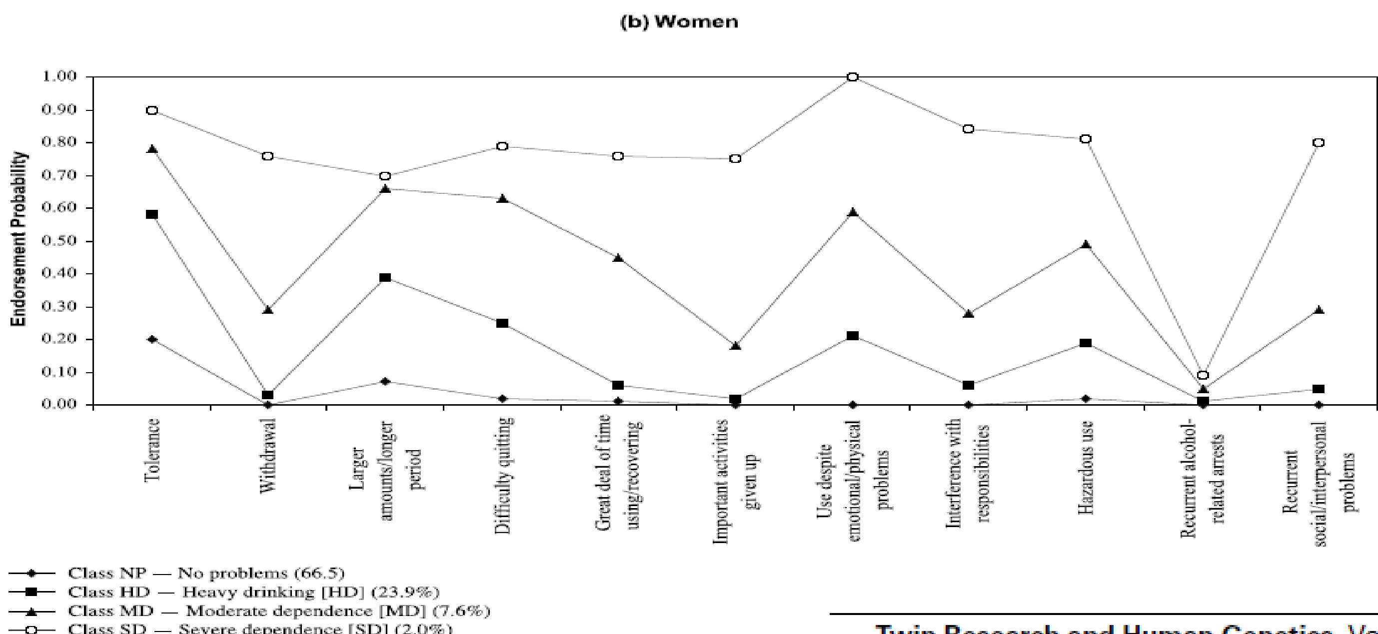
EXAMPLE 7.21: MIXTURE MODELING WITH KNOWN CLASSES (MULTIPLE GROUP ANALYSIS)



(Source: Muthén & Muthén, 1998-2010)



Lynskey et al. (2005)



Twin Research and Human Genetics Volume 8 Number 6 pp. 574–584

LC models with covariates

- Exploring nature of LCs in relation to covariates is common (see Clark & Muthen, 2011)

- Helps to define characteristics of classes (e.g. in terms of demographic factors)

- Two main methods:
  - Regression of most likely class membership on the covariates (e.g. Shevlin et al. 2007; Reboussin & Anthony, 2008)
  - Including the covariates in the analysis while forming the latent classes (e.g. Clark & Muthen, 2011)

Mplus saves information – as outlined in output

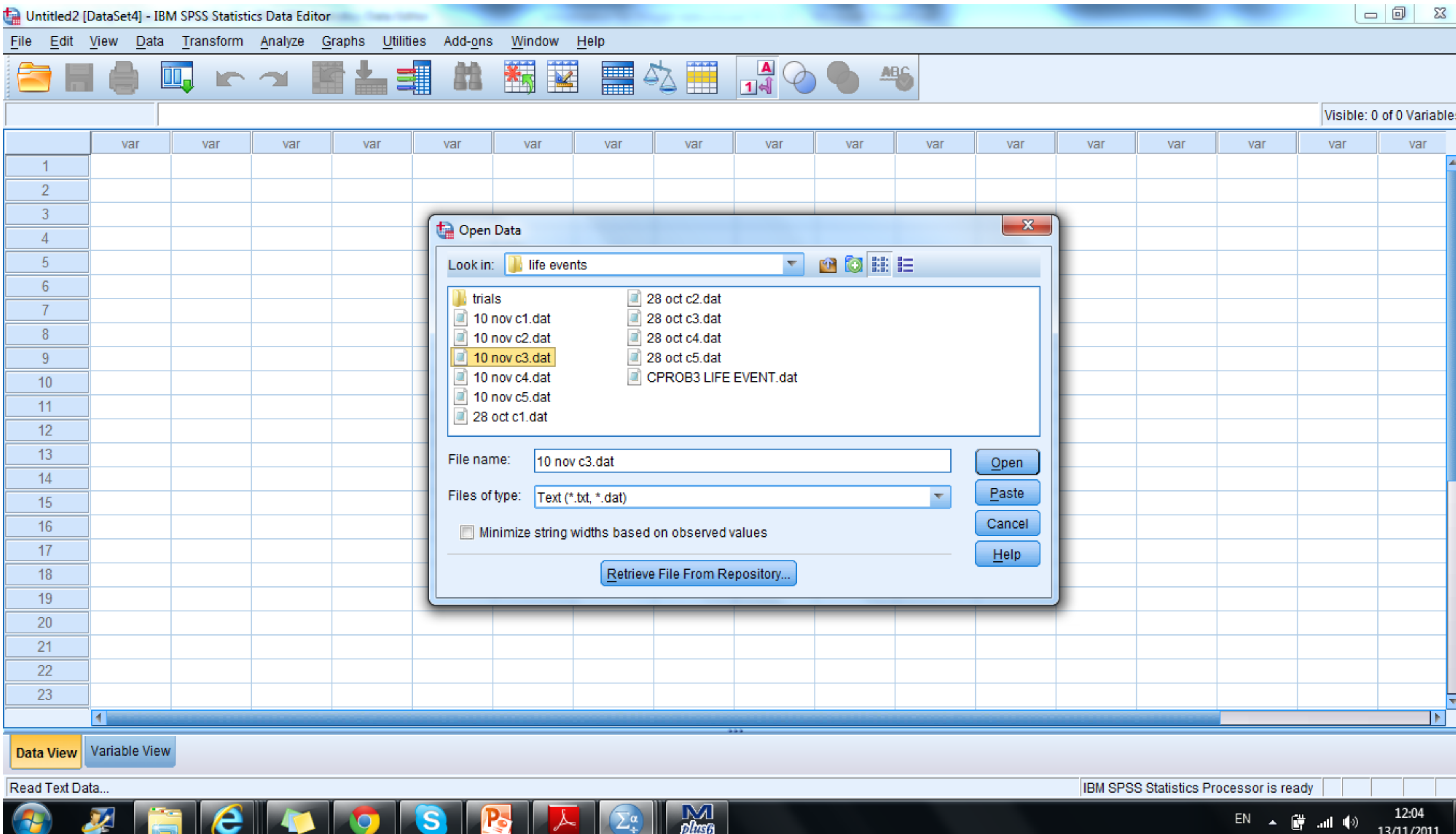
```
SAVEDATA INFORMATION

 Order and format of variables

   INJURY         F10.3
   SEPARATE       F10.3
   SACKED         F10.3
   POLICE         F10.3
   BULLIED        F10.3
   ABUSED         F10.3
   WEIGHT         F10.3
   IDNUM          I9
   CPROB1         F10.3
   CPROB2         F10.3
   CPROB3         F10.3
   C              F10.3
   STRATA         I4
   CLUSTER        I4

 Save file
   10 nov c3.dat

 Save file format
   7F10.3 I9 4F10.3 2I4
```

Should be located near the end of the output

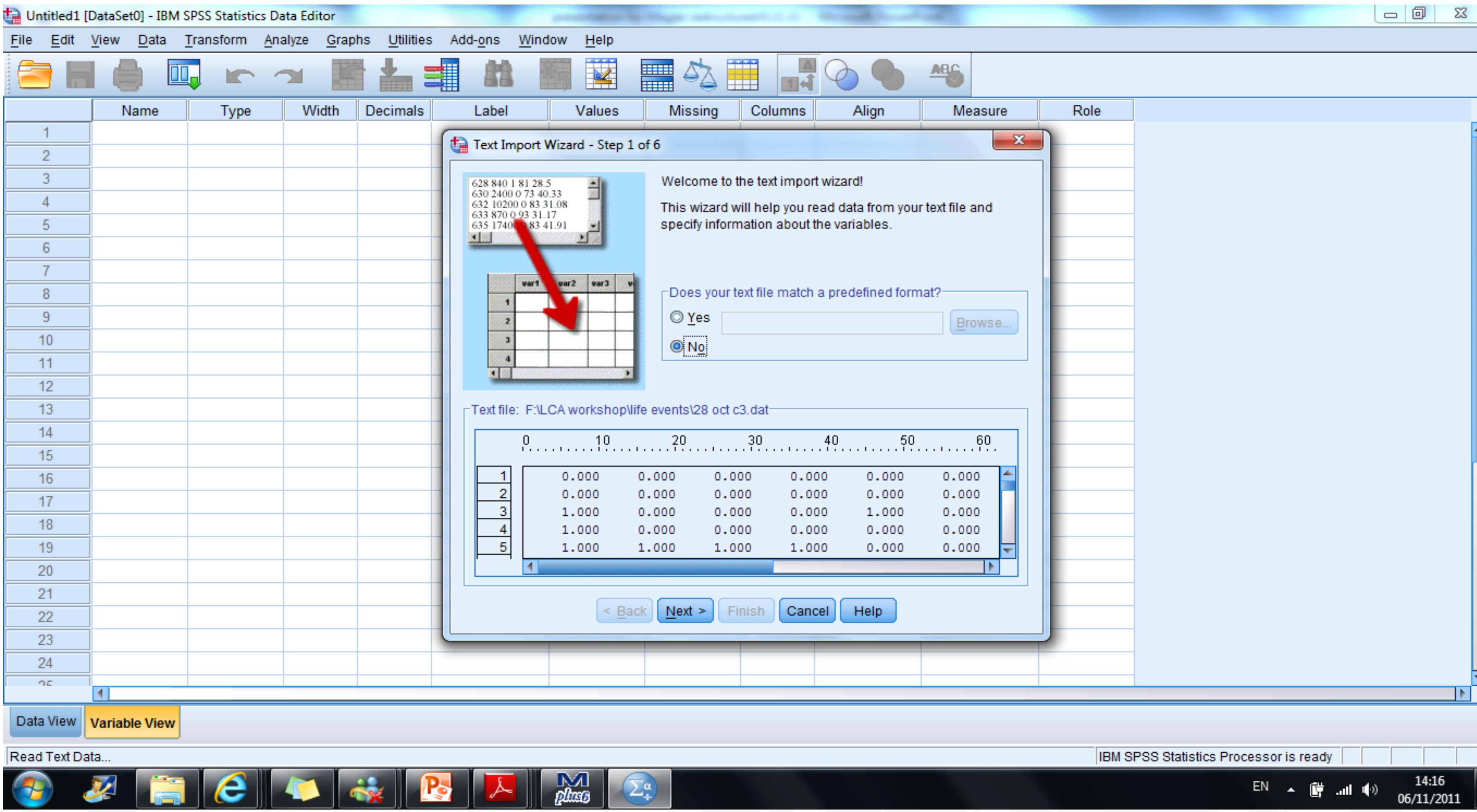
Mplus assigns each person in the sample to their most likely class

This is the file name to look for (it will be saved where you saved the input file)

Read text data saved in Mplus into SPSS



Series of stages…



The finished data file in SPSS

Class to which each case is assigned

*LC membership 3 class.sav [DataSet1] - PASW Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : INJURY 0 Visible: 14 of 14 Variables

	INJURY	SEPARATE	SACKED	POLICE	BULLIED	ABUSED	WEIGHT	IDNUM	CPROB1	CPROB2	CPROB3	C	STRATA	CLUSTER	var	var
1	0	0	0	0	0	0	1.204	11103072	.001	.959	.040	Low stress	94	103		
2	0	0	0	0	0	0	.779	11103102	.001	.959	.040	Low stress	94	103		
3	1	0	0	0	1	0	.602	11103112	.058	.672	.270	Low stress	94	103		
4	1	0	0	0	0	0	1.087	11103132	.005	.886	.109	Low stress	94	103		
5	1	1	1	1	0	0	.660	11103162	.022	.012	.966	Moderate stress	94	103		
6	0	0	0	0	0	0	1.126	11103172	.001	.959	.040	Low stress	94	103		
7	0	0	0	0	0	0	.779	11103192	.001	.959	.040	Low stress	94	103		
8	1	1	0	0	1	0	.768	11103202	.228	.317	.454	Moderate stress	94	103		
9	0	1	0	0	0	0	1.205	11103212	.004	.867	.129	Low stress	94	103		
10	1	0	1	0	0	0	1.087	11103222	.006	.534	.459	Low stress	94	103		
11	0	0	0	0	0	0	1.846	11103232	.001	.959	.040	Low stress	94	103		
12	0	0	0	0	1	0	1.535	11103242	.008	.873	.119	Low stress	94	103		
13	0	0	1	0	0	0	1.689	11103262	.001	.774	.225	Low stress	94	103		
14	0	0	0	0	0	0	.661	11103282	.001	.959	.040	Low stress	94	103		
15	0	1	1	0	0	1	.884	11109012	.294	.200	.506	Moderate stress	230	109		
16	0	1	0	0	1	0	.473	11109022	.048	.641	.311	Low stress	230	109		
17	1	0	1	0	1	0	1.129	11109032	.043	.251	.705	Moderate stress	230	109		
18	1	0	1	0	0	0	.486	11109042	.006	.534	.459	Low stress	230	109		
19	0	1	0	0	0	0	.884	11109052	.004	.867	.129	Low stress	230	109		
20	0	0	0	0	0	0	1.356	11109062	.001	.959	.040	Low stress	230	109		
21	0	1	0	1	0	0	.564	11109072	.019	.201	.780	Moderate stress	230	109		
22	1	0	1	0	0	1	.443	11109082	.355	.208	.437	Moderate stress	230	109		
23	0	1	0	0	0	0	.885	11109092	.004	.867	.129	Low stress	230	109		
24	0	0	0	0	0	0	2.117	11109122	.001	.959	.040	Low stress	230	109		
25	0	1	0	0	1	0	.898	11109132	.048	.641	.311	Low stress	230	109		
26	0	0	0	0	1	0	2.108	11109152	.008	.873	.119	Low stress	230	109		
27	0	0	1	0	0	0	.798	11109162	.001	.774	.225	Low stress	230	109		
28	0	1	1	0	0	0	.898	11109182	.005	.488	.507	Moderate stress	230	109		
29	1	1	1	0	0	0	.798	11109192	.023	.240	.736	Moderate stress	230	109		

Data View Variable View

PASW Statistics Processor is ready

Exploring relationship between class-membership and covariates

- Use the ‘C’ variable to conduct multinomial logistic regression analysis (if you have 3 or more classes)
- Choose a Reference class (usually largest class, but can be theory driven)
- Create ‘dummy variables’ for covariates (if necessary)
- Can run regression analysis in SPSS or Mplus

LC models including covariates

Mplus Discussion

Anonymous posted

4) When predictors of the latent class membership are included into LCA, model results, including latent class probabilities and conditional probabilities, often change. Does this mean we should always make model selection and interpret model results based on LCA model with predictors, and not report the model results without predictors?

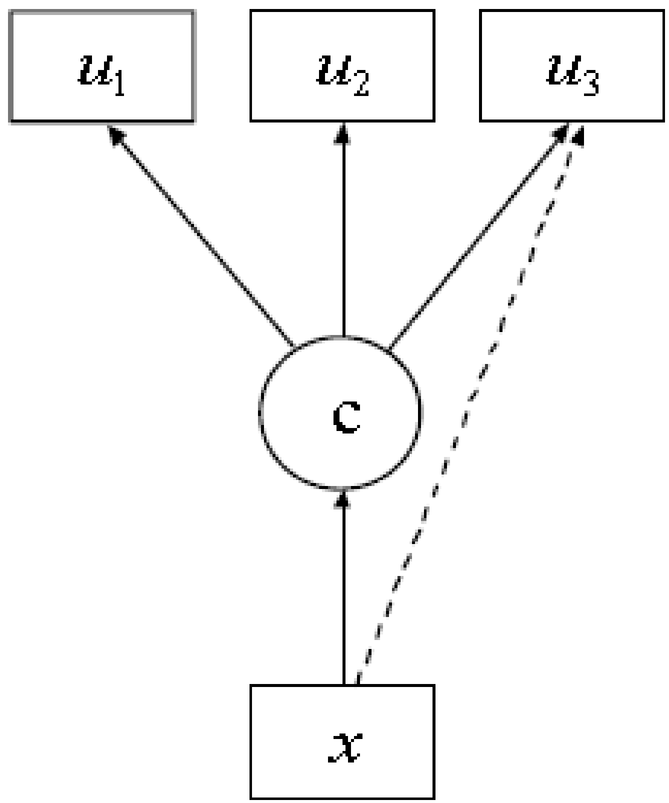
bmuthen posted

4) Yes, I think the final model should include covariates. I have a draft paper that discusses this issue, which I will be happy to send to you later this summer (remind me). For example, if the true model has covariates x influencing a latent class variable c and also have direct influence on the indicators u, then excluding x in the analysis will produce incorrect classification because of the direct influence of x on u.

Relating Latent Class Analysis Results to Variables not Included in the Analysis

Shaunna L. Clark & Bengt Muthén
University of California, Los Angeles

LC model with a covariate



Source: Muthén & Muthén, 2009

Stressful life events (3c) – with one covariate

```
TITLE: 2007 APMS Latent class analysis of stressful life events
3-class model, 6 life events, binary coded items
Two covariates (sex, marital) included

DATA:
file is
F:\LCA workshop\life events practical.dat;

VARIABLE:
NAMES are idnum weight cluster strata sex marital
injury separate sacked police bullied abused;

USEVARIABLES are injury separate sacked police bullied abused sex marital;
CATEGORICAL are injury separate sacked police bullied abused;

IDVARIABLE= idnum;
CLASSES = c(3);
WEIGHT IS weight;
CLUSTER IS cluster;
STRATIFICATION IS strata;

Missing are all (-9);

ANALYSIS:
type = complex mixture;
starts = 500 20;

MODEL:
%Overall%
c#1 c#2 on sex marital;

OUTPUT:
tech10 tech11 cinterval;

PLOT:
type = plot3;
series =  injury separate sacked police bullied abused(*);
```

Scored as 1=Married/Cohabiting; 0 = Other

Scored as 1=Male; 2=Female

LC3 is the reference group

LC model with covariate (sex) – 3c life events example

Output truncated

Women more likely than men to be assigned to LC 1, compared to LC3, controlling for marital status

CONFIDENCE INTERVALS FOR THE LOGISTIC REGRESSION ODDS RATIO RESULTS

Categorical Latent Variables

	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
C#1 ON							
SEX	1.196	1.470	1.634	$\boxed{2.834}$	4.917	5.463	6.715
MARITAL	0.353	0.382	0.398	0.494	0.612	0.638	0.691
C#2 ON							
SEX	0.004	0.007	0.008	0.024	0.072	0.088	0.132
MARITAL	1.281	1.407	1.476	$\boxed{1.897}$	2.437	2.557	2.808

Married/cohabiting respondents more likely than men to be assigned to LC 2, compared to LC3, controlling for sex

ANY FINALLY…

SOME COMMON ERROR MESSAGES IN MPLUS

Error/warning messages (not serious)

```
*** ERROR in DATA command
 The file specified for the FILE option cannot be found.  Check that this
 file exists:  F:\LCA workshop\stressful events (practical.dat
```

```
*** WARNING
 Variable name contains more than 8 characters.
 Only the first 8 characters will be printed in the output.
 Variable: S11AQ1A10
```

```
WARNING:  THE VARIANCE CONTRIBUTION FROM A STRATUM WITH A SINGLE CLUSTER
(PSU) IS BASED ON THE DIFFERENCE BETWEEN THE SINGLE CLUSTER VALUE AND
THE OVERALL CLUSTER MEAN.
```

```
 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 CHI-SQUARE TEST CANNOT BE COMPUTED BECAUSE THE FREQUENCY TABLE FOR THE
LATENT CLASS INDICATOR MODEL PART IS TOO LARGE.
```

Error/warning messages (may require attention!)

IN THE OPTIMIZATION, ONE OR MORE LOGIT THRESHOLDS APPROACHED AND WERE SET AT THE EXTREME VALUES. EXTREME VALUES ARE -15.000 AND 15.000.

WARNING: THE BEST LOGLIKELIHOOD VALUE WAS NOT REPLICATED. THE SOLUTION MAY NOT BE TRUSTWORTHY DUE TO LOCAL MAXIMA. INCREASE THE NUMBER OF RANDOM STARTS.

WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IN CLASS 2 IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES. CHECK THE TECH4 OUTPUT FOR MORE INFORMATION. PROBLEM INVOLVING VARIABLE ANG.

http://www.statmodel.com/cgi-bin/discus/discus.cgi?pg=topics

Mplus Discussion

Any questions???