Data Analysis and Statistical Software

Latent Class Analysis Using Stata

Chuck Huber
StataCorp
chuber@stata.com

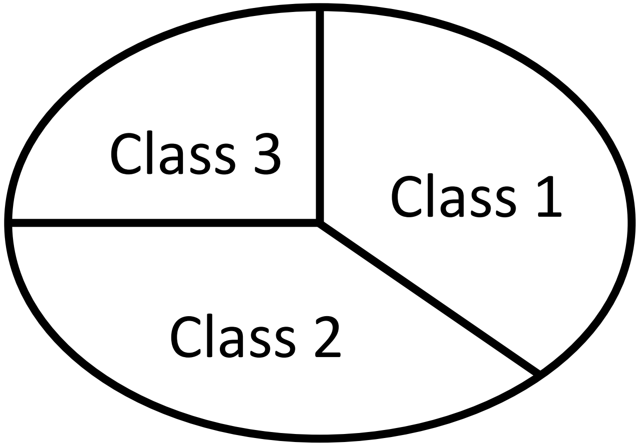
University College London
October 16, 2019

Outline

- Latent class analysis (LCA)
  - Estimation and postestimation options
  - **`margins`** and **`marginsplot`**
- Latent class analysis with covariates
- Latent class analysis by groups
- Latent profile analysis

Latent Class Analysis

- A latent class model is characterized by having a categorical latent variable and categorical observed variables.
- The levels of the categorical latent variable represent groups in the population and are called classes.
- We are interested in identifying and understanding these unobserved classes.



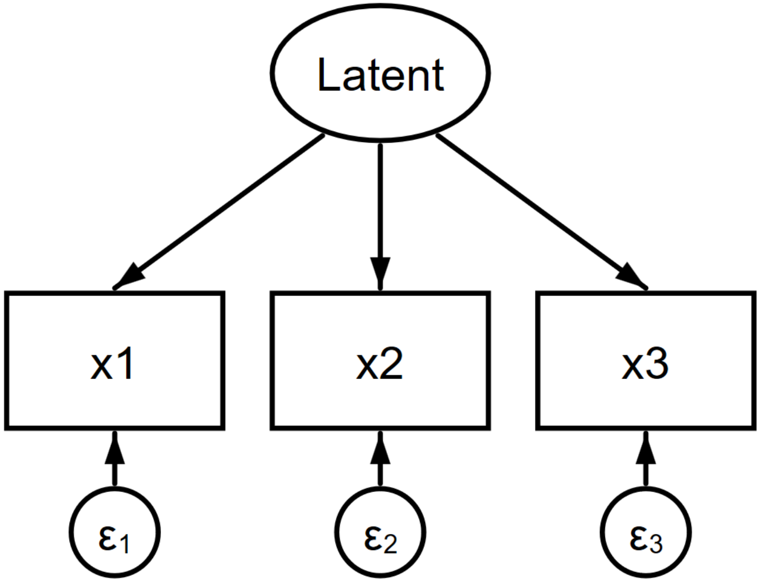
Latent Class Analysis

- Behavioral Research
  - Classify people who are more likely to exhibit specific behaviors
  - Different kinds of social phobias
  - Categories of eating disorders
- Medicine and Health
  - Identify patients with different disease risk profiles
- Marketing Research
  - Differentiate subsets of customers and their buying habits
- And many, many more…

Latent and Observed Variables

- <u>Latent variables</u> are hypothetical constructs that we cannot measure directly (e.g. “intelligence” or “depression”)

- <u>Observed variables</u> are variables that we observe (e.g. “SAT scores” or “Feel sad”)



Latent and Observed Variables

	Continuous Latent	Categorical Latent
Continuous Observed	Factor analysis	Latent Profile Analysis
Categorical Observed	Latent trait analysis or Item Response Theory	Latent Class Analysis

Reproduced from Table 1.1 in Collins and Lanza (2010)

Example Data

AGGRESSIVE BEHAVIOR
Volume 37, pages 19–35 (2011)

The Three Latent Classes of Adolescent Delinquency and the Risk Factors for Membership in Each Class

Penelope Anne Hasking[1*], Lawrence M. Scheier[2,3], and Arbi ben Abdallah[3]

[1]*School of Psychology and Psychiatry, Monash University, Clayton, Victoria, Australia*
[2]*LARS Research Institute, Inc., Las Vegas, Nevada*
[3]*Department of Psychiatry, Epidemiology and Prevention Research Group, Washington University School of Medicine, St. Louis, Missouri*

This study used latent class analysis to examine subpopulation membership based on self-reports of delinquent behaviors obtained from Australian youth. Three discrete identifiable classes were derived based on 51 indicators of physical violence, property damage, minor infractions, drug use, and social delinquency. One class of youth engaged in primarily rule breaking and norm violations including underage alcohol use, typical of this age period. A second class was more actively delinquent emphasizing drug use, trespassing, and various forms of disobedience. A third class of highly delinquent youth differed from their counterparts by endorsing drug use, thievery that involved stealing money, goods, and cars, property damage, gambling, precocious sexual experiences, involvement with pornographic materials, and fighting. Multinomial logistic regression predicting class membership indicated highly delinquent youth were more likely to be older males, use venting coping strategies, and be fun or novelty seeking compared with rule breakers. Findings are discussed in terms of refining current taxonomic arguments regarding the structure of delinquency and implications for prevention of early-stage antisocial behavior. Aggr. Behav. 37:19–35, 2011.

Example Data

- Class 1
  - "Possible rule breaking and norm violations"
  - May include underage drinking
- Class 2
  - "More actively delinquent"
  - Occasional drug use, skipping school, and non-violent crime
- Class 3
  - "Highly delinquent"
  - Regular drug use, theft, and violent crime

Example Data

```
. use atrisk.dta

. describe

Contains data from atrisk.dta
  obs:        10,000
 vars:             8                          20 Mar 2018 15:17
 size:        90,000
------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------
id              int     %9.0g                 Student identification number
age             byte    %9.0g                 Age (years)
male            byte    %9.0g      male       Male
alcohol         byte    %9.0g                 Ever consumed alcohol
truant          byte    %9.0g                 >10 unexcused absences from school
vandalism       byte    %9.0g                 Ever engaged in an act of vandalism
theft           byte    %9.0g                 Ever stolen something worth more than $25
weapon          byte    %9.0g                 Ever used a weapon in a fight
------------------------------------------------------------------------------------
Sorted by: id
```

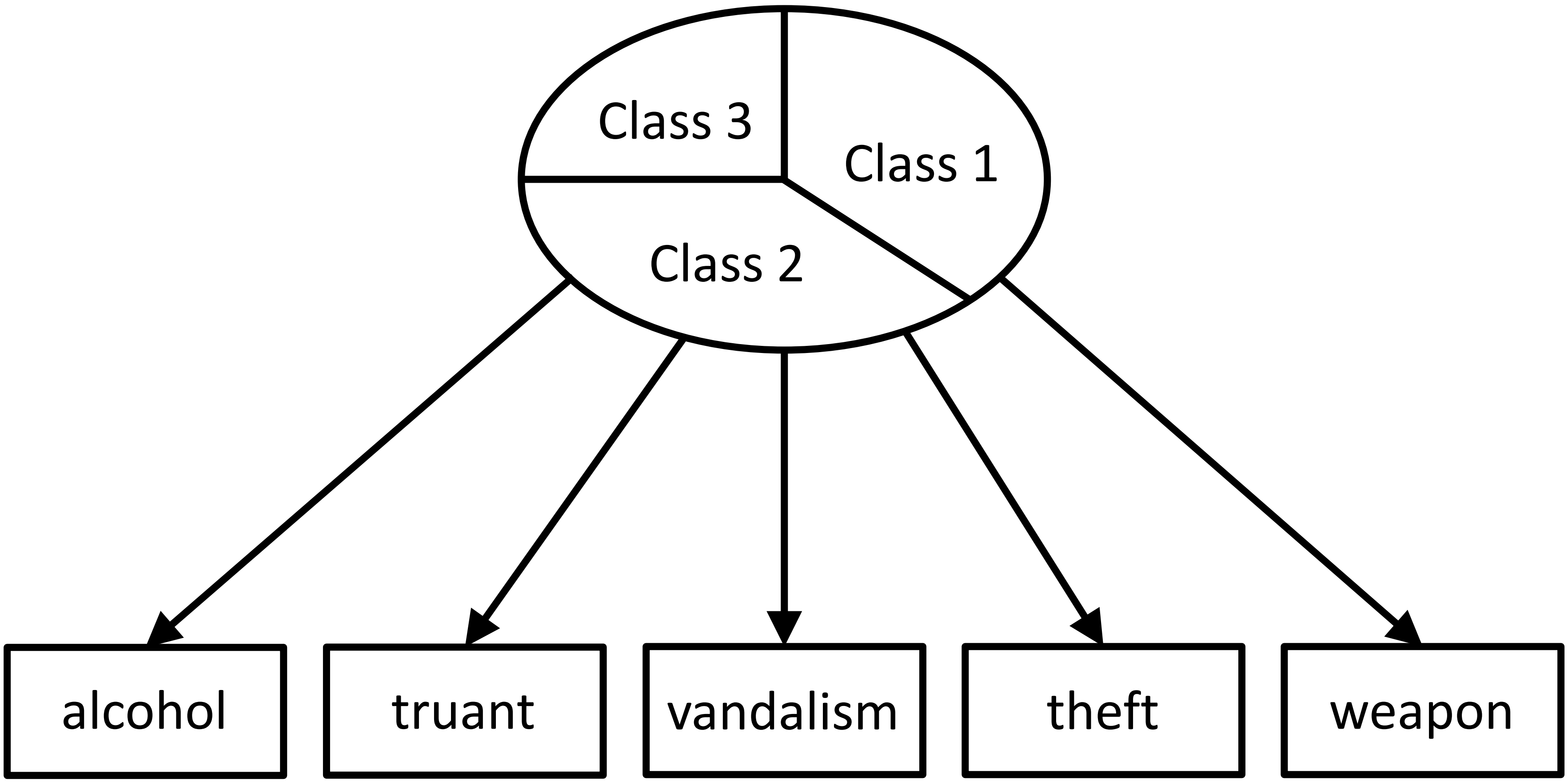
Note that we do not have a variable for class membership.

Example Data

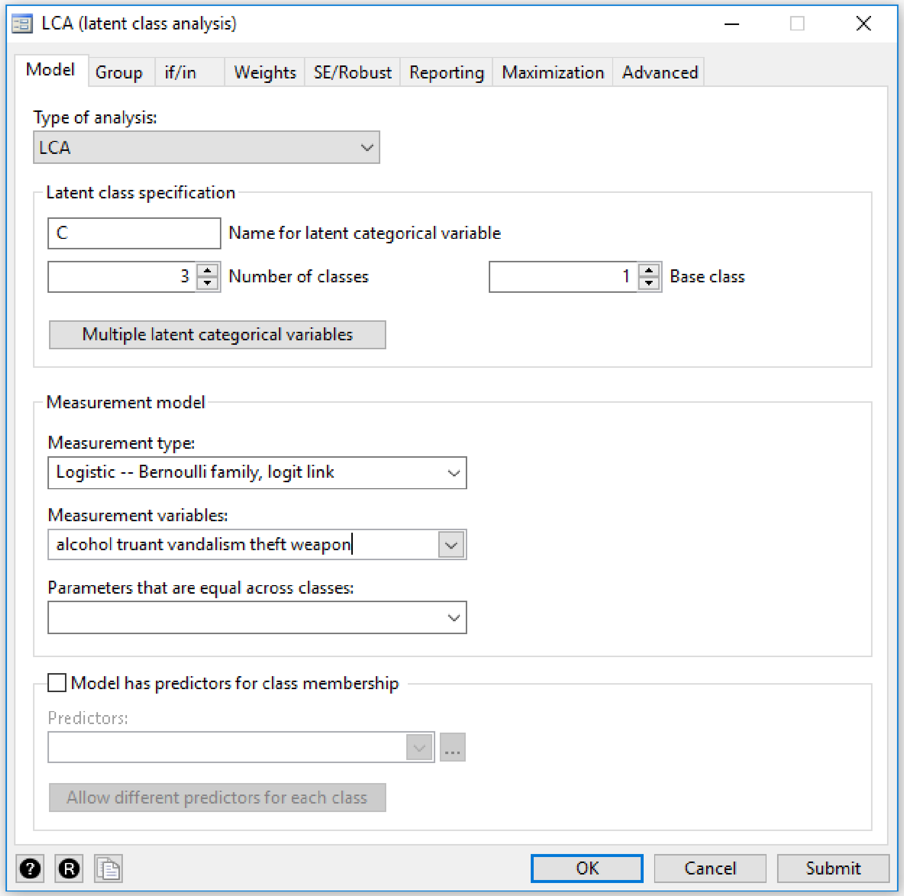
```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
id	10,000	5000.5	2886.896	1	10000
age	10,000	15.4826	1.70452	13	18
male	10,000	.5011	.5000238	0	1
alcohol	10,000	.3548	.4784766	0	1
truant	10,000	.094	.2918433	0	1
vandalism	10,000	.077	.2666048	0	1
theft	10,000	.0288	.1672524	0	1
weapon	10,000	.0265	.1606248	0	1

Conceptual Path Diagram



Statistics > LCA (latent class analysis)




```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     , family(bernoulli) link(logit) lclass(C 3)
```

The coefficients for class

```
Generalized structural equation model              Number of obs     =     10,000
Log likelihood = -14411.803
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.C	(base outcome)					
2.C						
_cons	-2.413129	.5429566	-4.44	0.000	-3.477304	-1.348954
3.C						
_cons	-3.679325	.4864832	-7.56	0.000	-4.632815	-2.725835

LCA Postestimation

```
. estat lcprob

Latent class marginal probabilities       Number of obs = 10,000
```

	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
C				
1	.8970422	.0358031	.8029742	.9490482
2	.0803164	.0405096	.0289462	.2037259
3	.0226414	.0112419	.0084863	.0590022

```
Class            : 1

Response         : alcohol
Family           : Bernoulli
Link             : logit

Response         : truant
Family           : Bernoulli
Link             : logit

Response         : vandalism
Family           : Bernoulli
Link             : logit

Response         : theft
Family           : Bernoulli
Link             : logit

Response         : weapon
Family           : Bernoulli
Link             : logit
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
alcohol						
_cons	-.7861047	.0489882	-16.05	0.000	-.8821198	-.6900895
truant						
_cons	-2.93729	.1906323	-15.41	0.000	-3.310922	-2.563657
vandalism						
_cons	-2.952349	.1000292	-29.51	0.000	-3.148402	-2.756295
theft						
_cons	-4.516317	.1614524	-27.97	0.000	-4.832758	-4.199876
weapon						
_cons	-4.427085	.1749759	-25.30	0.000	-4.770032	-4.084139

```
Class          : 2

Response       : alcohol
Family         : Bernoulli
Link           : logit

Response       : truant
Family         : Bernoulli
Link           : logit

Response       : vandalism
Family         : Bernoulli
Link           : logit

Response       : theft
Family         : Bernoulli
Link           : logit

Response       : weapon
Family         : Bernoulli
Link           : logit
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
alcohol						
_cons	.8948312	.4756056	1.88	0.060	-.0373386	1.827001
truant						
_cons	-.0699578	.4886625	-0.14	0.886	-1.027719	.887803
vandalism						
_cons	-1.155054	.371044	-3.11	0.002	-1.882286	-.4278206
theft						
_cons	-3.309095	2.03741	-1.62	0.104	-7.302346	.6841566
weapon						
_cons	-2.167636	.4837753	-4.48	0.000	-3.115818	-1.219454

```
Class          : 3

Response       : alcohol
Family         : Bernoulli
Link           : logit

Response       : truant
Family         : Bernoulli
Link           : logit

Response       : vandalism
Family         : Bernoulli
Link           : logit

Response       : theft
Family         : Bernoulli
Link           : logit

Response       : weapon
Family         : Bernoulli
Link           : logit
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
alcohol						
_cons	1.104349	.2815121	3.92	0.000	.5525958	1.656103
truant						
_cons	-.2185809	.2018425	-1.08	0.279	-.614185	.1770231
vandalism						
_cons	.3428592	.501954	0.68	0.495	-.6409527	1.326671
theft						
_cons	.938004	1.133686	0.83	0.408	-1.283979	3.159987
weapon						
_cons	-.6708939	.4134047	-1.62	0.105	-1.481152	.1393644

LCA Postestimation

```
. estat lcmean

Latent class marginal means                    Number of obs = 10,000
```

	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
1				
alcohol	.3130057	.0105341	.2927387	.3340132
truant	.0503407	.0091135	.0351984	.0715143
vandalism	.0496256	.0047177	.0411543	.0597321
theft	.010811	.0017266	.0079016	.0147758
weapon	.0118082	.0020417	.0084088	.0165588
2				
alcohol	.7098862	.0979499	.4906664	.8614041
truant	.4825177	.1220163	.2635266	.7084366
vandalism	.2395672	.0675949	.1321265	.3946469
theft	.0352605	.069307	.0006735	.6646658
weapon	.1026947	.0445792	.0424595	.2280326
3				
alcohol	.7510741	.052632	.6347376	.8397142
truant	.4455713	.0498627	.3511051	.5441406
vandalism	.5848849	.1218717	.3450312	.7902895
theft	.7186963	.2291994	.2168736	.9593004
weapon	.3382967	.0925415	.1852534	.5347848

Estimation and Postestimation

- First you estimate the parameters (fit the model)

```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     , family(bernoulli) link(logit) lclass(C 3)
```

- Then you calculate quantities such as goodness-of-fit statistics, model predictions, residuals, etc

```
estat lcmean
estat lcprob
```

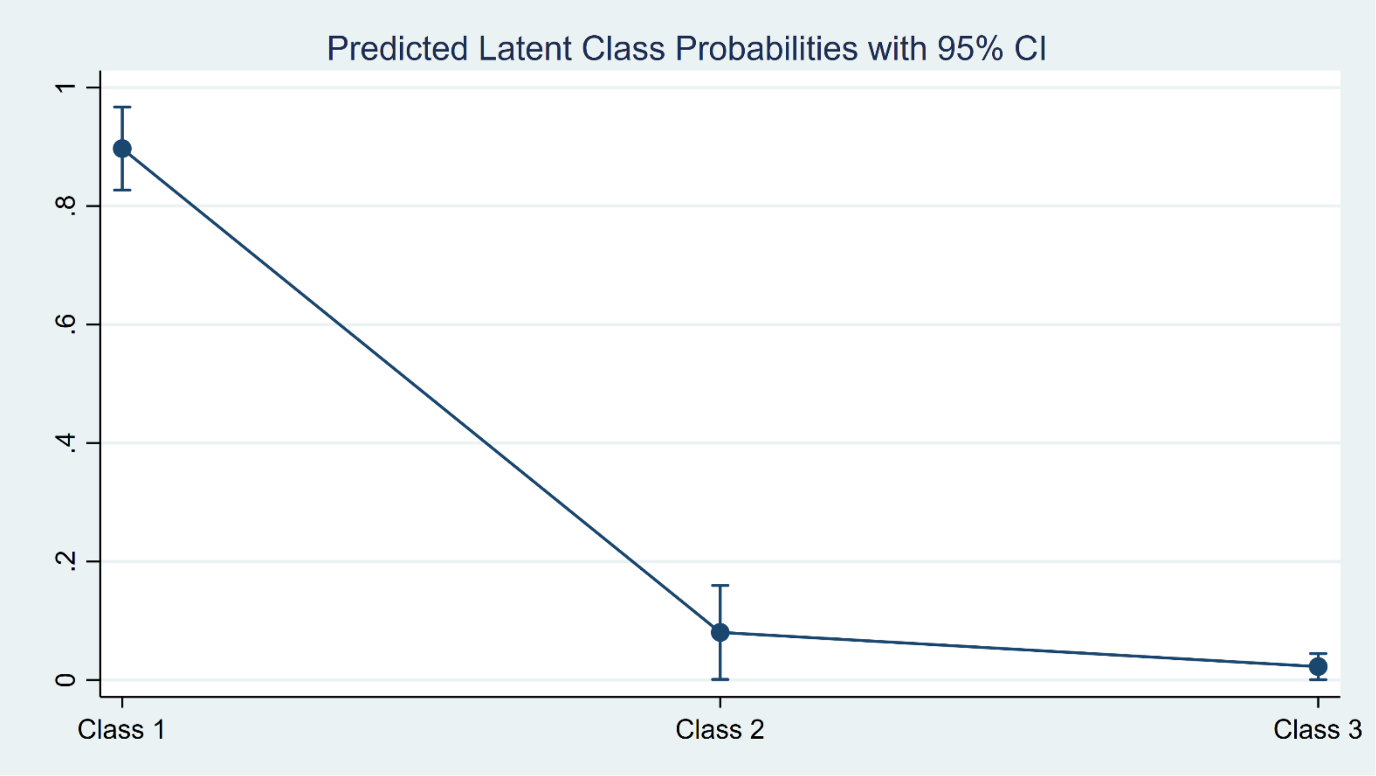
`margins` and `marginsplot`

- You can use **`margins`** to calculate marginal predictions after you fit almost any kind of model in Stata.

- You can use **`marginsplot`** to graph those marginal predictions.

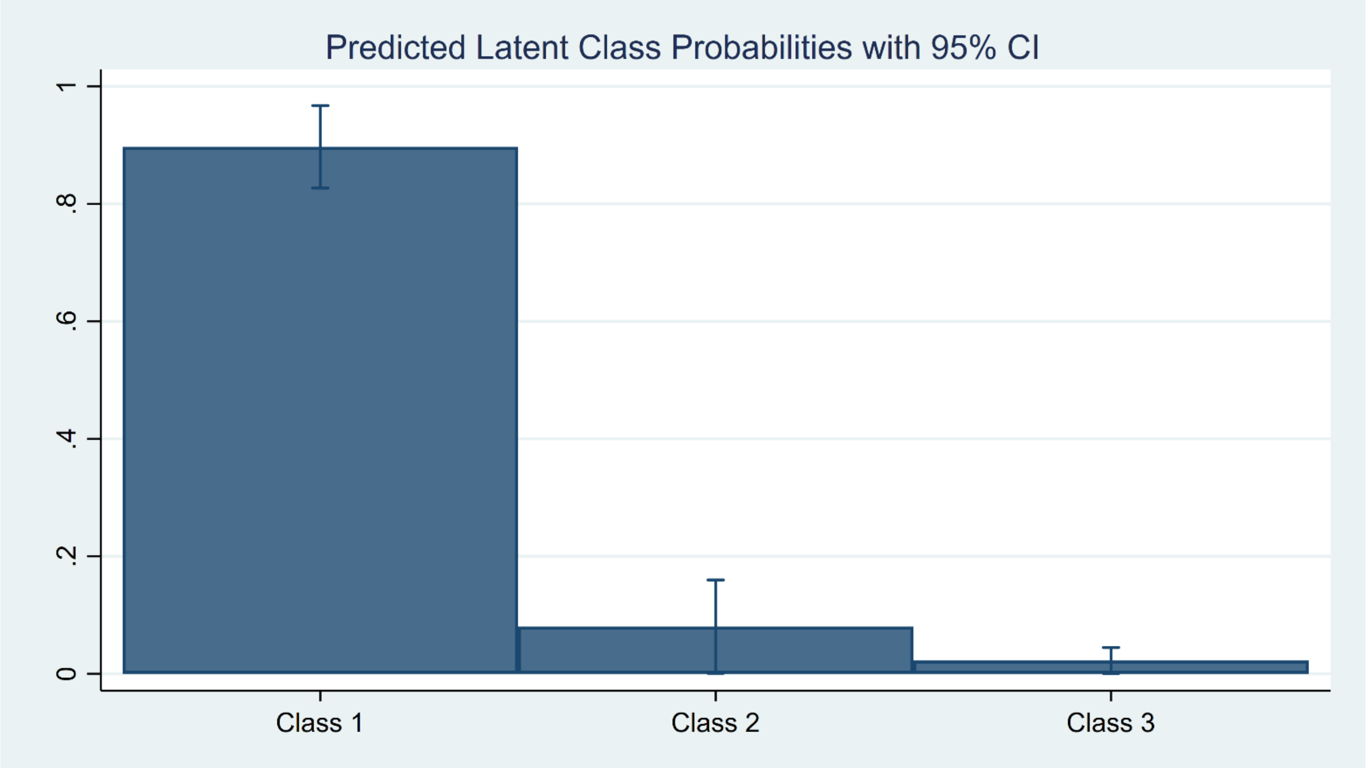
`margins` and `marginsplot`

```
margins, predict(classpr class(1))    ///
         predict(classpr class(2))    ///
         predict(classpr class(3))

marginsplot, xtitle("") ytitle("")                                          ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")                    ///
             title("Predicted Latent Class Probabilities with 95% CI")
```



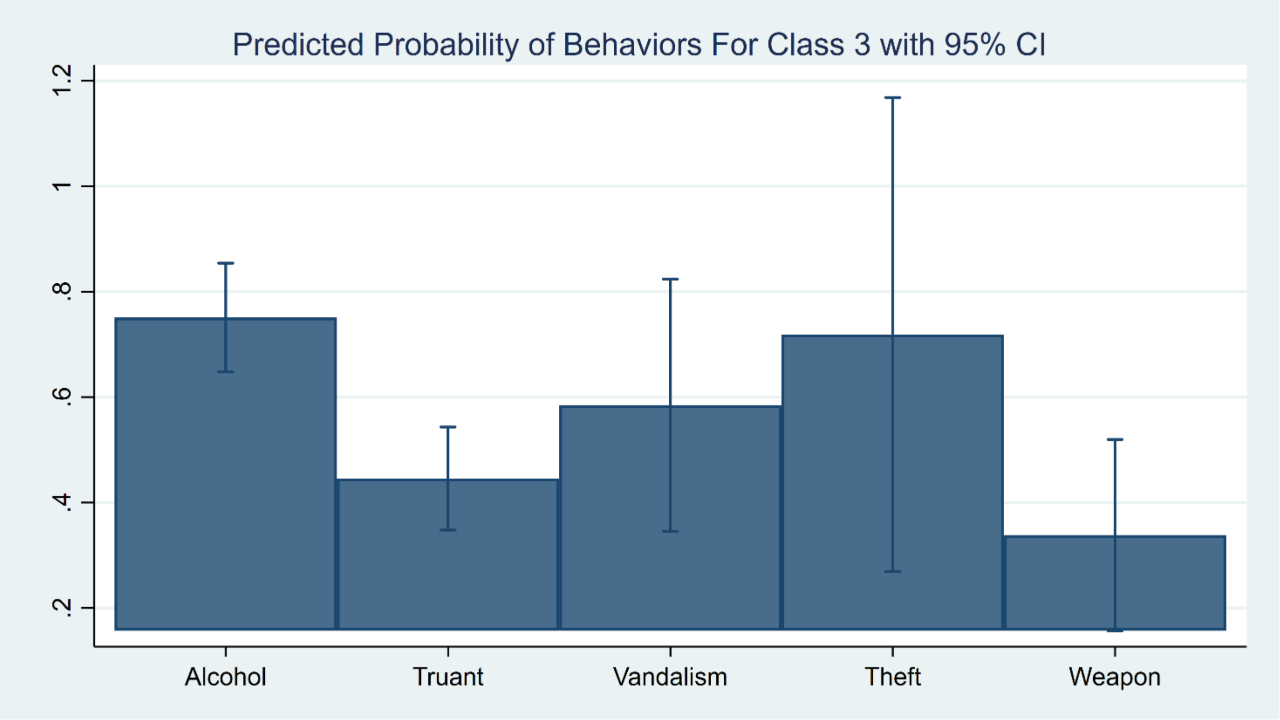
margins and marginsplot

```
margins, predict(classpr class(1))    ///
         predict(classpr class(2))    ///
         predict(classpr class(3))

marginsplot, recast(bar) xtitle("") ytitle("")                     ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")           ///
             title("Predicted Latent Class Probabilities with 95% CI")
```



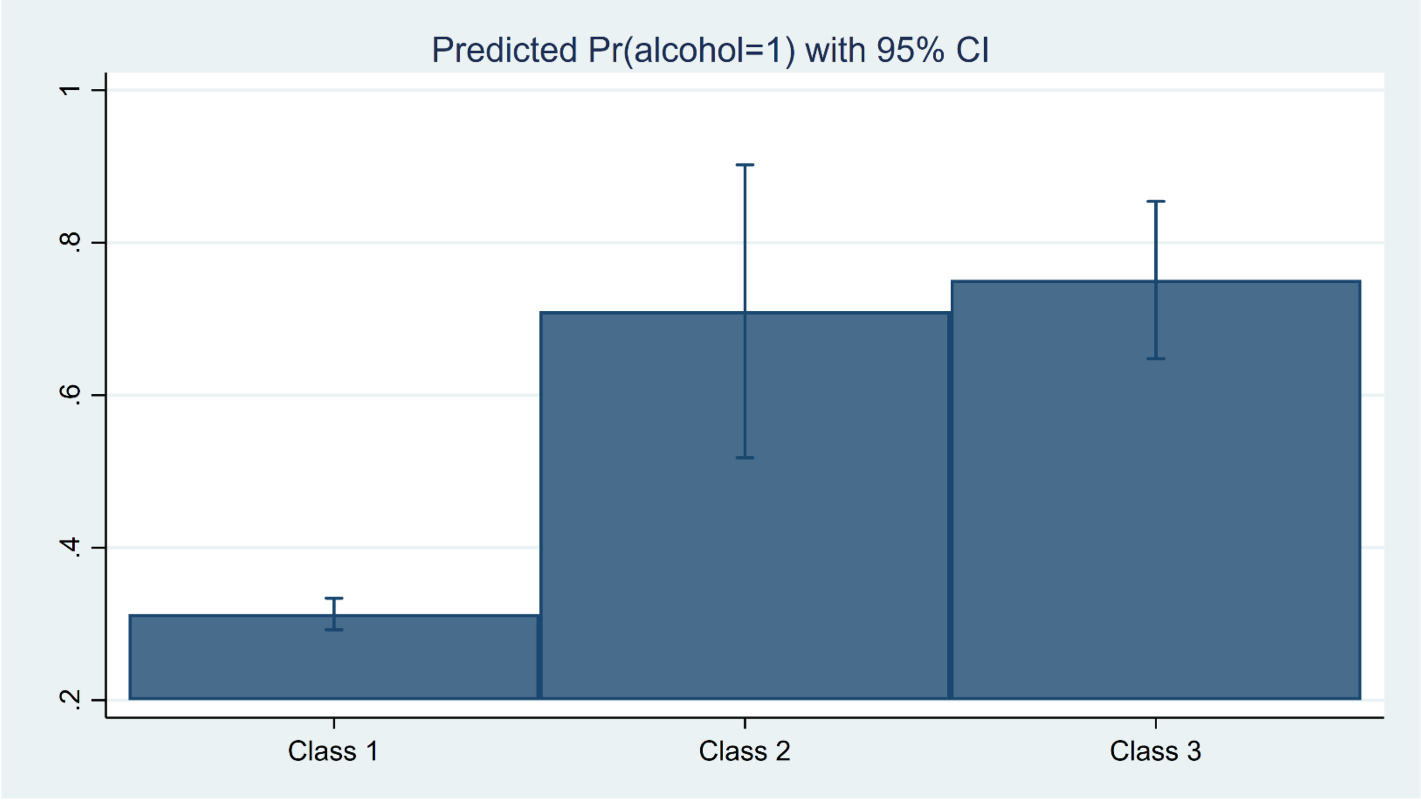
margins and **marginsplot**

```
margins, predict(outcome(alcohol)   class(3))       ///
         predict(outcome(truant)    class(3))       ///
         predict(outcome(vandalism) class(3))       ///
         predict(outcome(theft)     class(3))       ///
         predict(outcome(weapon)    class(3))

marginsplot, recast(bar) xtitle("") ytitle("")                          ///
             xlabel(1 "Alcohol" 2 "Truant" 3 "Vandalism" 4 "Theft" 5 "Weapon")   ///
             title("Predicted Probability of Behaviors For Class 3 with 95% CI")
```



`margins` and `marginsplot`

```
margins, predict(outcome(alcohol) class(1))       ///
         predict(outcome(alcohol) class(2))       ///
         predict(outcome(alcohol) class(3))

marginsplot, recast(bar) xtitle("") ytitle("")              ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")  ///
             title("Predicted Pr(alcohol=1) with 95% CI")
```



LCA Postestimation

```
. predict cpr*, classposteriorpr

. format %9.4f cpr1 cpr2 cpr3

. list id alcohol truant vandalism theft weapon cpr1 cpr2 cpr3    ///
>       in 1/10, abbreviate(10) separator(0) noobs
```

id	alcohol	truant	vandalism	theft	weapon	cpr1	cpr2	cpr3
1	0	0	0	0	0	0.9852	0.0144	0.0004
2	0	0	1	0	1	0.4245	0.3579	0.2176
3	0	0	0	0	0	0.9852	0.0144	0.0004
4	1	0	0	0	0	0.9248	0.0725	0.0027
5	1	1	0	0	1	0.0621	0.8196	0.1184
6	0	0	0	0	0	0.9852	0.0144	0.0004
7	0	0	0	0	0	0.9852	0.0144	0.0004
8	0	0	0	0	0	0.9852	0.0144	0.0004
9	0	0	0	0	0	0.9852	0.0144	0.0004
10	1	0	0	0	0	0.9248	0.0725	0.0027

LCA Postestimation

```
. estat lcgof
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(14)	20.646	model vs. saturated
p > chi2	0.111	
Information criteria		
AIC	28857.607	Akaike's information criterion
BIC	28980.183	Bayesian information criterion

How many classes?

```
quietly gsem (alcohol truant vandalism theft weapon <- ), logit lclass(C 1)
estimates store oneclass

quietly gsem (alcohol truant vandalism theft weapon <- ), logit lclass(C 2)
estimates store twoclass

quietly gsem (alcohol truant vandalism theft weapon <- ), logit lclass(C 3)
estimates store threeclass

. estimates stats oneclass twoclass threeclass
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
oneclass	10,000	.	-14863.47	5	29736.93	29772.98
twoclass	10,000	.	-14430.33	11	28882.66	28961.97
threeclass	10,000	.	-14411.8	17	28857.61	28980.18

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Options for Starting Values

Starting values options for models with categorical latent variables are as follows:

`startvalues(factor` [, *maxopts*]`)` specifies that starting values be computed by assigning each observation to an initial latent class that is determined by running a `factor` analysis on all the observed variables in the specified model. This is the default for models with categorical latent variables.

`startvalues(classid` *varname*[, *maxopts*]`)` specifies that starting values be computed by assigning each observation to an initial latent class specified in *varname*. *varname* is required to have each class represented in the estimation sample.

`startvalues(classpr` *varlist*[, *maxopts*]`)` specifies that starting values be computed using the initial class probabilities specified in *varlist*. *varlist* is required to contain k variables for a model with k latent classes. The values in *varlist* are normalized to sum to 1 within each observation.

`startvalues(randomid` [, `draws(`#`)` `seed(`#`)` *maxopts*]`)` specifies that starting values be computed by randomly assigning observations to initial classes.

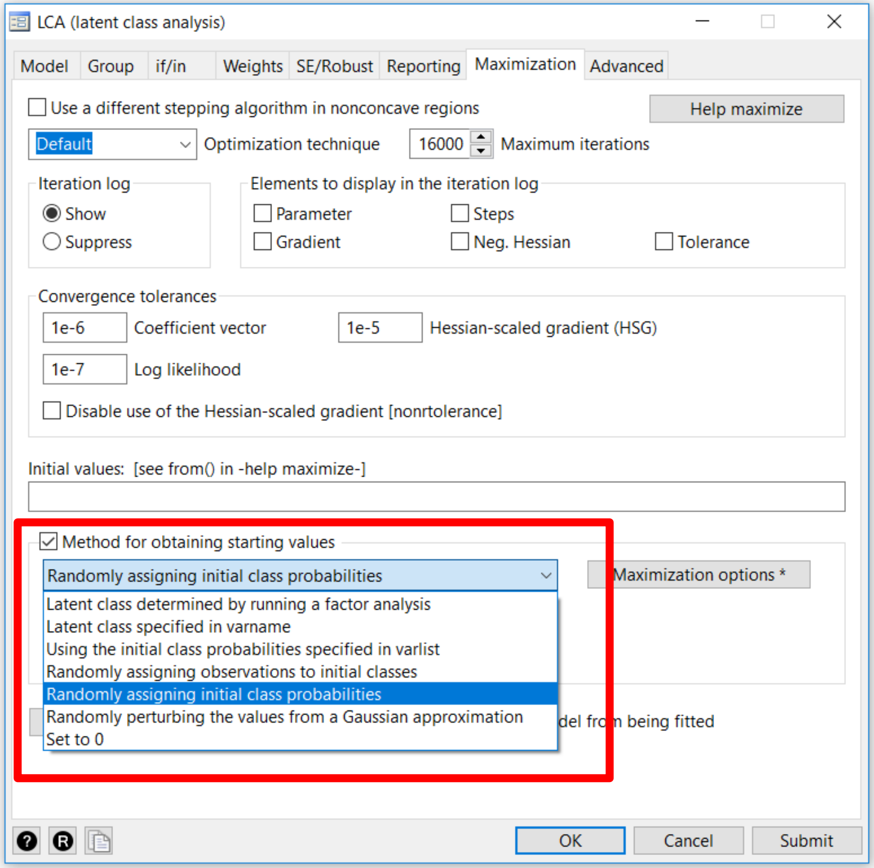
`startvalues(randompr` [, `draws(`#`)` `seed(`#`)` *maxopts*]`)` specifies that starting values be computed by randomly assigning initial class probabilities.

`startvalues(jitter` [$\#_c$ [$\#_v$]], `draws(`#`)` `seed(`#`)` *maxopts*]`)` specifies that starting values be constructed by randomly perturbing the values from a Gaussian approximation to each outcome.

$\#_c$ is the magnitude for randomly perturbing coefficients, intercepts, cutpoints, and scale parameters; the default value is 1.

$\#_v$ is the magnitude for randomly perturbing variances for Gaussian outcomes; the default value is 1.

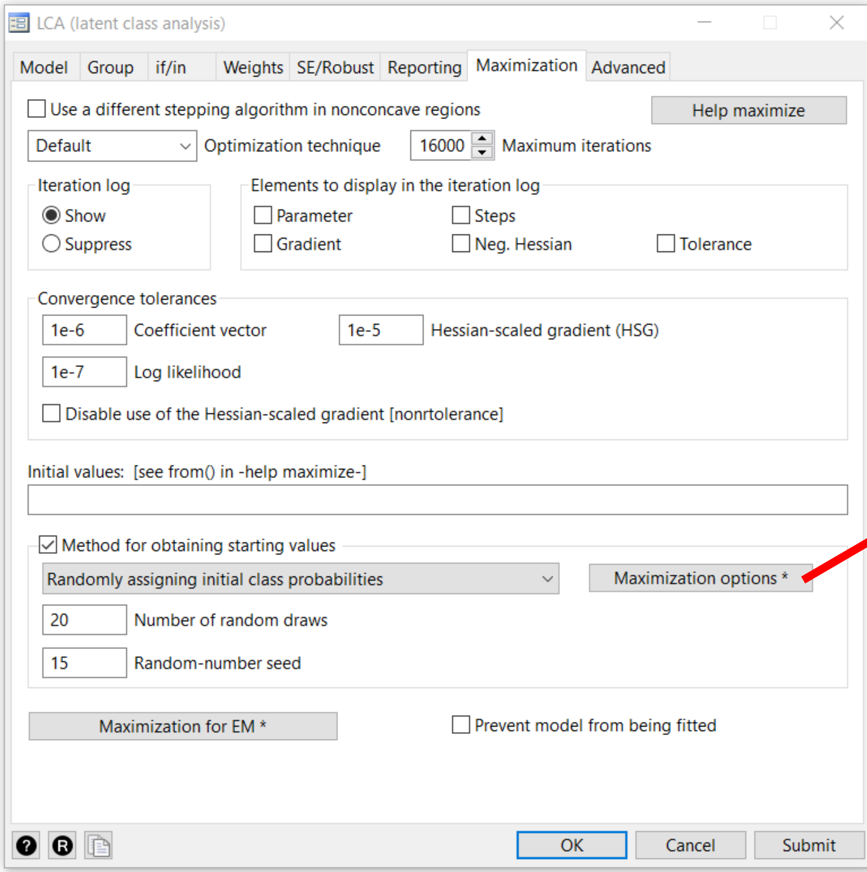
Options for Starting Values

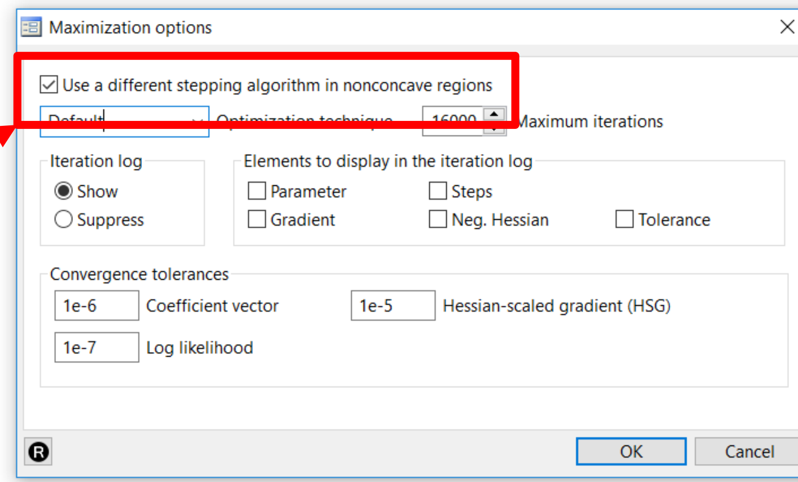



```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     , logit lclass(C 3)                                  ///
     startvalues(randompr, draws(20) seed(15))
```

Data Analysis and Statistical Software

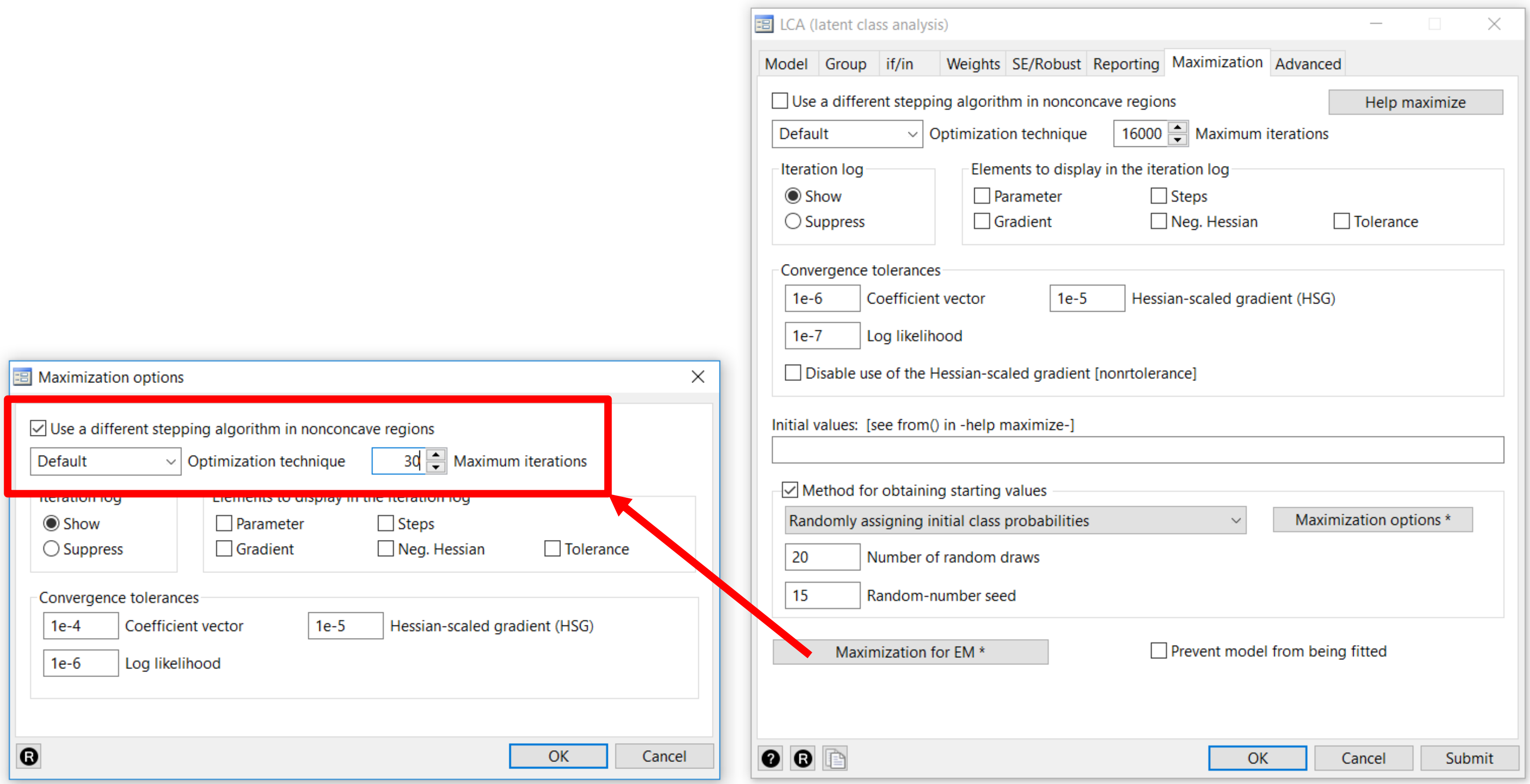
Options for Starting Values






```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     , logit lclass(C 3)                                 ///
     startvalues(randompr, draws(20) seed(15) difficult)
```

Options for the EM Algorithm




```
gsem (alcohol truant vandalism theft weapon <- _cons)     ///
     , logit lclass(C 3)                                  ///
     startvalues(randompr, draws(20) seed(15) difficult) ///
     emopts(iterate(30) difficult)
```

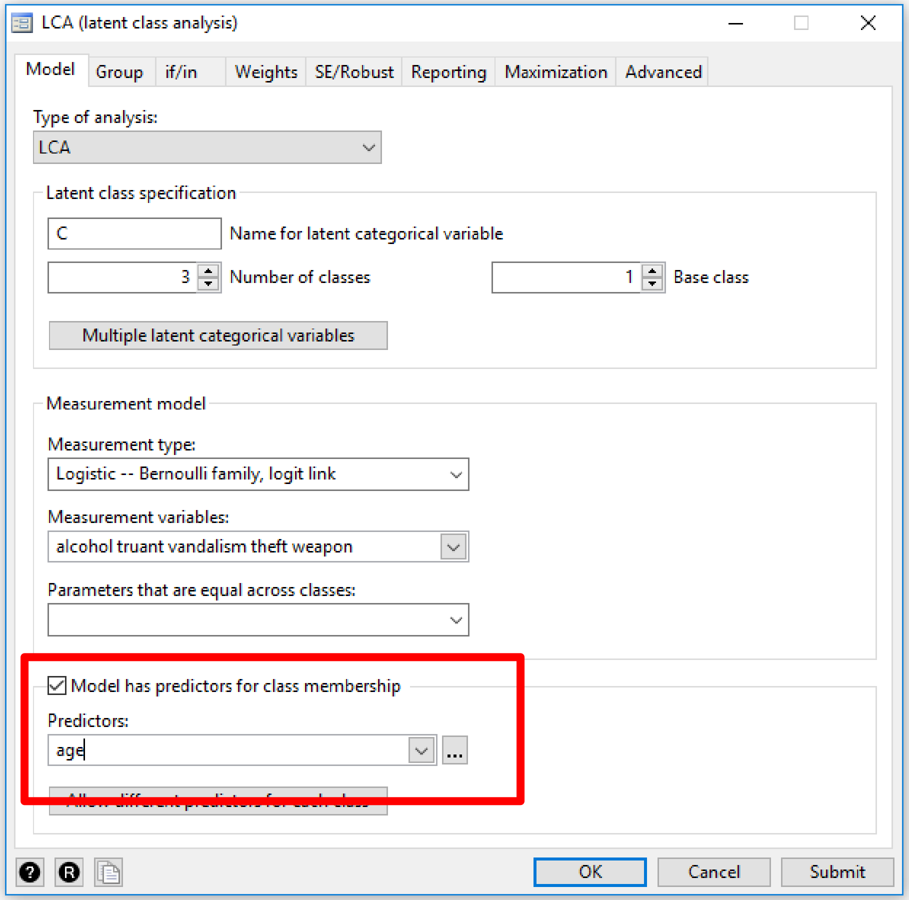
Outline

- Latent class analysis (LCA)
  - Estimation and postestimation options
  - **`margins`** and **`marginsplot`**
- Latent class analysis with covariates
- Latent class analysis by groups
- Latent profile analysis

LCA With Covariates

```
. use atrisk.dta

. describe

Contains data from atrisk.dta
  obs:        10,000
 vars:             8                          20 Mar 2018 15:17
 size:        90,000
--------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------------------
id              int     %9.0g                 Student identification number
age             byte    %9.0g                 Age (years)
male            byte    %9.0g      male       Male
alcohol         byte    %9.0g                 Ever consumed alcohol
truant          byte    %9.0g                 >10 unexcused absences from school
vandalism       byte    %9.0g                 Ever engaged in an act of vandalism
theft           byte    %9.0g                 Ever stolen something worth more than $25
weapon          byte    %9.0g                 Ever used a weapon in a fight
--------------------------------------------------------------------------------------------
Sorted by: id
```

LCA With Covariates




```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     (C <- age), family(bernoulli) link(logit)           ///
     lclass(C 3)
```

LCA With Covariates

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.C	(base outcome)					
2.C						
age	.2719666	.0540686	5.03	0.000	.165994	.3779391
_cons	-6.981905	.9669605	-7.22	0.000	-8.877113	-5.086698
3.C						
age	.2890307	.0506078	5.71	0.000	.1898412	.3882202
_cons	-8.29407	.9687997	-8.56	0.000	-10.19288	-6.395258

LCA With Covariates

```
. estat lcprob

Latent class marginal probabilities          Number of obs = 10,000
```

	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
C				
1	.9152694	.0130999	.8858127	.9376619
2	.0625325	.016597	.036884	.1040892
3	.0221981	.0116365	.0078758	.0609655

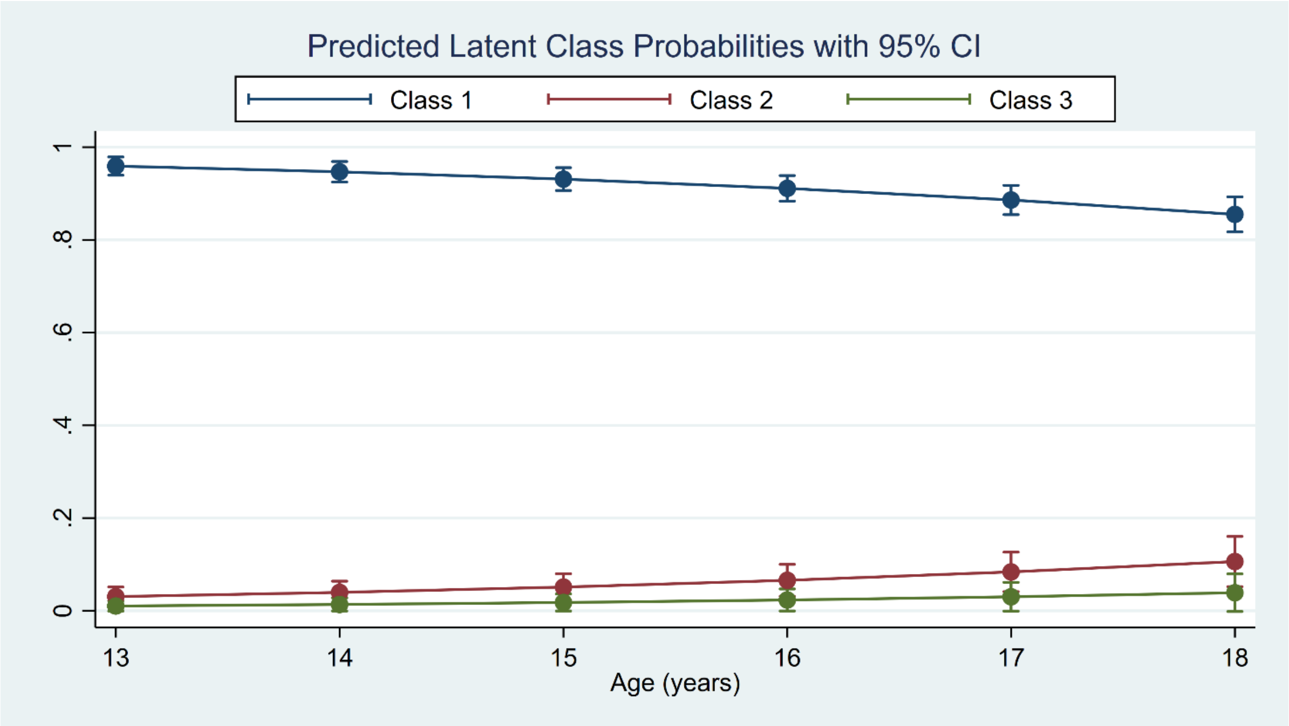
LCA With Covariates

```
. estat lcmean

Latent class marginal means                    Number of obs = 10,000
```

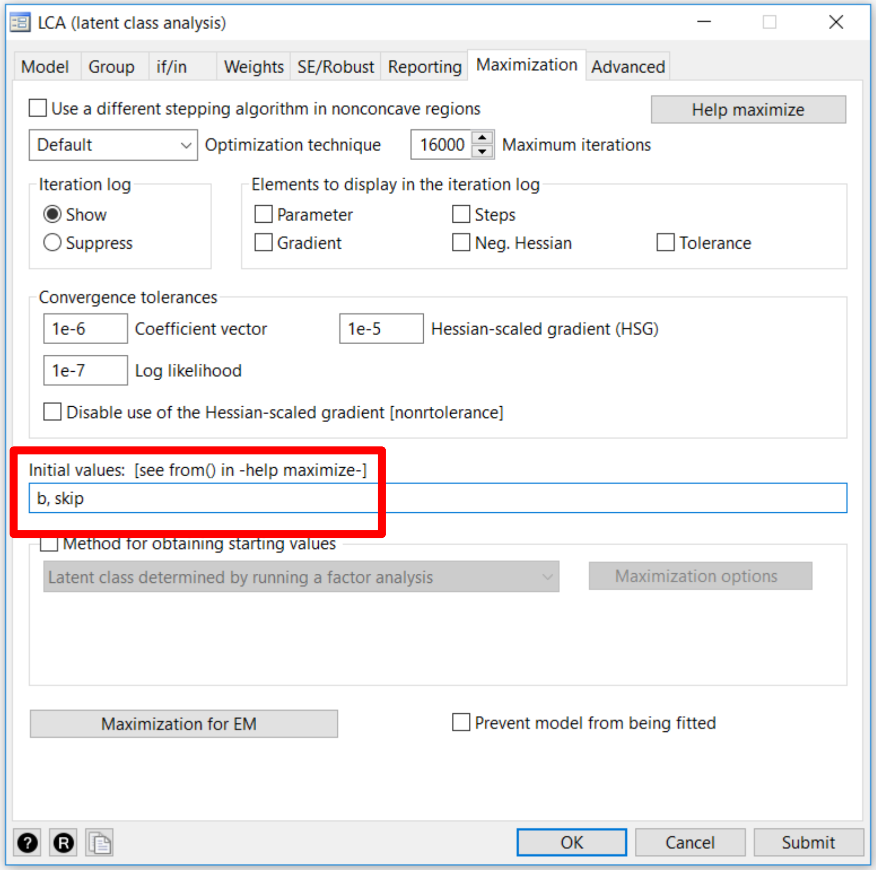
	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
1				
alcohol	.3177198	.0067361	.3046662	.3310664
truant	.055649	.0048065	.0469451	.0658553
vandalism	.0509588	.0032663	.0449244	.0577547
theft	.0108852	.0015597	.0082168	.0144076
weapon	.0125271	.0016357	.0096948	.0161733
2				
alcohol	.7622907	.0497867	.6517565	.8460291
truant	.530044	.0724128	.3894933	.6659856
vandalism	.2915297	.0641079	.1829752	.4305484
theft	3.46e-06	.002712	0	1
weapon	.1313218	.042565	.0678219	.2390298
3				
alcohol	.7357761	.0373358	.6564975	.8022686
truant	.4469371	.0417226	.3672857	.5294088
vandalism	.5463925	.0447149	.4582233	.6317444
theft	.8486219	.4398189	.0067831	.9997827
weapon	.3073444	.0374031	.2392267	.3850415

`margins` and `marginsplot`

```
margins, predict(classpr class(1))    ///
         predict(classpr class(2))    ///
         predict(classpr class(3))    ///
         at(age=(13(1)18))

marginsplot, title("Predicted Latent Class Probabilities with 95% CI")  ///
             legend(order(1 "Class 1" 2 "Class 2" 3 "Class 3")          ///
             rows(1) position(12) ring(1))   ytitle("")
```



Initial Values with the **`from()`** Option




```
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
     (C <- age), family(bernoulli) link(logit)            ///
     lclass(C 3) from(b, skip)
```

Initial Values with the `from()` Option

```
// Fit a model without age
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
      , family(bernoulli) link(logit)                     ///
      lclass(C 3)

// Store the parameter estimates in a matrix named b
matrix b = e(b)

// Fit a model including age using matrix b for the starting values
gsem (alcohol truant vandalism theft weapon <- _cons)    ///
      (C <- age), family(bernoulli) link(logit)           ///
      lclass(C 3)                                         ///
      from(b, skip)
```

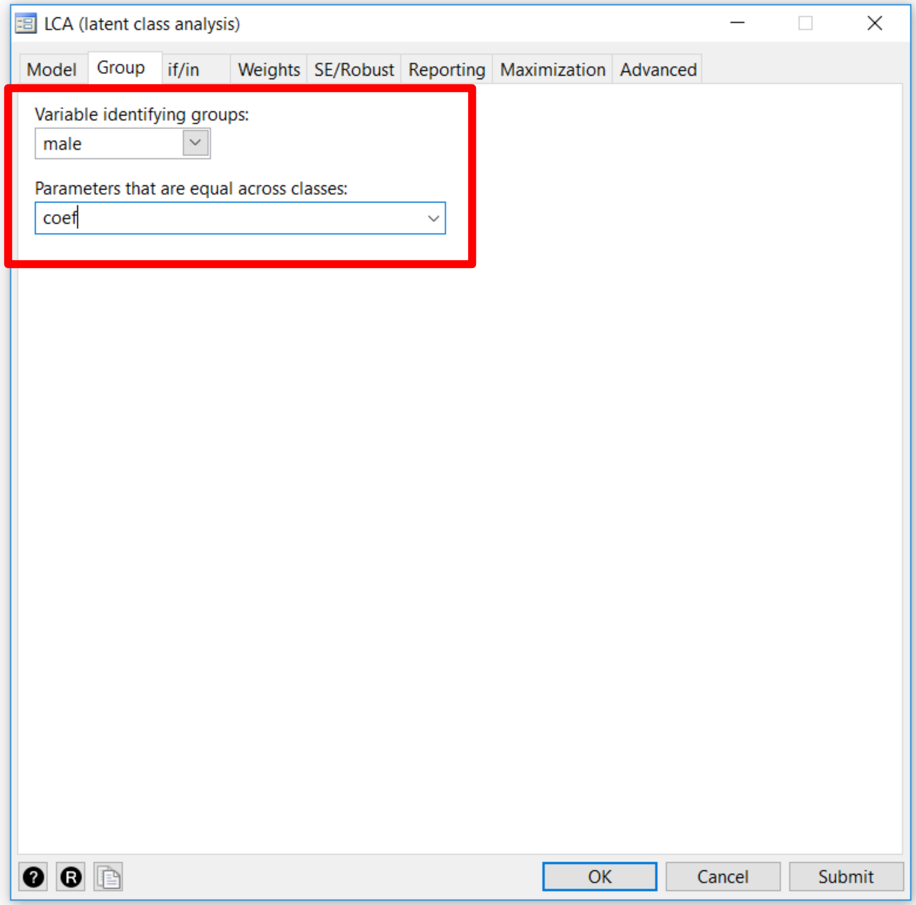
Outline

- Latent class analysis (LCA)
  - Estimation and postestimation options
  - **`margins`** and **`marginsplot`**
- Latent class analysis with covariates
- Latent class analysis by groups
- Latent profile analysis

LCA By Groups

```
. use atrisk.dta

. describe

Contains data from atrisk.dta
  obs:        10,000
 vars:             8                          20 Mar 2018 15:17
 size:        90,000
--------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------
id              int     %9.0g                 Student identification number
age             byte    %9.0g                 Age (years)
male            byte    %9.0g      male       Male
alcohol         byte    %9.0g                 Ever consumed alcohol
truant          byte    %9.0g                 >10 unexcused absences from school
vandalism       byte    %9.0g                 Ever engaged in an act of vandalism
theft           byte    %9.0g                 Ever stolen something worth more than $25
weapon          byte    %9.0g                 Ever used a weapon in a fight
--------------------------------------------------------------------------------
Sorted by: id
```

LCA By Groups




```
gsem (alcohol truant vandalism theft weapon <- _cons)     ///
     (C <- age), family(bernoulli) link(logit)            ///
     lclass(C 3) group(male) ginvariant(coef)
```

LCA By Groups

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.C	(base outcome)					
2.C						
male#c.age						
female	.1996802	.0642796	3.11	0.002	.0736945	.3256659
male	.1996802	.0642796	3.11	0.002	.0736945	.3256659
male						
female	-7.764133	1.374316	-5.65	0.000	-10.45774	-5.070524
male	-5.166688	1.259612	-4.10	0.000	-7.635483	-2.697893
3.C						
male#c.age						
female	.3405638	.0548289	6.21	0.000	.2331011	.4480265
male	.3405638	.0548289	6.21	0.000	.2331011	.4480265
male						
female	-8.4182	1.083083	-7.77	0.000	-10.541	-6.295396
male	-8.763309	.9569931	-9.16	0.000	-10.63898	-6.887637

LCA By Groups

```
. estat lcprob

Latent class marginal probabilities         Number of obs = 10,000
```

male	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
0				
C				
1	.9433078	.0113769	.9164295	.9619008
2	.00929	.010251	.0010557	.0768111
3	.0474022	.017409	.0228395	.0957914
1				
C				
1	.8587708	.0407295	.7589482	.9215296
2	.1115619	.0384299	.0554715	.2116591
3	.0296674	.0099256	.0153168	.0566891

```
. estat lcmean
```

Latent class marginal means Number of obs = 10,000

	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
0.male#1.C				
alcohol	.3024592	.0078545	.28729	.3180719
truant	.0467973	.0047443	.0383293	.0570251
vandalism	.0518493	.0039382	.0446519	.0601339
theft	.008876	.0019226	.0058015	.0135578
weapon	.009861	.0019113	.00674	.0144063
0.male#2.C				
alcohol	.8587083	.1022934	.5379601	.9694413
truant	.3125136	.1537263	.1005828	.6488477
vandalism	.6237619	.2253174	.20162	.9158532
theft	.811037	.366611	.0379878	.997861
weapon	.4970692	.2013821	.1693018	.8273751
0.male#3.C				
alcohol	.7052357	.0726249	.5467674	.825936
truant	.5974203	.1603976	.2865114	.8457754
vandalism	.3259337	.0611469	.2188929	.4548385
theft	.2014416	.0843749	.0827666	.4135572
weapon	.1799448	.057378	.0928923	.3198147
1.male#1.C				
alcohol	.326503	.0138507	.2999583	.3542082
truant	.0558936	.0115078	.0371758	.083221
vandalism	.0473091	.0058805	.0370294	.0602639
theft	.01436	.0024698	.0102431	.0200978
weapon	.0151411	.0028376	.0104769	.0218358
1.male#2.C				
alcohol	.7256983	.0832924	.5381264	.8572958
truant	.4556383	.091347	.2891031	.6327225
vandalism	.2130356	.0651992	.1121517	.3671416
theft	3.78e-07	.000225	0	1
weapon	.0785246	.0324125	.034207	.1701435
1.male#3.C				
alcohol	.7215414	.0501339	.6137411	.8086366
truant	.4624785	.0570091	.3543796	.5742239
vandalism	.6546292	.0627237	.523907	.7655226
theft	.6673	.2103064	.2385516	.9277502
weapon	.2956006	.0507453	.2065255	.4035547

Outline

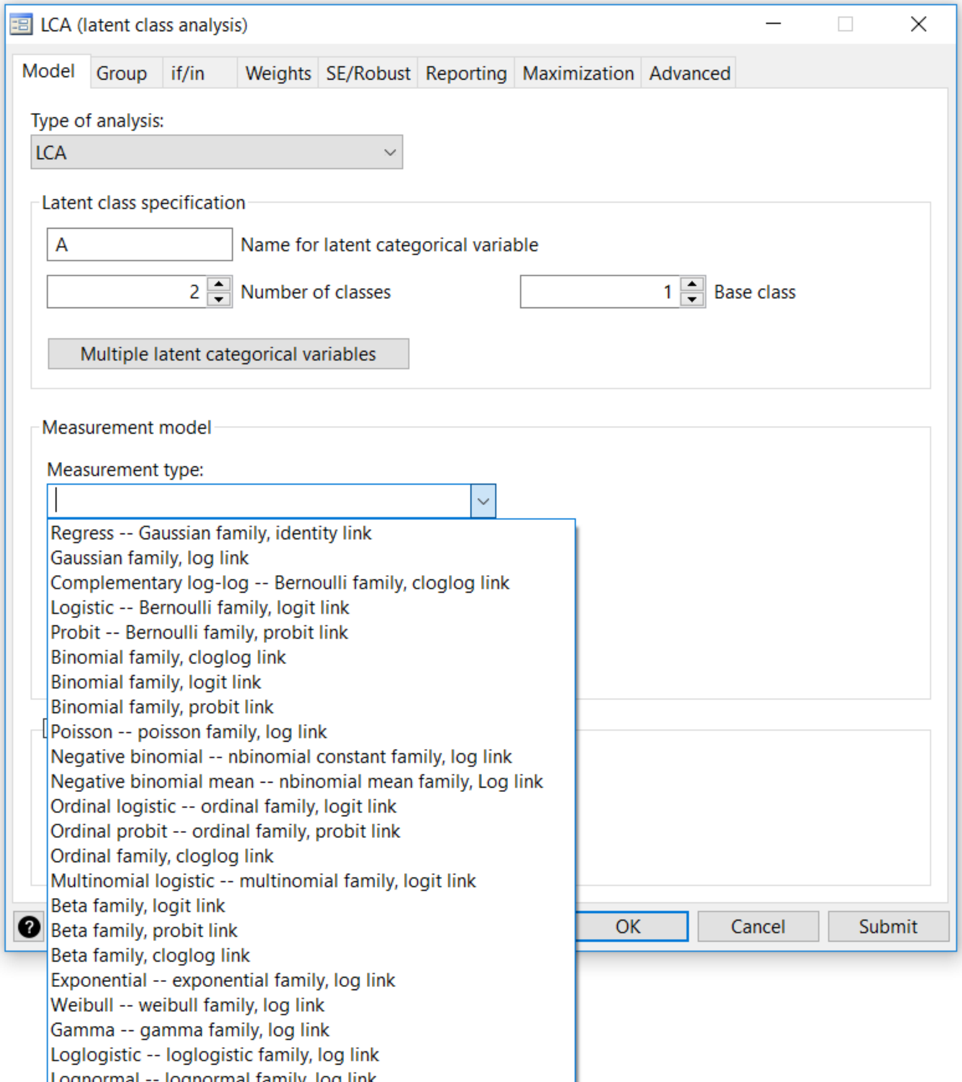
- Latent class analysis (LCA)
  - Estimation and postestimation options
  - **`margins`** and **`marginsplot`**
- Latent class analysis with covariates
- Latent class analysis by groups
- Latent profile analysis

Latent and Observed Variables

	Continuous Latent	Categorical Latent
Continuous Observed	Factor analysis	Latent Profile Analysis
Categorical Observed	Latent trait analysis or Item Response Theory	Latent Class Analysis

Reproduced from Table 1.1 in Collins and Lanza (2010)

Latent Profile Analysis



Latent Profile Analysis

Diabetologia 16, 17–24 (1979)

Diabetologia

An Attempt to Define the Nature of Chemical Diabetes Using a Multidimensional Analysis

G. M. Reaven and R. G. Miller

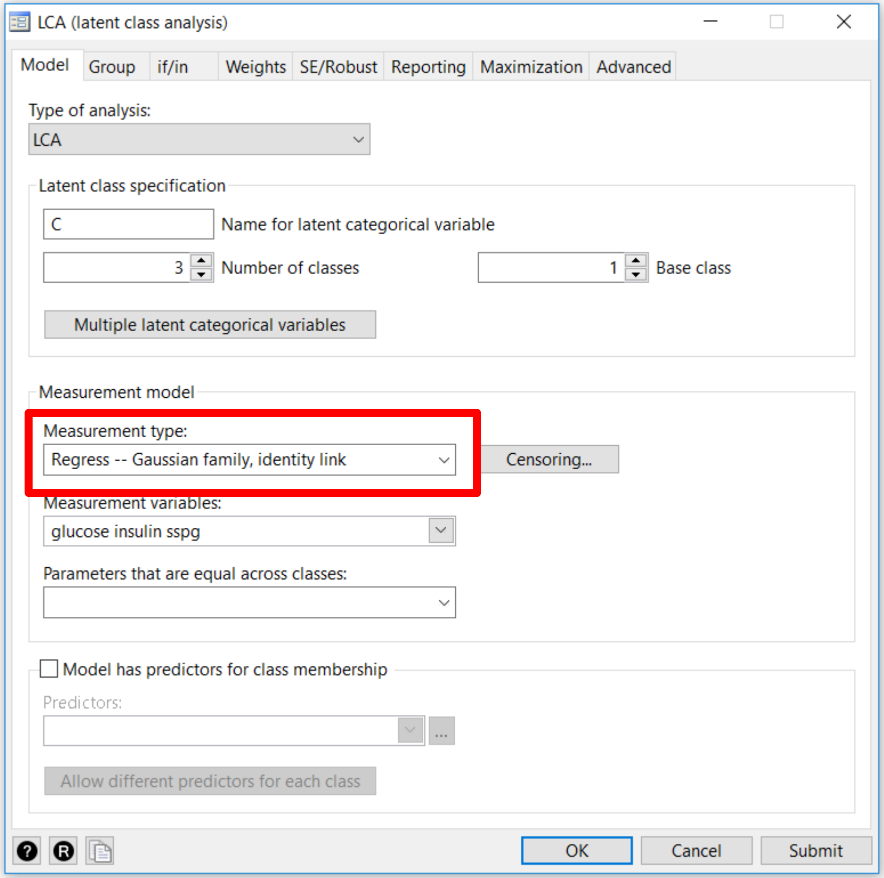
Departments of Medicine and Statistics, Stanford University and Veterans Administration Hospital, Palo Alto, California, USA

Summary. The relationship between chemical diabetes and overt diabetes in 145 nonobese adult subjects has been examined. The degree of glucose intolerance, insulin response to oral glucose, and insulin resistance in normal subjects and patients with nonketotic diabetes were first determined. These variables were then analyzed by a computer program which permits direct visualization of the three-dimensional shape of the data set. The picture obtained was that of a boomerang with two wings and a fat middle. The patients were then divided into three groups (normal, chemical diabetes, overt diabetes) on the basis of their oral glucose tolerance. Two-dimensional views of the relationship between the metabolic variables in these subjects resembled the initial three-dimensional image, i. e., a central core with two wings going off in opposite directions. The two wings represented patients with chemical diabetes and overt diabetes, respectively. Following this, the patients were reclassified on the basis of all three metabolic variables (plasma glucose response to oral glucose, plasma insulin response to oral glucose, and degree of insulin resistance) by means of a computer classification which employed a cluster analysis technique. This again resulted in the definition of three groups, in which there was a divergence between subjects classified as having chemical diabetes as contrasted to overt diabetes. This apparent separation between subjects with chemical and overt diabetes may explain why patients with chemical diabetes rarely develop overt diabetes.

Latent Profile Analysis

```
. webuse gsem_lca2
(Latent profile analysis)


. describe

Contains data from http://www.stata-press.com/data/r15/gsem_lca2.dta
  obs:           145                          Latent profile analysis
 vars:             7                          18 Jan 2017 12:39
 size:         3,045                          (_dta has notes)
--------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------
patient         int     %9.0g                 Patient ID
relwgt          float   %9.0g                 Relative weight
fglucose        int     %9.0g                 Fasting plasma glucose
glucose         float   %9.0g                 Glucose area (mg/10mL/hr)
insulin         float   %9.0g                 Insulin area (mIU/10mL/hr)
sspg            float   %9.0g                 Steady-state plasma glucose
cclass          byte    %17.0g     class      Clinical classification
--------------------------------------------------------------------------
```

Latent Profile Analysis




```
gsem (glucose insulin sspg <- _cons),     ///
     family(gaussian) link(identity)      ///
     lclass(C 3)
```

Latent Profile Analysis

```
. estat lcprob

Latent class marginal probabilities             Number of obs = 145
```

	Margin	Delta-method Std. Err.	[95% Conf.	Interval]
C				
1	.5289099	.0481727	.4345859	.6212135
2	.2899368	.045273	.2097056	.3858759
3	.1811533	.0322278	.126267	.2529898

Latent Profile Analysis

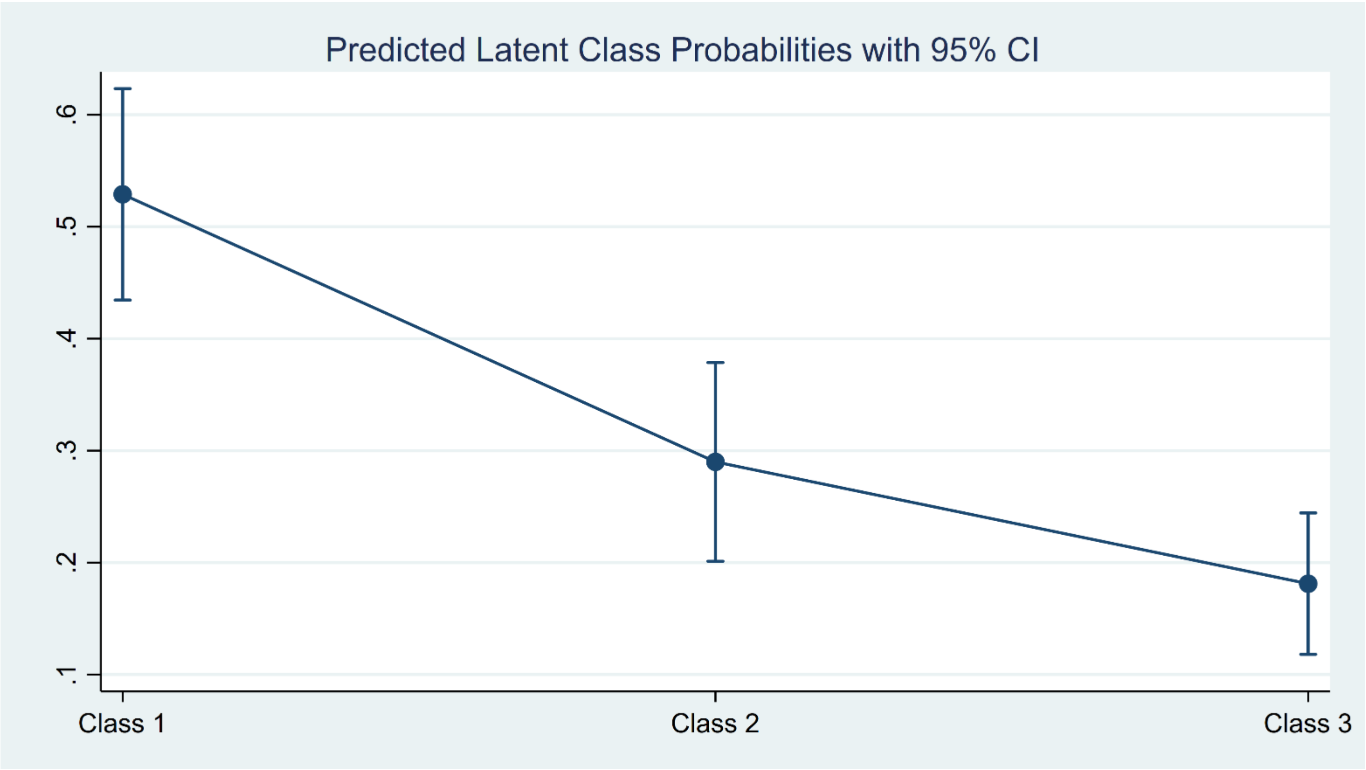
```
. estat lcmean
```

Latent class marginal means   Number of obs   =   145

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1						
glucose	35.6516	.5603517	63.62	0.000	34.55333	36.74987
insulin	16.36279	.6088787	26.87	0.000	15.16941	17.55617
sspg	10.68416	.6630563	16.11	0.000	9.38459	11.98372
2						
glucose	50.75437	1.754034	28.94	0.000	47.31653	54.19221
insulin	29.61048	2.578857	11.48	0.000	24.55602	34.66495
sspg	22.69286	.8822253	25.72	0.000	20.96373	24.42199
3						
glucose	114.761	5.00265	22.94	0.000	104.956	124.566
insulin	7.574333	.8903958	8.51	0.000	5.829189	9.319477
sspg	34.17127	1.573196	21.72	0.000	31.08786	37.25468

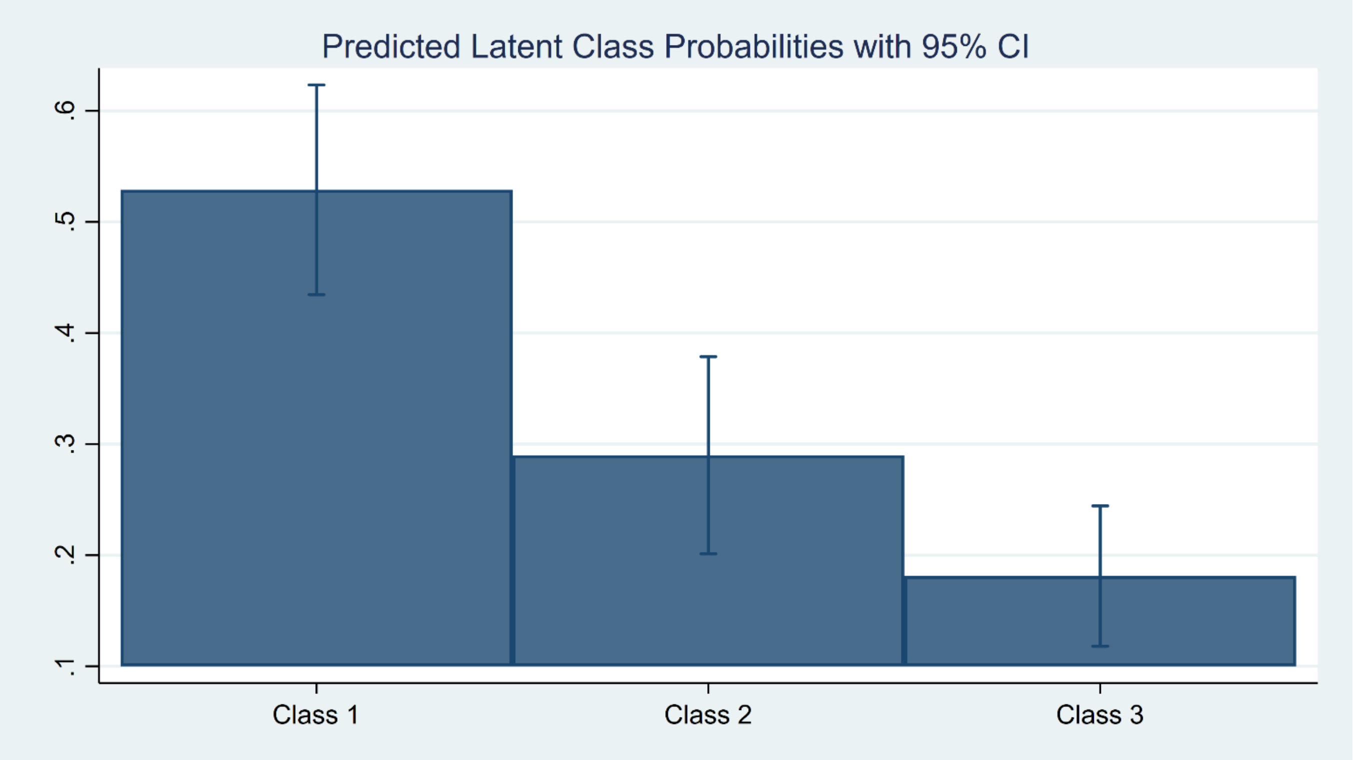
Latent Profile Analysis

```
margins, predict(classpr class(1))    ///
         predict(classpr class(2))    ///
         predict(classpr class(3))

marginsplot, xtitle("") ytitle("")                                  ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")            ///
             title("Predicted Latent Class Probabilities with 95% CI")
```



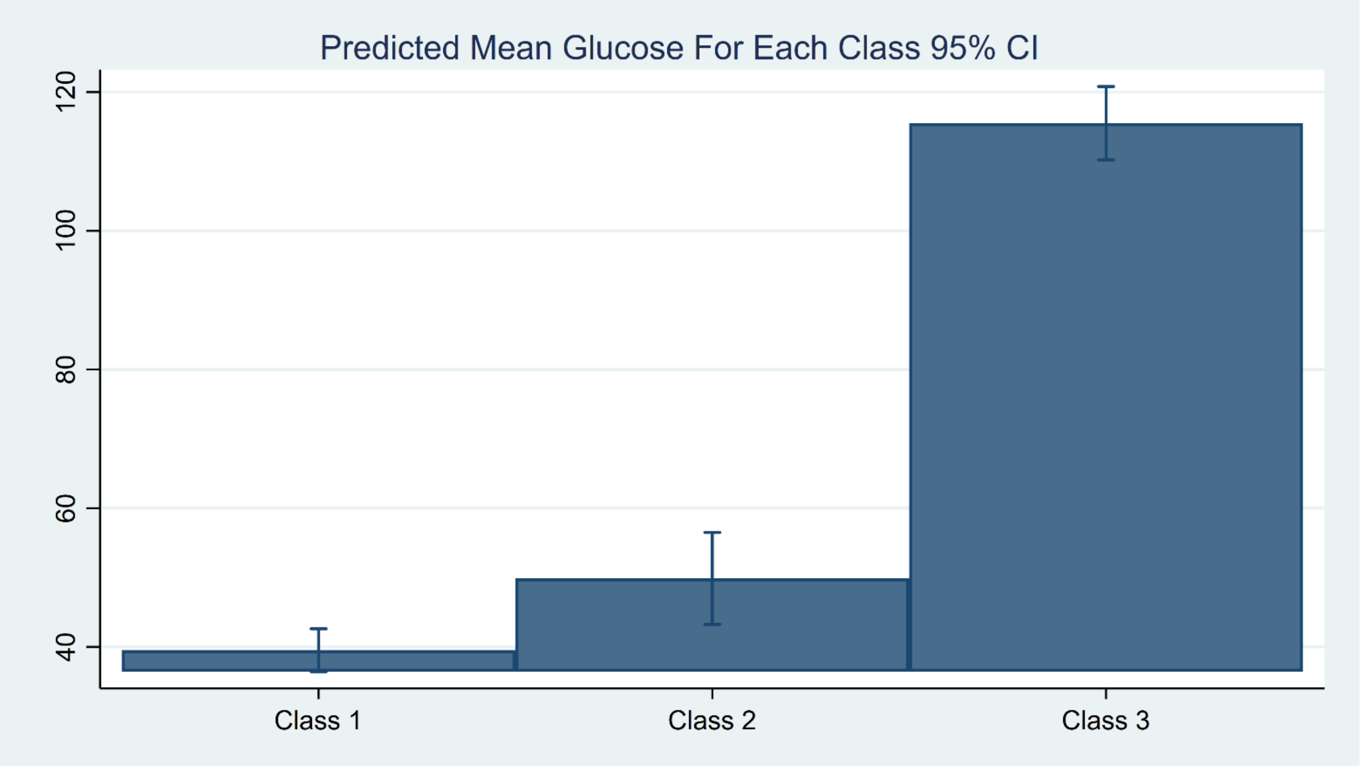
Latent Profile Analysis

```
margins, predict(classpr class(1))   ///
         predict(classpr class(2))   ///
         predict(classpr class(3))

marginsplot, recast(bar) xtitle("") ytitle("")                        ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")              ///
             title("Predicted Latent Class Probabilities with 95% CI")
```



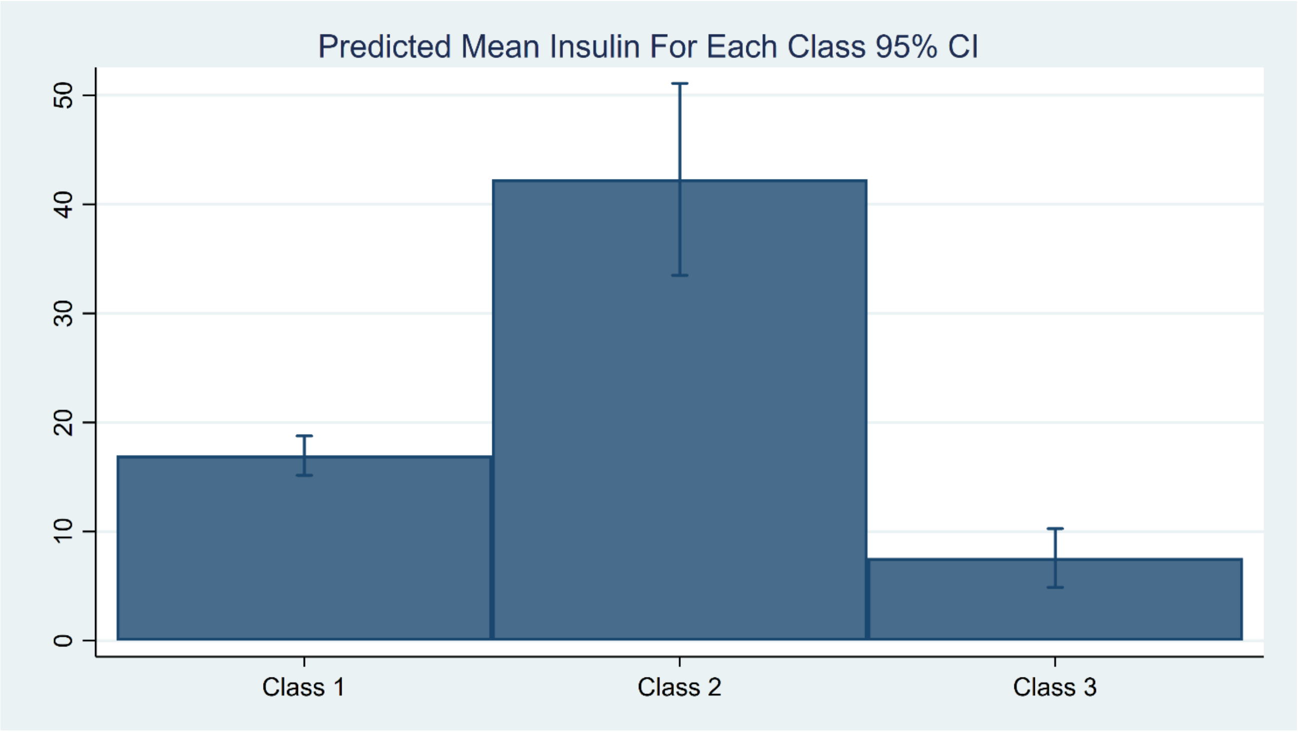
Latent Profile Analysis

```
margins, predict(outcome(glucose)   class(1))        ///
         predict(outcome(glucose)   class(2))        ///
         predict(outcome(glucose)   class(3))

marginsplot, recast(bar) xtitle("") ytitle("")                ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")     ///
             title("Predicted Mean Glucose For Each Class 95% CI")
```



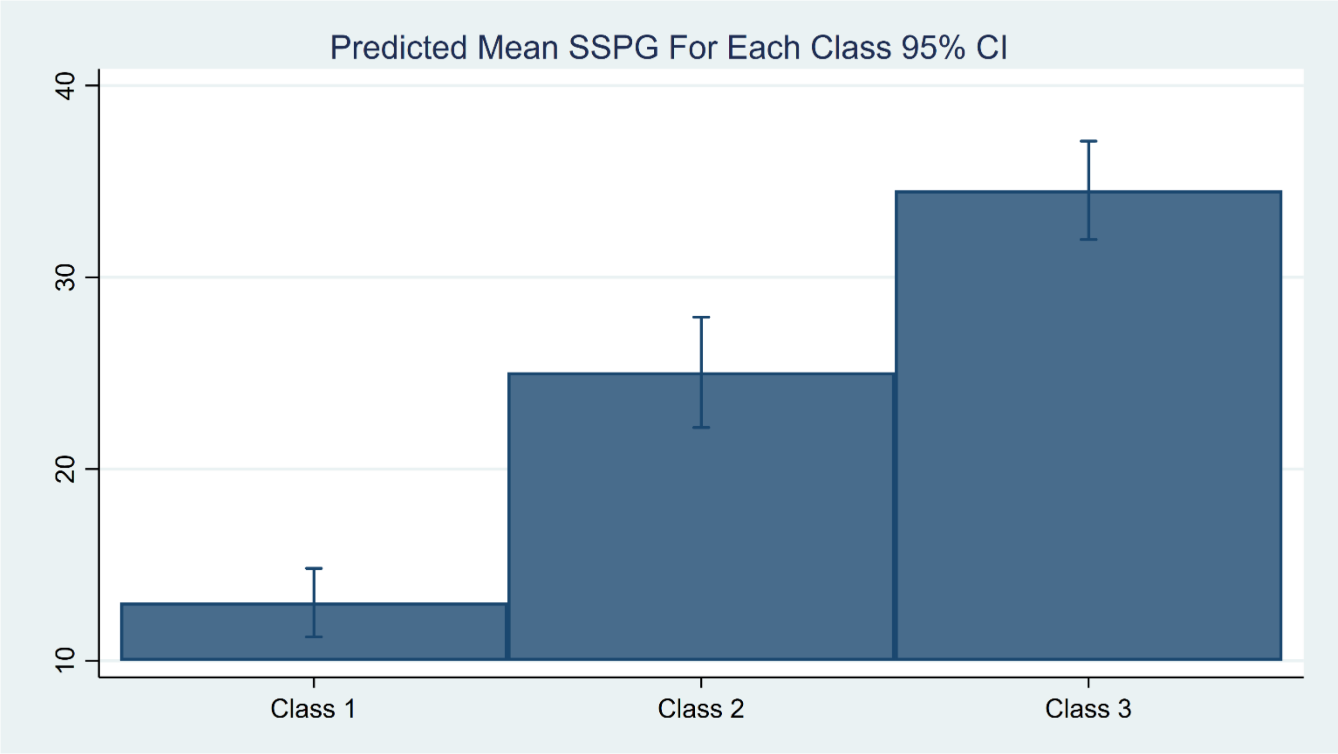
Latent Profile Analysis

```
margins, predict(outcome(insulin)   class(1))       ///
         predict(outcome(insulin)   class(2))       ///
         predict(outcome(insulin)   class(3))

marginsplot, recast(bar) xtitle("") ytitle("")               ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")     ///
             title("Predicted Mean Insulin For Each Class 95% CI")
```



Latent Profile Analysis

```
margins, predict(outcome(sspg)   class(1))       ///
         predict(outcome(sspg)   class(2))       ///
         predict(outcome(sspg)   class(3))

marginsplot, recast(bar) xtitle("") ytitle("")              ///
             xlabel(1 "Class 1" 2 "Class 2" 3 "Class 3")    ///
             title("Predicted Mean SSPG For Each Class 95% CI")
```



Summary

- Latent class analysis (LCA)
  - Estimation and postestimation options
  - **`margins`** and **`marginsplot`**
- Latent class analysis with covariates
- Latent class analysis by groups
- Latent profile analysis

Thanks for Coming!

Questions?

You can download the slides, dataset, and do-file here:
https://tinyurl.com/2019LCA

You can contact me at:
chuber@stata.com