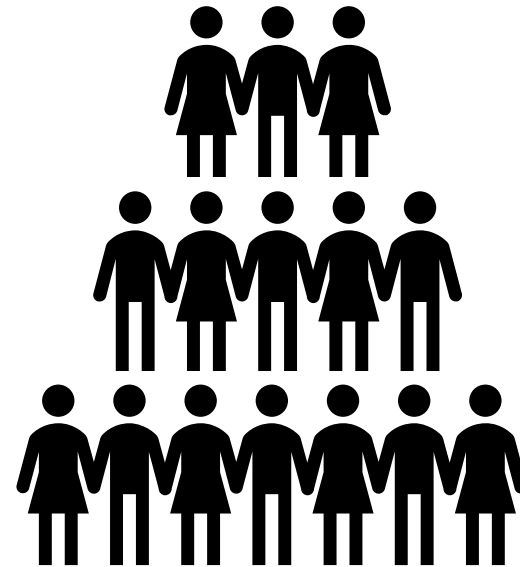


Latent class analysis in Stata

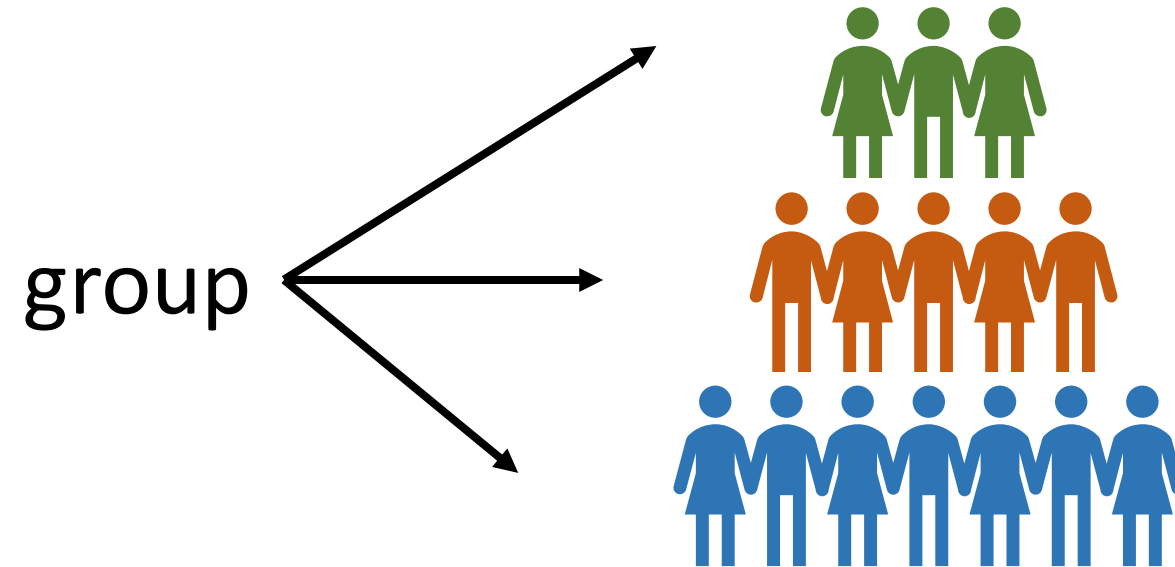
Meghan K. Cain, Ph.D.

April 29, 2025

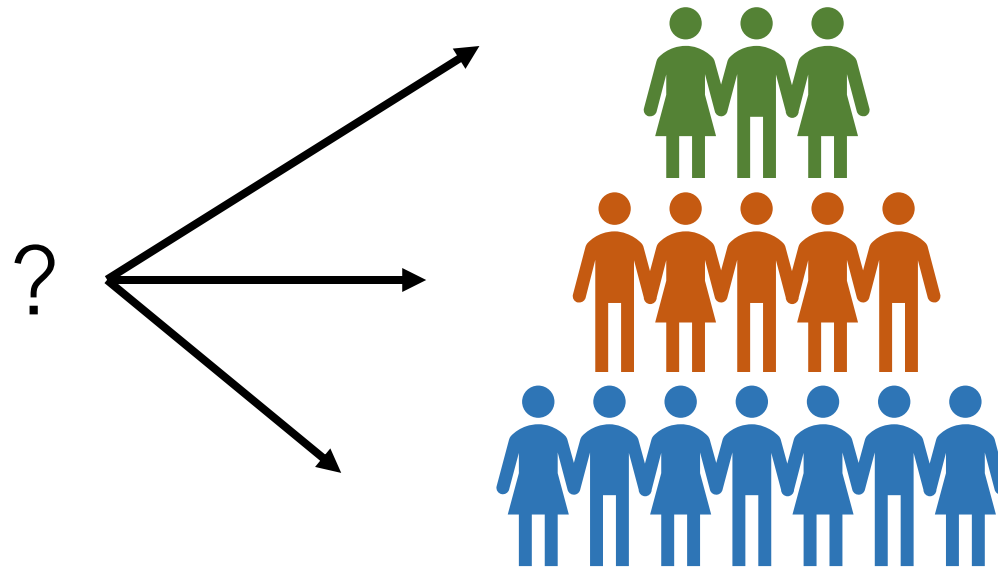
Population characteristics



Population characteristics

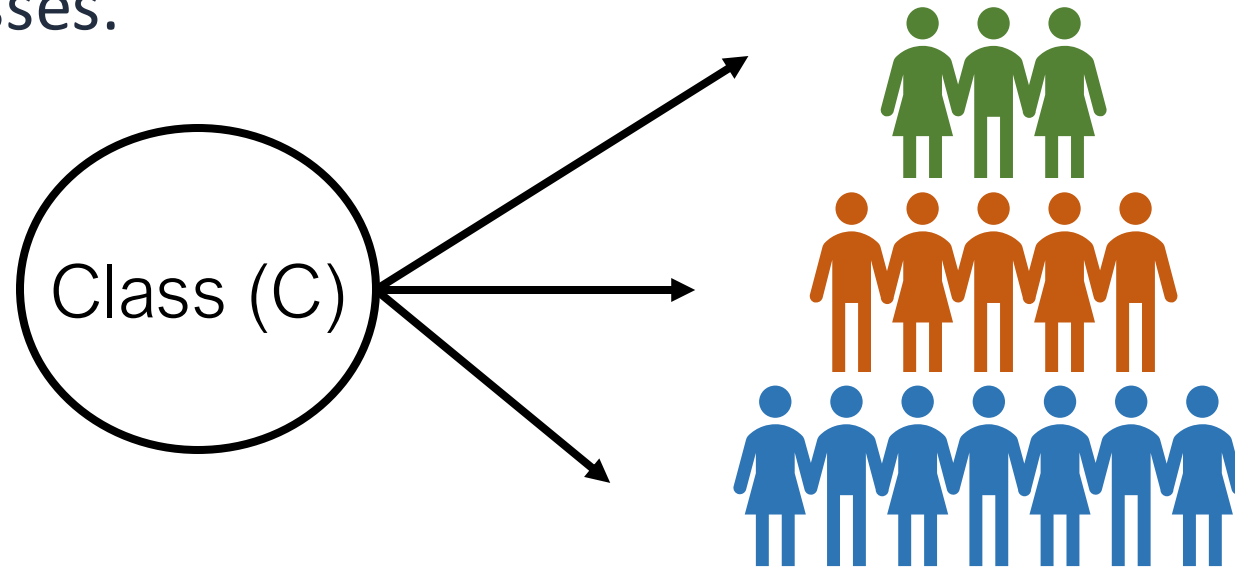


Population characteristics



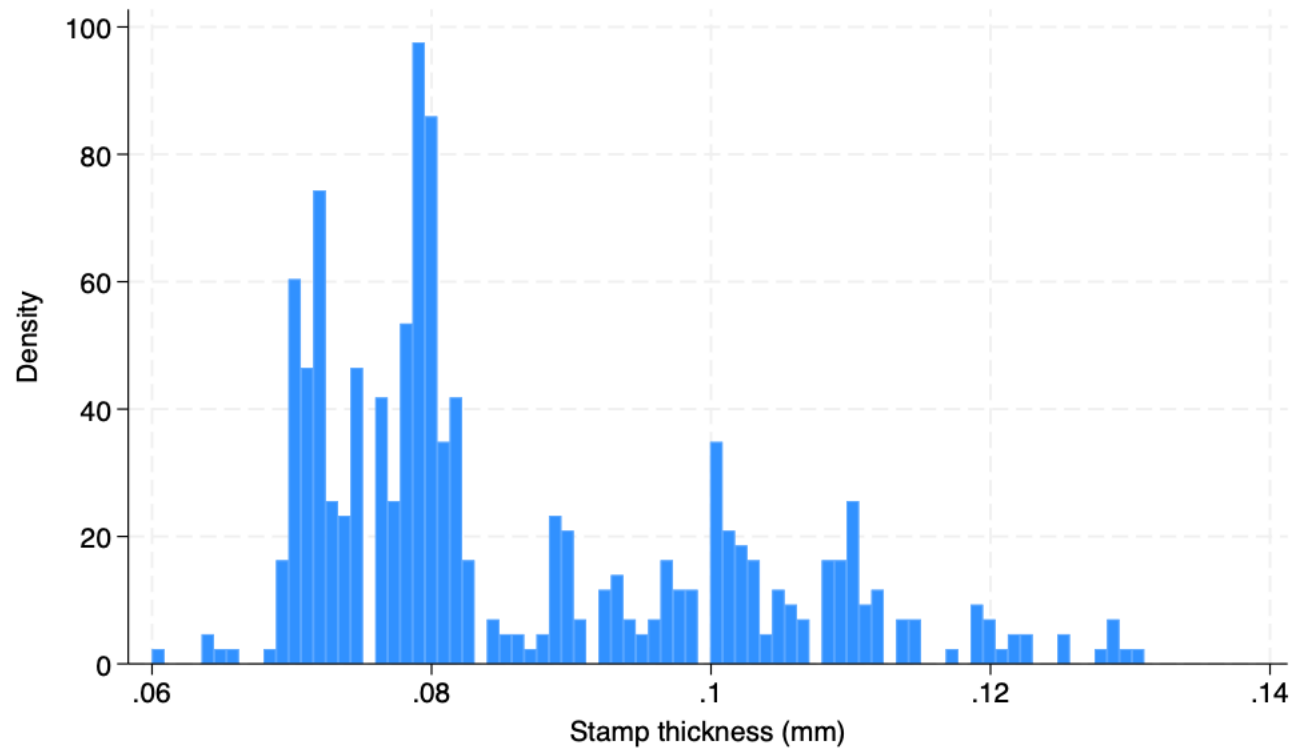
Latent class analysis (LCA)

- We use a categorical latent variable to represent unobserved groups in the population that we call classes.



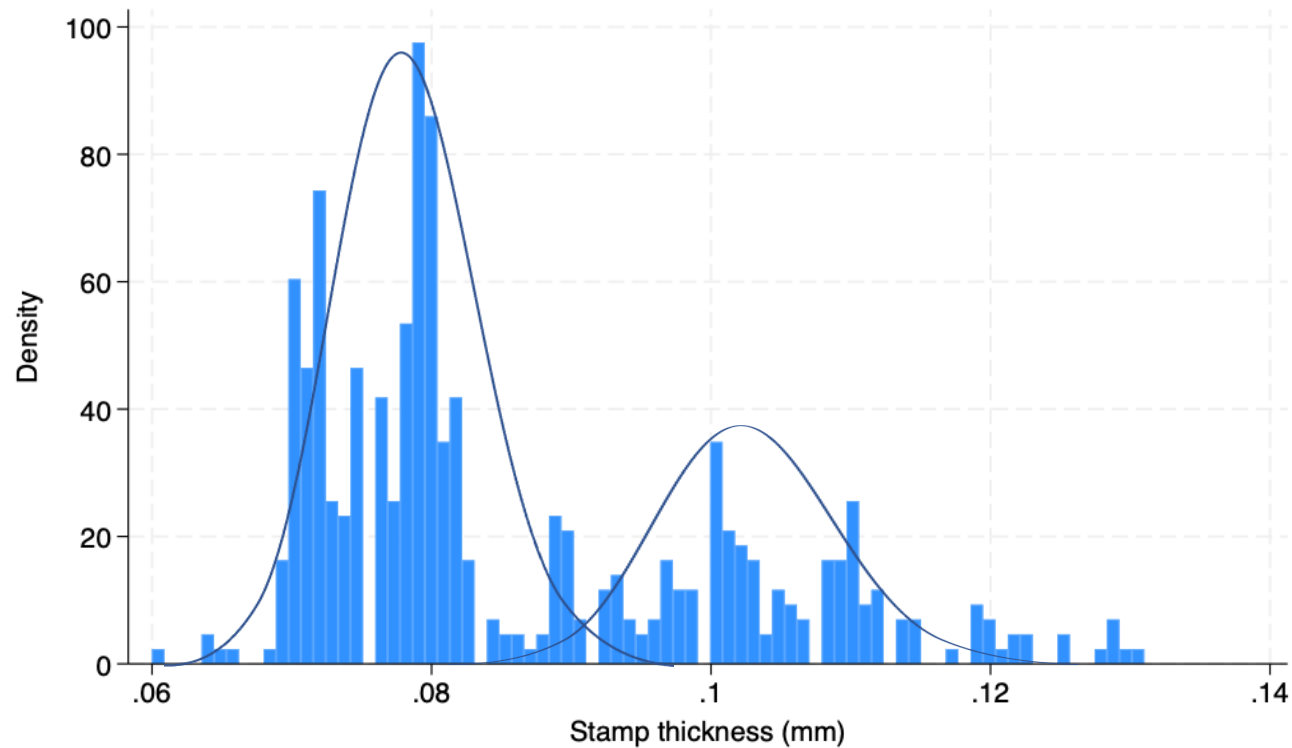
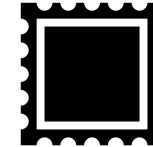
Basic mixture model

```
. use stamp, clear  
(1872 Hidalgo stamp of Mexico)  
. histogram thickness, bin(80)  
(bin=80, start=.06, width=.0008875)
```



Basic mixture model

```
. use stamp, clear  
(1872 Hidalgo stamp of Mexico)  
  
. histogram thickness, bin(80)  
(bin=80, start=.06, width=.0008875)
```



Mixture model in Stata

```
. fmm 2: regress thickness
```


Mixture model in Stata

```
. fmm 2: regress thickness  
  
. gsem (thickness <- ), lclass (C 2)
```

Class: 1

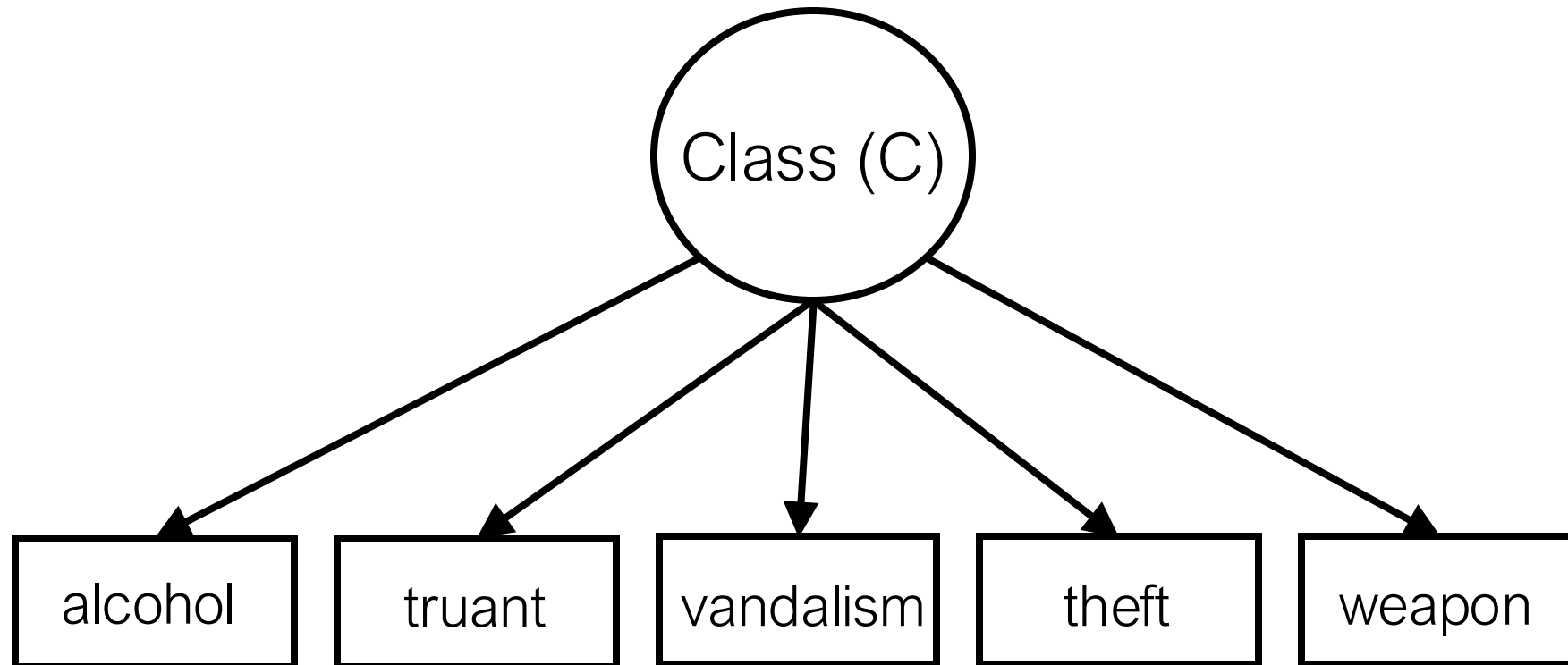
	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
thickness						
_cons	.0776762	.00043	180.65	0.000	.0768334	.0785189
var(e.thickness)	.0000518	3.72e-06			.000045	.0000596

Class: 2

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
thickness						
_cons	.1065861	.0007066	150.84	0.000	.1052011	.107971
var(e.thickness)	.0000518	3.72e-06			.000045	.0000596

Latent class analysis

- LCA identifies subpopulations (latent classes) by finding patterns in a set of variables (indicators).



Latent class analysis

- LCA identifies subpopulations (latent classes) by finding patterns in a set of variables (indicators).
- While factor analysis uses continuous latent variables to group indicators together, LCA uses categorical latent variables to group people together.

Latent class analysis

- LCA identifies subpopulations (latent classes) by finding patterns in a set of variables (indicators).
- While factor analysis uses **continuous** latent variables to group indicators together, LCA uses **categorical** latent variables to group people together.

Latent class analysis

- LCA identifies subpopulations (latent classes) by finding patterns in a set of variables (indicators).
- While factor analysis uses continuous latent variables to **group indicators** together, LCA uses categorical latent variables to **group people** together.

Latent class analysis

- LCA identifies subpopulations (latent classes) by finding patterns in a set of variables (indicators).
- While factor analysis uses continuous latent variables to group indicators together, LCA uses categorical latent variables to group people together.
- Unlike cluster analysis, LCA is a model that can be evaluated for fit.

Adolescent delinquency

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 Abdallah³

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. © 2010 Wiley-Liss, Inc.

Adolescent delinquency

```
. use atrisk, clear
```

```
. describe
```

Contains data from **atrisk.dta**

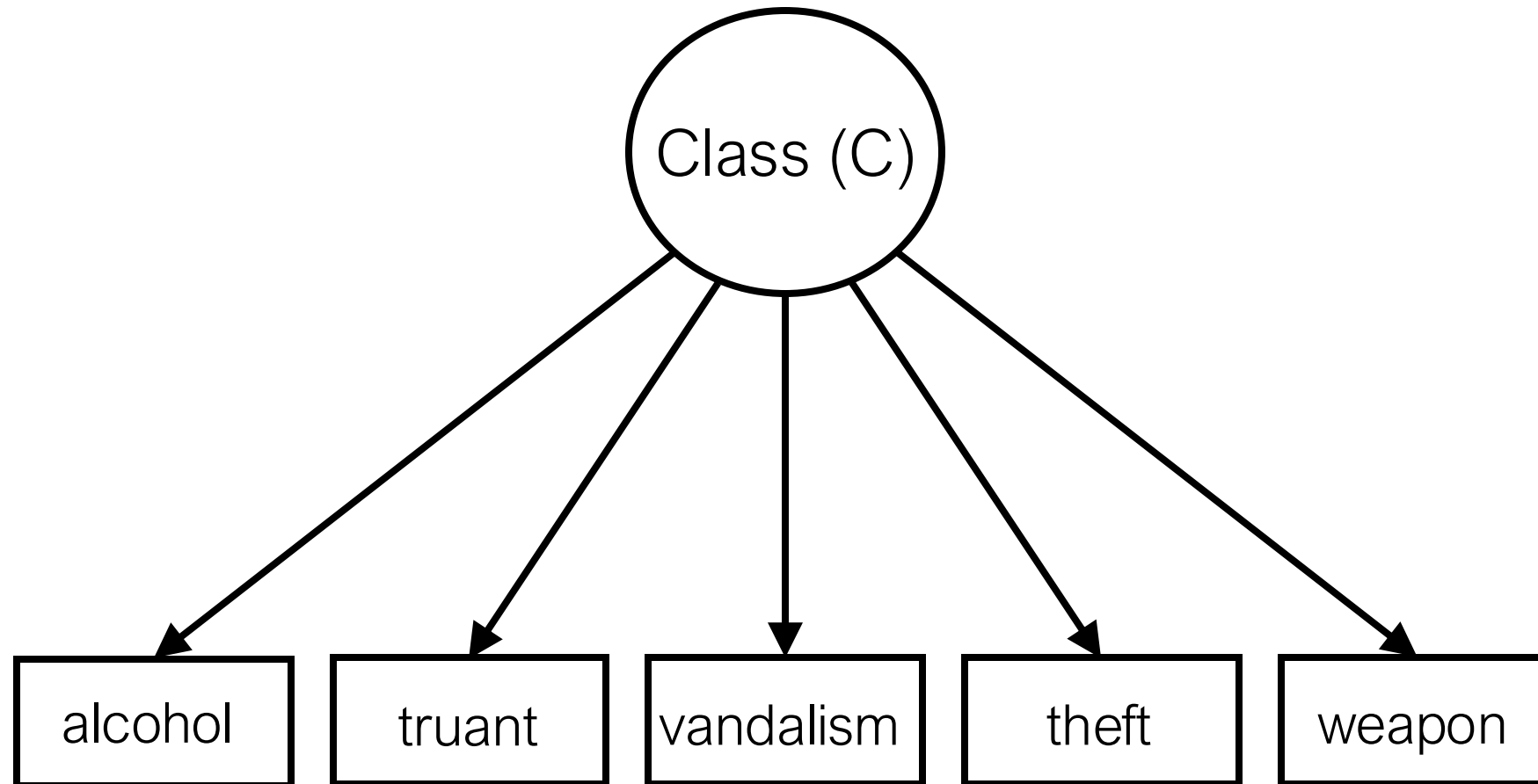
Observations: **10,000**

Variables: **8**

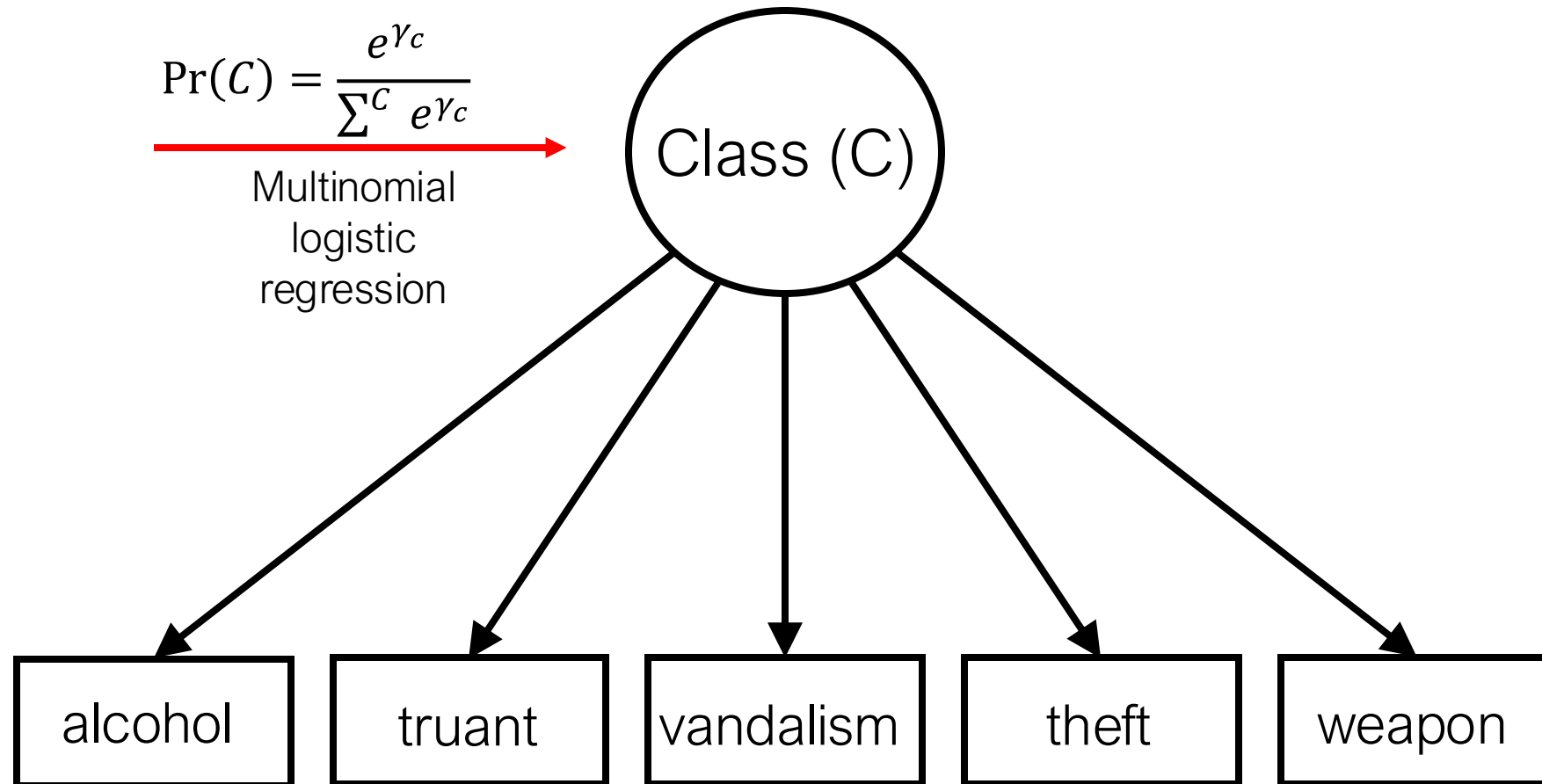
20 Mar 2018 15:17

Variable name	Storage type	Display format	Value 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

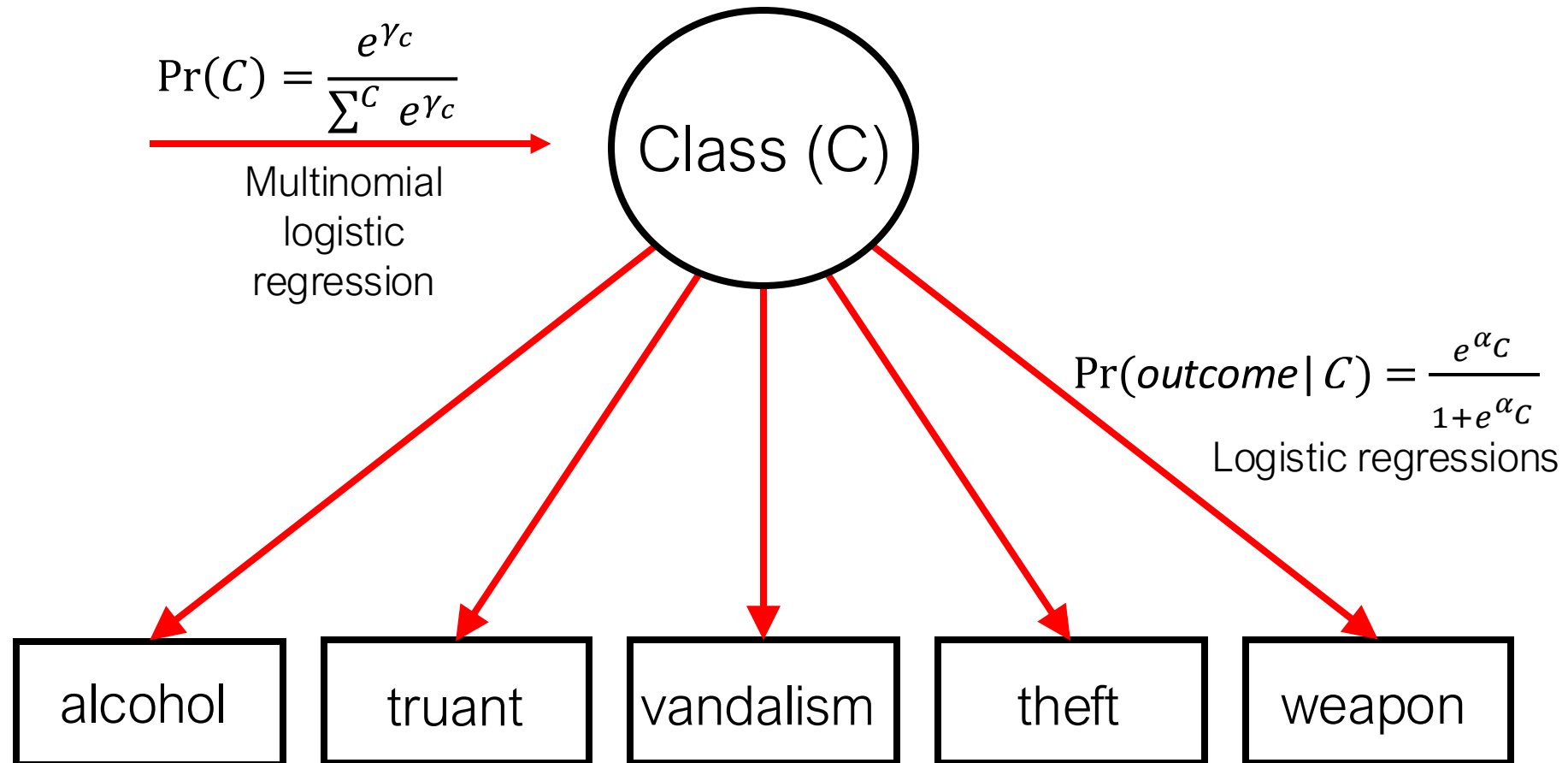
Path diagram



Model formulation



Model formulation



LCA with gsem

```
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 3)
```

Generalized structural equation model

Number of obs = **10,000**

Log likelihood = **-14411.803**

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1.C	(base outcome)					
2.C						
_cons	-2.412874	.542322	-4.45	0.000	-3.475806	-1.349943
3.C						
_cons	-3.679467	.486031	-7.57	0.000	-4.63207	-2.726864

Class probabilities

```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = **10,000**

	Delta-method			
	Margin	std. err.	[95% conf. interval]	
C				
1	.8970265	.0359702	.8024021	.9492062
2	.0803361	.0407568	.0287766	.2047962
3	.0226374	.011278	.0084565	.0591794

Class 1 coefficients

Class: 1

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
alcohol _cons	-.7861217	.0489699	-16.05	0.000	-.882101	-.6901424
truant _cons	-2.937357	.1906028	-15.41	0.000	-3.310931	-2.563782
vandalism _cons	-2.952382	.0999923	-29.53	0.000	-3.148363	-2.756401
theft _cons	-4.516348	.1614547	-27.97	0.000	-4.832793	-4.199902
weapon _cons	-4.427131	.1749286	-25.31	0.000	-4.769985	-4.084277

Class 2 coefficients

Class: 2

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
alcohol _cons	.8946583	.4753518	1.88	0.060	-.0370141	1.826331
truant _cons	-.0701591	.4882486	-0.14	0.886	-1.027109	.8867906
vandalism _cons	-1.155141	.3710583	-3.11	0.002	-1.882402	-.4278802
theft _cons	-3.3081	2.030692	-1.63	0.103	-7.288184	.671984
weapon _cons	-2.167753	.4838215	-4.48	0.000	-3.116026	-1.21948

Class 3 coefficients

Class: 3

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
alcohol _cons	1.104437	.2814714	3.92	0.000	.5527633	1.656111
truant _cons	-.2185509	.201888	-1.08	0.279	-.6142441	.1771422
vandalism _cons	.3430655	.5011581	0.68	0.494	-.6391862	1.325317
theft _cons	.9379999	1.133868	0.83	0.408	-1.28434	3.16034
weapon _cons	-.6707211	.412806	-1.62	0.104	-1.479806	.1383638

Class marginal means

```
. estat lcmean
```

Latent class marginal means

Number of obs = **10,000**

		Delta-method		
		Margin	std. err.	[95% conf. interval]
1	alcohol	.3130014	.0105677	.292671 .3340769
	truant	.0503374	.0091374	.0351619 .0715765
	vandalism	.0496237	.0047319	.0411291 .0597635
	theft	.0108104	.001729	.0078975 .0147815
	weapon	.0118075	.0020457	.0084026 .0165691
2	alcohol	.7098509	.0982563	.4899118 .8617235
	truant	.4824624	.1225526	.262666 .7092613
	vandalism	.2395505	.0676341	.1320636 .394734
	theft	.0352751	.0698457	.0006543 .671288
	weapon	.1026828	.0446039	.0424289 .2281197
3	alcohol	.7510949	.0527702	.6344201 .8399282
	truant	.4455853	.0498903	.3510688 .5442083
	vandalism	.5849456	.1227232	.3435264 .7914747
	theft	.7186878	.2287942	.2175413 .9591436
	weapon	.3383428	.0931645	.1844707 .5361794

Class enumeration

- When deciding on the number of classes, it's important to consider statistical fit as well as substantive interpretability.
- Start with one class, then increase the number of classes until you can't estimate any more.

Options for starting values

startvalues() option	Description
factor	runs a factor analysis on all observed variables to obtain preliminary class predictions
randomid, draws(#)	randomly assigns observations to initial classes
randompr, draws(#)	randomly assigns initial class probabilities
jitter, draws(#)	randomly perturbs starting values from a Gaussian approximation to each outcome
classid <i>varname</i>	specifies a variable that identifies the initial class membership for each case
classpr <i>varlist</i>	specifies a list of variables that give the probability of membership in each class

Options for starting values

startvalues() option	Description
factor	runs a factor analysis on all observed variables to obtain preliminary class predictions
randomid, draws(#)	randomly assigns observations to initial classes
randompr, draws(#)	randomly assigns initial class probabilities
jitter, draws(#)	randomly perturbs starting values from a Gaussian approximation to each outcome
classid <i>varname</i>	specifies a variable that identifies the initial class membership for each case
classpr <i>varlist</i>	specifies a list of variables that give the probability of membership in each class

iterate(#) - set the maximum number of iterations

Class enumeration

```
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 1)
. estimates store c1
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 2)
. estimates store c2
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 3)
. estimates store c3
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 4)
. estimates store c4
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 5) startvalues(randompr, draws(20)) iter(1000)
convergence not achieved
```

Class enumeration

```
. lcstats c*
```

Latent class statistics

	Classes	N	ll	Rank	Entropy	df	LMR	P>LMR
c1	1	10,000	-14,863.47	5				
c2	2	10,000	-14,430.33	11	0.8152	6	850.88	<0.001
c3	3	10,000	-14,411.80	17	0.7926	6	36.39	0.012
c4	4	10,000	-14,404.05	23	0.7206	6	15.22	0.004

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.
Likelihood-ratio tests compare the given model versus the same model with one less latent class.

Class enumeration

```
. lcstats c*
```

Latent class statistics

	Classes	N	ll	Rank	Entropy	df	LMR	P>LMR
c1	1	10,000	-14,863.47	5				
c2	2	10,000	-14,430.33	11	0.8152	6	850.88	<0.001
c3	3	10,000	-14,411.80	17	0.7926	6	36.39	0.012
c4	4	10,000	-14,404.05	23	0.7206	6	15.22	0.004

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.
Likelihood-ratio tests compare the given model versus the same model with one less latent class.

. lcstats c*, all split

Latent class statistics

	N	Rank	AIC	BIC	AICc	CAIC	Entropy
c1	10,000	5	29,736.93	29,772.98	29,736.94	29,777.98	
c2	10,000	11	28,882.66	28,961.97	28,882.68	28,972.97	0.8152
c3	10,000	17	28,857.61	28,980.18	28,857.67	28,997.18	0.7926
c4	10,000	23	28,854.11	29,019.94	28,854.22	29,042.94	0.7206

AIC is the Akaike information criterion.

BIC is the Bayesian information criterion.

AICc is the corrected Akaike information criterion.

CAIC is the consistent Akaike information criterion.

BIC, AICc, and CAIC use N = number of observations.

	Classes	ll	df	VLMR	P>VLMR	LMR	P>LMR
c1	1	-14,863.47					
c2	2	-14,430.33	6	866.28	<0.001	850.88	<0.001
c3	3	-14,411.80	6	37.05	0.012	36.39	0.012
c4	4	-14,404.05	6	15.50	0.004	15.22	0.004

VLMR is the Vuong-Lo-Mendell-Rubin likelihood-ratio test statistic.

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.

Likelihood-ratio tests compare the given model versus the same




```
. lcstats c*, all split
```

Latent class statistics

	N	Rank	AIC	BIC	AICc	CAIC	Entropy
c1	10,000	5	29,736.93	29,772.98	29,736.94	29,777.98	
c2	10,000	11	28,882.66	28,961.97	28,882.68	28,972.97	0.8152
c3	10,000	17	28,857.61	28,980.18	28,857.67	28,997.18	0.7926
c4	10,000	23	28,854.11	29,019.94	28,854.22	29,042.94	0.7206

AIC is the Akaike information criterion.

BIC is the Bayesian information criterion.

AICc is the corrected Akaike information criterion.

CAIC is the consistent Akaike information criterion.

BIC, AICc, and CAIC use N = number of observations.

	Classes	ll	df	VLMR	P>VLMR	LMR	P>LMR
c1	1	-14,863.47					
c2	2	-14,430.33	6	866.28	<0.001	850.88	<0.001
c3	3	-14,411.80	6	37.05	0.012	36.39	0.012
c4	4	-14,404.05	6	15.50	0.004	15.22	0.004

VLMR is the Vuong-Lo-Mendell-Rubin likelihood-ratio test statistic.

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.

Likelihood-ratio tests compare the given model versus the same



lcstats builds a collection

```
. lcstats c*, results(k_classes bic lmr df p_lmr entropy) noshownames noicnotes
```

Latent class statistics

Classes	BIC	LMR	df	P>LMR	Entropy
1	29,772.98				
2	28,961.97	850.88	6	<0.001	0.8152
3	28,980.18	36.39	6	0.012	0.7926
4	29,019.94	15.22	6	0.004	0.7206

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.

Likelihood-ratio tests compare the given model versus the same model with one less latent class.

```
. collect export lcstats.pdf
```

lcstats builds a collection

Acrobat File Edit View Window Help

lcstats.pdf + Create

All tools Edit Convert E-Sign Find text or tools AI Assistant

Classes	BIC	LMR	df	P>LMR	Entropy
1	29,772.98				
2	28,961.97	850.88	6	<0.001	0.8152
3	28,980.18	36.39	6	0.012	0.7926
4	29,019.94	15.22	6	0.004	0.7206

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic.
Likelihood-ratio tests compare the given model versus the same model with one less latent class.

STATISTICS • VISUALIZATION • DATA MAANIPULATION • REPORTING

Probabilities for class

```
. estimates restore c2  
(results c2 are active now)
```

```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = 10,000

	Delta-method			
	Margin	std. err.	[95% conf. interval]	
C				
1	.9306413	.0083098	.9124771	.945262
2	.0693587	.0083098	.054738	.0875229

```
. estimates restore c4  
(results c4 are active now)
```

```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = 10,000



```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = 10,000

	Delta-method			
	Margin	std. err.	[95% conf. interval]	
C				
1	.9306413	.0083098	.9124771	.945262
2	.0693587	.0083098	.054738	.0875229

```
. estimates restore c4
```

(results c4 are active now)

```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = 10,000

	Delta-method			
	Margin	std. err.	[95% conf. interval]	
C				
1	.0342343	.0048704	.0258711	.0451757
2	.7991788	.0957705	.5527033	.9276233
3	.1582926	.0923625	.0461028	.4225552
4	.0082942	.0025278	.0045586	.015045

Class means

```
. estimates restore c2  
(results c2 are active now)
```

```
. estat lcmean
```

Latent class marginal means

Number of obs = 10,000

		Delta-method		
		Margin	std. err.	[95% conf. interval]
1	alcohol	.3247101	.0057798	.313486 .3361392
	truant	.0652179	.0036996	.058331 .072855
	vandalism	.0502722	.0030228	.0446675 .0565385
	theft	.0094947	.0015693	.0068643 .0131198
	weapon	.0120166	.0015064	.0093956 .0153574
2	alcohol	.7585413	.0292529	.6967064 .811186
	truant	.4801932	.0335947	.4150662 .5460003
	vandalism	.4356286	.0359824	.3668456 .5069817
	theft	.2878347	.0302229	.232384 .3504757

```
. estimates restore c2
(results c2 are active now)
```

```
. estat lcmean
```

Latent class marginal means

Number of obs = 10,000

		Delta-method		
		Margin	std. err.	[95% conf. interval]
1	alcohol	.3247101	.0057798	.313486 .3361392
	truant	.0652179	.0036996	.058331 .072855
	vandalism	.0502722	.0030228	.0446675 .0565385
	theft	.0094947	.0015693	.0068643 .0131198
	weapon	.0120166	.0015064	.0093956 .0153574
2	alcohol	.7585413	.0292529	.6967064 .811186
	truant	.4801932	.0335947	.4150662 .5460003
	vandalism	.4356286	.0359824	.3668456 .5069817
	theft	.2878347	.0302229	.232384 .3504757
	weapon	.220836	.0247446	.1761491 .2731014

Class marginal means

```
. estimates restore c4  
(results c4 are active now)
```

```
. estat lcmean
```

Latent class marginal means

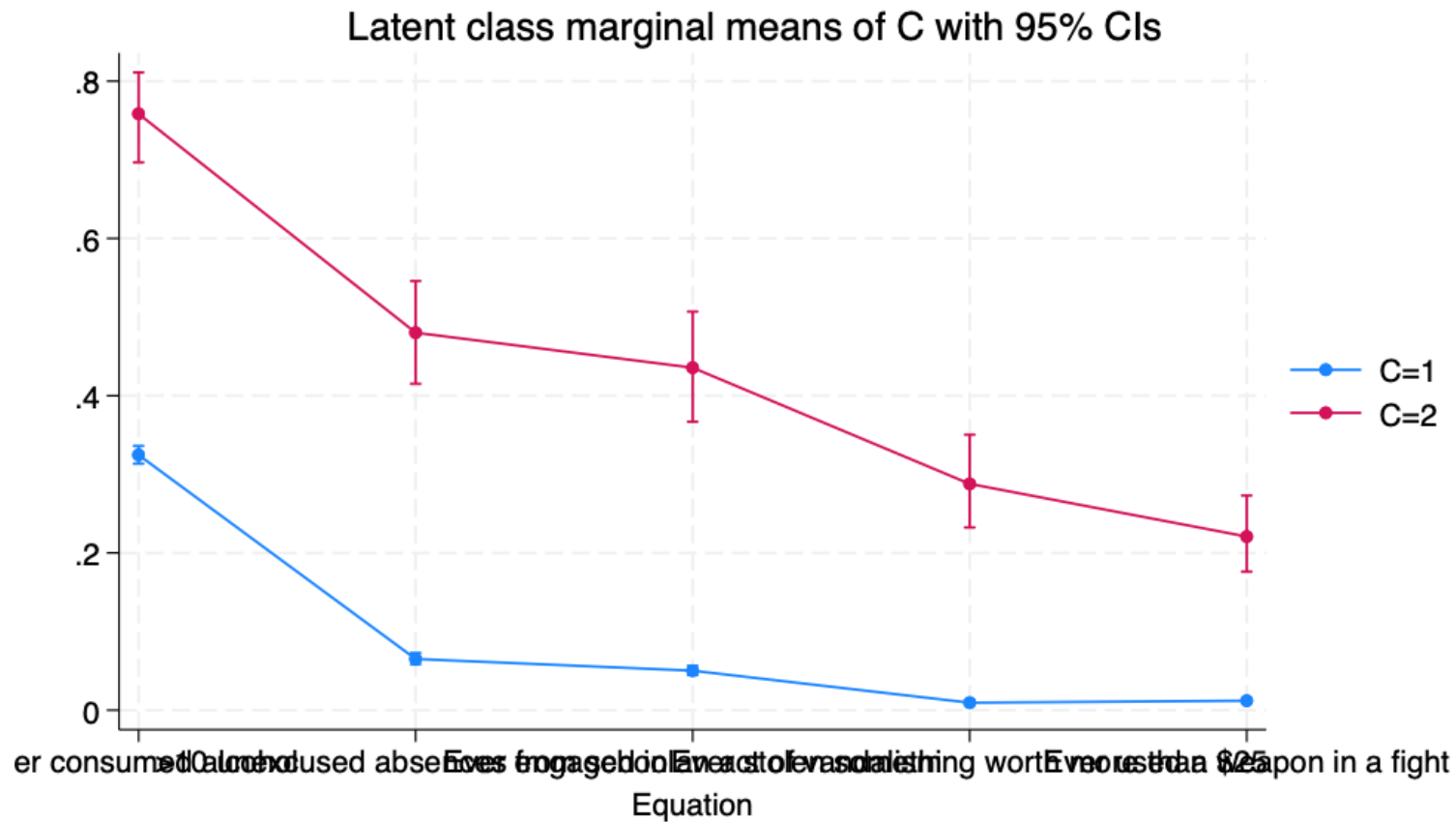
Number of obs = 10,000

		Delta-method		
		Margin	std. err.	[95% conf. interval]
1	alcohol	.638615	.0407451	.5556011 .7141015
	truant	.3939359	.0451283	.3097544 .4849227
	vandalism	.9999989	.0016705	0 1
	theft	.2232881	.0339505	.1637912 .296728
	weapon	.0454772	.0431023	.0067589 .2501362
2	alcohol	.2874116	.022061	.2461765 .332507
	truant	.0388241	.0149248	.0181103 .0812687
	vandalism	.047335	.0048535	.0386808 .057809

1					
	alcohol	.638615	.0407451	.5556011	.7141015
	truant	.3939359	.0451283	.3097544	.4849227
	vandalism	.9999989	.0016705	0	1
	theft	.2232881	.0339505	.1637912	.296728
	weapon	.0454772	.0431023	.0067589	.2501362
2					
	alcohol	.2874116	.022061	.2461765	.332507
	truant	.0388241	.0149248	.0181103	.0812687
	vandalism	.047335	.0048535	.0386808	.057809
	theft	.0062453	.0041495	.0016921	.0227711
	weapon	.0096807	.0030227	.0052416	.0178117
3					
	alcohol	.6043986	.0974356	.4073636	.7725082
	truant	.2814262	.0838081	.1480896	.4687575
	vandalism	3.53e-06	.0016931	0	1
	theft	.0719507	.024336	.0365648	.1367219
	weapon	.0563028	.0259933	.0223592	.1346777
4					
	alcohol	.9129382	.067945	.6625281	.982459
	truant	.5954166	.081372	.4315202	.7404787
	vandalism	.5950442	.0887795	.4164769	.7515636
	theft	.575756	.1131815	.3537036	.7709264
	weapon	1	.000257	0	1

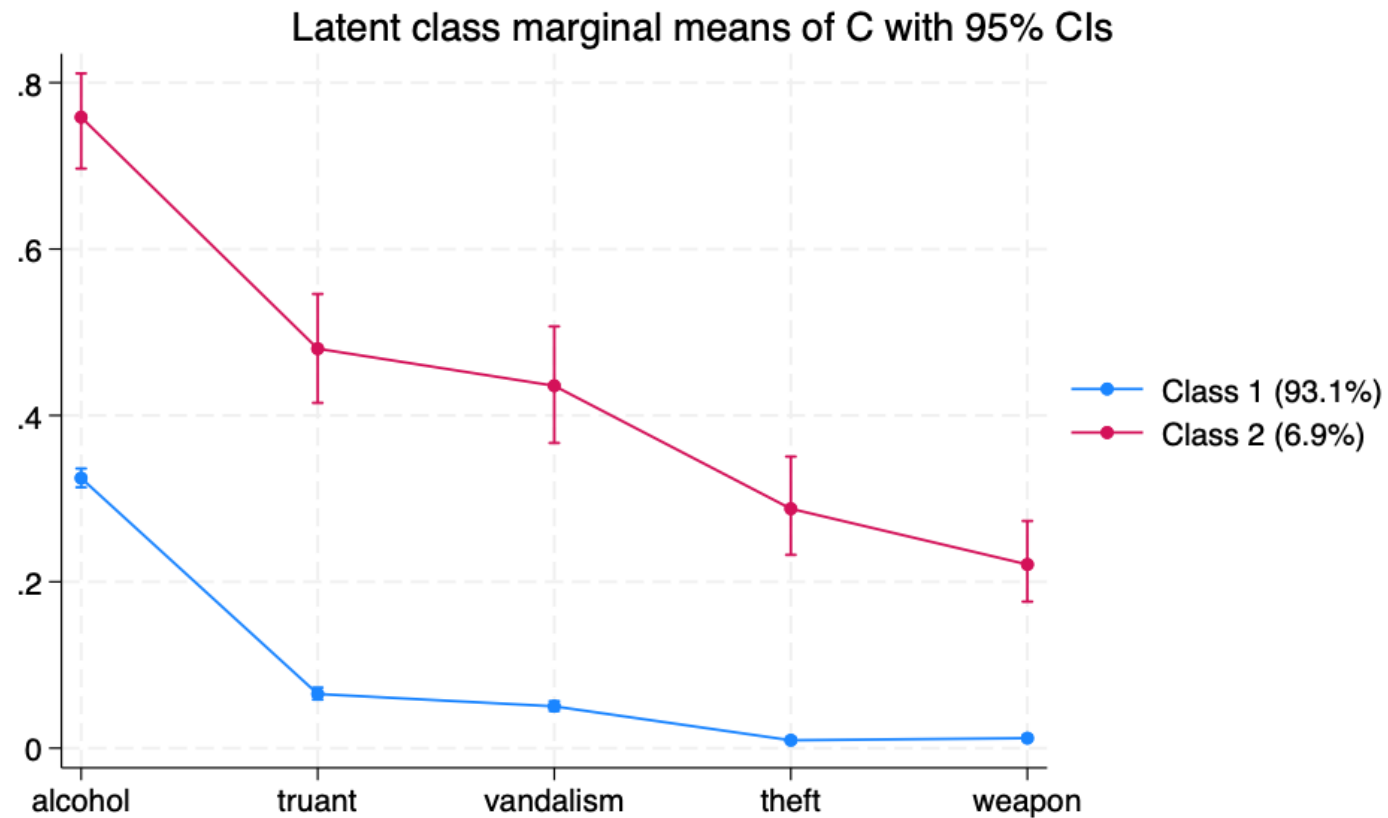
Class characteristics plot

```
. marginsplot
```



Class characteristics plot

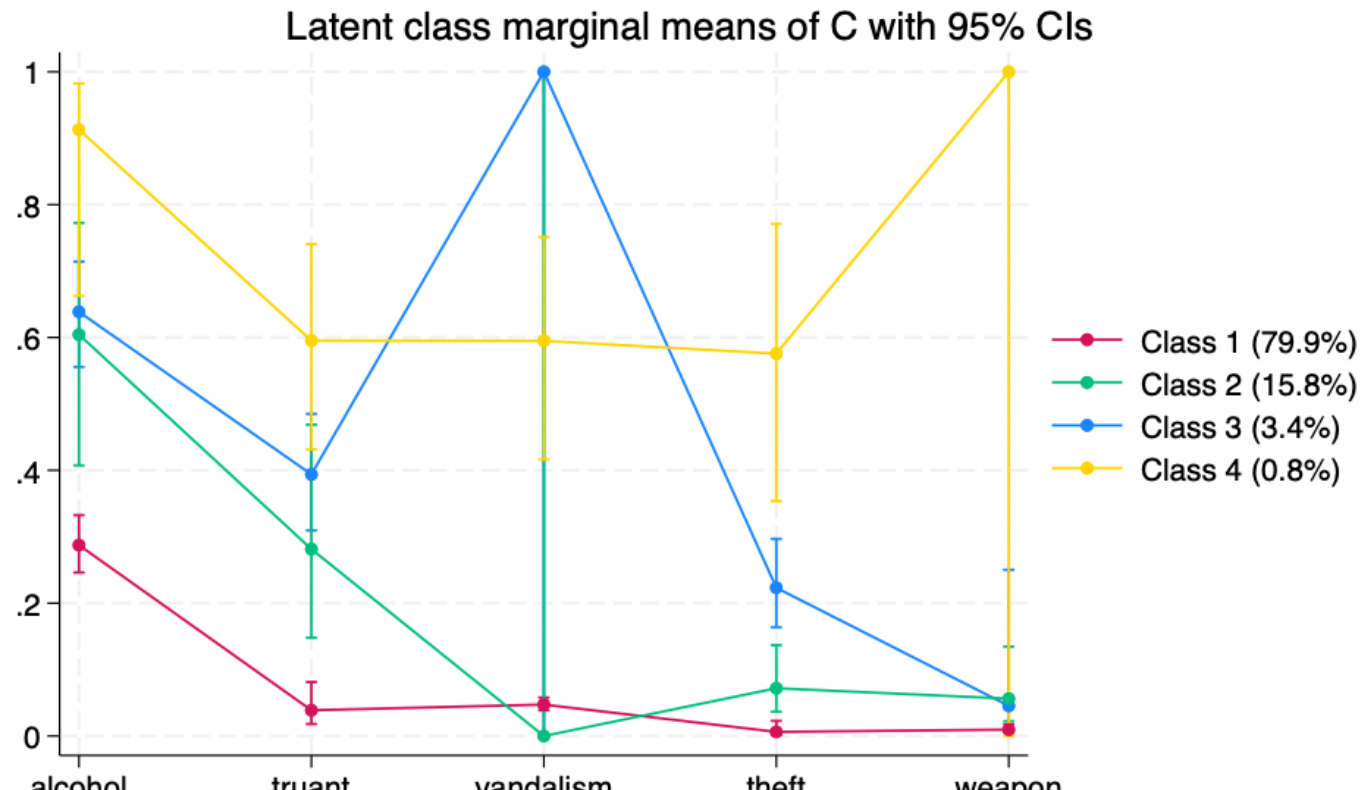
```
. marginsplot, xtitle("") xlabel(1 "alcohol" 2 "truant" 3 "vandalism" 4 "theft" 5 "weapon")  
> legend(order(3 "Class 1 (93.1%)" 4 "Class 2 (6.9%)"))
```



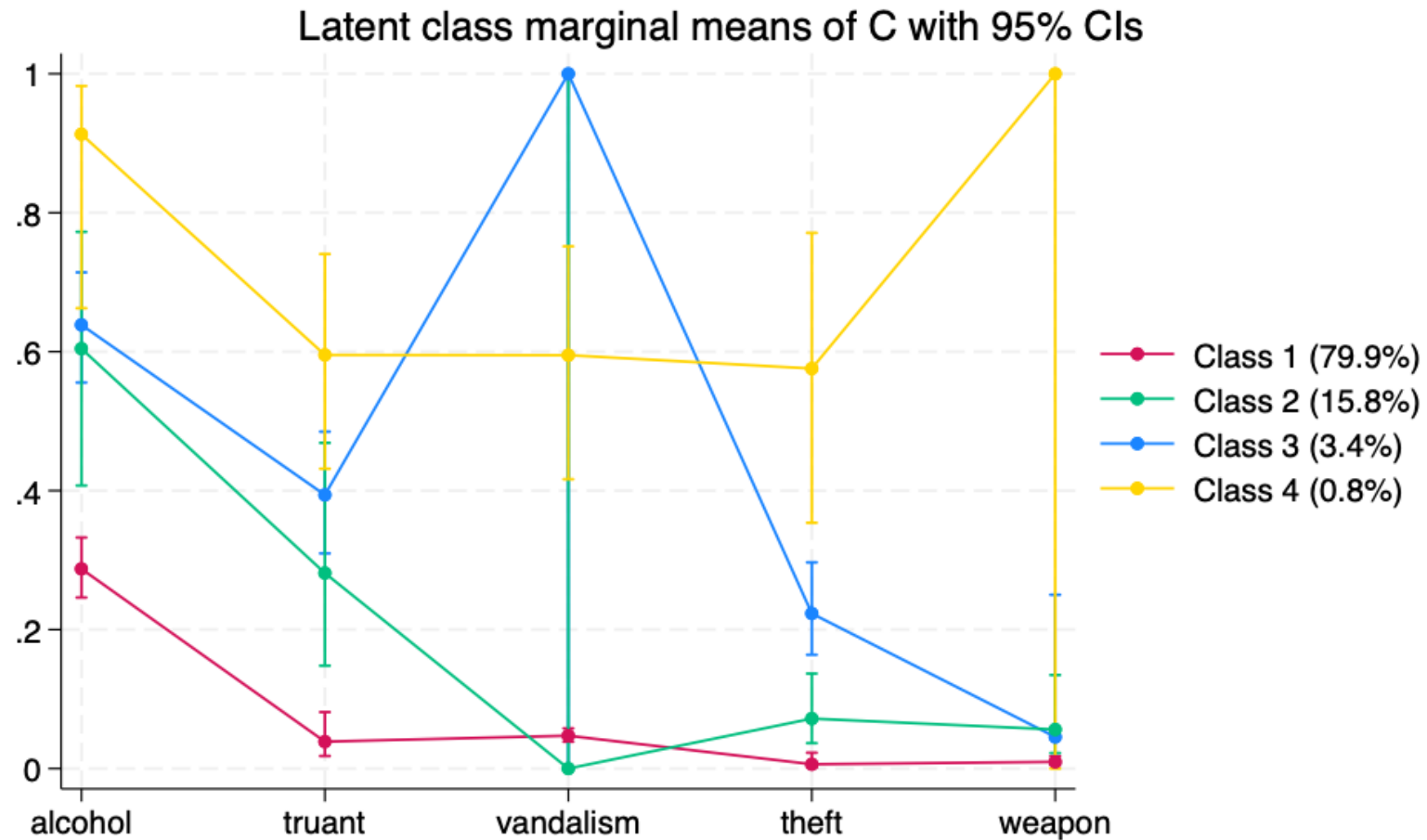
Class characteristics plot

```
. estimates restore c4  
(results c4 are active now)
```

```
. estat lcmean  
. marginsplot, xtitle("") xlabel(1 "alcohol" 2 "truant" 3 "vandalism" 4 "theft" 5 "weapon")  
> legend(order(6 "Class 1 (79.9%)" 7 "Class 2 (15.8%)" 5 "Class 3 (3.4%)" 8 "Class 4 (0.8%)"))
```



Class characteristics plot

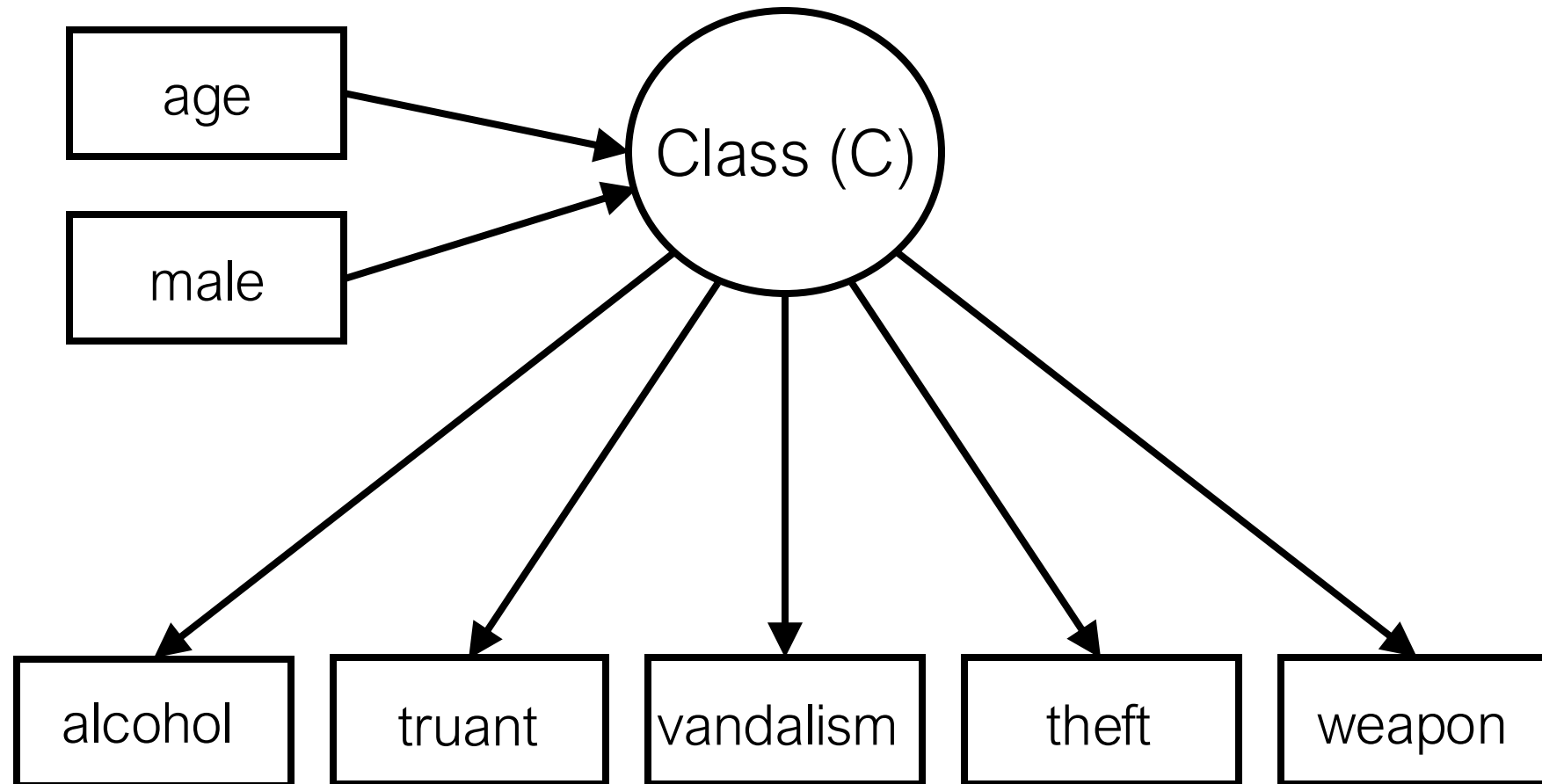


Predicted probabilities

```
. predict cpr*, classposteriorpr  
. sort id  
. list id alcohol truant vandalism theft weapon cp* in 1/10, clean
```

	id	alcohol	truant	vandalism	theft	weapon	cpr1	cpr2	cpr3	cpr4
1.	1	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
2.	2	0	0	1	0	1	.4505261	.4240032	.0000141	.1254566
3.	3	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
4.	4	1	0	0	0	0	3.93e-08	.7746661	.2253338	6.26e-11
5.	5	1	1	0	0	1	1.44e-07	.0361128	.6216332	.3422539
6.	6	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
7.	7	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
8.	8	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
9.	9	0	0	0	0	0	1.07e-08	.9286851	.0713149	2.88e-12
10.	10	1	0	0	0	0	3.93e-08	.7746661	.2253338	6.26e-11

Adding predictors



Adding predictors

```
. gsem (alcohol truant vandalism theft weapon <-, logit) (C<- i.male age), lclass(C 4) startvalues(jitter, draws(20))
```

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1.C		(base outcome)					
2.C							
	male						
	Male	.4024039	.4678719	0.86	0.390	-.5146082	1.319416
	age	.4382565	.2054551	2.13	0.033	.0355719	.8409411
	_cons	-10.57885	4.290153	-2.47	0.014	-18.9874	-2.170306
3.C							
	male						
	Male	.4229449	.1875035	2.26	0.024	.0554448	.790445
	age	.2966482	.0545834	5.43	0.000	.1896668	.4036296
	_cons	-8.587772	1.206789	-7.12	0.000	-10.95304	-6.222508
4.C							
	male						
	Male	2.308748	2.028106	1.14	0.255	-1.666267	6.283763
	age	.0385397	.0947323	0.41	0.684	-.1471322	.2242116
	_cons	-4.257064	2.513254	-1.69	0.090	-9.182951	.668823

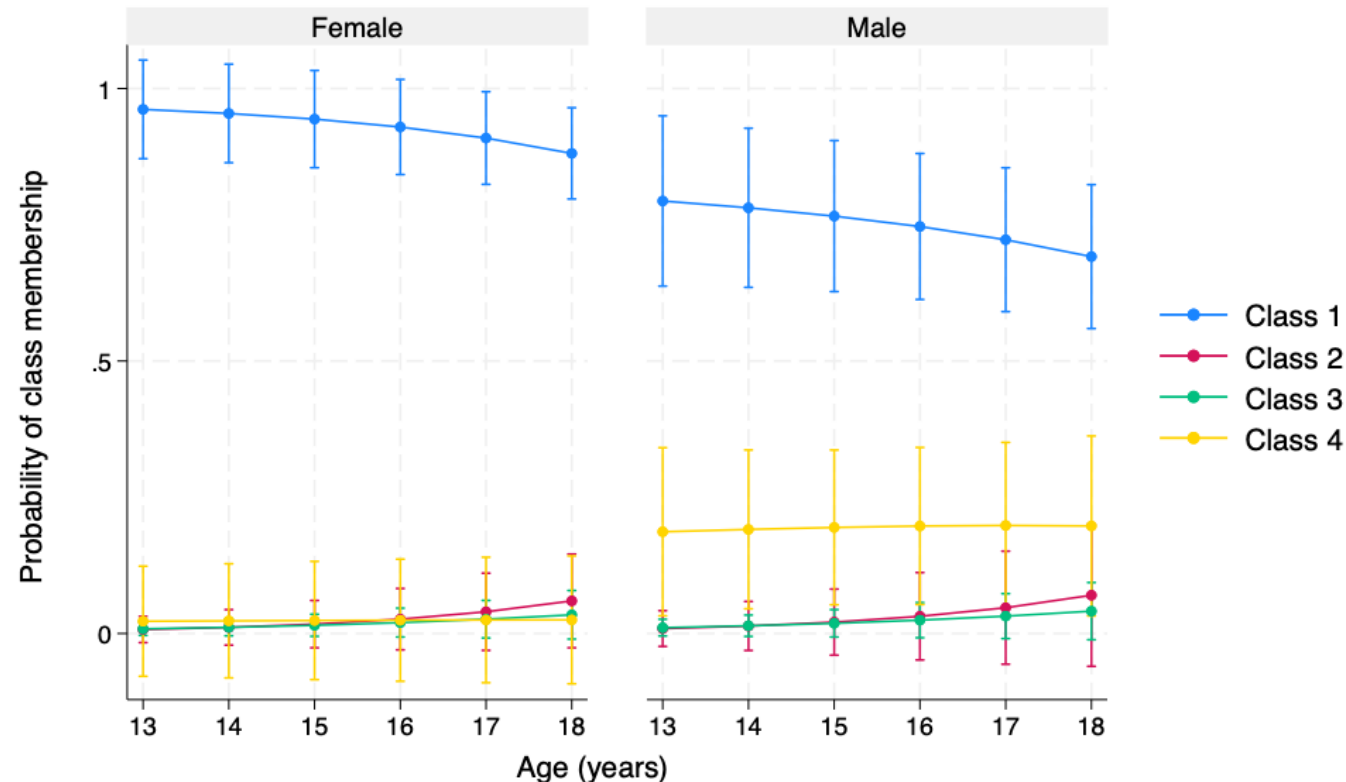
Adding predictors (odds ratios)

```
. estat eform C
```

	exp(b)	Std. err.	z	P> z	[95% conf. interval]	
1.C	(base outcome)					
2.C						
male						
Male	1.495415	.6996628	0.86	0.390	.5977348	3.741236
age	1.550002	.3184559	2.13	0.033	1.036212	2.318548
_cons	.0000254	.0001092	-2.47	0.014	5.67e-09	.1141427
3.C						
male						
Male	1.52645	.2862147	2.26	0.024	1.057011	2.204377
age	1.345342	.0734333	5.43	0.000	1.208847	1.497249
_cons	.0001864	.0002249	-7.12	0.000	.0000175	.0019843
4.C						
male						
Male	10.06182	20.40644	1.14	0.255	.1889511	535.8013
age	1.039292	.0984545	0.41	0.684	.8631798	1.251336
_cons	.0141638	.0355973	-1.69	0.090	.0001028	1.951939

Class probabilities by covariates

```
. margins male, at(age=(13/18))  
  predict(classpr class(1)) predict(classpr class(2)) predict(classpr class(3)) predict(classpr class(4))  
  
. marginsplot, by(male) ytitle("Probability of class membership")  
  legend(order (5 "Class 1" 6 "Class 2" 7 "Class 3" 8 "Class 4")) byopts(title(""))
```



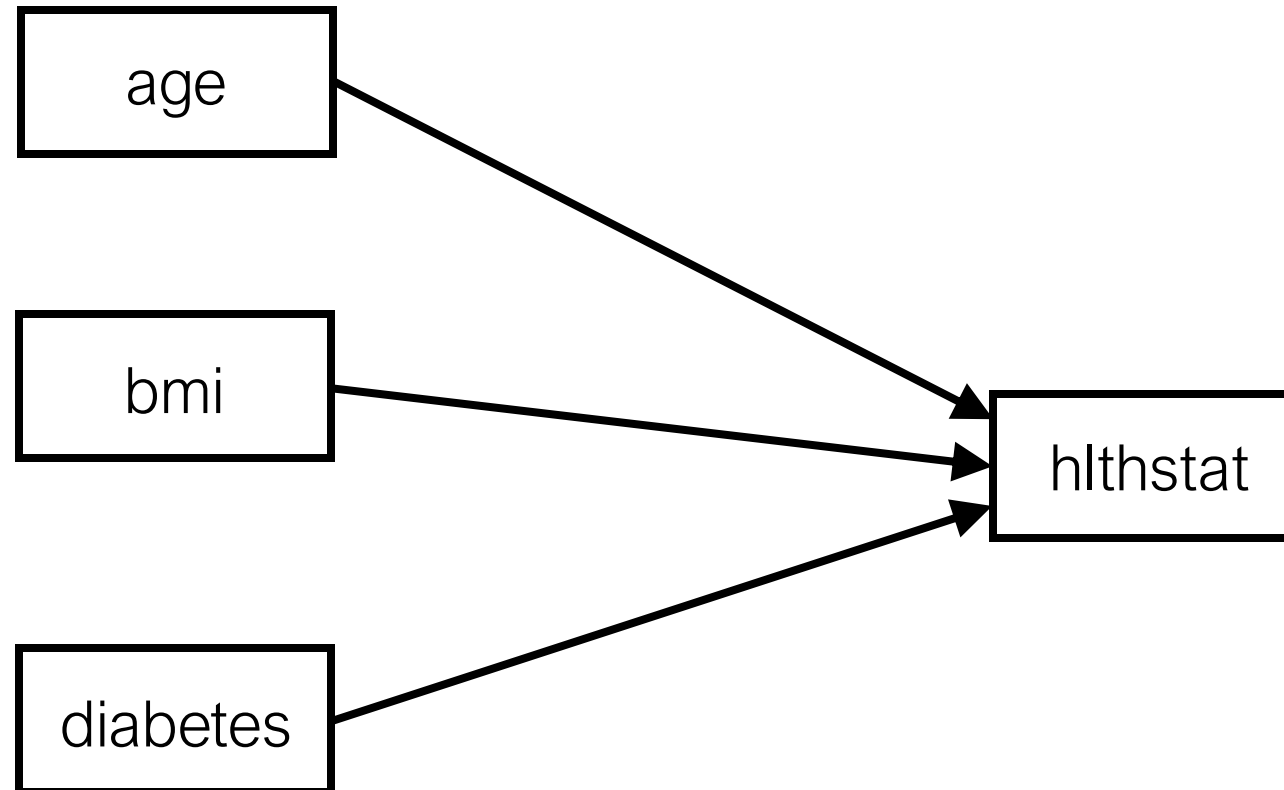
NHANES example data

```
. use nhanes, clear
```

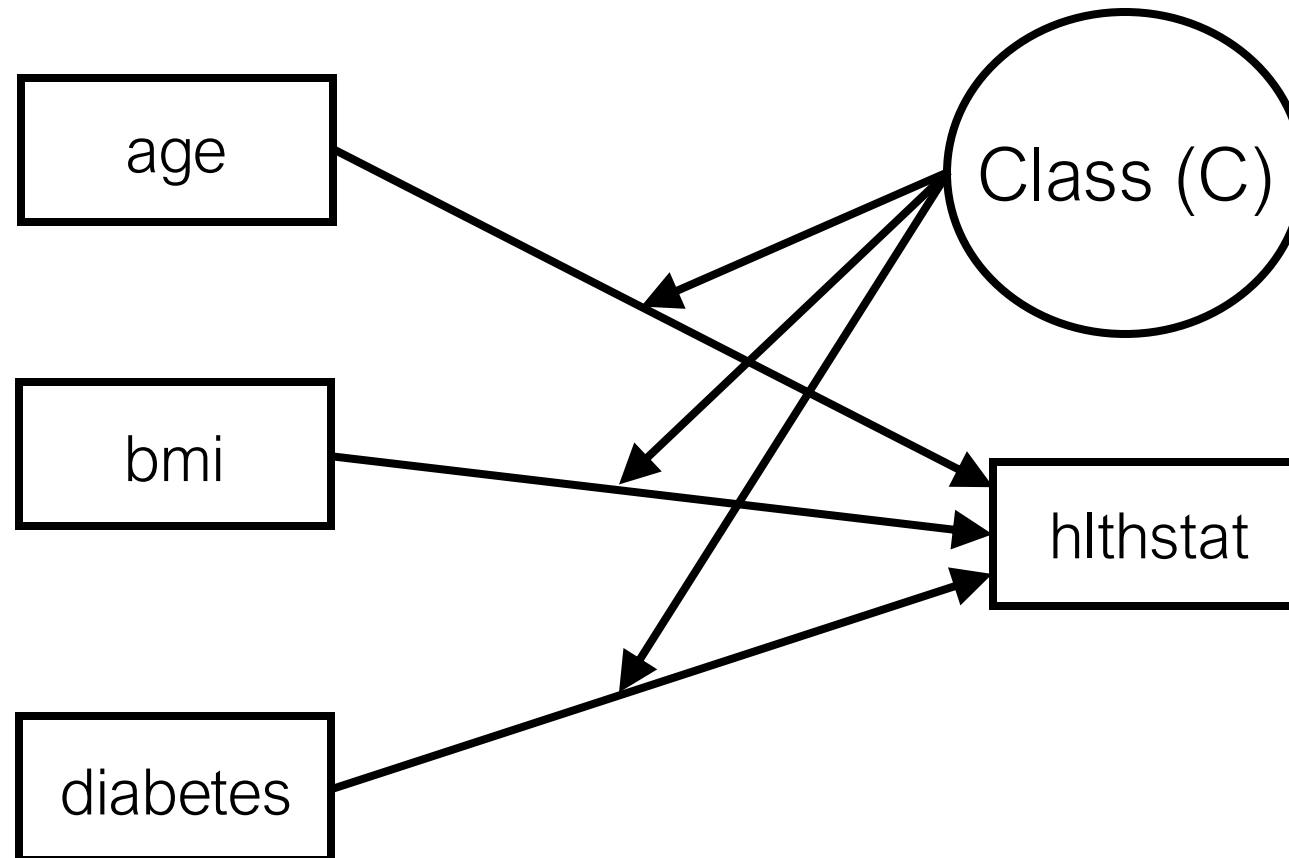
```
. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
smsa	10335	3	2.65612	1	4	Standard metropolitan statistic...
age	10335	55	47.56584	20	74	Age (years)
bmi	10335	9927	25.53897	12.3856	61.1297	Body mass index (BMI)
diabetes	10335	2	.0482825	0	1	Diabetes status
bpsystol	10335	108	130.8876	65	300	Systolic blood pressure
bpdiast	10335	68	81.71959	35	150	Diastolic blood pressure
hlthstat	10335	5	2.586164	1	5	Health status

Linear regression



Linear regression mixture



Linear regression mixture

```
. gsem (hlthstat <- age i.diabetes bmi), lclass(C 2)
```

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1.C	(base outcome)					
2.C						
_cons	-.160437	.0469464	-3.42	0.001	-.2524502	-.0684237

```
. estat lcprob
```

	Delta-method		
	Margin	std. err.	[95% conf. interval]
C			
1	.5400234	.0116614	.5170993 .5627795
2	.4599766	.0116614	.4372205 .4829007

Class coefficients

```
. gsem (hlthstat <- age i.diabetes bmi), lclass(C 2)
```

Class: 1

	Coefficient	Std. err.
hlthstat		
age	.0154491	.0007959
diabetes		
Diabetic	.8468695	.0689841
bmi	.0187634	.0029626
_cons	.5881338	.0761477
var(e.hlthstat)	.5730265	.0137406

Class: 2

	Coefficient	Std. err.
hlthstat		
age	.0266895	.0008866
diabetes		
Diabetic	.6404264	.0598009
bmi	.0109067	.0026019
_cons	1.869163	.082183
var(e.hlthstat)	.5730265	.0137406

Class coefficients

```
. gsem (hlthstat <- age i.diabetes bmi), lclass(C 2)
```

Class: 1			
		Coefficient	Std. err.
hlthstat			
	age	.0154491	.0007959
	diabetes		
	Diabetic	.8468695	.0689841
	bmi	.0187634	.0029626
	_cons	.5881338	.0761477
var(e.hlthstat)		.5730265	.0137406

Class: 2			
		Coefficient	Std. err.
hlthstat			
	age	.0266895	.0008866
	diabetes		
	Diabetic	.6404264	.0598009
	bmi	.0109067	.0026019
	_cons	1.869163	.082183
var(e.hlthstat)		.5730265	.0137406

Class coefficients

```
. gsem (hlthstat <- age i.diabetes bmi@b), lclass(C 2)
```

Class: 1			Class: 2		
	Coefficient	Std. err.		Coefficient	Std. err.
hlthstat			hlthstat		
age	.0156473	.0007907	age	.0266047	.0008824
diabetes			diabetes		
Diabetic	.8507712	.0693641	Diabetic	.6327083	.059429
bmi	.0142115	.0021331	bmi	.0142115	.0021331
_cons	.690505	.0605036	_cons	1.784947	.0726737
var(e.hlthstat)	.572821	.0137438	var(e.hlthstat)	.572821	.0137438

The lcinvariant() option

`lcinvariant(pclassname)` – specify parameters that are equal across latent classes

<i>pclassname</i>	Description
cons	intercepts and cutpoints
coef	fixed coefficients
errvar	covariances of errors
scale	scaling parameters
all	all the above
none	none of the above

Constraining all coefficients

```
. gsem (hlthstat <- age diabetes#c.bmi), lclass(C 2) lcinvariant(coef) ///  
> startvalues(jitter, draws(10) seed(3289))
```

Class: 1			Class: 2		
	Coefficient	Std. err.		Coefficient	Std. err.
hlthstat			hlthstat		
age	.0239003	.0002114	age	.0239003	.0002114
diabetes			diabetes		
Diabetic	.9596019	.0978575	Diabetic	.9596019	.0978575
bmi	.0015937	.001128	bmi	.0015937	.001128
diabetes#c.bmi			diabetes#c.bmi		
Diabetic	.0004875	.0034765	Diabetic	.0004875	.0034765
_cons	1.487675	.0305586	_cons	.3520712	.0270075
var(e.hlthstat)	1.236106	.018291	var(e.hlthstat)	.0102907	.0007992



Class-specific models

```
. gsem (1: hlthstat <- age bmi i.diabetes) ///  
> (2: hlthstat <- diabetes##c.age), lclass(C 2) startvalues(classpr cprob1 cprob2)
```

Class: 1			Class: 2		
	Coefficient	Std. err.		Coefficient	Std. err.
hlthstat			hlthstat		
age	.0156067	.0007903	age	.0276636	.0009063
bmi	.0170333	.0030314			
diabetes			diabetes		
Diabetic	.8491393	.0678729	Diabetic	2.095236	.3610216
_cons	.6315143	.0765613	diabetes#c.age		
			Diabetic	-.0238014	.0058113
var(e.hlthstat)	.5740851	.0139179	_cons	2.107465	.0518677
			var(e.hlthstat)	.5740851	.0139179



Comparing models (different models)

```
. gsem (hlthstat <- age diabetes##c.bmi), lclass(C 2)
. estimates store m1
. gsem (1: hlthstat <- age bmi i.diabetes) ///
> (2: hlthstat <- diabetes##c.age), lclass(C 2) startvalues(classpr cprob1 cprob2)
. estimates store m2
. estimates stats m*
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
m1	10,335	.	-15465.39	10	30950.77	31023.2
m2	10,335	.	-15464.49	10	30948.99	31021.42

Comparing models (different classes)

```
. gsem (hlthstat <- age diabetes##c.bmi), lclass(C 2) startvalue(randompr, draws(20)) log
. estimates store m1
. gsem (hlthstat <- age diabetes##c.bmi), lclass(C 3) startvalue(randompr, draws(20)) log
. estimates store m3
. lcstats m1 m3
```

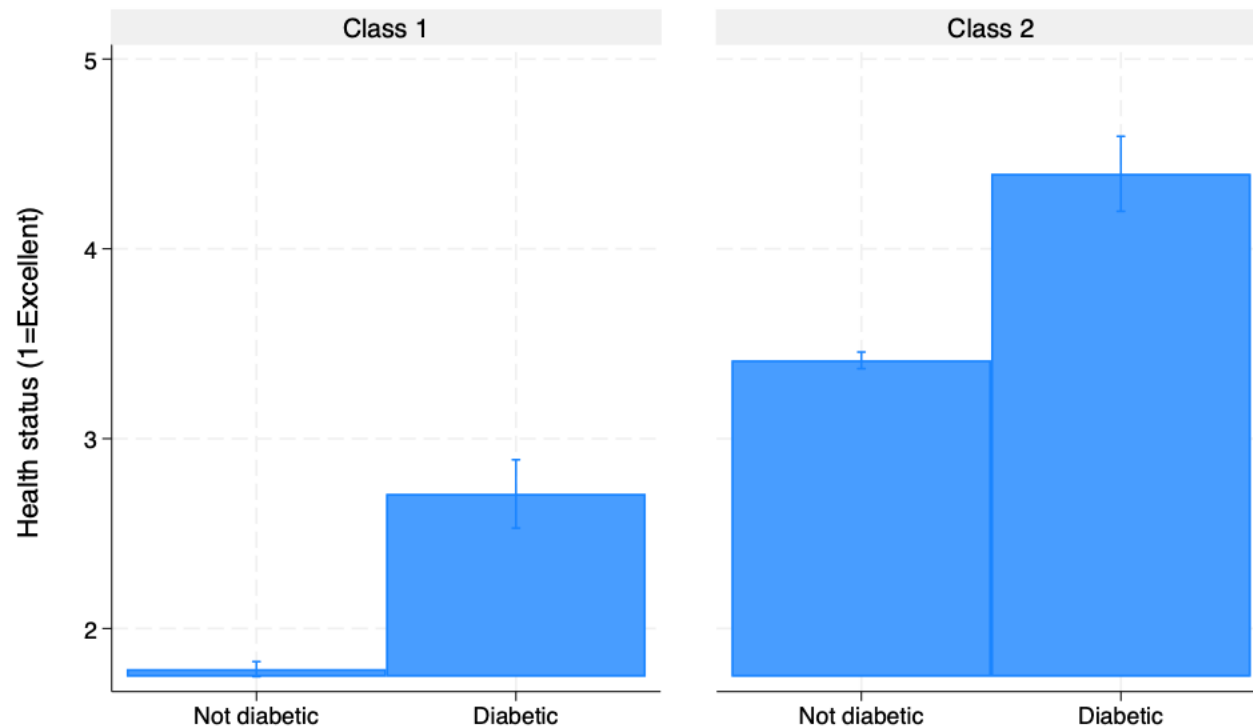
Latent class statistics

	Classes	N	ll	Rank	Entropy	df	LMR	P>LMR
m1	2	10,335	-15,465.06	12	0.5185			
m3	3	10,335	-15,184.06	18	0.7195	6	552.06	<0.001

LMR is the Lo-Mendell-Rubin-adjusted likelihood-ratio test statistic. Likelihood-ratio tests compare the given model versus the same model with one less latent class.

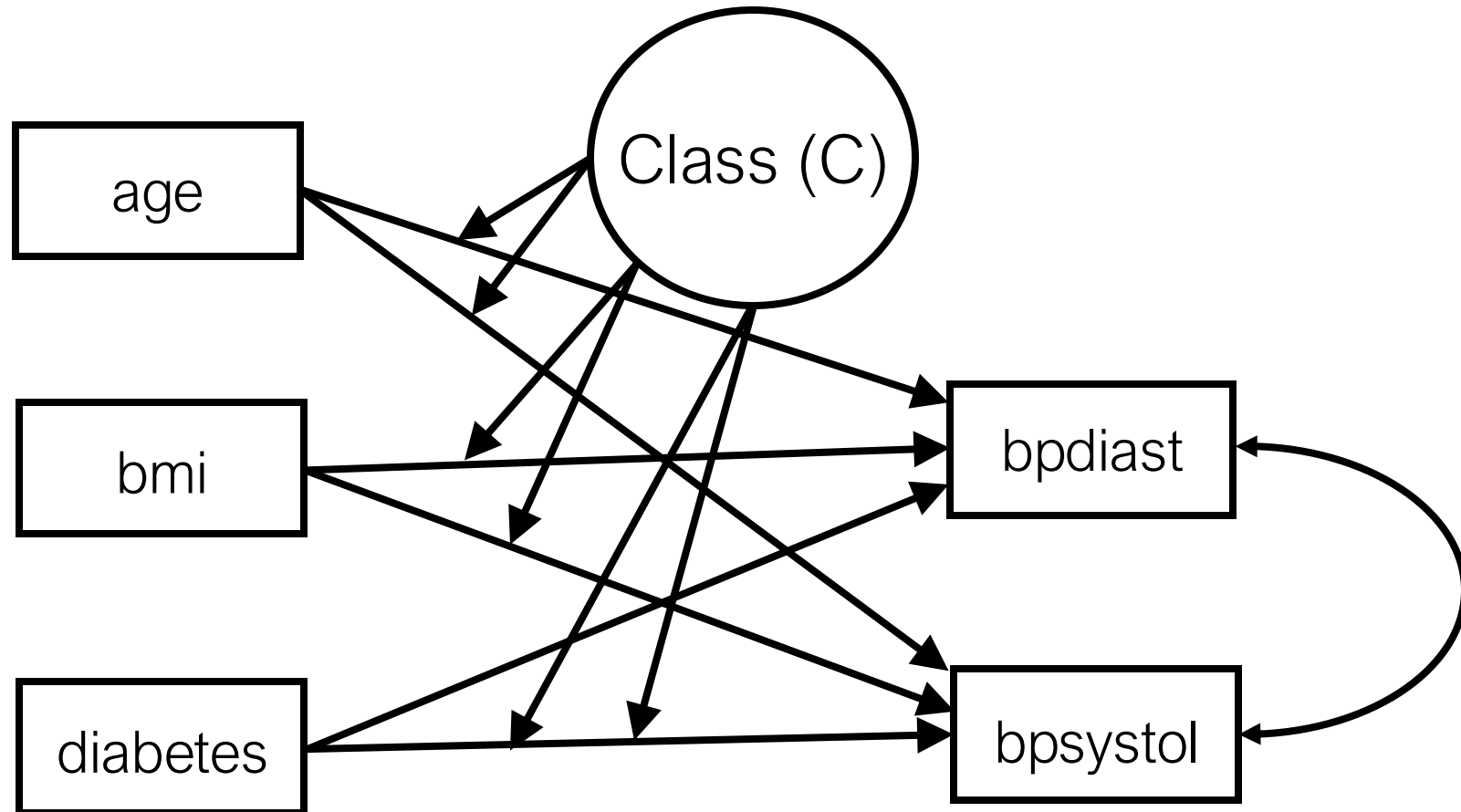
Marginal means by class

```
. margins diabetes, predict(class(1)) predict(class(2))  
. marginsplot, by(_predict, label("Class 1" "Class 2")) recast(bar)  
  ytitle("Health status (1=Excellent)") xtitle("") byopts(title(""))
```



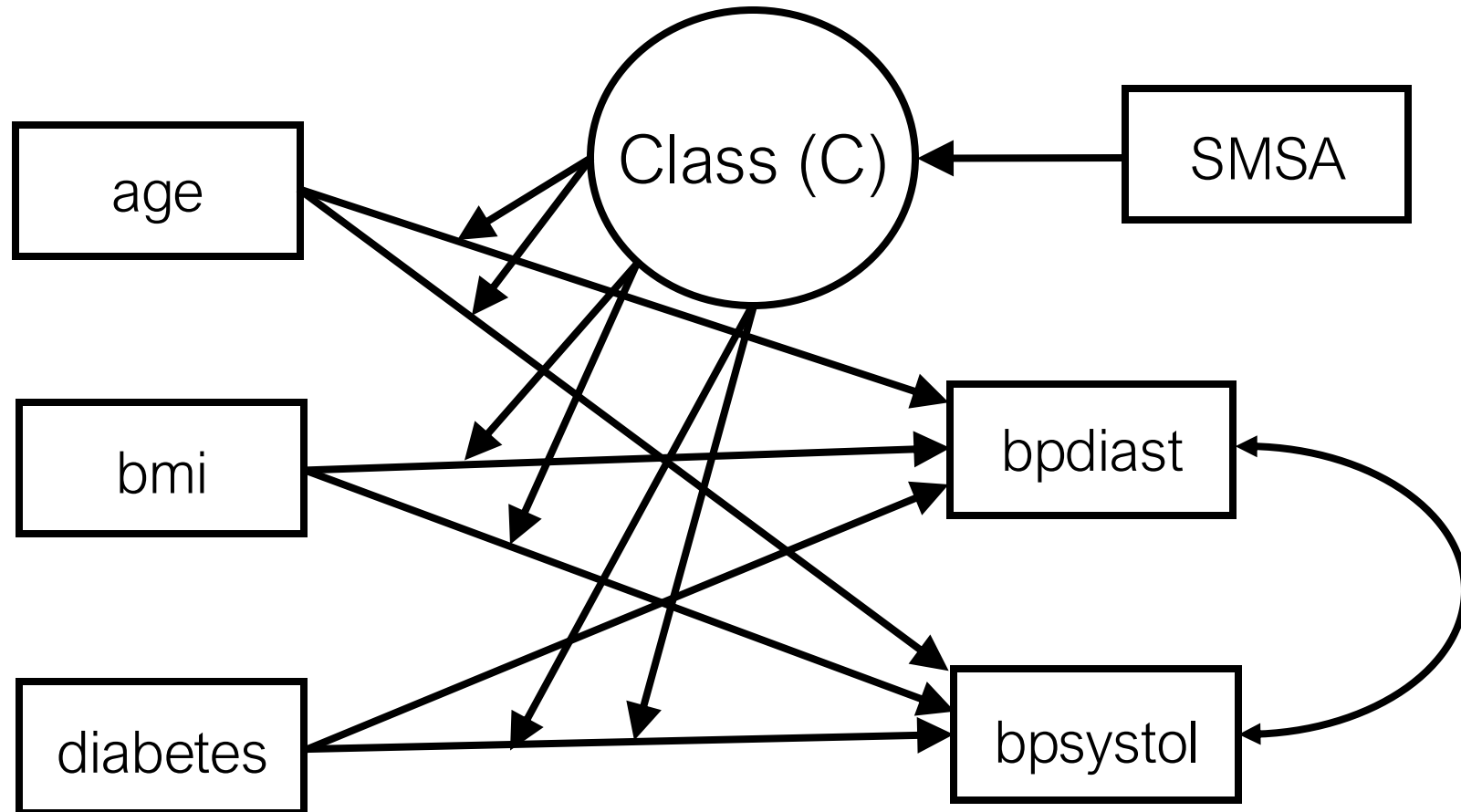
Multivariate regression mixture

```
. gsem (bpdiastr bpsystol <- age i.diabetes bmi), lclass(C 2)
```



Multivariate regression mixture

```
. gsem (bpdiastr bpsystol <- age i.diabetes bmi) (C <- i.smsa), lclass(C 2)
```





Thank you!

Questions?