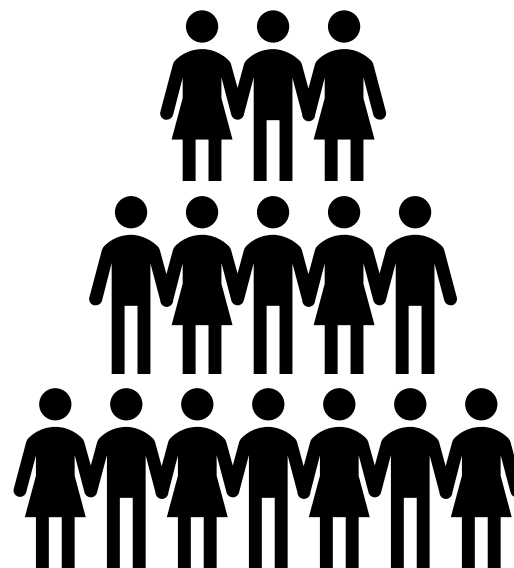


Latent class analysis in Stata

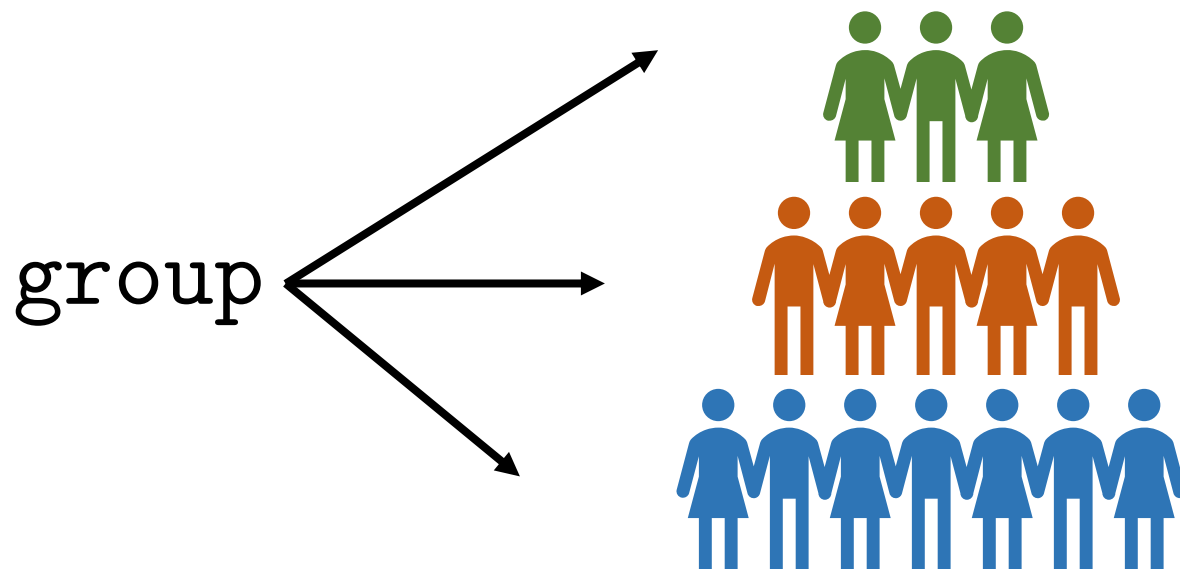
Meghan K. Cain, Ph.D. | June 21, 2022

You can download the dataset, do-file, and slides here:
<https://tinyurl.com/StataLCA>

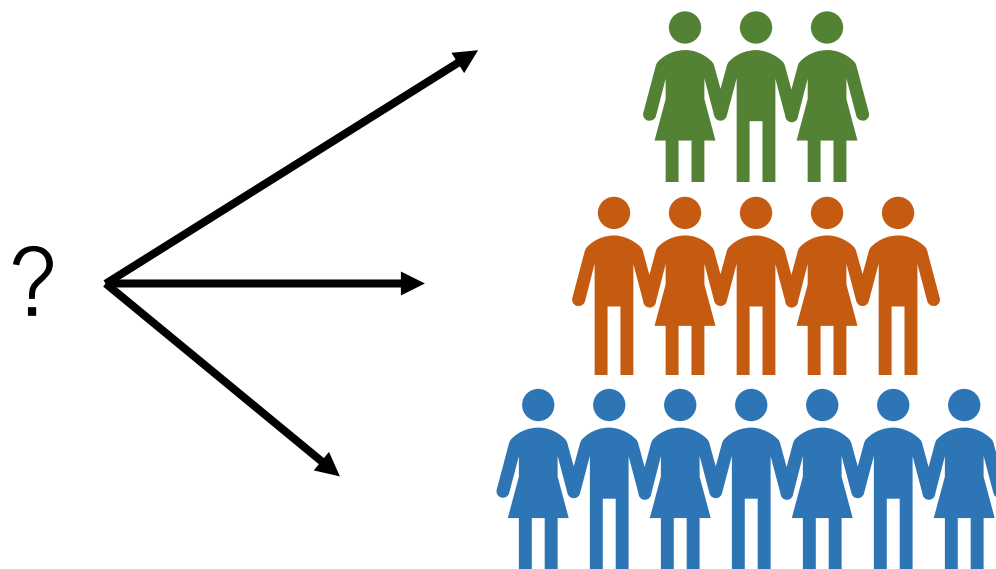
Population characteristics



Population characteristics

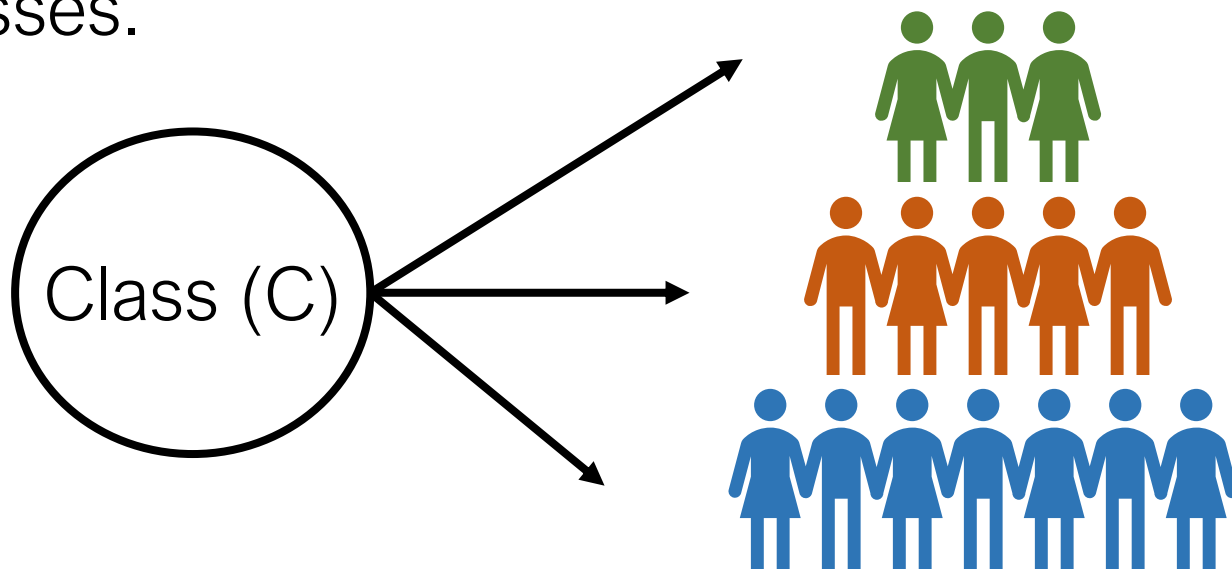


Population characteristics



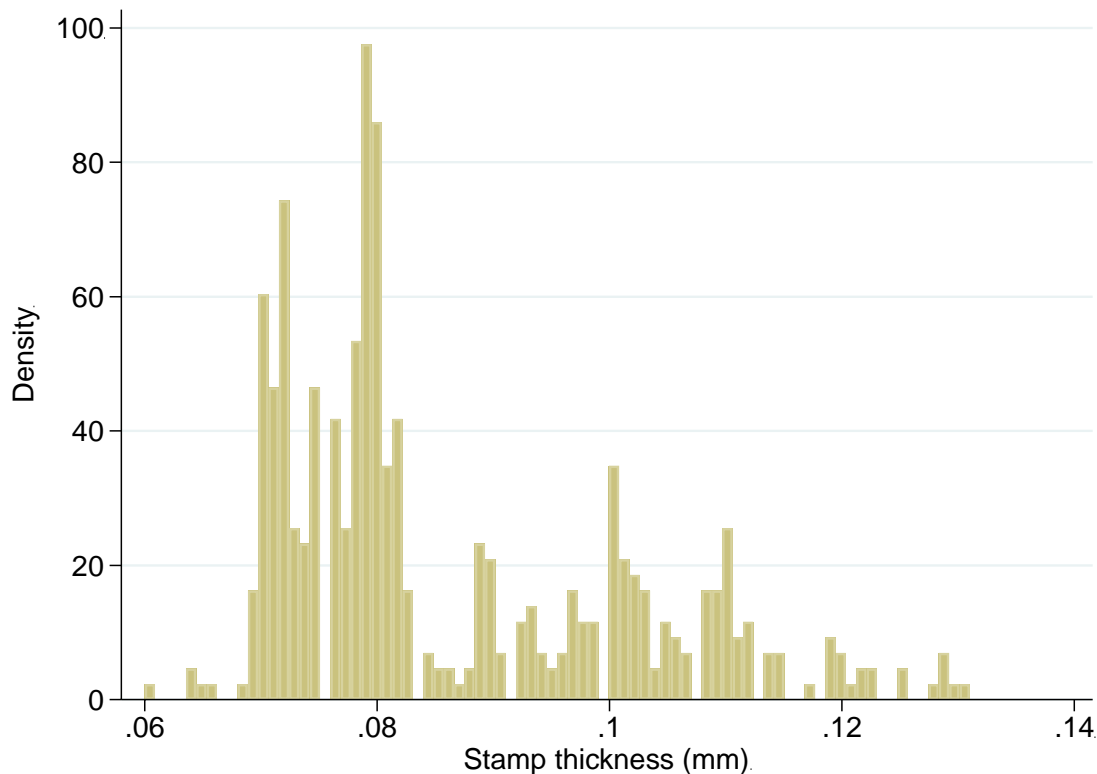
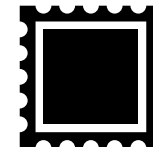
Latent class analysis (LCA)

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



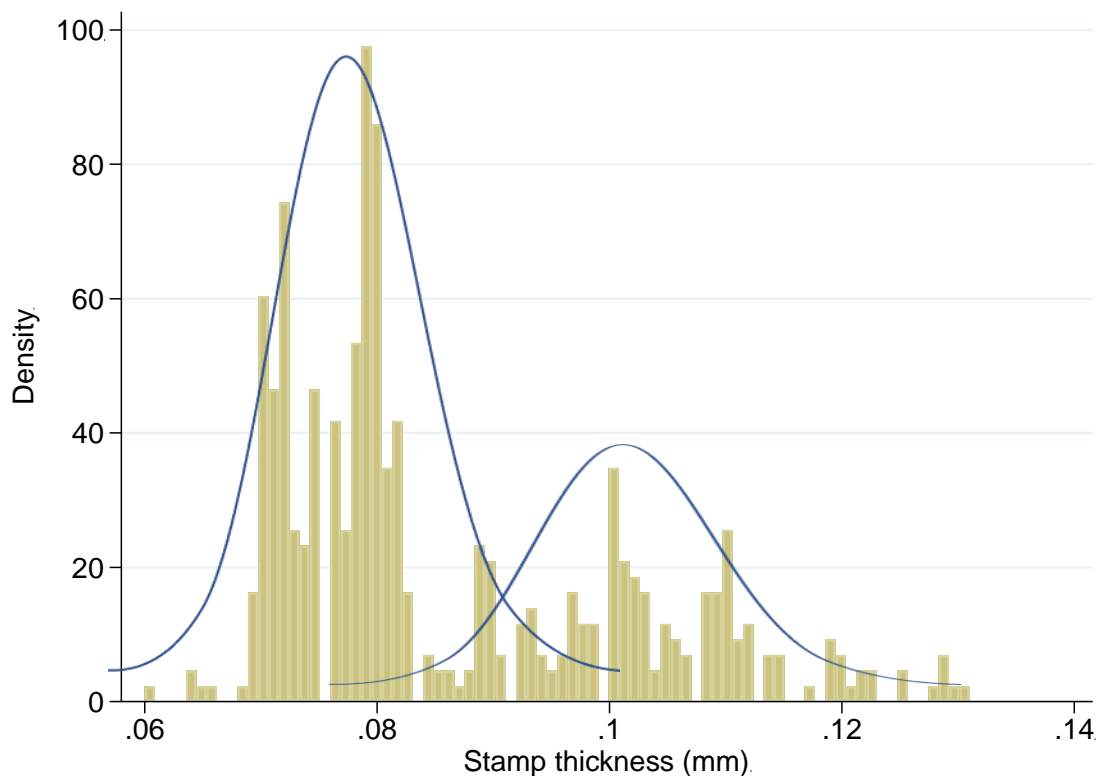
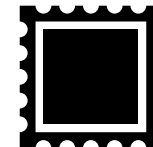
Basic mixture model

```
. 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

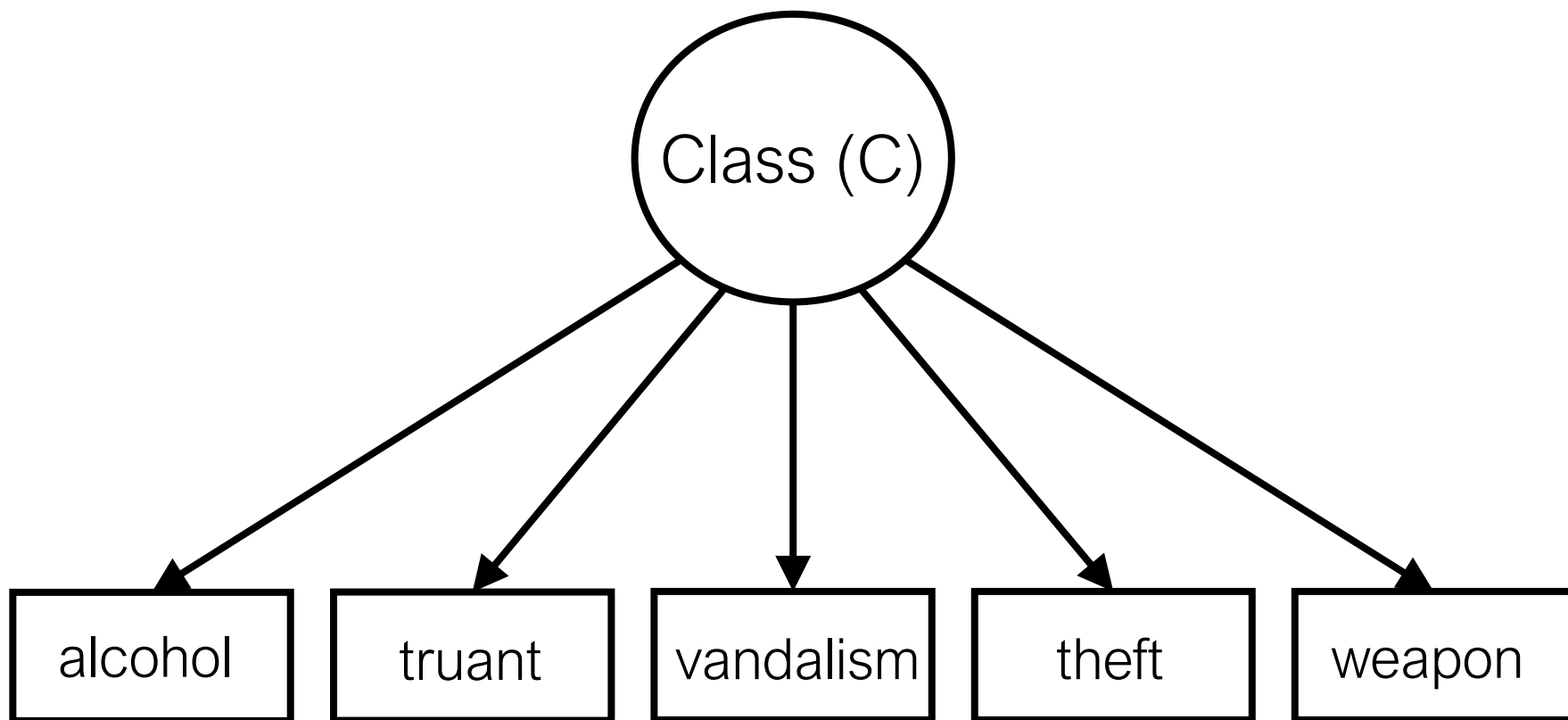
	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
thickness						
_cons	.0776762	.00043	180.65	0.000	.0768334	.0785189
var(e.thickness)	.0000518	3.72e-06			.000045	.0000596

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.
- Unlike cluster analysis, LCA is a model that can be evaluated for fit.

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.

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.

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.

Types of LCA

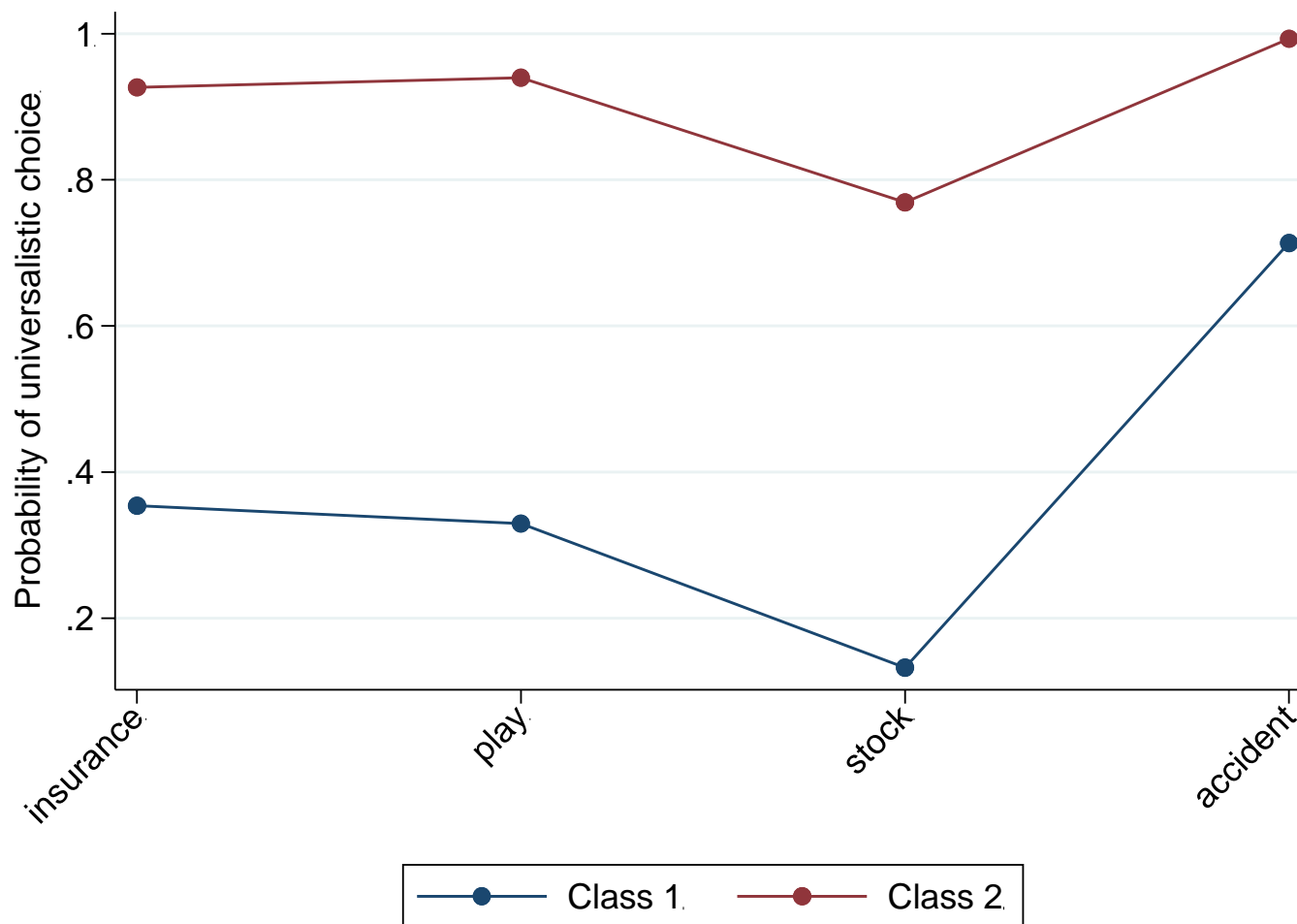
- **Latent class analysis (LCA):**
categorical/binary indicators
- **Latent profile analysis (LPA):**
continuous indicators
- **Latent transition analysis (LTA):**
one indicator that changes over time
- All these models can be analyzed in the same way with minor modifications.

Example:

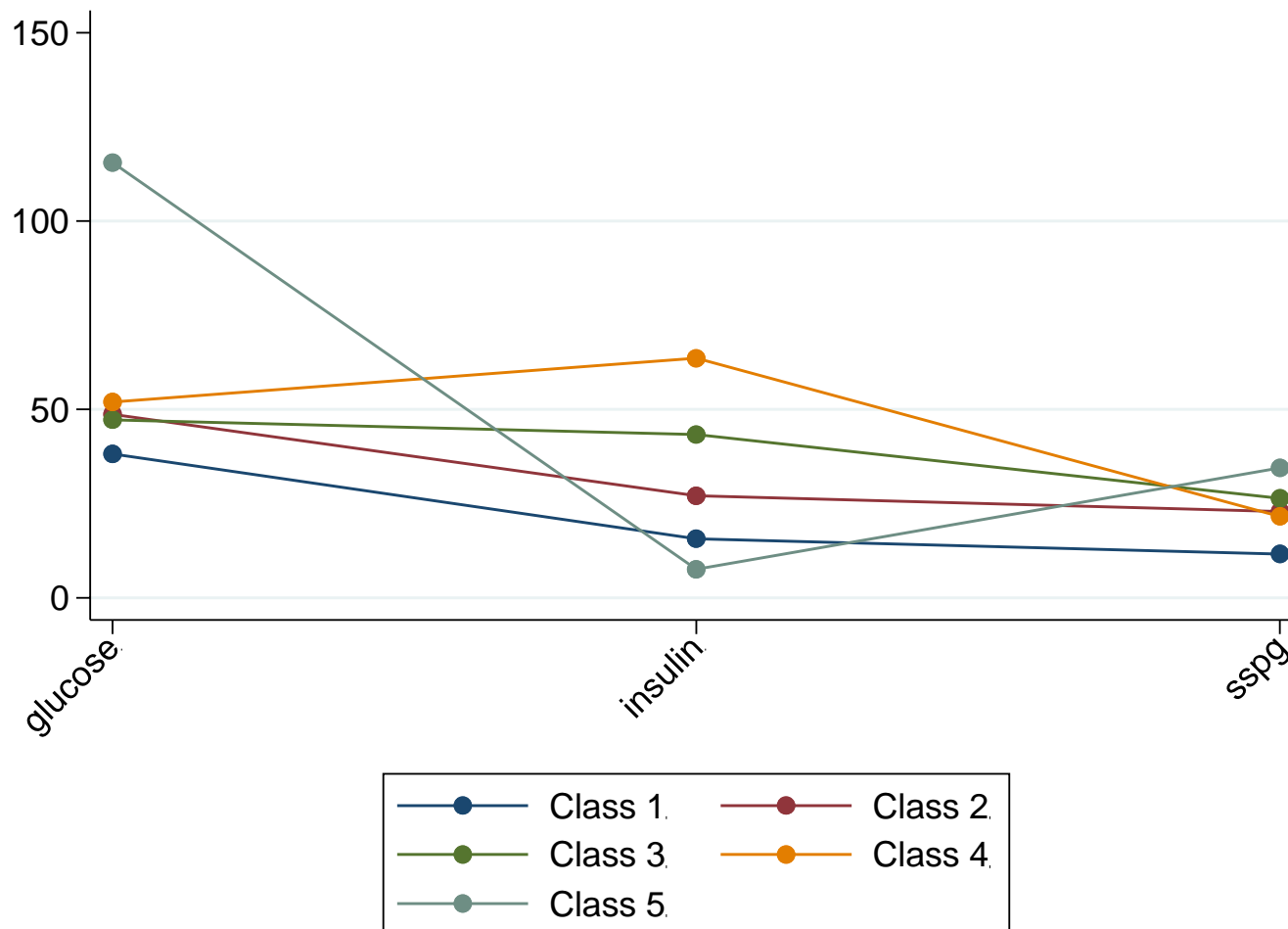
Universalistic vs particularistic

1. Would you disclose health concerns to a friend's insurance company?
2. Would you give a negative review of a friend's play?
3. Would you keep a company secret from a friend?
4. Would you testify against a friend in an accident case?

Universalistic vs particularistic



Health profiles



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**

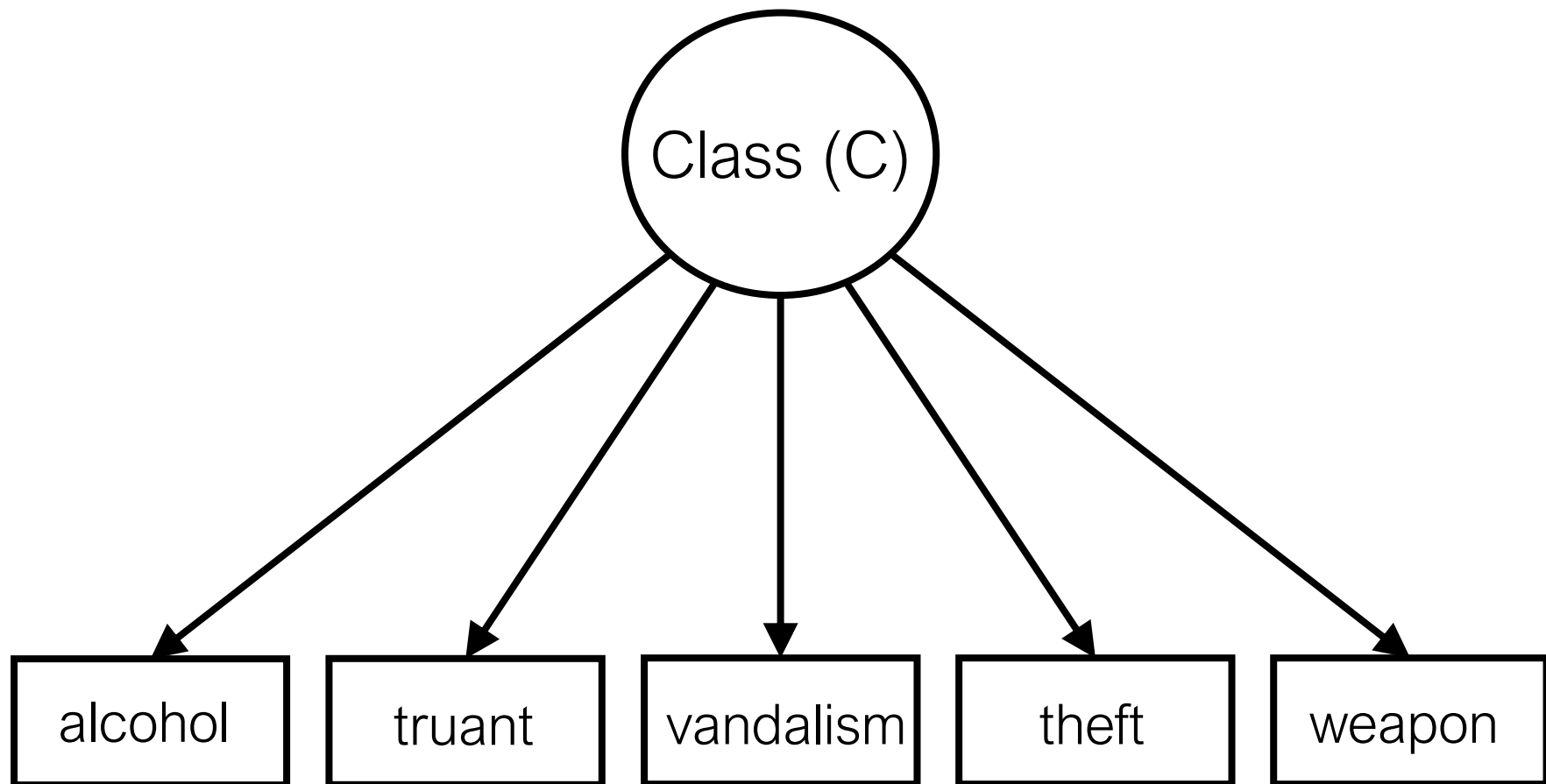
Observations:10,000

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

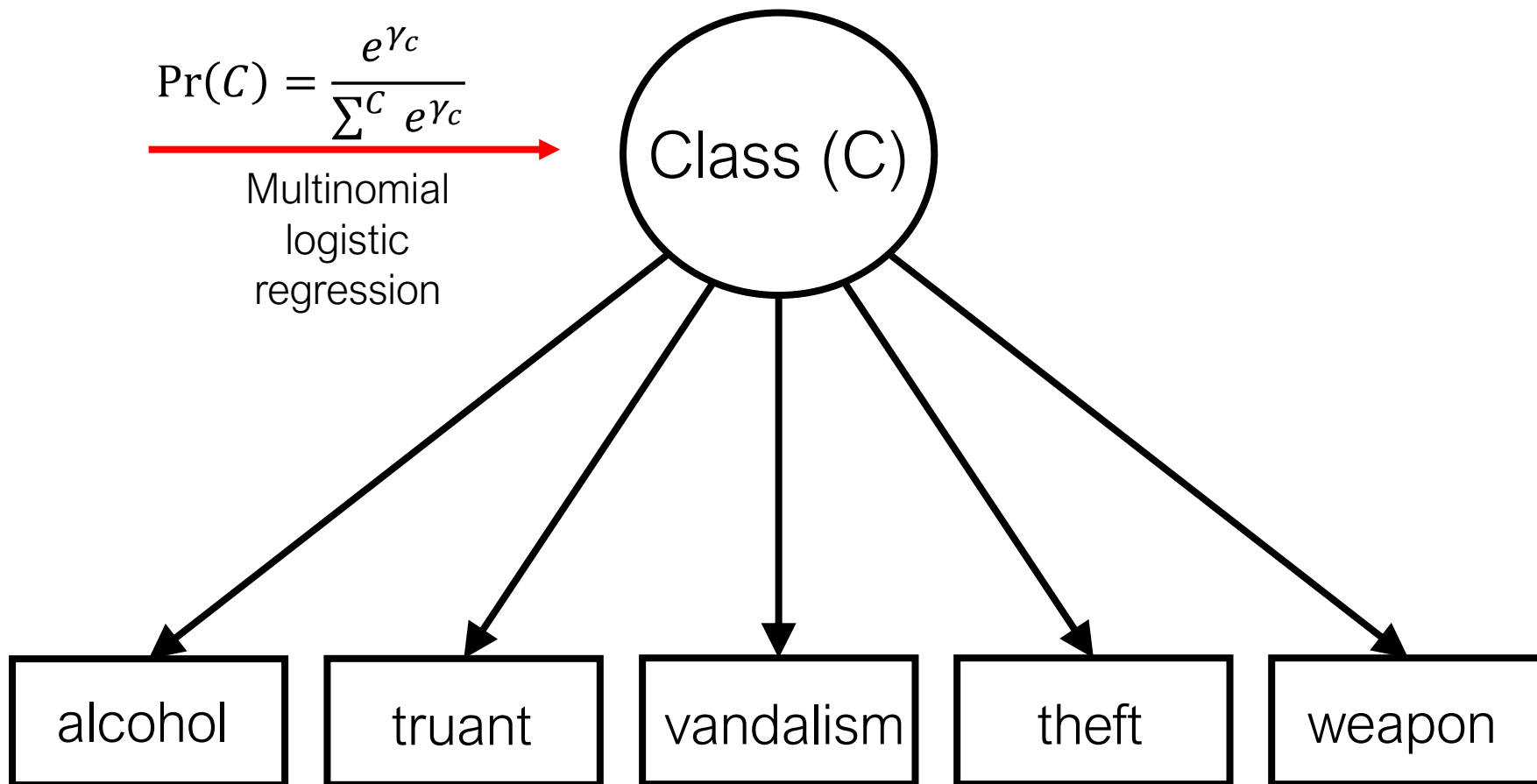
Path diagram



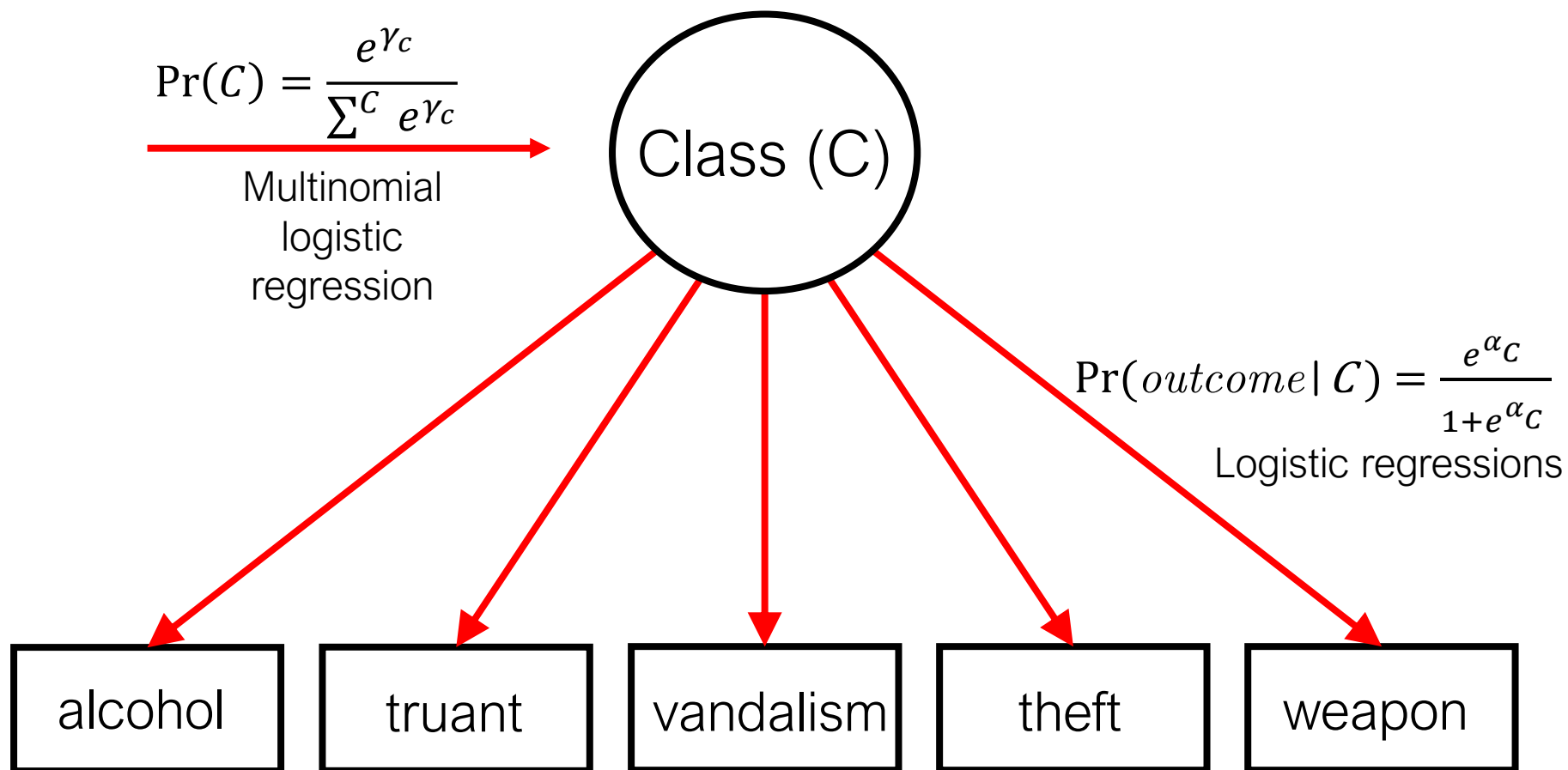
Model formulation

$$\Pr(C) = \frac{e^{\gamma_c}}{\sum^C e^{\gamma_c}}$$

Multinomial
logistic
regression



Model formulation



LCA in Stata

```
. 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) iter(500)
. estimates store c3
. gsem (alcohol truant vandalism theft weapon <- , logit), lclass(C 4) ///
> iter(1000) startvalues(randompr, draws(10) seed(27823))
  convergence not achieved
  r(430);
```

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) iter(500)
. estimates store c3

. estimates stats c*
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
c1	10,000	.	-14863.47	5	29736.93	29772.98
c2	10,000	.	-14430.33	11	28882.66	28961.97
c3	10,000	.	-14411.8	17	28857.61	28980.18

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) iter(500)
. estimates store c3

. estimates stats c*
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
c1	10,000	.	-14863.47	5	29736.93	29772.98
c2	10,000	.	-14430.33	11	28882.66	28961.97
c3	10,000	.	-14411.8	17	28857.61	28980.18

Probabilities for class

```
. estimates restore c2
```

```
. 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 c3
```

```
. estat lcprob
```

Latent class marginal probabilities

Number of obs = **10,000**

	Delta-method			
	Margin	std. err.	[95% conf. interval]	
C				
1	.8970379	.035944	.8024977	.949189
2	.0803194	.0407247	.0287884	.2046539
3	.0226427	.0112753	.0084624	.0591667

Class means

```
. estimates restore c2
. 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 c3
```

```
. estat lcmean
```

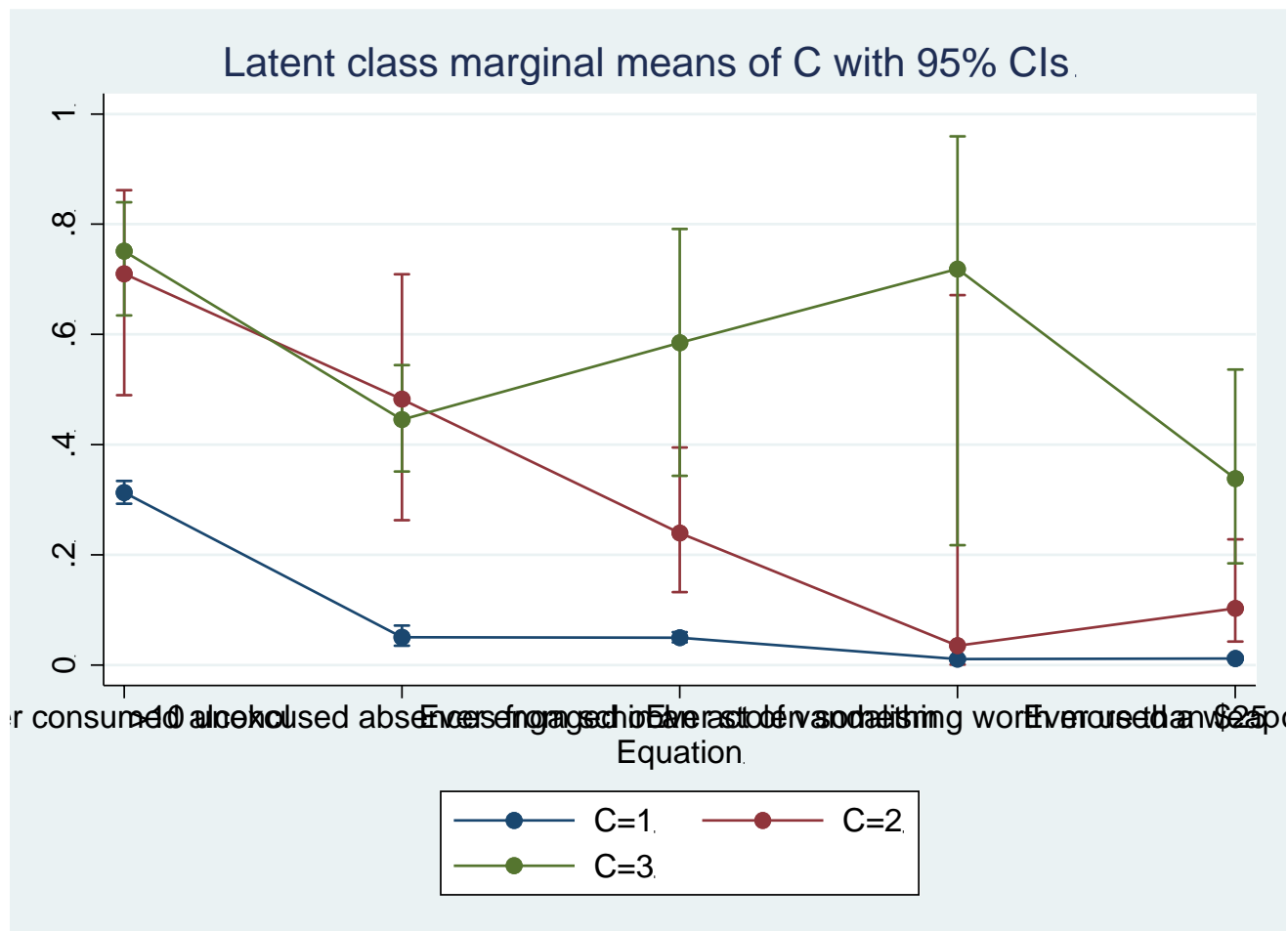
Latent class marginal means

Number of obs = **10,000**

		Delta-method		
		Margin	std. err.	[95% conf. interval]
1	alcohol	.313002	.0105301	.2927426 .3340014
	truant	.0503375	.0091115	.0351981 .071506
	vandalism	.049624	.0047158	.0411558 .0597262
	theft	.0108107	.0017266	.0079013 .0147755
	weapon	.0118076	.0020411	.0084092 .0165566
2	alcohol	.7098505	.0979048	.4907475 .861324
	truant	.4824674	.1219121	.263645 .7082274
	vandalism	.2395513	.0675944	.1321132 .3946326
	theft	.0352944	.0691424	.0006831 .6619473
	weapon	.1026839	.0445793	.042451 .2280279
3	alcohol	.7510906	.0526221	.6347765 .8397153
	truant	.4455787	.0498741	.3510917 .5441701
	vandalism	.584935	.1216742	.3454305 .790065
	theft	.7186955	.2292366	.2168124 .9593142
	weapon	.3383354	.0924126	.1854567 .5345359

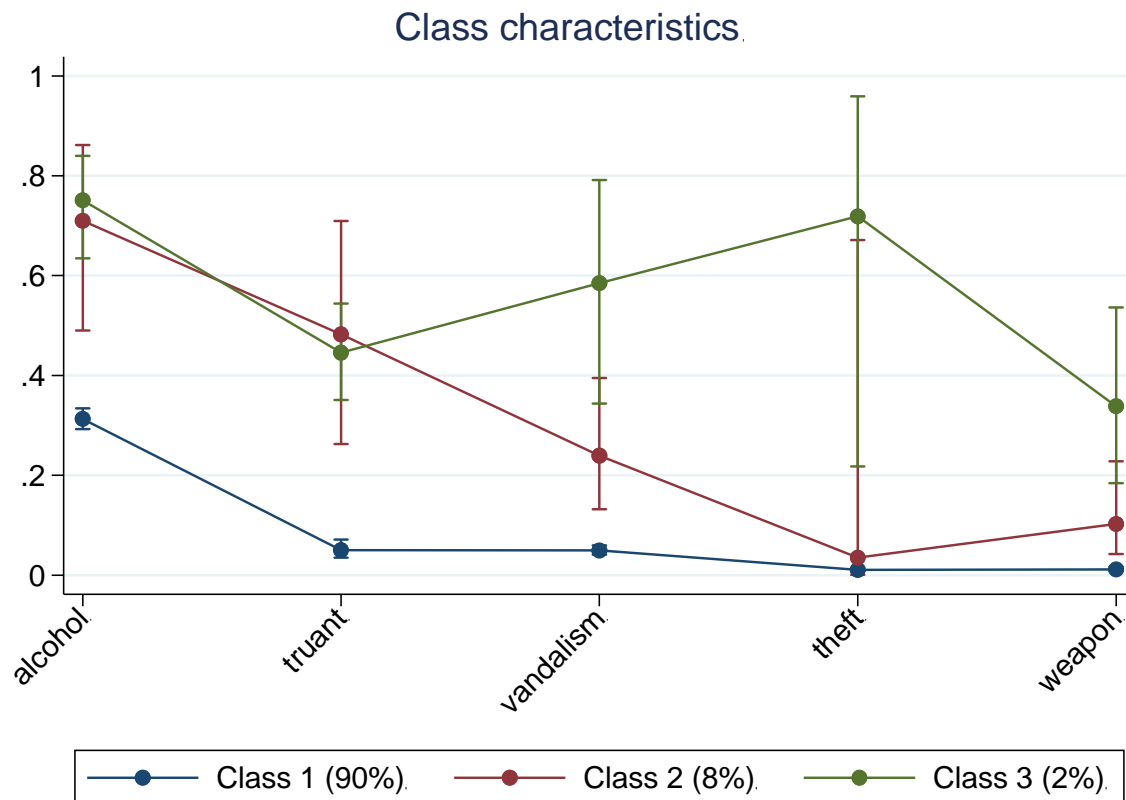
Class characteristics plot

```
. estat lcmean
. marginsplot
```



Class characteristics plot

```
. marginsplot, xtitle("") title("Class characteristics") scheme(S2flyer) ///  
> xlabel(1 "alcohol" 2 "truant" 3 "vandalism" 4 "theft" 5 "weapon", angle(45)) ///  
> legend(order(4 "Class 1 (90%)" 5 "Class 2 (8%)" 6 "Class 3 (2%)") rows(1))
```



Goodness of fit

```
. 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

Predicted probabilities

```
. predict cpr*, classposteriorpr
```

```
. list id alcohol truant vandalism theft weapon cp* in 1/10
```

	id	alcohol	truant	vandal~m	theft	weapon	cpr1	cpr2	cpr3
1.	1	0	0	0	0	0	.9851736	.014389	.0004373
2.	2	0	0	1	0	1	.4243329	.3580979	.2175692
3.	3	0	0	0	0	0	.9851736	.014389	.0004373
4.	4	1	0	0	0	0	.9247524	.0725288	.0027189
5.	5	1	1	0	0	1	.0620399	.8196018	.1183583
6.	6	0	0	0	0	0	.9851736	.014389	.0004373
7.	7	0	0	0	0	0	.9851736	.014389	.0004373
8.	8	0	0	0	0	0	.9851736	.014389	.0004373
9.	9	0	0	0	0	0	.9851736	.014389	.0004373
10.	10	1	0	0	0	0	.9247524	.0725288	.0027189

Predicted class membership

```
. egen maxpr = rowmax(cpr*)
. generate predclass = 1 if cpr1==maxpr
. replace predclass = 2 if cpr2==maxpr
. replace predclass = 3 if cpr3==maxpr
. list id cpr* maxpr predclass in 1/10
```

	id	cpr1	cpr2	cpr3	maxpr	predcl~s
1.	1	.9851727	.0143899	.0004373	.9851727	1
2.	2	.4243248	.3581067	.2175684	.4243248	1
3.	3	.9851727	.0143899	.0004373	.9851727	1
4.	4	.9247497	.0725315	.0027187	.9247497	1
5.	5	.0620393	.8196039	.1183568	.8196039	2
6.	6	.9851727	.0143899	.0004373	.9851727	1
7.	7	.9851727	.0143899	.0004373	.9851727	1
8.	8	.9851727	.0143899	.0004373	.9851727	1
9.	9	.9851727	.0143899	.0004373	.9851727	1
10.	10	.9247497	.0725315	.0027187	.9247497	1

Classification matrix

```
. table predclass, statistic(mean cpr1 cpr2 cpr3)
```

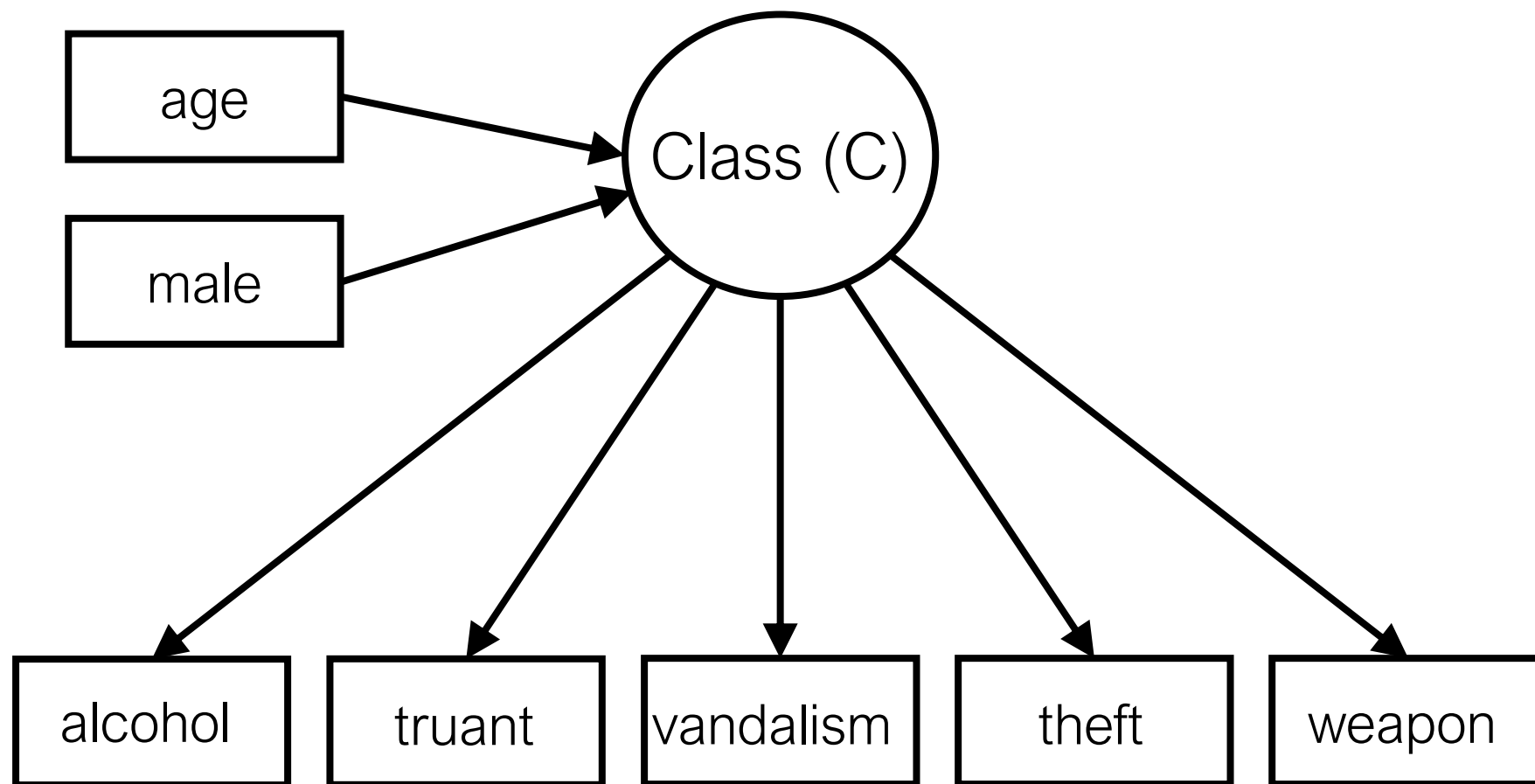
	Class1pr	Class2pr	Class3pr
predclass			
1	.9424795	.0512143	.0063062
2	.3219311	.6143815	.0636874
3	.0720893	.106033	.8218777
Total	.8970286	.0803335	.022638

Classification matrix

```
. table predclass, statistic(mean cpr1 cpr2 cpr3)
```

	Class1pr	Class2pr	Class3pr
predclass			
1	.9424795	.0512143	.0063062
2	.3219311	.6143815	.0636874
3	.0720893	.106033	.8218777
Total	.8970286	.0803335	.022638

Adding predictors



Adding predictors

```
. gsem (alcohol truant vandalism theft weapon <- , logit) (C <- i.male age), ///
> lclass(C 3) startvalues(classpr cpr1 cpr2 cpr3)
```

Generalized structural equation model
Log likelihood = **-14336.506**

Number of obs = **10,000**

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1.C	(base outcome)					
2.C						
male						
Male	1.578438	.4376642	3.61	0.000	.7206319	2.436244
age	.1241829	.0565678	2.20	0.028	.0133122	.2350537
_cons	-4.877966	1.164192	-4.19	0.000	-7.15974	-2.596192
3.C						
male						
Male	.3741327	.1726691	2.17	0.030	.0357076	.7125579
age	.3351213	.0485855	6.90	0.000	.2398954	.4303472
_cons	-8.531909	.8530293	-10.00	0.000	-10.20382	-6.860002

Adding predictors

```
. gsem (alcohol truant vandalism theft weapon <- , logit) (C <- i.male age), ///
> lclass(C 3) startvalues(classpr cpr1 cpr2 cpr3)
```

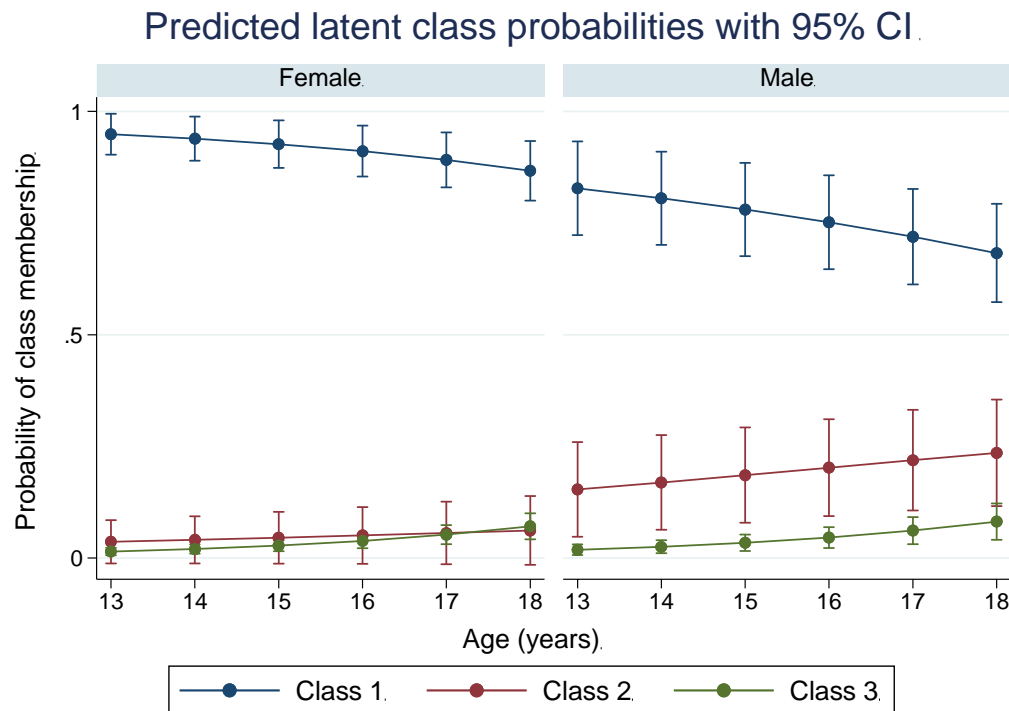
Generalized structural equation model
Log likelihood = **-14336.506**

Number of obs = **10,000**

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
1.C	(base outcome)					
2.C						
male						
Male	1.578438	.4376642	3.61	0.000	.7206319	2.436244
age	.1241829	.0565678	2.20	0.028	.0133122	.2350537
_cons	-4.877966	1.164192	-4.19	0.000	-7.15974	-2.596192
3.C						
male						
Male	.3741327	.1726691	2.17	0.030	.0357076	.7125579
age	.3351213	.0485855	6.90	0.000	.2398954	.4303472
_cons	-8.531909	.8530293	-10.00	0.000	-10.20382	-6.860002

Class probabilities by covariates

```
. margins male, at(age=(13/18)) predict(classpr class(1))  
> predict(classpr class(2)) predict(classpr class(3))  
. marginsplot, by(male) ytitle("Probability of class membership")  
> byopts(title("Predicted latent class probabilities with 95% CI"))  
> legend(order(4 "Class 1" 5 "Class 2" 6 "Class 3") rows(1))
```



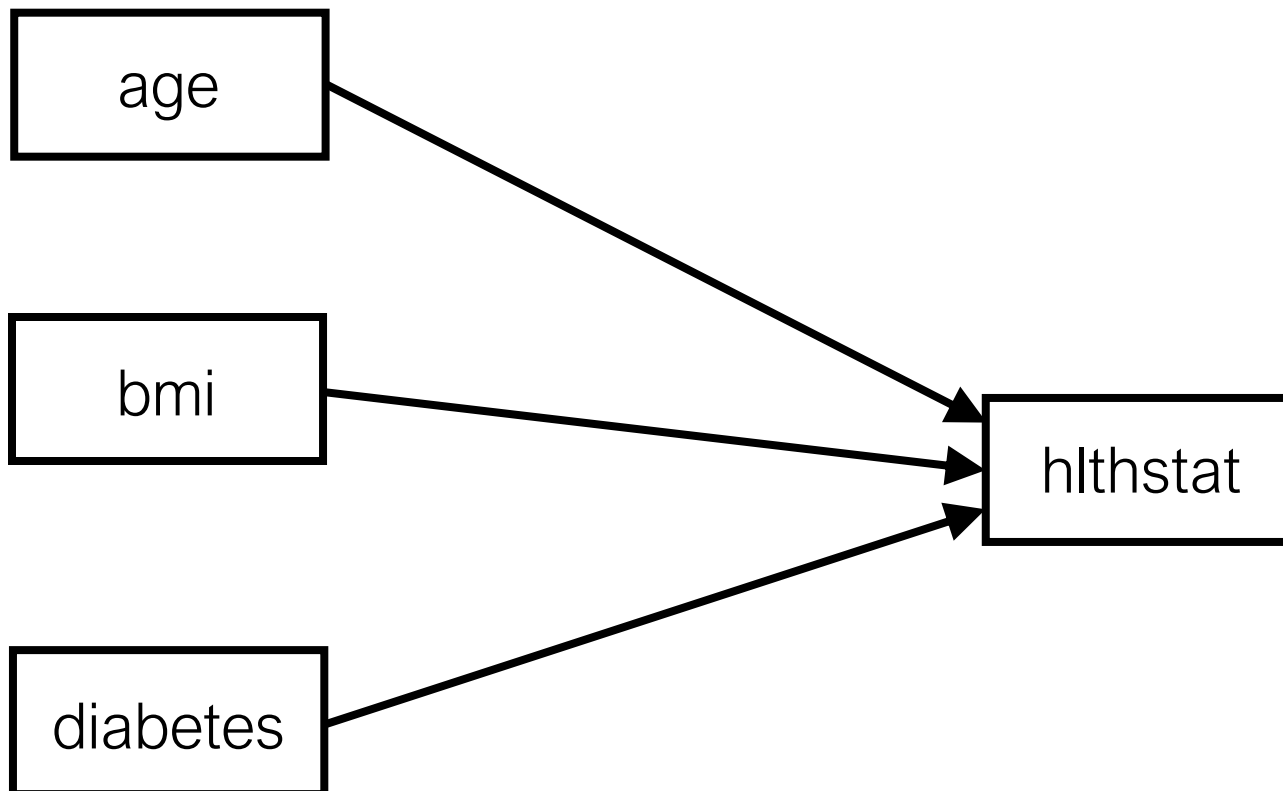
NHANES example data

```
. use nhanes, clear
```

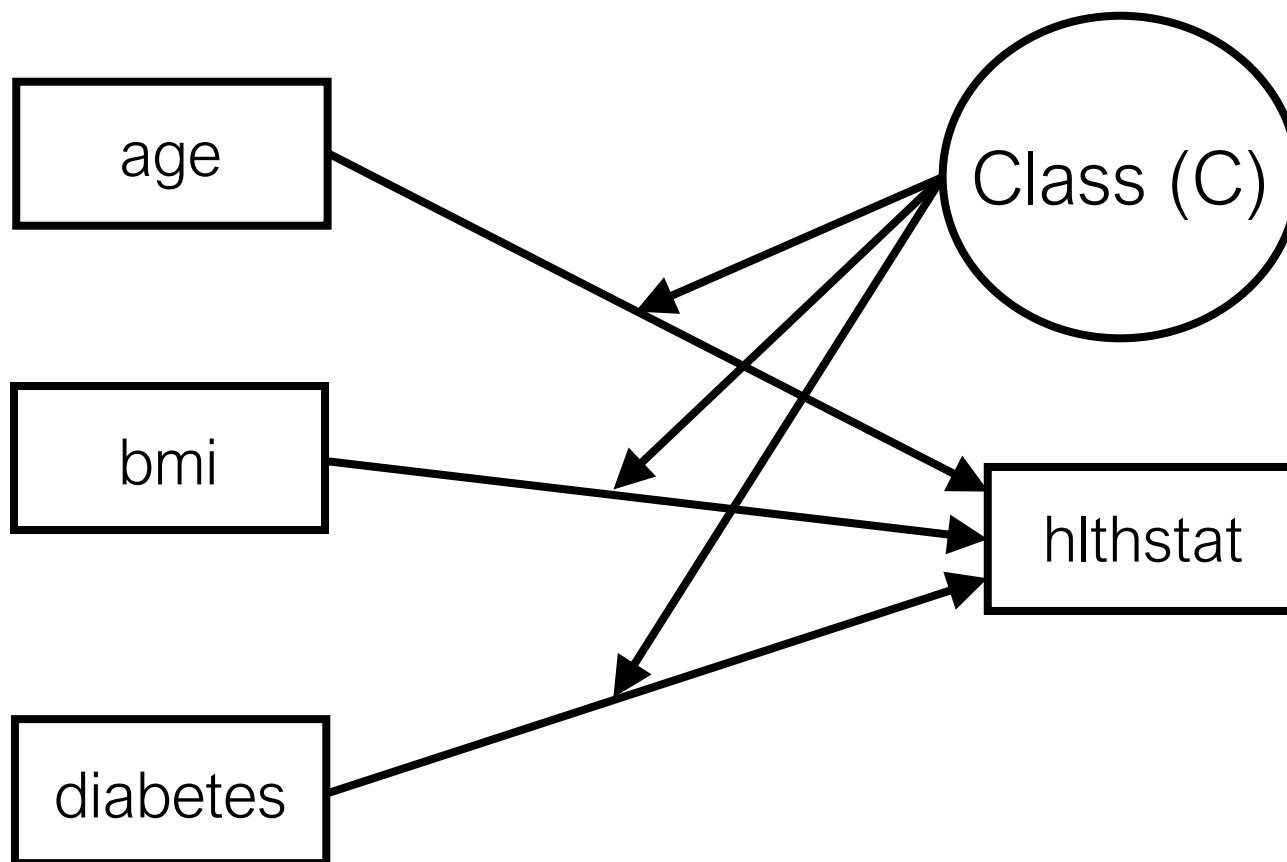
```
. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
smsa	10335	3	2.65612	1	4	Standard metropolitan statistical area
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

	Coefficient	Std. err.
hlthstat		
age	.0156473	.0007907
diabetes		
Diabetic	.8507712	.0693641
bmi	.0142115	.0021331
_cons	.690505	.0605036
var(e.hlthstat)	.572821	.0137438

Class: 2

	Coefficient	Std. err.
hlthstat		
age	.0266047	.0008824
diabetes		
Diabetic	.6327083	.059429
bmi	.0142115	.0021331
_cons	1.784947	.0726737
var(e.hlthstat)	.572821	.0137438

The `lcinvariant()` option

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

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

	Coefficient	Std. err.
hlthstat		
age	.0239003	.0002114
diabetes		
Diabetic	.9596019	.0978575
bmi	.0015937	.001128
diabetes#c.bmi		
Diabetic	.0004875	.0034765
_cons	1.487675	.0305586
var(e.hlthstat)	1.236106	.018291

Class: 2

	Coefficient	Std. err.
hlthstat		
age	.0239003	.0002114
diabetes		
Diabetic	.9596019	.0978575
bmi	.0015937	.001128
diabetes#c.bmi		
Diabetic	.0004875	.0034765
_cons	.3520712	.0270075
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

	Coefficient	Std. err.
hlthstat		
age	.0156067	.0007903
bmi	.0170333	.0030314
diabetes		
Diabetic	.8491393	.0678729
_cons	.6315143	.0765613
var(e.hlthstat)	.5740851	.0139179

Class: 2

	Coefficient	Std. err.
hlthstat		
age	.0276636	.0009063
diabetes		
Diabetic	2.095236	.3610216
diabetes#c.age		
Diabetic	-.0238014	.0058113
_cons	2.107465	.0518677
var(e.hlthstat)	.5740851	.0139179

Comparing 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*
```

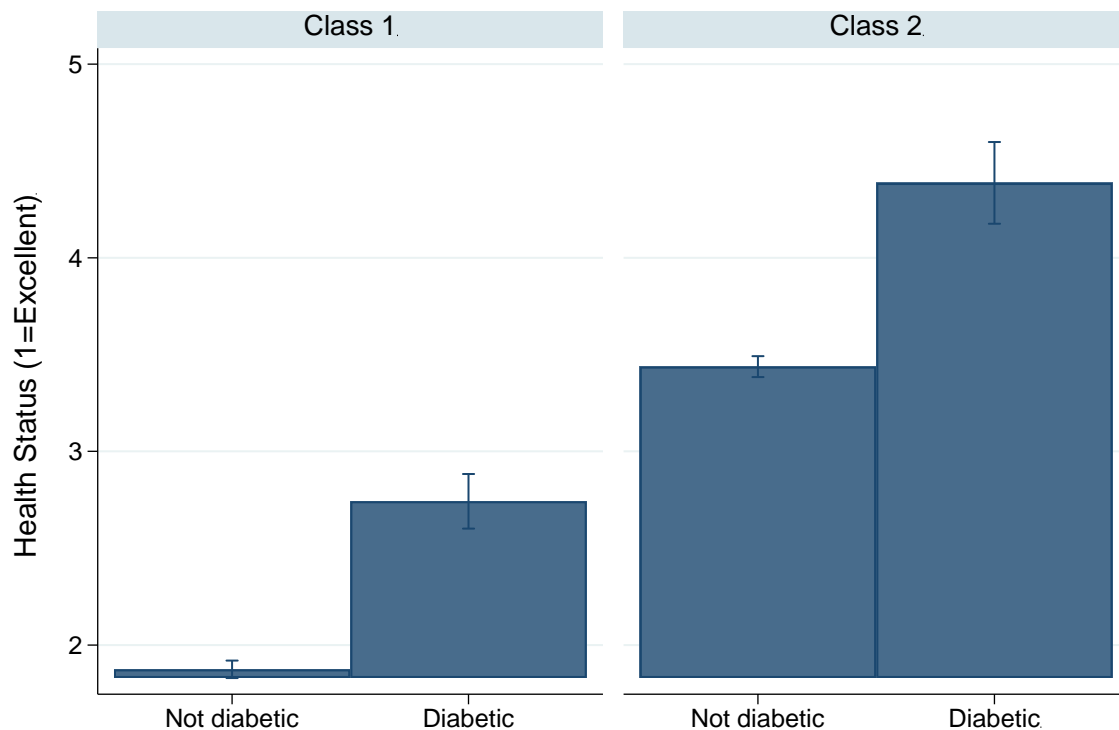
Akaike's information criterion and Bayesian information criterion

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

Marginal means by class

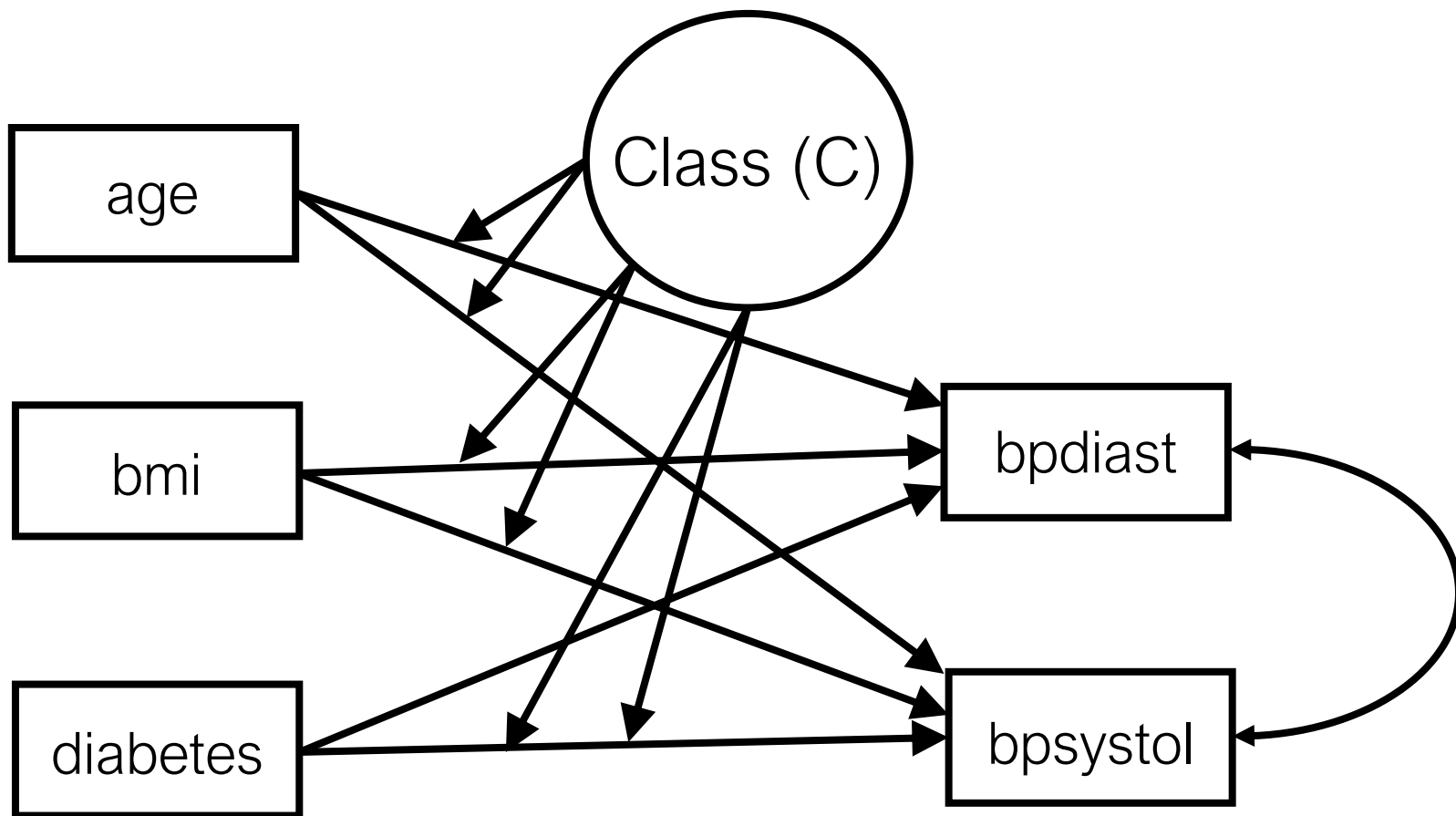
```
. margins diabetes, predict(class(1)) predict(class(2))  
. marginsplot, by(_predict, label("Class 1" "Class 2")) recast(bar) ///  
> ytitle("Health Status (1=Excellent)") xtitle("")
```

Predictive margins of diabetes with 95% CIs



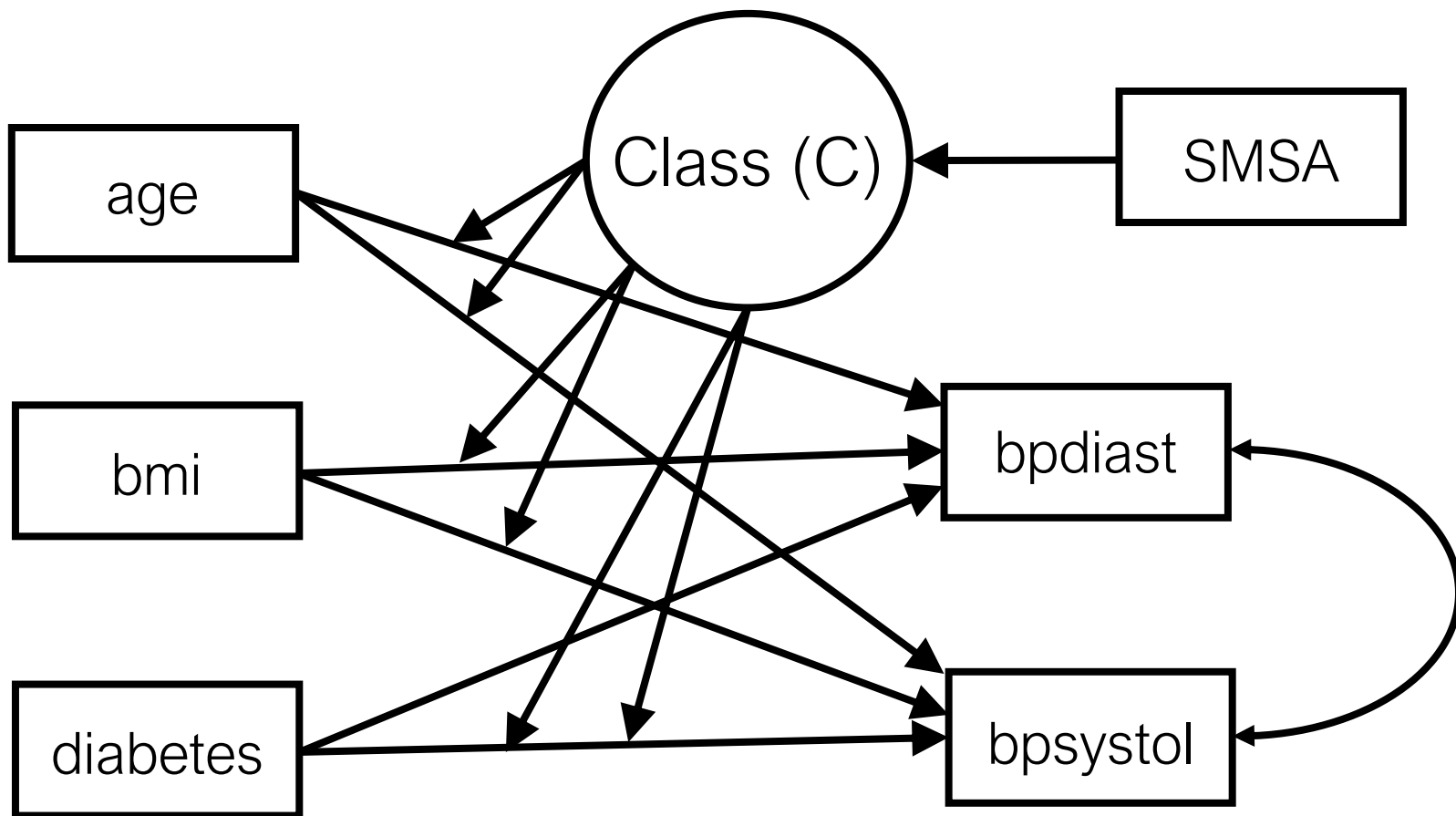
Multivariate regression mixture

```
. 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?

You can download the dataset, do-file, and slides here:

<https://tinyurl.com/StataLCA>

You can contact tech support at tech-support@stata.com