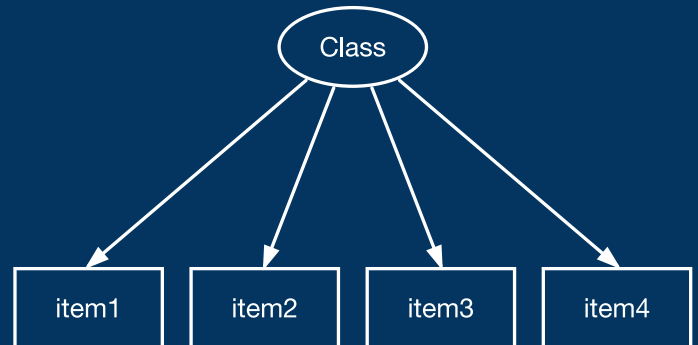


Latent class analysis (LCA)

Discover and understand the unobserved groupings in your data—consumers with different buying preferences, adolescents with different patterns of behavior, individuals with different health status classifications ...

Determine who is likely to be in each group and how that group's characteristics differ from other groups.



• Fit

- Latent class models
- Latent profile models
- Path models with categorical latent variables
- Multiple-group models with known groups

• Categorical latent variables measured by

- Binary items
- Ordinal items
- Continuous items
- Count items
- Categorical items
- Fractional items
- Even survival items

• Model-based method of classification

• Estimate probabilities, means, and counts for items in each class

• Estimate proportion of population in each class

• Predict class membership

• Goodness of fit

- G^2
- AIC
- BIC

• Multiple options for obtaining starting values

• Support for complex survey data

• Point and click to fit any model

We could fit a latent class model for adolescent behaviors using variables that indicate whether an individual consumed alcohol, was truant from school, used a weapon in a fight, engaged in vandalism, or committed theft. We simply type

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

Our model assumes that there are three latent classes—three unobserved groups of adolescents.

Based on the fitted model, we can estimate the proportion of adolescents in the population belonging to each class.

Latent class marginal probabilities		Number of obs = 10,000		
C	Margin	Delta-method std. err.	[95% conf. interval]	
1	.1631459	.0390464	.1001516	.2545542
2	.7979467	.0389126	.7110459	.8637217
3	.0389074	.0165519	.0167174	.0879179

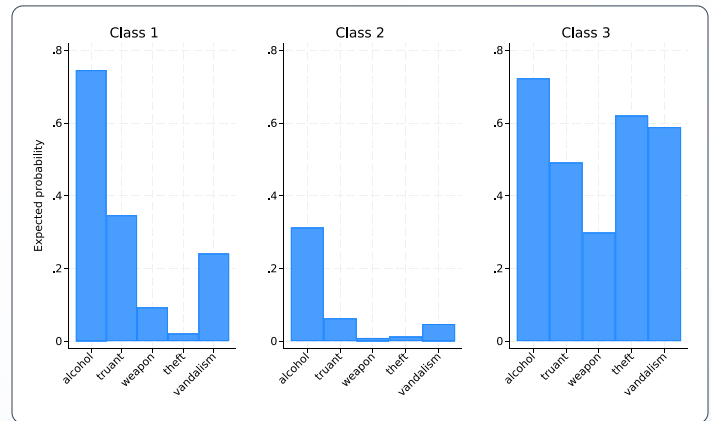
About 16% of adolescents are expected to be in the first class, 80% in the second, and 4% in the third.

How do these classes differ? **estat lcmean** estimates the mean—in this case, a probability—of the observed variables in each class.

Latent class marginal means		Number of obs = 10,000			
	Margin	Delta-method std. err.	[95% conf. interval]		
1	alcohol	.7453054	.055844	.6217857	.8389347
	truant	.3461541	.0511504	.2537076	.4518892
	weapon	.0928717	.0273733	.0513735	.162161
	theft	.0207514	.0341545	.0007855	.3635619
	vandalism	.2407638	.0519997	.1536777	.3564169
2	alcohol	.3120356	.0150696	.2832886	.3423065
	truant	.0626883	.0076641	.0492432	.0794975
	weapon	.0089407	.0023358	.0053525	.0148983
	theft	.0123995	.002113	.0088731	.0173028
	vandalism	.0471581	.005303	.0377877	.0587103
3	alcohol	.7227077	.0346378	.6500293	.7852786
	truant	.4910226	.0426644	.4084192	.5741192
	weapon	.2985074	.0498658	.2106265	.4042764
	theft	.6199426	.1870201	.2560825	.8854454
	vandalism	.5883387	.0735654	.4407243	.7216029

Probabilities of drinking alcohol, being truant, etc., are the lowest for individuals in the second class. The third class has higher probabilities of engaging in each of these behaviors.

We can use **margins** and **marginsplot** to visually compare the probabilities of participating in these activities across classes.



Did our model fit well?

Fit statistic	Value	Description
Likelihood ratio	chi2_ms(14)	model vs. saturated
	p > chi2	0.949
Information criteria		
	AIC	32510.523 Akaike's information criterion
	BIC	32633.099 Bayesian information criterion

The likelihood-ratio test indicates that our model has reasonable fit. We could use AIC and BIC to compare the fit of this model with other models.

We are not limited to this basic latent class model.

Want to use continuous instead of binary observed variables?

```
. gsem (y1 y2 y3 y4 y5 <-), regress lclass(C 3)
```

Or use ordinal observed variables?

```
. gsem (y1 y2 y3 y4 y5 <-), ologit lclass(C 3)
```

Or even mix types of observed variables?

```
. gsem (y1 <-, regress)  
      (y2 <-, poisson)  
      (y3 <-, logit)  
      (y4 <-, logit)  
      (y5 <-, ologit),  
      lclass(C 3)
```

Want to include a predictor of class membership?

```
. gsem (y1 y2 y3 y4 y5 <-) (C <- income),  
      logit lclass(C 3)
```

Want to fit a path model with class-specific parameter estimates?

```
. gsem (y1 <- y2 x1 x2)  
      (y2 <- y3 x1 x3)  
      (y3 <- x2 x3 x4),  
      lclass(C 3)
```

You can do all of this and much more.