Title

> **example 51g —** Latent class goodness-of-fit statistics

[Description](#)   [Remarks and examples](#)   [Reference](#)   [Also see](#)

Description

Here we demonstrate how to obtain goodness-of-fit statistics for latent class models.

We continue with [SEM] **example 50g**, where we fit a two-class model:

```
. use http://www.stata-press.com/data/r15/gsem_lca1
. gsem (accident play insurance stock <- ), logit lclass(C 2)
```

See *Latent class models* in [SEM] **intro 5** for background.

Remarks and examples

Remarks are presented under the following headings:

> *Likelihood-ratio (G^2) test*
> *Comparing models*

Likelihood-ratio (G^2) test

For standard latent class models with observed variables that are all categorical, one way to evaluate model fit is to compare the model we have just fit with a saturated model. We can use the `estat lcgof` command to perform a likelihood-ratio test of whether our model fits as well as the saturated model. The corresponding likelihood-ratio statistic is sometimes referred to as G^2 in latent class analysis literature.

```
. estat lcgof
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(6)	2.720	model vs. saturated
p > chi2	0.843	
Information criteria		
AIC	1026.935	Akaike's information criterion
BIC	1057.313	Bayesian information criterion

We fail to reject the null hypothesis that our model fits as well as the saturated model.

`estat lcgof` also reports Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC). These are useful for comparing models but not useful for determining goodness-of-fit for a single model.

Comparing models

In latent class analysis, we often compare models that have different numbers of classes. Following Goodman (2002), we compare models that allow for one, two, and three latent classes. We have already fit the two-class model using the gsem command above. Before we move on, we will store the results of this model.

```
. estimates store twoclass
```

Next, we fit the one-class model, store the results, and perform the likelihood-ratio test comparing it with the saturated model.

```
. quietly gsem (accident play insurance stock <- ), logit lclass(C 1)
. estimates store oneclass
. estat lcgof
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(11)	81.084	model vs. saturated
p > chi2	0.000	
Information criteria		
AIC	1095.300	Akaike's information criterion
BIC	1108.801	Bayesian information criterion

We reject the null hypothesis in this case. The one-class model does not fit well.

We also fit the three-class model.

```
. quietly gsem (accident play insurance stock <- ), logit lclass(C 3)
. estimates store threeclass
. estat lcgof
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(1)	0.387	model vs. saturated
p > chi2	0.534	
Information criteria		
AIC	1034.602	Akaike's information criterion
BIC	1081.856	Bayesian information criterion

Based on this test, the three-class model, like the two-class model, does not fit worse than the saturated model.

We will compare our three models using AIC and BIC. Smaller values of AIC and BIC are better. We could look back at the AIC and BIC values reported by our three estat lcgof commands, but we will instead create a table that reports these information criteria for all three models using estimates stats.

```
. estimates stats oneclass twoclass threeclass
```
Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
oneclass	216	.	-543.6498	4	1095.3	1108.801
twoclass	216	.	-504.4677	9	1026.935	1057.313
threeclass	216	.	-503.3011	14	1034.602	1081.856

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

The two-class model has both the smallest AIC and the smallest BIC.

Reference

Goodman, L. A. 2002. Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In *Applied Latent Class Analysis*, ed. J. A. Hagenaars and A. L. McCutcheon, 3–55. Cambridge: Cambridge University Press.

Also see

[SEM] **example 50g** — Latent class model

[SEM] **gsem** — Generalized structural equation model estimation command

[SEM] **intro 5** — Tour of models

[SEM] **estat lcgof** — Latent class goodness-of-fit statistics