Models in STATA

# Clogit

clogit — Conditional (fixed-effects) logistic regression

### Description

clogit fits a conditional logistic regression model for matched case –control data, also known as a fixed-effects logit model for panel data. clogit can compute robust and cluster–robust standard errors and adjust results for complex survey designs.

Results: <https://toioslo.sharepoint.com/:f:/r/sites/YoungTalentroadpricing/Delte%20dokumenter/General/WP2/Analysis/Sara/Analysis%20second%20pilot/clogit/Tables/CE%20policy?csf=1&web=1&e=RZpWmh>

# Cmlogit

Conditional logit (McFadden’s) choice model

### Description

cmclogit fits McFadden’s choice model, which is a specific case of the more general conditional logistic regression model fit by clogit. The command requires multiple observations for each case (representing one individual or decision maker), where each observation represents an alternative that may be chosen. cmclogit allows two types of independent variables: alternative-specific variables, which vary across both cases and alternatives, and case-specific variables, which vary across only cases.

Results:

<https://toioslo.sharepoint.com/:f:/r/sites/YoungTalentroadpricing/Delte%20dokumenter/General/WP2/Analysis/Askill/Tables?csf=1&web=1&e=lvP9gn>

# Multinomial logit

# mlogit — Multinomial (polytomous) logistic regression

### Description

mlogit fits a multinomial logit (MNL) model for a categorical dependent variable with outcomes that have no natural ordering. The actual values taken by the dependent variable are irrelevant. The MNL model is also known as the polytomous logistic regression model. Some people refer to conditional logistic regression as multinomial logistic regression.

Results: <https://toioslo.sharepoint.com/:f:/r/sites/YoungTalentroadpricing/Delte%20dokumenter/General/WP2/Analysis/Sara/Analysis%20second%20pilot/mlogit?csf=1&web=1&e=bjciZX>

# Latent class model

<https://www.stata.com/meeting/uk18/slides/uk18_MacDonald.pdf>

gsem lclass options — Fitting models with latent classes

### Description

gsem can fit models with categorical latent variables having specified numbers of latent classes. Some parameters can vary across classes while others are constrained to be equal across classes. gsem performs such estimation when the lclass() option is specified. The lcinvariant(pclassname) option specifies which parameters are to be constrained to be equal across the latent classes.

## Procedure on STATA

##### 1. Fit Models with Different Numbers of Classes

* Begin by fitting models with a range of possible numbers of classes (e.g., 1 to 5 or more). In Stata, you can use the gsem command or an LCA-specific package like lclogit or gllamm.

### 2. Evaluate Model Fit Criteria

Use these statistical criteria to compare models with different numbers of classes:

a. Akaike Information Criterion (AIC)

* A measure of model fit penalized for complexity. Lower values indicate a better model.

b. Bayesian Information Criterion (BIC)

* Similar to AIC but imposes a stronger penalty for the number of parameters. It often favors simpler models.
* **BIC is commonly preferred for selecting the number of classes.**

c. Sample-Adjusted BIC (SABIC)

* Adjusts BIC for small sample sizes.

### 3. Likelihood Ratio Test (LRT)

* **Lo-Mendell-Rubin Adjusted LRT** or **Vuong-Lo-Mendell-Rubin (VLMR) Test**:
  + Compares a model with kkk classes to one with k−1k-1k−1 classes.
  + A significant p-value indicates that the model with kkk classes fits better than the model with k−1k-1k−1 classes.

Stata Command for LRT:

* You can use the estat lcfit command after fitting a model in Stata to compare fit indices for different numbers of classes.

### 4. Entropy

* Entropy ranges from 0 to 1 and indicates the classification quality:
  + Higher values suggest better separation of classes.
  + Entropy is not a selection criterion by itself but can help interpret class distinctiveness.

### 5. Interpretability and Practicality

* Even if the statistical criteria suggest a certain number of classes, consider:
  + Whether the classes make sense substantively.
  + The size of each class (very small classes may be unreliable or uninformative).

### 6. Visualize Results

* Use item-response probabilities, class membership probabilities, and graphs to explore the characteristics of the identified classes.

### 

# finite mixture models

Populations are often divided into groups or subpopulations—age groups, income brackets, levels of education. Regression models or distributions likely differ across these groups. But sometimes we don't have a variable that identifies the groups. Perhaps the identifying variable is simply missing. Perhaps it is hard to collect—honest reporting of drug use, sex of goldfish, etc. Perhaps it is inherently unobservable—penchant for risky behavior, high propensity to save money, etc. In such cases, we can use finite mixture models (FMMs) to model the probability of belonging to each unobserved group, to estimate distinct parameters of a regression model or distribution in each group, to classify individuals into the groups, and to draw inferences about how each group behaves.

Latent class models that have one dependent variable, can be seen as finite mixture models. The fmm prefix allows us to easily fit finite mixture models for a variety of distributions

Finite mixture model Number of obs = 11,532

Log likelihood = -7010.1513

------------------------------------------------------------------------------

| Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

1.Class | (base outcome)

-------------+----------------------------------------------------------------

2.Class |

\_cons | -.5812296 .3906016 -1.49 0.137 -1.346795 .1843354

-------------+----------------------------------------------------------------

3.Class |

\_cons | -.2938017 .8092783 -0.36 0.717 -1.879958 1.292355

------------------------------------------------------------------------------

Class: 1

Response: choice

Model: mlogit

------------------------------------------------------------------------------

| Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

0.choice | (base outcome)

-------------+----------------------------------------------------------------

1.choice |

price\_ur | -1.041368 1.004537 -1.04 0.300 -3.010225 .9274887

price\_un | -2.706258 1.282536 -2.11 0.035 -5.219982 -.1925339

price\_o | -2.782835 1.624328 -1.71 0.087 -5.966459 .4007888

ev | .7970199 .4456856 1.79 0.074 -.0765077 1.670548

d\_rev5 | 1.952859 .4789825 4.08 0.000 1.014071 2.891648

\_cons | -.4377032 .4342471 -1.01 0.313 -1.288812 .4134056

------------------------------------------------------------------------------

Class: 2

Response: choice

Model: mlogit

------------------------------------------------------------------------------

| Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

0.choice | (base outcome)

-------------+----------------------------------------------------------------

1.choice |

price\_ur | 2.651815 2.159385 1.23 0.219 -1.580503 6.884133

price\_un | 6.885244 5.515632 1.25 0.212 -3.925197 17.69568

price\_o | 3.258045 1.993092 1.63 0.102 -.6483441 7.164434

ev | 122.0279 2079776 0.00 1.000 -4076165 4076409

d\_rev5 | .5870813 .6438397 0.91 0.362 -.6748213 1.848984

\_cons | -1.926604 .6375585 -3.02 0.003 -3.176196 -.677012

------------------------------------------------------------------------------

Class: 3

Response: choice

Model: mlogit

------------------------------------------------------------------------------

| Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

0.choice | (base outcome)

-------------+----------------------------------------------------------------

1.choice |

price\_ur | -6.072383 2.939334 -2.07 0.039 -11.83337 -.3113943

price\_un | -5.160603 3.978974 -1.30 0.195 -12.95925 2.638043

price\_o | 4.268362 2.859609 1.49 0.136 -1.336369 9.873093

ev | -.1765528 .9046262 -0.20 0.845 -1.949588 1.596482

d\_rev5 | 1.116295 .5847468 1.91 0.056 -.0297878 2.262378

\_cons | -1.830122 1.314267 -1.39 0.164 -4.406039 .7457952

------------------------------------------------------------------------------

The intercept and the coefficients on x1 and x2 will be estimated separately for the two classes. In addition, we estimate the probability of being in each class. If we have a variable z that predicts class membership, the command syntax becomes:

(y <- x1 x2) (C <- z), lclass(C 2)

gsem (choice <- price\_ur price\_un price\_o ev d\_rev2-d\_rev5, regress) (C <- policy), regress lclass(

> ariant(coef) nonrtolerance

with:

dependent variable: choice

covariates: price\_ur price\_un price\_o ev d\_rev2-d\_rev5

C <- policy: classes are defined with reference to the choice of option 1,2 or status quo

lcinvariant(coef): it states all coefficients constrained to be equal across classes.

lclass(C 6): this option specifies that the name of our categorical latent variable is C and that it has six latent classes

nonrtolerance: is a maximize option.

