Description   Remarks and examples   Reference

Description

Seven commands for fitting choice models (CM) are documented in this manual. These commands are used almost exclusively with choice data. Many other commands can also be useful for modeling choice data but are often used with other types of data as well. In this entry, we give you an overview of estimation commands available in Stata for modeling choice data.

Remarks and examples

Remarks are presented under the following headings:

Specialized choice model commands

The following commands are documented in this manual and are designed specifically for fitting choice models:

Estimators for discrete choices

`cmclogit`	Conditional logit (McFadden's) choice model
`cmmixlogit`	Mixed logit choice model
`cmmprobit`	Multinomial probit choice model
`nlogit`	Nested logit regression

See [CM] **Intro 5** for details on these estimators.

Estimators for rank-ordered choices

`cmrologit`	Rank-ordered logit choice model
`cmroprobit`	Rank-ordered probit choice model

See [CM] **Intro 6** for details on these estimators.

Estimator for discrete choices with panel data

`cmxtmixlogit`	Panel-data mixed logit choice model

See [CM] **Intro 7** for details on this estimator.

What is special about the seven estimators listed here is that they all require your data to be in long form. That is, each case consists of multiple Stata observations, one for each of its available alternatives. All of these estimators allow you to include alternative-specific variables as covariates in your model. In addition, each of these estimators handles unbalanced choice sets. In [CM] **Intro 5**, [CM] **Intro 6**, and [CM] **Intro 7**, we provide more in-depth introductions to these estimators for discrete choices, rank-ordered alternatives, and discrete choices in panel data, respectively.

Other commands for choice models

Many other commands in Stata can be used for choice modeling. When you use these general commands with choice data, it is important to consider the restrictions or limitations of the model to make sure that it is the best command for modeling your choice data. For instance, the `mlogit` command fits multinomial logit models. When you use multinomial logit to fit a choice model, you are required to have only case-specific variables as predictors. Multinomial logit also requires balanced choice sets (that is, every decision maker must have the same available alternatives). Another example is the `clogit` command. You can use it to fit the same McFadden's choice model fit by `cmclogit`. In fact, `cmclogit` calls `clogit` to produce its estimates. However, because `cmclogit` is specifically designed for choice models, it gives you appropriate handling of missing values for choice data, and the postestimation command `margins` gives you options for unbalanced data after `cmclogit`. Nonetheless, when your choice data meet the requirements for one of Stata's commands for binary or categorical outcomes, they are useful for choice modeling.

Models for cross-sectional data

Stata has many commands for fitting binary and categorical outcome models that can be applied to some types of choice data. When a decision maker chooses from only two possible alternatives, the commands for binary outcomes may be useful. When a decision maker chooses from more than two outcomes, the commands for categorical outcomes may be appropriate. In addition to commands for common models such as logistic and probit, you can select from commands that address problems including heteroskedasticity, endogenous covariates, and sample selection. You can use the `fmm:` prefix to fit mixtures of choice models. If you want to simultaneously model more than one outcome variable or if you want to include latent variables, you can use `gsem` to fit a generalized structural equation model that allows binary and categorical outcomes. You can also use Bayesian estimation. The `bayes:` prefix allows you to fit Bayesian regression models. `bayesmh` is a flexible command that allows you to specify your own Bayesian model.

The following commands can fit choice models for cross-sectional data:

Estimators for binary choices

`cloglog`	Complementary log–log regression
`logistic`	Logistic regression, reporting odds ratios
`logit`	Logistic regression, reporting coefficients
`probit`	Probit regression

Exact statistics

`exlogistic`	Exact logistic regression

With endogenous sample selection

`heckprobit`	Probit model with sample selection

With heteroskedasticity

`hetprobit`	Heteroskedastic probit model

With endogenous covariates

`ivprobit`	Probit model with continuous endogenous covariates

With endogenous covariates and sample selection

`eprobit`	Extended probit regression

Finite mixture models

fmm: cloglog	Finite mixtures of complementary log–log regression models
fmm: logit	Finite mixtures of logistic regression models
fmm: probit	Finite mixtures of probit regression models

Multiple outcome variables and latent variables

gsem	Generalized structural equation model estimation command

Bayesian estimation

bayes: cloglog	Bayesian complementary log–log regression
bayes: logistic	Bayesian logistic regression, reporting odds ratios
bayes: logit	Bayesian logistic regression, reporting coefficients
bayes: probit	Bayesian probit regression
bayes: hetprobit	Bayesian heteroskedastic probit regression
bayes: heckprobit	Bayesian probit model with sample selection
bayesmh	Bayesian models using Metropolis–Hastings algorithm

Estimators for categorical outcomes

clogit	Conditional (fixed-effects) logistic regression
mlogit	Multinomial (polytomous) logistic regression
mprobit	Multinomial probit regression

Finite mixture models

fmm: mlogit	Finite mixtures of multinomial (polytomous) logistic regression models

Systems of equations and latent variables

gsem	Generalized structural equation model estimation command

Bayesian estimation

bayes: clogit	Bayesian conditional logistic regression
bayes: mlogit	Bayesian multinomial logistic regression
bayes: mprobit	Bayesian multinomial probit regression
bayesmh	Bayesian models using Metropolis–Hastings algorithm

Models for panel data

If you are working with panel data, you may be interested in standard panel-data commands for binary outcomes. For categorical outcomes, the `gsem` command can fit a multinomial logit model with random effects in addition to accommodating multiple outcome variables and latent variables for both binary and categorical outcomes. Bayesian estimation is also available. The `bayes:` prefix provides support for some of the `xt` commands, and it can be used with `me` commands to fit random-effects models. The following commands can be useful for choice models with panel data:

Estimators for binary choices

`xtcloglog`	Random-effects and population-averaged cloglog models
`xtlogit`	Fixed-effects, random-effects, and population-averaged logit models
`xtprobit`	Random-effects and population-averaged probit models

With endogenous covariates and sample selection

`xteprobit`	Extended random-effects probit regression

Multiple outcome variables and latent variables

`gsem`	Generalized structural equation models

Bayesian estimation

`bayes: mecloglog`	Bayesian multilevel complementary log–log regression
`bayes: xtlogit`	Bayesian random-effects logit model
`bayes: xtprobit`	Bayesian random-effects probit model
`bayesmh`	Bayesian models using Metropolis–Hastings algorithm

Estimators for categorical choices

Multinomial logistic regression

`xtmlogit`	Fixed-effects and random-effects multinomial logit models

Multiple outcome variables and latent variables

`gsem`	Generalized structural equation model estimation command

Bayesian estimation

`bayes: xtmlogit`	Bayesian random-effects multinomial logit model
`bayesmh`	Bayesian models using Metropolis–Hastings algorithm

In addition to the commands listed here, commands listed in the previous section that fit models for cross-sectional data can be used with panel data provided that they allow the `vce(cluster)` option. The point estimates from these commands have a population-averaged interpretation and are consistent but less efficient than the estimates from an appropriate panel-data estimator. The default standard errors reported by commands for cross-sectional data are inappropriate for panel or otherwise clustered data because they assume that observations are independent. However, by including the `vce(cluster)` option, you will get standard errors that relax this assumption and provide valid inference for this type of data.

Multilevel models for clustered data

You can also use multilevel modeling commands for choice models when your data are clustered or grouped and observations within the clusters are not independent. Perhaps your observations are students, and those students come from multiple classrooms. You might even have students within classrooms and classrooms within schools. The `me` commands fit multilevel models that account for the correlation within clusters. For categorical outcomes, the `gsem` command can fit a multilevel multinomial logit model in addition to accommodating multiple outcome variables and latent variables for both binary and categorical outcomes. Bayesian estimation is also available.

The following commands can be useful for multilevel choice models for clustered data:

Estimators for binary choices

`mecloglog`	Multilevel mixed-effects complementary log–log regression
`melogit`	Multilevel mixed-effects logistic regression
`meprobit`	Multilevel mixed-effects probit regression

Multiple outcome variables and latent variables

`gsem`	Generalized structural equation model estimation command

Bayesian estimation

`bayes: mecloglog`	Bayesian multilevel complementary log–log regression
`bayes: melogit`	Bayesian multilevel logistic regression
`bayes: meprobit`	Bayesian multilevel probit regression
`bayesmh`	Bayesian models using Metropolis–Hastings algorithm

Estimators for categorical choices

Multilevel multinomial logistic regression

`gsem`	Generalized structural equation model estimation command

Multiple outcome variables and latent variables

`gsem`	Generalized structural equation model estimation command

Bayesian estimation

`bayesmh`	Bayesian models using Metropolis–Hastings algorithm

Reference

Islam, F., J. F. Thrasher, F. Xiao, R. R. Moran, and J. W. Hardin. 2023. Data management and techniques for best–worst discrete choice experiments. *Stata Journal* 23: 1020–1044.