Postestimation commands   predict   margins   Remarks and examples
Methods and formulas   Reference   Also see

Postestimation commands

The following standard postestimation commands are available after `clogit`:

Command	Description
contrast	contrasts and ANOVA-style joint tests of parameters
estat ic	Akaike's, consistent Akaike's, corrected Akaike's, and Schwarz's Bayesian information criteria (AIC, CAIC, AICc, and BIC, respectively)
estat summarize	summary statistics for the estimation sample
estat vce	variance–covariance matrix of the estimators (VCE)
estat (svy)	postestimation statistics for survey data
estimates	cataloging estimation results
etable	table of estimation results
* hausman	Hausman's specification test
lincom	point estimates, standard errors, testing, and inference for linear combinations of parameters
linktest	link test for model specification
* lrtest	likelihood-ratio test
margins	marginal means, predictive margins, marginal effects, and average marginal effects
marginsplot	graph the results from margins (profile plots, interaction plots, etc.)
nlcom	point estimates, standard errors, testing, and inference for nonlinear combinations of parameters
predict	probabilities, influence statistics, residuals, etc.
predictnl	point estimates, standard errors, testing, and inference for generalized predictions
pwcompare	pairwise comparisons of parameters
suest	seemingly unrelated estimation
test	Wald tests of simple and composite linear hypotheses
testnl	Wald tests of nonlinear hypotheses

*`hausman` and `lrtest` are not appropriate with `svy` estimation results.

predict

Description for predict

predict creates a new variable containing predictions such as probabilities, linear predictions, standard errors, influence statistics, lack-of-fit statistics, Hosmer and Lemeshow leverages, Pearson residuals, and the equation-level score.

Menu for predict

Statistics > Postestimation

Syntax for predict

> predict [*type*] *newvar* [*if*] [*in*] [, *statistic* <u>nooff</u>set]

statistic	Description
Main	
pc1	probability of a positive outcome; the default
pu0	probability of a positive outcome, assuming fixed effect is zero
xb	linear prediction
stdp	standard error of the linear prediction
* <u>dbeta</u>	Delta-β influence statistic
* <u>dx2</u>	Delta-χ^2 lack-of-fit statistic
* <u>gdbeta</u>	Delta-β influence statistic for each group
* <u>gdx2</u>	Delta-χ^2 lack-of-fit statistic for each group
* <u>hat</u>	Hosmer and Lemeshow leverage
* <u>residuals</u>	Pearson residuals
* <u>rstandard</u>	standardized Pearson residuals
<u>score</u>	first derivative of the log likelihood with respect to $\mathbf{x}_j\boldsymbol{\beta}$

Unstarred statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when if e(sample) is not specified.

Starred statistics are available for multiple controls per case–matching design only. They are not available if vce(robust), vce(cluster *clustvar*), or pweights were specified with clogit.

dbeta, dx2, gdbeta, gdx2, hat, and rstandard are not available if constraints() was specified with clogit.

Options for predict

> [Main]

pc1, the default, calculates the probability of a positive outcome conditional on one positive outcome within group.

pu0 calculates the probability of a positive outcome, assuming that the fixed effect is zero.

xb calculates the linear prediction.

stdp calculates the standard error of the linear prediction.

dbeta calculates the Delta-β influence statistic, a standardized measure of the difference in the coefficient vector that is due to deletion of the observation.

dx2 calculates the Delta-χ^2 influence statistic, reflecting the decrease in the Pearson χ^2 that is due to deletion of the observation.

gdbeta calculates the approximation to the Pregibon stratum-specific Delta-β influence statistic, a standardized measure of the difference in the coefficient vector that is due to deletion of the entire stratum.

gdx2 calculates the approximation to the Pregibon stratum-specific Delta-χ^2 influence statistic, reflecting the decrease in the Pearson χ^2 that is due to deletion of the entire stratum.

hat calculates the Hosmer and Lemeshow leverage or the diagonal element of the hat matrix.

residuals calculates the Pearson residuals.

rstandard calculates the standardized Pearson residuals.

score calculates the equation-level score, $\partial \ln L / \partial(\mathbf{x}_{it}\boldsymbol{\beta})$.

nooffset is relevant only if you specified offset(*varname*) for clogit. It modifies the calculations made by predict so that they ignore the offset variable; the linear prediction is treated as $\mathbf{x}_j\mathbf{b}$ rather than as $\mathbf{x}_j\mathbf{b} + \text{offset}_j$. This option cannot be specified with dbeta, dx2, gdbeta, gdx2, hat, and rstandard.

margins

Description for margins

margins estimates margins of response for probabilities and linear predictions.

Menu for margins

Statistics > Postestimation

Syntax for margins

margins [*marginlist*] [, *options*]

margins [*marginlist*] , <u>pred</u>ict(*statistic* ...) [<u>pred</u>ict(*statistic* ...) ...] [*options*]

statistic	Description
pu0	probability of a positive outcome, assuming fixed effect is zero; the default
xb	linear prediction
pc1	not allowed with margins
stdp	not allowed with margins
<u>d</u>beta	not allowed with margins
<u>d</u>x2	not allowed with margins
<u>g</u>dbeta	not allowed with margins
<u>g</u>dx2	not allowed with margins
hat	not allowed with margins
<u>res</u>iduals	not allowed with margins
<u>rs</u>tandard	not allowed with margins
<u>sc</u>ore	not allowed with margins

Statistics not allowed with margins are functions of stochastic quantities other than e(b).

For the full syntax, see [R] **margins**.

Remarks and examples

predict may be used after clogit to obtain predicted values of the index $\mathbf{x}_{it}\boldsymbol{\beta}$. Predicted probabilities for conditional logistic regression must be interpreted carefully. Probabilities are estimated for each group as a whole, not for individual observations. Furthermore, the probabilities are conditional on the number of positive outcomes in the group (that is, the number of cases and the number of controls), or it is assumed that the fixed effect is zero. predict may also be used to obtain influence and lack-of-fit statistics for an individual observation and for the whole group, to compute Pearson, standardized Pearson residuals, and leverage values.

predict may be used for both within-sample and out-of-sample predictions.

▷ Example 1

Suppose that we have $1 : k_{2i}$ matched data and that we have previously fit the following model:

```
. use https://www.stata-press.com/data/r19/clogitid
. clogit y x1 x2, group(id)
(output omitted)
```

To obtain the predicted values of the index, we could type `predict idx, xb` to create a new variable called `idx`. From `idx`, we could then calculate the predicted probabilities. Easier, however, would be to type

```
. predict phat
(option pc1 assumed; probability of success given one success within group)
```

`phat` would then contain the predicted probabilities.

As noted previously, the predicted probabilities are really predicted probabilities for the group as a whole (that is, they are the predicted probability of observing $y_{it} = 1$ *and* $y_{it'} = 0$ for all $t' \neq t$). Thus, if we want to obtain the predicted probabilities for the estimation sample, it is important that, when we make the calculation, predictions be restricted to the same sample on which we estimated the data. We cannot predict the probabilities and then just keep the relevant ones because the entire sample determines each probability. Thus, assuming that we are not attempting to make out-of-sample predictions, we type

```
. predict phat2 if e(sample)
(option pc1 assumed; probability of success given one success within group)
```

◁

Methods and formulas

Recall that $i = 1, \ldots, n$ denote the groups and $t = 1, \ldots, T_i$ denote the observations for the ith group.

`predict` produces probabilities of a positive outcome within group conditional on there being one positive outcome (`pc1`),

$$\Pr\left(y_{it} = 1 \ \middle| \ \sum_{t=1}^{T_i} y_{it} = 1 \right) = \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta})}{\sum_{t=1}^{T_i} \exp(\mathbf{x}_{it}\boldsymbol{\beta})}$$

or `predict` calculates the unconditional `pu0`:

$$\Pr(y_{it} = 1) = \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{it}\boldsymbol{\beta})}$$

Let $N = \sum_{j=1}^{n} T_j$ denote the total number of observations, p denote the number of covariates, and $\hat{\theta}_{it}$ denote the conditional predicted probabilities of a positive outcome (`pc1`).

For the multiple control per case ($1 : k_{2i}$) matching, Hosmer, Lemeshow, and Sturdivant (2013, 248–251) propose the following diagnostics:

The Pearson residual is

$$r_{it} = \frac{(y_{it} - \hat{\theta}_{it})}{\sqrt{\hat{\theta}_{it}}}$$

The leverage (hat) value is defined as

$$h_{it} = \hat{\theta}_{it} \tilde{\mathbf{x}}_{it}^T (\widetilde{\mathbf{X}}^T \mathbf{U} \widetilde{\mathbf{X}})^{-1} \tilde{\mathbf{x}}_{it}$$

where $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \sum_{j=1}^{T_i} \mathbf{x}_{ij} \hat{\theta}_{ij}$ is the $1 \times p$ row vector of centered by a weighted stratum-specific mean covariate values, $\mathbf{U}_N = \text{diag}\{\hat{\theta}_{it}\}$, and the rows of $\widetilde{\mathbf{X}}_{N \times p}$ are composed of $\tilde{\mathbf{x}}_{it}$ values.

The standardized Pearson residual is

$$r_{sit} = \frac{r_{it}}{\sqrt{1 - h_{it}}}$$

The lack-of-fit and influence diagnostics for an individual observation are (respectively) computed as

$$\Delta \chi_{it}^2 = r_{sit}^2$$

and

$$\Delta \hat{\beta}_{it} = \Delta \chi_{it}^2 \frac{h_{it}}{1 - h_{it}}$$

The lack-of-fit and influence diagnostics for the groups are the group-specific totals of the respective individual diagnostics shown above.

Reference

Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

Also see

[R] **clogit** — Conditional (fixed-effects) logistic regression

[U] **20 Estimation and postestimation commands**