Title

stata.com

exlogistic postestimation — Postestimation tools for exlogistic

Description

The following postestimation commands are of special interest after `exlogistic`:

Command	Description
`estat predict`	single-observation prediction
`estat se`	report ORs or coefficients and their asymptotic standard errors

The following standard postestimation command is also available:

Command	Description
`estat summarize`	summary statistics for the estimation sample

`estat summarize` is not allowed if the `binomial()` option was specified in `exlogistic`.

See [R] **estat summarize** for details.

Special-interest postestimation commands

`estat predict` computes a predicted probability (or linear predictor), its asymptotic standard error, and its exact confidence interval for 1 observation. Predictions are carried out by estimating the constant coefficient after shifting the independent variables and conditioned variables by the values specified in the `at()` option or by their medians. Therefore, predictions must be done with the estimation sample in memory. If a different dataset is used or if the dataset is modified, then an error will result.

`estat se` reports odds ratio or coefficients and their asymptotic standard errors. The estimates are stored in the matrix `r(estimates)`.

Syntax for estat

Single-observation prediction

```
estat predict [ , options ]
```

Report ORs or coefficients and their asymptotic standard errors

```
estat se [ , coef ]
```

options	Description
pr	probability; the default
xb	linear effect
at(*atspec*)	use the specified values for the *indepvars* and condvars()
level(#)	set confidence level for the predicted value; default is level(95)
memory(#[b|k|m|g])	set limit on memory usage; default is memory(10m)
nolog	do not display the enumeration log

These statistics are available only for the estimation sample.

Menu for estat

Statistics > Postestimation > Reports and statistics

Options for estat predict

pr, the default, calculates the probability.

xb calculates the linear effect.

at(*varname* = # [[*varname* = #] [...]]) specifies values to use in computing the predicted value. Here *varname* is one of the independent variables, *indepvars*, or the conditioned variables, condvars(). The default is to use the median of each independent and conditioned variable.

level(#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see [U] **20.7 Specifying the width of confidence intervals**.

memory(#[b|k|m|g]) sets a limit on the amount of memory estat predict can use when generating the conditional distribution of the constant parameter sufficient statistic. The default is memory(10m), where m stands for megabyte, or 1,048,576 bytes. The following are also available: b stands for byte; k stands for kilobyte, which is equal to 1,024 bytes; and g stands for gigabyte, which is equal to 1,024 megabytes. The minimum setting allowed is 1m and the maximum is 512m or 0.5g, but do not attempt to use more memory than is available on your computer. Also see *Remarks and examples* in [R] **exlogistic** for details on enumerating the conditional distribution.

nolog prevents the display of the enumeration log. By default, the enumeration log is displayed showing the progress of enumerating the distribution of the observed successes conditioned on the independent variables shifted by the values specified in at() (or by their medians). See *Methods and formulas* in [R] **exlogistic** for details of the computations.

Option for estat se

`coef` requests that the estimated coefficients and their asymptotic standard errors be reported. The default is to report the odds ratios and their asymptotic standard errors.

Remarks and examples

stata.com

Predictions must be done using the estimation sample. This is because the prediction is really an estimated constant coefficient (the intercept) after shifting the independent variables and conditioned variables by the values specified in `at()` or by their medians. The justification for this approach can be seen by rewriting the model as

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = (\alpha + \mathbf{x}_0\boldsymbol{\beta}) + (\mathbf{x}_i - \mathbf{x}_0)\boldsymbol{\beta}$$

where $\mathbf{x}_0$ are the specified values for the *indepvars* (Mehta and Patel 1995). Because the estimation of the constant term is required, this technique is not appropriate for stratified models that used the `group()` option.

▷ Example 1

To demonstrate, we return to the example 2 in [R] **exlogistic** using data from a prospective study of perinatal infection and HIV-1. Here there was an investigation into whether the blood serum levels of CD4 and CD8 measured in infants at 6 months of age might predict their development of HIV infection. The blood serum levels are coded as ordinal values 0, 1, and 2. These data are used by Mehta and Patel (1995) as an exposition of exact logistic.

```
. use http://www.stata-press.com/data/r13/hiv_n
(prospective study of perinatal infection of HIV-1; binomial form)
. generate byte cd4_0 = (cd4==0)
. generate byte cd4_1 = (cd4==1)
. generate byte cd8_0 = (cd8==0)
. generate byte cd8_1 = (cd8==1)
. exlogistic hiv cd4_0 cd4_1 cd8_0 cd8_1, terms(cd4=cd4_0 cd4_1,
> cd8=cd8_0 cd8_1) binomial(n) test(probability) saving(dist)
  (output omitted)
```

```
. estat predict
Enumerating sample-space combinations:
observation 1:   enumerations =          3
observation 2:   enumerations =         12
observation 3:   enumerations =          5
observation 4:   enumerations =          5
observation 5:   enumerations =          5
observation 6:   enumerations =         35
observation 7:   enumerations =         15
observation 8:   enumerations =         15
observation 9:   enumerations =          9
observation 10:  enumerations =          9
observation 11:  enumerations =          5
observation 12:  enumerations =         18
note: CMLE estimate for _cons is -inf; computing MUE
Predicted value at cd4_0 = 0, cd4_1 = 0, cd8_0 = 0, cd8_1 = 1
```

hiv	Predicted	Std. Err.	[95% Conf.	Interval]
Probability	0.0390*	N/A	0.0000	0.1962

```
(*) identifies median unbiased estimates (MUE); because an MUE
    is computed, there is no SE estimate
```

Because we did not specify values by using the `at()` option, the median values of the *indepvars* are used for the prediction. By default, medians are used instead of means because we want to use values that are observed in the dataset. If the means of the binary variables `cd4_0`–`cd8_1` were used, we would have created floating point variables in $(0, 1)$ that not only do not properly represent the indicator variables but also would be a source of computational inefficiency in generating the conditional distribution. Because the MUE is computed for the predicted value, there is no standard-error estimate.

From the example discussions in [R] **exlogistic**, the infants at highest risk are those with a CD4 level of 0 and a CD8 level of 2. Below we use the `at()` option to make a prediction at these blood serum levels.

```
. estat predict, at(cd4_0=1 cd4_1=0 cd8_0=0 cd8_1=0) nolog
note: CMLE estimate for _cons is +inf; computing MUE
Predicted value at cd4_0 = 1, cd4_1 = 0, cd8_0 = 0, cd8_1 = 0
```

hiv	Predicted	Std. Err.	[95% Conf.	Interval]
Probability	0.9063*	N/A	0.4637	1.0000

```
(*) identifies median unbiased estimates (MUE); because an MUE
    is computed, there is no SE estimate
```

◁

Stored results

`estat predict` stores the following in `r()`:

Scalars

`r(imue)`	`1` if `r(pred)` is an MUE and `0` if a CMLE
`r(pred)`	estimated probability or the linear effect
`r(se)`	asymptotic standard error of `r(pred)`

Macros

`r(estimate)`	prediction type: `pr` or `xb`
`r(level)`	confidence level

Matrices

`r(ci)`	confidence interval
`r(x)`	*indepvars* and `condvars()` values

Reference

Mehta, C. R., and N. R. Patel. 1995. Exact logistic regression: Theory and examples. *Statistics in Medicine* 14: 2143–2160.

Also see

[R] **exlogistic** — Exact logistic regression

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