Title

nbreg postestimation — Postestimation tools for nbreg and gnbreg

Description

The following postestimation commands are available for `nbreg` and `gnbreg`:

command	description
`estat`	AIC, BIC, VCE, and estimation sample summary
`estat` (svy)	postestimation statistics for survey data
`estimates`	cataloging estimation results
`lincom`	point estimates, standard errors, testing, and inference for linear combinations of coefficients
`linktest`	link test for model specification
`lrtest`[1]	likelihood-ratio test
`margins`	marginal means, predictive margins, marginal effects, and average marginal effects
`nlcom`	point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients
`predict`	predictions, residuals, influence statistics, and other diagnostic measures
`predictnl`	point estimates, standard errors, testing, and inference for generalized predictions
`suest`	seemingly unrelated estimation
`test`	Wald tests of simple and composite linear hypotheses
`testnl`	Wald tests of nonlinear hypotheses

[1] `lrtest` is not appropriate with `svy` estimation results.

See the corresponding entries in the *Base Reference Manual* for details, but see [SVY] **estat** for details about `estat` (svy).

Syntax for predict

`predict` [*type*] *newvar* [*if*] [*in*] [, *statistic* `nooffset`]

`predict` [*type*] { *stub*`*` | $newvar_{\text{reg}}$ $newvar_{\text{disp}}$ } [*if*] [*in*] , `scores`

statistic	description
Main	
`n`	number of events; the default
`ir`	incidence rate (equivalent to `predict ..., n nooffset`)
`pr(`*n*`)`	unconditional probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	unconditional probability $\Pr(a \le y_j \le b)$
`xb`	linear prediction
`stdp`	standard error of the linear prediction

In addition, relevant only after `gnbreg` are the following:

statistic	description
Main	
`alpha`	predicted values of α_j
`lnalpha`	predicted values of $\ln\alpha_j$
`stdplna`	standard error of predicted $\ln\alpha_j$

These statistics are available both in and out of sample; type `predict ... if e(sample) ...` if wanted only for the estimation sample.

Menu

Statistics > Postestimation > Predictions, residuals, etc.

Options for predict

Main

`n`, the default, calculates the predicted number of events, which is $\exp(\mathbf{x}_j\boldsymbol{\beta})$ if neither `offset(`*varname*$_o$`)` nor `exposure(`*varname*$_e$`)` was specified when the model was fit; $\exp(\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j)$ if `offset()` was specified; or $\exp(\mathbf{x}_j\boldsymbol{\beta}) \times \text{exposure}_j$ if `exposure()` was specified.

`ir` calculates the incidence rate $\exp(\mathbf{x}_j\boldsymbol{\beta})$, which is the predicted number of events when exposure is 1. This is equivalent to specifying both the `n` and the `nooffset` options.

`pr(`*n*`)` calculates the unconditional probability $\Pr(y_j = n)$, where *n* is a nonnegative integer that may be specified as a number or a variable.

`pr(`*a*`,`*b*`)` calculates the unconditional probability $\Pr(a \leq y_j \leq b)$, where *a* and *b* are nonnegative integers that may be specified as numbers or variables;

b missing ($b \geq .$) means $+\infty$;
`pr(20,.)` calculates $\Pr(y_j \geq 20)$;
`pr(20,`*b*`)` calculates $\Pr(y_j \geq 20)$ in observations for which $b \geq .$ and calculates
$\Pr(20 \leq y_j \leq b)$ elsewhere.

`pr(.,`*b*`)` produces a syntax error. A missing value in an observation of the variable *a* causes a missing value in that observation for `pr(`*a*`,`*b*`)`.

`xb` calculates the linear prediction, which is $\mathbf{x}_j\boldsymbol{\beta}$ if neither `offset()` nor `exposure()` was specified; $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j$ if `offset()` was specified; or $\mathbf{x}_j\boldsymbol{\beta} + \ln(\text{exposure}_j)$ if `exposure()` was specified; see `nooffset` below.

`stdp` calculates the standard error of the linear prediction.

`alpha`, `lnalpha`, and `stdplna` are relevant after `gnbreg` estimation only; they produce the predicted values of α_j, $\ln\alpha_j$, and the standard error of the predicted $\ln\alpha_j$, respectively.

`nooffset` is relevant only if you specified `offset()` or `exposure()` when you fit the model. It modifies the calculations made by `predict` so that they ignore the offset or exposure variable; the linear prediction is treated as $\mathbf{x}_j\boldsymbol{\beta}$ rather than as $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j$ or $\mathbf{x}_j\boldsymbol{\beta} + \ln(\text{exposure}_j)$. Specifying `predict ..., nooffset` is equivalent to specifying `predict ..., ir`.

scores calculates equation-level score variables.

The first new variable will contain $\partial \ln L / \partial(\mathbf{x}_j \boldsymbol{\beta})$.

The second new variable will contain $\partial \ln L / \partial(\ln \alpha_j)$ for dispersion(mean) and gnbreg.

The second new variable will contain $\partial \ln L / \partial(\ln \delta)$ for dispersion(constant).

Remarks

After nbreg and gnbreg, predict returns the expected number of deaths per cohort and the probability of observing the number of deaths recorded or fewer.

```
. use http://www.stata-press.com/data/r11/rod93

. nbreg deaths i.cohort, nolog

Negative binomial regression                      Number of obs   =         21
                                                  LR chi2(2)      =       0.14
Dispersion     = mean                             Prob > chi2     =     0.9307
Log likelihood = -108.48841                       Pseudo R2       =     0.0007

------------------------------------------------------------------------------
      deaths |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      cohort |
          2  |   .0591305   .2978419     0.20   0.843    -.5246289      .64289
          3  |  -.0538792   .2981621    -0.18   0.857    -.6382662    .5305077
             |
       _cons |   4.435906   .2107213    21.05   0.000       4.0229    4.848912
-------------+----------------------------------------------------------------
    /lnalpha |  -1.207379   .3108622                     -1.816657   -.5980999
-------------+----------------------------------------------------------------
       alpha |     .29898   .0929416                      .1625683    .5498555
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =  434.62 Prob>=chibar2 = 0.000

. predict count
(option n assumed; predicted number of events)

. predict p, pr(0, deaths)

. summarize deaths count p

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      deaths |        21    84.66667    48.84192         10        197
       count |        21    84.66667     4.00773         80   89.57143
           p |        21    .4991542    .2743702   .0070255   .9801285
```

The expected number of deaths ranges from 80 to 90. The probability $\Pr(y_i \leq \texttt{deaths})$ ranges from 0.007 to 0.98.

Methods and formulas

All postestimation commands listed above are implemented as ado-files.

In the following, we use the same notation as in [R] **nbreg**.

Methods and formulas are presented under the following headings:

Mean-dispersion model
Constant-dispersion model

Mean-dispersion model

The equation-level scores are given by

$$\begin{aligned} \text{score}(\mathbf{x}\boldsymbol{\beta})_j &= p_j(y_j - \mu_j) \\ \text{score}(\tau)_j &= -m\left\{\frac{\alpha_j(\mu_j - y_j)}{1+\alpha_j\mu_j} - \ln(1+\alpha_j\mu_j) + \psi(y_j+m) - \psi(m)\right\} \end{aligned}$$

where $\tau_j = \ln\alpha_j$ and $\psi(z)$ is the digamma function.

Constant-dispersion model

The equation-level scores are given by

$$\begin{aligned} \text{score}(\mathbf{x}\boldsymbol{\beta})_j &= m_j\left\{\psi(y_j+m_j) - \psi(m_j) + \ln(p)\right\} \\ \text{score}(\tau)_j &= y_j - (y_j+m_j)(1-p) - \text{score}(\mathbf{x}\boldsymbol{\beta})_j \end{aligned}$$

where $\tau_j = \ln\delta_j$.

Also see

[R] **nbreg** — Negative binomial regression

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