Postestimation commands

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

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>contrast</td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td>estat ic</td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td>estat summarize</td>
<td>summary statistics for the estimation sample</td>
</tr>
<tr>
<td>estat vce</td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estat (svy)</td>
<td>postestimation statistics for survey data</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td>*forecast</td>
<td>dynamic forecasts and simulations</td>
</tr>
<tr>
<td>*hausman</td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td>lincom</td>
<td>point estimates, standard errors, testing, and inference for linear combinations of coefficients</td>
</tr>
<tr>
<td>linktest</td>
<td>link test for model specification</td>
</tr>
<tr>
<td>*lrtest</td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td>margins</td>
<td>marginal means, predictive margins, marginal effects, and average marginal effects</td>
</tr>
<tr>
<td>marginsplot</td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td>nlcom</td>
<td>point estimates, standard errors, testing, and inference for nonlinear combinations of coefficients</td>
</tr>
<tr>
<td>predict</td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td>predictnl</td>
<td>point estimates, standard errors, testing, and inference for generalized predictions</td>
</tr>
<tr>
<td>pwcompare</td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td>suest</td>
<td>seemingly unrelated estimation</td>
</tr>
<tr>
<td>test</td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td>testnl</td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

* forecast, hausman, and lrtest are not appropriate with svy estimation results. forecast is also not appropriate with mi estimation results.
predict

Description for predict

predict creates a new variable containing predictions such as numbers of events, incidence rates, probabilities, linear predictions, standard errors, and predicted values.

Menu for predict

Statistics > Postestimation

Syntax for predict

```
predict [type] newvar [if] [in] [, statistic nooffset]
predict [type] { stub* | newvar|reg newvar|disp } [if] [in], scores
```

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>number of events; the default</td>
</tr>
<tr>
<td>n</td>
<td>incidence rate (equivalent to predict ... , n nooffset)</td>
</tr>
<tr>
<td>pr(n)</td>
<td>probability Pr(y_j = n)</td>
</tr>
<tr>
<td>pr(a,b)</td>
<td>probability Pr(a ≤ y_j ≤ b)</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the linear prediction</td>
</tr>
</tbody>
</table>

In addition, relevant only after gnbreg are the following:

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>predicted values of α_j</td>
</tr>
<tr>
<td>lnalp</td>
<td>predicted values of lnα_j</td>
</tr>
<tr>
<td>stdplna</td>
<td>standard error of predicted lnα_j</td>
</tr>
</tbody>
</table>

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

Options for predict

n, the default, calculates the predicted number of events, which is exp(x_jβ) if neither `offset(varname_o)` nor `exposure(varname_e)` was specified when the model was fit; `exp(x_jβ + offset_j)` if `offset()` was specified; or `exp(x_jβ) × exposure_j` if `exposure()` was specified.

ir calculates the incidence rate exp(x_jβ), which is the predicted number of events when exposure is 1. This is equivalent to specifying both the n and the nooffset options.
\texttt{pr}(n) \text{ calculates the probability } \Pr(y_j = n), \text{ where } n \text{ is a nonnegative integer that may be specified as a number or a variable.}

\texttt{pr}(a,b) \text{ calculates the probability } \Pr(a \leq y_j \leq b), \text{ where } a \text{ and } b \text{ are nonnegative integers that may be specified as numbers or variables;}

\texttt{pr}(20, .) \text{ calculates } \Pr(y_j \geq 20); 
\texttt{pr}(20, b) \text{ calculates } \Pr(y_j \geq 20) \text{ in observations for which } b \geq . \text{ and calculates } 
\Pr(20 \leq y_j \leq b) \text{ elsewhere.}

\texttt{pr}(., b) \text{ produces a syntax error. A missing value in an observation of the variable } a \text{ causes a}
\text{ missing value in that observation for } \texttt{pr}(a,b).

\texttt{xb} \text{ calculates the linear prediction, which is } x_j \delta \text{ if neither } \texttt{offset()} \text{ nor } \texttt{exposure()} \text{ was specified;}
\texttt{x}_j \delta + \texttt{offset}_j \text{ if } \texttt{offset()} \text{ was specified; or } x_j \delta + \ln(\texttt{exposure}_j) \text{ if } \texttt{exposure()} \text{ was specified;}
\text{ see } \texttt{nooffset} \text{ below.}

\texttt{stdp} \text{ calculates the standard error of the linear prediction.}

\texttt{alpha}, \texttt{lnalpha}, \text{ and } \texttt{stdplna} \text{ are relevant after } \texttt{gnbreg} \text{ estimation only; they produce the predicted}
\text{ values of } \alpha_j, \ln\alpha_j, \text{ and the standard error of the predicted } \ln\alpha_j, \text{ respectively.}

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

\texttt{scores} \text{ calculates equation-level score variables.}

\text{The first new variable will contain } \partial \ln L / \partial (x_j \delta).
\text{The second new variable will contain } \partial \ln L / \partial (\ln \alpha_j) \text{ for } \texttt{dispersion(mean)} \text{ and } \texttt{gnbreg}.
\text{The second new variable will contain } \partial \ln L / \partial (\ln \delta) \text{ for } \texttt{dispersion(constant)}. 
margins

Description for margins

margins estimates margins of response for numbers of events, incidence rates, probabilities, linear predictions, and predicted values.

Menu for margins

Statistics > Postestimation

Syntax for margins

margins [marginlist] [ , options ]
margins [marginlist] , predict(statistic ...) [predict(statistic ...) ...] [options ]

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>number of events; the default</td>
</tr>
<tr>
<td>ir</td>
<td>incidence rate (equivalent to predict ..., n nooffset)</td>
</tr>
<tr>
<td>pr(n)</td>
<td>probability Pr((y_j = n))</td>
</tr>
<tr>
<td>pr((a, b))</td>
<td>probability Pr((a ≤ y_j ≤ b))</td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction</td>
</tr>
<tr>
<td>stdp</td>
<td>not allowed with margins</td>
</tr>
</tbody>
</table>

In addition, relevant only after gnbreg are the following:

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_j)</td>
<td>predicted values of (\alpha_j)</td>
</tr>
<tr>
<td>(\ln\alpha_j)</td>
<td>predicted values of (\ln\alpha_j)</td>
</tr>
<tr>
<td>stdplna</td>
<td>not allowed with margins</td>
</tr>
</tbody>
</table>

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

For the full syntax, see [R] margins.
Remarks and examples

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.

```stata
. use https://www.stata-press.com/data/r16/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

                     Coef.  Std. Err.     z  P>|z|     [95% Conf. Interval]
----------------------- -------- -------- -------- -------- ------------------------
cohort                  
1960-1967              .0591305  .2978419  0.20  0.843    -.5246289    .6428976
1968-1976              -.0538792  .2981621  -.18  0.857    -.6382662    .5305077
_cons                   4.435906  .2107213  21.05  0.000     4.0229      4.848912
/lnalpha               -1.207379  .3108622 -1.816657 -0.5980999
alpha                  .29898   .0929416   .1625683  .5498555
```

```
LR test of alpha=0:  chibar2(01) = 434.62  Prob >= chibar2 = 0.000
```

```stata
. predict count, n
. predict p, pr(0, deaths)
. summarize deaths count p
```

```
Variable |    Obs  Mean    Std. Dev. | Min | Max
---------|--------|----------------|-----|-----
deads    |   21   |  84.66667  48.84192 |  10 | 197
    count |   21   |  84.66667  4.00773 |  80 | 89.57143
       p   |   21   |  0.4991542 0.2743702 | 0.0070255 | 0.9801285
```

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

Methods and formulas

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

$$
score(x\beta)_j = p_j(y_j - \mu_j)
$$

$$
score(\tau)_j = -m \left\{ \frac{\alpha_j(\mu_j - y_j)}{1 + \alpha_j \mu_j} - \ln(1 + \alpha_j \mu_j) + \psi(y_j + m) - \psi(m) \right\}
$$

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

The equation-level scores are given by

\[
\begin{align*}
\text{score}(x\beta)_j &= m_j \{ \psi(y_j + m_j) - \psi(m_j) + \ln(p) \} \\
\text{score}(\tau)_j &= y_j - (y_j + m_j)(1 - p) - \text{score}(x\beta)_j
\end{align*}
\]

where \( \tau_j = \ln\delta_j \).

Reference


Also see

[R] nbreg — Negative binomial regression

[U] 20 Estimation and postestimation commands