# zip postestimation — Postestimation tools for zip

## Postestimation commands

The following postestimation commands are available after `zip`:

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

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

Description for predict

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

Menu for predict

Statistics > Postestimation

Syntax for predict

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

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main</strong></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>number of events; the default</td>
</tr>
<tr>
<td>ir</td>
<td>incidence rate</td>
</tr>
<tr>
<td>pr</td>
<td>probability of a degenerate zero</td>
</tr>
<tr>
<td>pr(n)</td>
<td>probability Pr(yj = n)</td>
</tr>
<tr>
<td>pr(a,b)</td>
<td>probability Pr(a ≤ yj ≤ 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>

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 \((1 - F_j) \exp(x_j \beta)\) if neither `offset()` nor `exposure()` was specified when the model was fit, where \(F_j\) is the predicted probability of a zero outcome; \((1 - F_j) \exp(x_j \beta + \text{offset}_j)\) if `offset()` was specified; or \((1 - F_j) \{ \exp(x_j \beta) \times \text{exposure}_j \}\) if `exposure()` was specified.

ir calculates the incidence rate, which is the predicted number of events when exposure is 1. This is equivalent to specifying both the n and the `nooffset` options.

pr calculates the probability of a degenerate zero, predicted from the fitted degenerate distribution \(F_j = F(z_j \gamma)\). If `offset()` was specified within the `inflate()` option, then \(F_j = F(z_j \gamma + \text{offset}_j)\) is calculated.

pr(n) calculates the probability \(Pr(yj = n)\), where n is a nonnegative integer that may be specified as a number or a variable. Note that pr is not equivalent to pr(0).

pr(a,b) calculates the probability \(Pr(a \leq yj \leq b)\), where a and b are nonnegative integers that may be specified as numbers or variables;
b missing (b ≥ .) means +∞;
pr(20, .) calculates Pr(y_j ≥ 20);
pr(20, b) calculates Pr(y_j ≥ 20) in observations for which b ≥ . and calculates
Pr(20 ≤ y_j ≤ 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 x_jβ if neither offset() nor exposure() was specified;
x_jβ + offset_jβ if offset() was specified; or x_jβ + ln(exposure_j) if exposure() was specified; see nooffset below.

stdp calculates the standard error of the linear prediction.

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 x_jβ rather than as x_jβ + offset_jβ or x_jβ + ln(exposure_j). Specifying
predict ..., nooffset is equivalent to specifying predict ..., ir.

scores calculates equation-level score variables.

The first new variable will contain ∂ln L/∂(x_jβ).

The second new variable will contain ∂ln L/∂(z_jγ).
margins

Description for margins

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

Menu for margins

Statistics > Postestimation

Syntax for margins

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

statistic Description
n number of events; the default
ir incidence rate
pr probability of a degenerate zero
pr(n) probability \( \Pr(y_j = n) \)
pr(a,b) probability \( \Pr(a \leq y_j \leq b) \)
xb linear prediction
stdp 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

Example 1: Obtaining predicted counts

Continuing with example 1 from [R] zip, we will use predict to compute the predicted number of fish captured by each individual.

. use https://www.stata-press.com/data/r16/fish
. zip count persons livebait, inflate(child camper)
   (output omitted)
. predict numfish
   (option n assumed; predicted number of events)
. summarize numfish
   
<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>numfish</td>
<td>250</td>
<td>2.770999</td>
<td>3.269588</td>
<td>.079269</td>
<td>13.55015</td>
</tr>
</tbody>
</table>
The average predicted number of fish caught by all visitors, regardless of whether or not they fished, is 2.77 fish.

Example 2: Obtaining predicted probabilities

predict with the pr option computes the probability that an individual does not fish.

```
predict pr, pr
```

On the other hand, predict with the pr(n) option computes the probability of catching n fish; particularly, the probability of catching zero fish will be

```
predict pr0, pr(0)
list pr pr0 in 1
```

<table>
<thead>
<tr>
<th>pr</th>
<th>pr0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3793549</td>
<td>0.8609267</td>
</tr>
</tbody>
</table>

Notice that pr0 is always equal to or greater than pr. For example, for the first individual, the probability of not fishing is 0.38; on the other hand, the probability of catching zero fish (0.86) is equal to the sum of the probability of not fishing and the probability of fishing but not catching any fish. pr0 can be also computed as one minus the probability of catching at least one fish, that is:

```
predict pr_catch, pr(1,.)
gen pr0b = 1-pr_catch
```

Methods and formulas

See Methods and formulas in [R] zip for the model definition and notation.

The probabilities calculated using the pr(n) option are the probability \( \Pr(y_j = n) \). These are calculated using

\[
\Pr(y_j = 0|x_j, z_j) = F_j + (1 - F_j) \exp(-\lambda_j)
\]

\[
\Pr(y_j = n|x_j, z_j) = (1 - F_j) \frac{\lambda_j^n \exp(-\lambda_j)}{n!}
\]

for \( n = 1, 2, \ldots \)

where \( F_j \) is the probability of obtaining an observation from the degenerate distribution whose mass is concentrated at zero. \( F_j \) can be obtained by using the pr option.

See Cameron and Trivedi (2013, sec. 4.6) for further details.

References


Also see

[R] zip — Zero-inflated Poisson regression
[U] 20 Estimation and postestimation commands