Title

zip postestimation — Postestimation tools for zip

Description

The following postestimation commands are available for `zip`:

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
`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*$_{\rm reg}$ *newvar*$_{\rm inflate}$ } [*if*] [*in*] `, scores`

statistic	description
Main	
`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`	standard error of the linear prediction

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 $(1-p_j)\exp(\mathbf{x}_j\boldsymbol{\beta})$ if neither `offset()` nor `exposure()` was specified when the model was fit, where p_j is the predicted probability of a zero outcome; $(1-p_j)\exp\{(\mathbf{x}_j\boldsymbol{\beta})+\text{offset}_j\}$ if `offset()` was specified; or $(1-p_j)\{\exp(\mathbf{x}_j\boldsymbol{\beta})\times\text{exposure}_j\}$ 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` calculates the probability $\Pr(y_j=0)$, where this zero was obtained from the degenerate distribution $F(\mathbf{z}_j\boldsymbol{\gamma})$. If `offset()` was specified within the `inflate()` option, then $F(\mathbf{z}_j\boldsymbol{\gamma}+\text{offset}_j^{\gamma})$ is calculated.

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

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

`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 $\partial\ln L/\partial(\mathbf{z}_j\boldsymbol{\gamma})$.

Methods and formulas

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

The probabilities calculated using the `pr(`*n*`)` option are the probability $\Pr(y_i = n)$. These are calculated using

$$\Pr(0|\mathbf{x}_i) = \omega_i + (1 - \omega_i)\,\exp(-\lambda_i)$$

$$\Pr(n|\mathbf{x}_i) = (1 - \omega_i)\frac{\lambda_i^n \exp(-\lambda_i)}{n!} \qquad \text{for } n = 1, 2, \ldots$$

where ω_i is the probability of obtaining an observation from the degenerate distribution whose mass is concentrated at zero. ω_i can be obtained from the `pr` option.

See Cameron and Trivedi (1998, sec. 4.7) for further details.

Reference

Cameron, A. C., and P. K. Trivedi. 1998. *Regression Analysis of Count Data.* Cambridge: Cambridge University Press.

Also see

[R] **zip** — Zero-inflated Poisson regression

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