tpoisson postestimation — Postestimation tools for tpoisson

Postestimation commands

The following postestimation commands are available after `tpoisson`:

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

* `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, conditional means, probabilities, conditional probabilities, linear predictions, standard errors, and equation-level scores.

Menu for predict

Statistics > Postestimation

Syntax for predict

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

statistic	Description
Main	
`n`	number of events; the default
`ir`	incidence rate
`cm`	conditional mean, $E(y_j \mid ll_j < y_j < ul_j)$
`pr(`*n*`)`	probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	probability $\Pr(a \leq y_j \leq b)$
`cpr(`*n*`)`	conditional probability $\Pr(y_j = n \mid ll_j < y_j < ul_j)$
`cpr(`*a*`,`*b*`)`	conditional probability $\Pr(a \leq y_j \leq b \mid ll_j < y_j < ul_j)$
`xb`	linear prediction
`stdp`	standard error of the linear prediction
`score`	first derivative of the log likelihood with respect to $\mathbf{x}_j\boldsymbol{\beta}$

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

Main

`n`, the default, calculates the predicted number of events, which is $\exp(\mathbf{x}_j\boldsymbol{\beta})$ if neither `offset()` nor `exposure()` 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.

`cm` calculates the conditional mean,

$$E(y_j \mid ll_j < y_j < ul_j) = \frac{E(y_j, ll_j < y_j < ul_j)}{\Pr(ll_j < y_j < ul_j)}$$

where ll_j is the left-truncation point specified at estimation and ul_j is the right-truncation point specified at estimation.

`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.

`pr(`*a*`,`*b*`)` calculates the 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*`)`.

`cpr(`*n*`)` calculates the conditional probability $\Pr(y_j = n \,|\, ll_j < y_j < ul_j)$, where n is a nonnegative integer that may be specified as a number or a variable. ll_j and ul_j are as defined in `cm`.

`cpr(`*a*`,`*b*`)` calculates the conditional probability $\Pr(a \leq y_j \leq b \,|\, ll_j < y_j < ul_j)$, where a and b are as defined in `pr(`*a*`,`*b*`)` with the additional restrictions that $a > ll_j$ and $b < ul_j$. ll_j and ul_j are as defined in `cm`.

`xb` calculates the linear prediction, which is $\mathbf{x}_j\boldsymbol{\beta}$ if neither `offset()` nor `exposure()` was specified when the model was fit; $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j$ 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.

`score` calculates the equation-level score, $\partial \ln L / \partial(\mathbf{x}_j\boldsymbol{\beta})$.

`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`.

margins

Description for margins

`margins` estimates margins of response for numbers of events, incidence rates, conditional means, probabilities, conditional 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
`cm`	conditional mean, $E(y_j \mid ll_j < y_j < ul_j)$
`pr(`*n*`)`	probability $\Pr(y_j = n)$
`pr(`*a*`,`*b*`)`	probability $\Pr(a \leq y_j \leq b)$
`cpr(`*n*`)`	conditional probability $\Pr(y_j = n \mid ll_j < y_j < ul_j)$
`cpr(`*a*`,`*b*`)`	conditional probability $\Pr(a \leq y_j \leq b \mid ll_j < y_j < ul_j)$
`xb`	linear prediction
`stdp`	not allowed with `margins`
`score`	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

stata.com

▷ Example 1: Obtaining margins of the conditional mean

In example 1 of [R] **tpoisson**, a truncated Poisson model is fit to the number of pairs of shoes owned on runs per week, miles run per week, gender, age, and marital status. We continue that example to determine the effect of miles run per week on the average number of pairs of shoes owned.

After reading in the data, we use `summarize` to obtain the 25th, 50th, and 75th percentiles for miles run per week.

```
. use http://www.stata-press.com/data/r15/runshoes
. summarize mpweek, detail
                        Miles per week
-------------------------------------------------------------
      Percentiles      Smallest
 1%            5              5
 5%            5              5
10%            5              5       Obs                  60
25%         12.5              5       Sum of Wgt.          60

50%         27.5                      Mean           24.71167
                        Largest       Std. Dev.      14.34934
75%         32.5           47.5
90%         47.5           47.5       Variance       205.9034
95%         47.5           47.5       Skewness       .1948568
99%         57.5           57.5       Kurtosis       2.065304
```

We fit the model from example 1 of [R] **tpoisson** again. We next specify these values for the percentiles to `margins` to estimate the conditional mean of the number of pairs of shoes at different quantiles of miles run per week. To do this, we use the `at()` option of `margins`.

```
. quietly tpoisson shoes rpweek mpweek male age married
. margins, at(mpweek=(12.5 27.5 32.5)) predict(cm)
Predictive margins                              Number of obs     =         60
Model VCE    : OIM

Expression   : Conditional mean of n > ll(0), predict(cm)

1._at        : mpweek          =        12.5

2._at        : mpweek          =        27.5

3._at        : mpweek          =        32.5

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   1.942149   .2111564     9.20   0.000      1.52829    2.356008
          2  |   2.376253   .1714522    13.86   0.000     2.040213    2.712293
          3  |   2.564339   .1948129    13.16   0.000     2.182513    2.946165
------------------------------------------------------------------------------
```

We see that people who run 12.5 miles per week are expected to own 1.94 pairs of shoes. The expected number of pairs of shoes owned increases as the average miles per week increases. We expect people who run 27.5 miles per week have 2.38 pairs of shoes and those who run 32.5 miles per week have 2.56 pairs of shoes.

◁

Methods and formulas

Continuing from *Methods and formulas* in [R] **tpoisson**, the equation-level scores are given by

$$\text{score}(\mathbf{x}\boldsymbol{\beta})_j = y_j - \lambda_j - \frac{\{\exp(-\lambda_j)\lambda_j^{ll_j}/ll_j! - \exp(-\lambda_j)\lambda_j^{(ul_j-1)}/(ul_j-1)!\}\lambda_j}{\Pr(ll_j < Y < ul_j \,|\, \xi_j)}$$

Also see

[R] **tpoisson** — Truncated Poisson regression

[U] 20 Estimation and postestimation commands