_predict — Obtain predictions, residuals, etc., after estimation programming command

Description Syntax Options Methods and formulas Reference Also see

Description

_predict is for use by programmers as a subroutine for implementing the predict command for use after estimation; see [R] **predict**.

Syntax

After regress

```
_predict [type] newvar [if] [in] [ , xb stdp stdf stdr hat cooksd
  residuals rstandard rstudent nolabel ]
```

After single-equation (SE) estimators

```
_predict [type] newvar [if] [in] [ , xb stdp nooffset nolabel ]
```

After multiple-equation (ME) estimators

```
_predict [type] newvar [if] [in] [ , xb stdp stddp nooffset nolabel
  equation(eqspec[ , eqspec]) ]
```

Options

xb calculates the linear prediction from the fitted model. That is, all models can be thought of as estimating a set of parameters $b_1, b_2, \ldots, b_k$, and the linear prediction is $\widehat{y}_j = b_1x_{1j} + b_2x_{2j} + \cdots + b_kx_{kj}$, often written in matrix notation as $\widehat{y}_j = \mathbf{x}_j\mathbf{b}$. For linear regression, the values $\widehat{y}_j$ are called the predicted values or, for out-of-sample predictions, the forecast. For logit and probit, for example, $\widehat{y}_j$ is called the logit or probit index.

It is important to understand that the $x_{1j}, x_{2j}, \ldots, x_{kj}$ used in the calculation are obtained from the data currently in memory and do not have to correspond to the data on the independent variables used in fitting the model (obtaining the $b_1, b_2, \ldots, b_k$).

stdp calculates the standard error of the prediction after any estimation command. Here the prediction is understood to mean the same thing as the "index", namely, $\mathbf{x}_j\mathbf{b}$. The statistic produced by stdp can be thought of as the standard error of the predicted expected value, or mean index, for the observation's covariate pattern. This is also commonly referred to as the standard error of the fitted value.

stdf calculates the standard error of the forecast, which is the standard error of the point prediction for 1 observation. It is commonly referred to as the standard error of the future or forecast value. By construction, the standard errors produced by stdf are always larger than those produced by stdp; see *Methods and formulas* in [R] **predict**.

stdr calculates the standard error of the residuals.

hat (or leverage) calculates the diagonal elements of the projection hat matrix.

`cooksd` calculates the Cook's D influence statistic (Cook 1977).

`residuals` calculates the residuals.

`rstandard` calculates the standardized residuals.

`rstudent` calculates the Studentized (jackknifed) residuals.

`nooffset` may be combined with most statistics and specifies that the calculation be made, ignoring any offset or exposure variable specified when the model was fit.

This option is available, even if not documented, for `predict` after a specific command. If neither the `offset(`*varname*`)` option nor the `exposure(`*varname*`)` option was specified when the model was fit, specifying `nooffset` does nothing.

`nolabel` prevents `_predict` from labeling the newly created variable.

`stddp` is allowed only after you have previously fit a multiple-equation model. The standard error of the difference in linear predictions ($\mathbf{x}_{1j}\mathbf{b} - \mathbf{x}_{2j}\mathbf{b}$) between equations 1 and 2 is calculated. Use the `equation()` option to get the standard error of the difference between other equations.

`equation(`*eqspec*[`,` *eqspec*]`)` is relevant only when you have previously fit a multiple-equation model. It specifies the equation to which you are referring. *eqspec* can be an equation number or an equation name. For example, if *eqspec* is a number, then `equation(#1)` would mean that the calculation is to be made for the first equation (that is, for the first dependent variable), `equation(#2)` would mean the second, and so on. If *eqspec* is a name, then `equation(income)` would refer to the equation named `income`, and `equation(hours)` would refer to the equation named `hours`.

If you do not specify `equation()`, results are the same as if you had specified `equation(#1)`.

Typically, specify one equation with, for instance, the `xb`, `stdp`, or `residuals` statistic.

Specify two equations (for example, `equation(#1, #2)` or `equation(income, hours)`) with other statistics that refer to between-equation concepts, such as `difference` or `stddp`. `equation()` is required with these statistics.

Methods and formulas

See *Methods and formulas* in [R] **predict** and [R] **regress**.

Reference

Cook, R. D. 1977. Detection of influential observation in linear regression. *Technometrics* 19: 15–18. https://doi.org/10.1080/00401706.1977.10489493.

Also see

[R] **predict** — Obtain predictions, residuals, etc., after estimation

[U] 20 Estimation and postestimation commands

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.