Title

estat classification — Classification statistics and table

Description

`estat classification` reports various summary statistics, including the classification table.

`estat classification` requires that the current estimation results be from `logistic`, `logit`, `probit`, or `ivprobit`; see [R] **logistic**, [R] **logit**, [R] **probit**, or [R] **ivprobit**.

Quick start

Display classification table and related statistics for current estimation results

```
estat classification
```

Change probability threshold for assignment to positive outcome to 75%

```
estat classification, cutoff(.75)
```

Classification for observations with `catvar` = 2

```
estat classification if catvar==2
```

Menu for estat

Statistics > Postestimation

Syntax

`estat classification` [*if*] [*in*] [*weight*] [, *options*]

options	Description
Main	
`all`	display summary statistics for all observations in the data
`cutoff(#)`	positive outcome threshold; default is `cutoff(0.5)`

`fweight`s are allowed; see **[U] 11.1.6 weight**.

`estat classification` is not appropriate after the `svy` prefix.

Options

Main

`all` requests that the statistic be computed for all observations in the data, ignoring any `if` or `in` restrictions specified by the estimation command.

`cutoff(#)` specifies the value for determining whether an observation has a predicted positive outcome. An observation is classified as positive if its predicted probability is $\geq$ #. The default is 0.5.

Remarks and examples

stata.com

`estat classification` presents the classification statistics and classification table after `logistic`, `logit`, `probit`, or `ivprobit`.

Statistics are produced either for the estimation sample (the default) or for any set of observations. When weights, `if`, or `in` is used with the estimation command, it is not necessary to repeat the qualifier when you want statistics computed for the estimation sample. Specify `if`, `in`, or the `all` option only when you want statistics computed for a set of observations other than the estimation sample. Specify weights only when you want to use a different set of weights.

▷ Example 1

We illustrate `estat classification` after `logistic`; see [R] **logistic**.

```
. use http://www.stata-press.com/data/r14/lbw
(Hosmer & Lemeshow data)

. logistic low age lwt i.race smoke ptl ht ui
  (output omitted)

. estat classification

Logistic model for low

              ──────── True ────────
Classified │         D            ~D  │      Total
───────────┼──────────────────────────┼───────────
     +     │        21            12  │         33
     -     │        38           118  │        156
───────────┼──────────────────────────┼───────────
   Total   │        59           130  │        189

Classified + if predicted Pr(D) >= .5
True D defined as low != 0
──────────────────────────────────────────────────
Sensitivity                     Pr( +| D)   35.59%
Specificity                     Pr( -|~D)   90.77%
Positive predictive value       Pr( D| +)   63.64%
Negative predictive value       Pr(~D| -)   75.64%
──────────────────────────────────────────────────
False + rate for true ~D        Pr( +|~D)    9.23%
False - rate for true D         Pr( -| D)   64.41%
False + rate for classified +   Pr(~D| +)   36.36%
False - rate for classified -   Pr( D| -)   24.36%
──────────────────────────────────────────────────
Correctly classified                        73.54%
──────────────────────────────────────────────────
```

The overall rate of correct classification is estimated to be 73.54, with 90.77% of the normal weight group correctly classified (specificity) and only 35.59% of the low weight group correctly classified (sensitivity). Classification is sensitive to the relative sizes of each component group, and always favors classification into the larger group. This phenomenon is evident here.

By default, `estat classification` uses a cutoff of 0.5, although you can vary this with the `cutoff()` option. You can use the `lsens` command to review the potential cutoffs; see [R] **lsens**.

◁

Stored results

`estat classification` stores the following in `r()`:

Scalars

`r(P_corr)`	percent correctly classified
`r(P_p1)`	sensitivity
`r(P_n0)`	specificity
`r(P_p0)`	false-positive rate given true negative
`r(P_n1)`	false-negative rate given true positive
`r(P_1p)`	positive predictive value
`r(P_0n)`	negative predictive value
`r(P_0p)`	false-positive rate given classified positive
`r(P_1n)`	false-negative rate given classified negative

Methods and formulas

Let j index observations. Define c as the `cutoff()` specified by the user or, if not specified, as 0.5. Let p_j be the predicted probability of a positive outcome and y_j be the actual outcome, which we will treat as 0 or 1, although Stata treats it as 0 and non-0, excluding missing observations.

A prediction is classified as *positive* if $p_j \geq c$ and otherwise is classified as *negative*. The classification is *correct* if it is *positive* and $y_j = 1$ or if it is *negative* and $y_j = 0$.

Sensitivity is the fraction of $y_j = 1$ observations that are correctly classified. *Specificity* is the percentage of $y_j = 0$ observations that are correctly classified.

References

Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

Kohler, U., and F. Kreuter. 2012. *Data Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press.

Also see

[R] **logistic** — Logistic regression, reporting odds ratios

[R] **logit** — Logistic regression, reporting coefficients

[R] **probit** — Probit regression

[R] **ivprobit** — Probit model with continuous endogenous covariates

[R] **lroc** — Compute area under ROC curve and graph the curve

[R] **lsens** — Graph sensitivity and specificity versus probability cutoff

[R] **estat gof** — Pearson or Hosmer–Lemeshow goodness-of-fit test

[R] **roc** — Receiver operating characteristic (ROC) analysis

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