Syntax

\texttt{estat classification} \textit{[if]} \textit{[in]} \textit{[weight]} \textit{[options]}

	extit{options} \textit{Description}

\begin{tabular}{ll}
\hline
Main & \\
\texttt{all} & display summary statistics for all observations in the data \\
\texttt{cutoff(\#)} & positive outcome threshold; default is \texttt{cutoff(0.5)} \\
\hline
\end{tabular}

weights are allowed; see \texttt{[U] 11.1.6 weight}.
\texttt{estat classification} is not appropriate after the \texttt{svy} prefix.

Menu for estat

Statistics \textgreater{} Postestimation \textgreater{} Reports and statistics

Description

\texttt{estat classification} reports various summary statistics, including the classification table.

\texttt{estat classification} requires that the current estimation results be from \texttt{logistic}, \texttt{logit}, \texttt{probit}, or \texttt{ivprobit}; see \texttt{[R] logistic}, \texttt{[R] logit}, \texttt{[R] probit}, or \texttt{[R] ivprobit}.

Options

Main

- \texttt{all} requests that the statistic be computed for all observations in the data, ignoring any \texttt{if} or \texttt{in} restrictions specified by the estimation command.

- \texttt{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

\texttt{estat classification} presents the classification statistics and classification table after \texttt{logistic}, \texttt{logit}, \texttt{probit}, or \texttt{ivprobit}.

Statistics are produced either for the estimation sample (the default) or for any set of observations. When weights, \texttt{if}, or \texttt{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 \texttt{if}, \texttt{in}, or the \texttt{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.
We illustrate `estat classification` after `logistic`; see `[R] logistic`.

```
. use http://www.stata-press.com/data/r13/lbw
   (Hosmer & Lemeshow data)
. logistic low age lwt i.race smoke ptl ht ui
   (output omitted)
. estat classification
```

<table>
<thead>
<tr>
<th>Classified</th>
<th>True</th>
<th>-D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>21</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>-</td>
<td>38</td>
<td>118</td>
<td>156</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>130</td>
<td>189</td>
</tr>
</tbody>
</table>

Classified + if predicted Pr(D) >= .5
True D defined as low != 0

<table>
<thead>
<tr>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>Pr(+</td>
</tr>
<tr>
<td>Specificity</td>
<td>Pr(-</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>Pr(D</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>Pr(D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rate</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>False + rate for true -D</td>
<td>Pr(+</td>
</tr>
<tr>
<td>False - rate for true D</td>
<td>Pr(-</td>
</tr>
<tr>
<td>False + rate for classified +</td>
<td>Pr(D</td>
</tr>
<tr>
<td>False - rate for classified -</td>
<td>Pr(D</td>
</tr>
</tbody>
</table>

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 \texttt{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


Also see

[R] \texttt{logistic} — Logistic regression, reporting odds ratios
[R] \texttt{logit} — Logistic regression, reporting coefficients
[R] \texttt{probit} — Probit regression
[R] \texttt{ivprobit} — Probit model with continuous endogenous regressors
[R] \texttt{lroc} — Compute area under ROC curve and graph the curve
[R] \texttt{lsens} — Graph sensitivity and specificity versus probability cutoff
[R] \texttt{estat gof} — Pearson or Hosmer–Lemeshow goodness-of-fit test
[R] \texttt{roc} — Receiver operating characteristic (ROC) analysis
[U] 20 Estimation and postestimation commands