Title

stata.com

lsens — Graph sensitivity and specificity versus probability cutoff

Syntax

```
lsens [depvar] [if] [in] [weight] [, options]
```

options	Description
Main	
`all`	graph all observations in the data
`genprob(`*varname*`)`	create variable containing probability cutoffs
`gensens(`*varname*`)`	create variable containing sensitivity
`genspec(`*varname*`)`	create variable containing specificity
`replace`	overwrite existing variables
`nograph`	suppress the graph
Advanced	
`beta(`*matname*`)`	row vector containing model coefficients
Plot	
connect_options	affect rendition of the plotted points connected by lines
Add plots	
`addplot(`*plot*`)`	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***

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

`lsens` is not appropriate after the `svy` prefix.

Menu

Statistics > Binary outcomes > Postestimation > Sensitivity/specificity plot

Description

`lsens` graphs sensitivity and specificity versus probability cutoff and optionally creates new variables containing these data.

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

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.

`genprob(`*varname*`)`, `gensens(`*varname*`)`, and `genspec(`*varname*`)` specify the names of new variables created to contain, respectively, the probability cutoffs and the corresponding sensitivity and specificity.

`replace` requests that existing variables specified for `genprob()`, `gensens()`, or `genspec()` be overwritten.

`nograph` suppresses graphical output.

Advanced

`beta(`*matname*`)` specifies a row vector containing model coefficients. The columns of the row vector must be labeled with the corresponding names of the independent variables in the data. The dependent variable *depvar* must be specified immediately after the command name. See *Models other than the last fitted model* later in this entry.

Plot

connect_options affect the rendition of the plotted points connected by lines; see *connect_options* in [G-2] **graph twoway scatter**.

Add plots

`addplot(`*plot*`)` provides a way to add other plots to the generated graph. See [G-3] ***addplot_option***.

Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding `by()`. These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Remarks and examples

stata.com

Remarks are presented under the following headings:

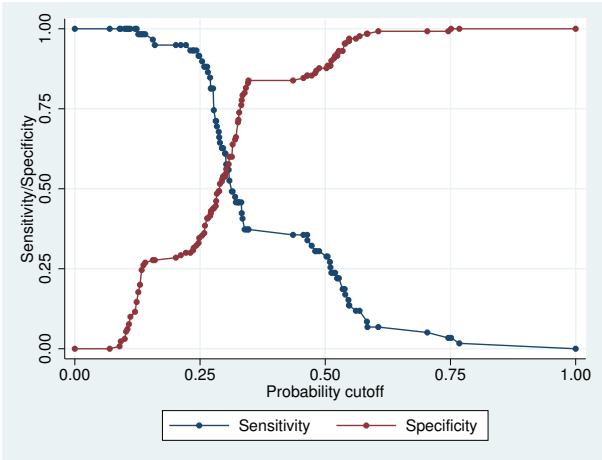
Introduction

`lsens` plots sensitivity and specificity; it plots both sensitivity and specificity versus probability cutoff c. The graph is equivalent to what you would get from `estat classification` (see [R] **estat classification**) if you varied the cutoff probability c from 0 to 1.

▷ Example 1

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

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



`lsens` optionally creates new variables containing the probability cutoff, sensitivity, and specificity.

```
. lsens, genprob(p) gensens(sens) genspec(spec) nograph
```

The variables created will have $M + 2$ distinct nonmissing values: one for each of the M covariate patterns, one for $c = 0$, and another for $c = 1$. Values are recorded for $p = 0$, for each of the observed predicted probabilities, and for $p = 1$. The total number of observations required to do this can be fewer than `_N`, the same as `_N`, or `_N` + 1, or `_N` + 2. If more observations are added, they are added at the end of the dataset and the values of the original variables are set to missing in the added observations. How the values added align with existing observations is irrelevant.

◁

❑ Technical note

`logistic`, `logit`, `probit`, or `ivprobit` and `lsens` keep track of the estimation sample. If you type, for instance, `logistic ... if x==1`, then when you type `lsens`, the statistics will be calculated on the `x==1` subsample of the data automatically.

You should specify `if` or `in` with `lsens` only when you wish to produce graphs and calculate statistics for a set of observations other than the estimation sample.

If the `logistic` model was fit with `fweights`, `lsens` properly accounts for the weights in its calculations. You do not have to specify the weights when you run `lsens`. Weights should be specified with `lsens` only when you wish to use a different set of weights.

❑

Models other than the last fitted model

By default, `lsens` uses the last model fit. You may also directly specify the model to `lsens` by inputting a vector of coefficients with the `beta()` option and passing the name of the dependent variable *depvar* to `lsens`.

▷ Example 2

Suppose that someone publishes the following logistic model of low birthweight:

$$\Pr(\texttt{low}=1) = F(-0.02\,\texttt{age} - 0.01\,\texttt{lwt} + 1.3\,\texttt{black} + 1.1\,\texttt{smoke} + 0.5\,\texttt{ptl} + 1.8\,\texttt{ht} + 0.8\,\texttt{ui} + 0.5)$$

where F is the cumulative logistic distribution. These coefficients are not odds ratios; they are the equivalent of what `logit` produces.

We can see whether this model fits our data. First we enter the coefficients as a row vector and label its columns with the names of the independent variables plus `_cons` for the constant (see [P] **matrix define** and [P] **matrix rownames**).

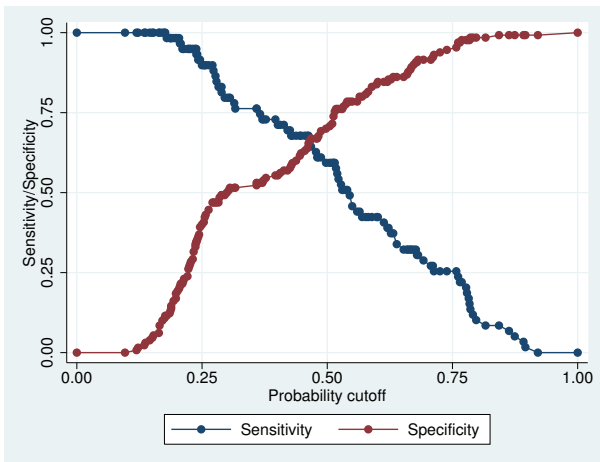
```
. use http://www.stata-press.com/data/r13/lbw3, clear
(Hosmer & Lemeshow data)
. matrix input b = (-0.02, -.01, 1.3, 1.1, .5, 1.8, .8, .5)
. matrix colnames b = age lwt black smoke ptl ht ui _cons
```

We can use `lroc` (see [R] **lroc**) to examine the predictive ability of the model:

```
. lroc low, beta(b) nograph
Logistic model for low
number of observations =       189
area under ROC curve   =    0.7275
```

The area under the curve indicates that this model does have some predictive power. We can obtain a graph of sensitivity and specificity as a function of the cutoff probability by typing

```
. lsens low, beta(b)
```



◁

Stored results

`lsens` stores the following in `r()`:

Scalars
`r(N)` number of observations

Methods and formulas

Let j index observations and c be the cutoff probability. 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.

Reference

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

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 regressors

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

[R] **estat classification** — Classification statistics and table

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

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

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