The Stata Journal (2002)
2, Number 4, pp. 351–357

Two-graph receiver operating characteristic

Michael E. Reichenheim
Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro, Brazil

Abstract. The command `roctg` allows visualizing sensitivity (Se) and specificity (Sp) curves according to the range of values of a new diagnostic test, given a "true" state of an event, the reference test. On request, several options for displaying Se and Sp estimates in, or enhancements for, the graphs are also available.

Keywords: st0025, concurrent validity, sensitivity, specificity, ROC analysis

1 Description

Receiver Operating Characteristic (ROC) analysis is often used to evaluate the accuracy of a new diagnostic test (NT) *vis-à-vis* a reference method supposedly capable of identifying the "true" state of an event—the *Reference Test* (RT) (Cleves 1999, Hanley and McNeil 1982, Tanner and Swets 1954). The analysis applies to a situation where NT is ordinal or continuous and RT, by definition, is dichotomous (normal/abnormal). Traditionally, the analysis uses the ROC curve, a graph of the sensitivity (Se) versus 1 minus specificity (Sp) of NT, Se being the fraction of $NT+$ among the $RT+$ and Sp, the fraction of $NT-$ among the $RT-$. Stata includes a series of procedures that carry out ROC analysis (`roctab`, `rocfit`, `rocplot`, etc.), yet none of those allows identifying an "optimal" score where the square of the difference between Se and Sp is minimized; i.e., where the respective curves cross. Note that the term "optimal" is meant only in the sense that it indicates the value of NT yielding the highest combination of Se and Sp.

`roctg` is a complement to Stata's `roc` commands and enables visualizing the sensitivity and specificity curves on a single graph, according to the range of values of NT, given the RT. On request, several options for displaying Se and Sp estimates in the Results window or enhancements for the graphs are also available. The procedure handles both ordinal/integer and continuous/noninteger NT variables and needs the insert `diagt` (Seed and Tobias 2001) to run properly.

2 Syntax

`roctg` *var_reftest var_newtest* [`if` *exp*] [`in` *range*] [`, cband optimal smooth lowess bwidth(#) abnormal(min | max) cont interval(#) display norank nograph saving(`*filename*`) replace level(#) symbol(`*symbol*`) xlabel(`*numlist*`) ylabel(`*numlist*`)`]

3 Options

`cband` requests that confidence bands be plotted for the sensitivity and specificity curves.

`optimal` requests that an `xline` be placed at the "optimal" cutoff point referred to above. The actual value is also shown on the graph.

`smooth` requests that the curves be smoothed using the `ksm` procedure.

`lowess` specifies that Cleveland's robust locally weighted regression is to be used in the smoothing procedure.

`bwidth(#)` specifies the bandwidth of the smoothing procedure. The default is 0.2.

`abnormal(min | max)` specifies whether, in variable NT, "abnormality" moves towards the minimum (`min`) or the maximum (`max`) value. The default is `max`. Note that this specification assumes that $RT+ = 1$ and $RT- = 0$.

`cont` specifies that the NT is to be handled as a continuous/noninteger variable rather than the default (ordinal/integer). The option requests that the cutoff points of NT used to calculate Se and Sp be based on intervals of $[(x_{\max} - x_{\min})/n_{\text{unique}}]$ rather than 1. In order to enhance computational efficiency, the number of cutoff points is automatically trimmed at 50 if the calculated interval leads to a value above this and provided that the sample size is above 100. If not, the number of cutoff points is set to half the sample size. Alternatively, the user may have full control over the desired number of cutoff points through the `interval(#)` option outlined below.

`interval(#)` specifies the intervals of the successive scores of NT for which Se and Sp estimates are calculated. Default values depend on whether or not option `cont` is requested. Note that the finer the interval, the longer `roctg` will take to run.

`display` outputs the cutoff point (score) of variable NT, for which the sensitivity and specificity curves cross, as well as both point-estimates and exact binomial confidence intervals. `display` also outputs in rank order, the five scores where $(Se - Sp)^2$ is lowest plus the point-estimates and exact binomial confidence intervals. Since the latter calculations are based on smoothed values, option `smooth` must also be requested to obtain the second part of the display.

`norank` suppresses the second part of the `display` option.

`nograph` supresses the graph when there is only interest in information from `display` or when saving results (see below). The option is ignored if none of those are requested.

`saving(`*filename*`)` requests that the calculated variables used for the graph be saved in *filename*. This enables the user to redraw new graphs at his/her own discretion. Note that saved values relate to the specified options.

`replace` indicates that the file specified by `saving()` may already exist, and, if it does, it should be overwritten.

`level(#)` specifies the confidence level (%) for the confidence interval. The default is 95%.

`symbol(`*symbol*`)` requests that symbols be placed on the point-estimate curves. All symbols available in `graxes` may be used plus "`[score]`", which specifically requests *NT* scores as symbols.

`xlabel(`*numlist*`)` is the usual graph option for customizing x-labels.

`ylabel(`*numlist*`)` is the usual graph option for customizing y-labels.

4 Example

To illustrate `roctg`, data relating to an evaluation of a "new" diagnostic method to assess gestational age (variable *GA_N*) is used. The *reference test* (e.g., ultrasonography) is represented by a binary variable called *GA_R*. A simple graph issuing `roctg` and using the `symbol()` option to display the scores on each curve is provided in Figure 1. The option `abnormal(min)` is also requested since the lowest *GA_N* value represents the most abnormal gestational age.

```
. roctg GA_R GA_N, symbol([score]) abnormal(min)
```

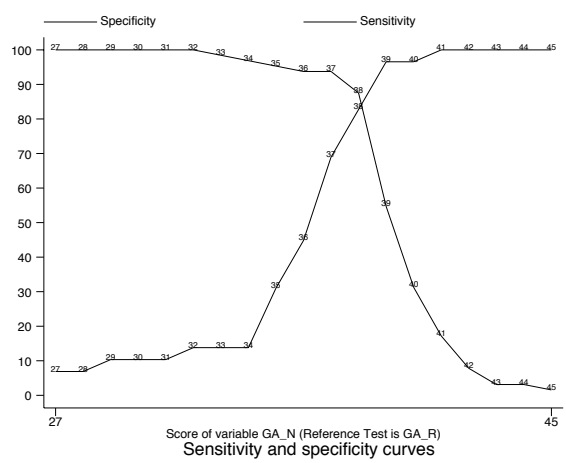

Figure 1: Graph provided by roctg with scores as symbols.

Issuing `roctg` with options `cband` and `smooth` produces the graph in Figure 2. The option `optimal` has also been requested and shows the position of the cutoff point that gives the highest combination of *Se* and *Sp*. For the sake of illustration, 99% confidence bands are shown, although `level()` is 95% by default. Note that this information is stated on the top left-hand side of the graph, along with the bandwidth used for smoothing on the right-hand side. Also note that `xlabel()` has been changed in order to improve the output.

```
. roctg GA_R GA_N, abnormal(min) cband smooth optimal level(99) xlabel(30 35 40 45)
```

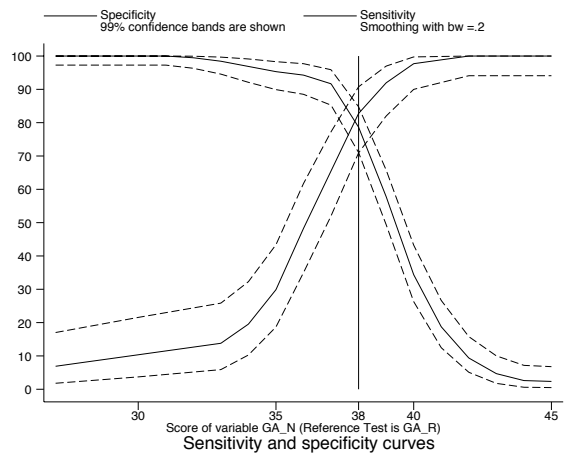

Figure 2: Graph provided by roctg with several options for enhancement.

Requesting `display` produces the following displays in the Results windows. For convenience, `nograph` has been issued. Note that in order to enable further analysis and/or new customized graphs, all the information used by `roctg` has been dumped by `saving()` to an external file. The option `replace` has also been requested, which implies that the file `external.dta` already existed. If `replace` was not specified, then an appropriate warning message would have been issued had the user inadvertently tried to overwrite the file.

```
. roctg GA_R GA_N, ab(min) cb sm display nograph saving(external) replace
                                                                  n = 279
-----------------------------------------------------------------------
  Given reference test GA_R, for variable GA_N, curves cross at
  score  38 and values (95% CI):
     Sensitivity (obs) = 82.76% (73.16% - 90.02%)
     Specificity (obs) = 87.50% (81.97% - 91.82%)
-----------------------------------------------------------------------

Note: score value and results dependent on interval size (1.00),
      which entails  18 cutoff points for variable GA_N.
-------------------------------------------------------------------------
  Rank   Score     Sensitivity (sm)              Specificity (sm)
-------------------------------------------------------------------------
  1       38       82.76 (73.85% - 89.25%)       78.65 (72.89% - 83.47%)
  2       37       65.52 (55.15% - 74.78%)       91.67 (86.88% - 95.09%)
  3       39       91.95 (84.56% - 96.19%)       57.81 (51.37% - 64.00%)
  4       36       48.28 (37.95% - 58.73%)       94.27 (89.99% - 97.10%)
  5       35       29.89 (21.01% - 40.19%)       95.31 (91.32% - 97.80%)
-------------------------------------------------------------------------
   Note: estimates are smoothed

Sensitivity and specificity values dumped
to external.dta on  8 Aug 2002 13:58
```

Assume that a variable (*GR_N*) is being envisaged as a surrogate for another "trully" tapping intra-uterine growth retardation (*IUGR*). In this example, birth weight is proposed, and, provided that sensitivity and specificity estimates are acceptable, it would be interesting to know the "optimal" cutoff point of maximal discrimination to be used in practice. Since *GR_N* is continuous and values range in the thousands, calculations, by default, would be based on too many points without much gain in accuracy regarding the identification of the desired cutoff. `roctg` perceives this and issues an appropriate warning.

```
. roctg IUGR GR_N, abnormal(min) nograph
Warning: Current specifications imply  3231 cutoff points for variable
         GR_N. Consider breaking and controlling options cont and
         int() to improve efficiency ... or wait ...
--Break--
r(1);
```

Following the suggestion, the option `cont` is then added to the statement:

```
. roctg IUGR GR_N, ab(min) nog cont
                                                              n = 227
-------------------------------------------------------------------
  Given reference test IUGR, for variable GR_N, curves cross at
  score 2718.36 and values (95% CI):
     Sensitivity (obs) = 65.31% (50.36% - 78.33%)
     Specificity (obs) = 67.42% (60.00% - 74.24%)
-------------------------------------------------------------------
Note: score value and results dependent on interval size (64.62),
      which entails  50 cutoff points for variable GR_N.
```

Note that the option `cont` with no further specification entailed 50 cutoff points placed at intervals of 64.62 grams (birth weight). Albeit quite computationally efficient, the analysis could be further improved by actively controlling the number of cutoff points. As the output below shows, doubling those by means of thinning the intervals with the option `interval(33)` provides a more accurate "optimal" cutoff point. Note that for the data at hand, further increasing the number of points did not make much difference, showing that the *Se* and *Sp* estimates have been stabilized. At any rate, it can be seen that for all practical purposes, the "optimal" cutoff point lies around 2720 g. Whether or not birth weight is worth using as a surrogate for *IUGR* on the basis of the *Se* and *Sp* estimates is debatable.

```
. roctg IUGR GR_N, ab(min) nog cont interval(33)
                                                              n = 227
-------------------------------------------------------------------
  Given reference test IUGR, for variable GR_N, curves cross at
  score 2724.00 and values (95% CI):
     Sensitivity (obs) = 65.31% (50.36% - 78.33%)
     Specificity (obs) = 66.85% (59.42% - 73.72%)
-------------------------------------------------------------------
Note: score value and results dependent on interval size (33.00),
      which entails  98 cutoff points for variable GR_N.
```

It would be quite relevant to explore the performance of *GR_N vis-à-vis IUGR* stratifying by gestational age (*GA_R*). Perhaps, the "optimal" cutoff points and *Se* and *Sp* estimates are different for the two domains, a finding that would clearly have practical implications. A sub-group analysis may be requested by using the `if` qualifier.

```
. roctg IUGR GR_N if GA_R==1, ab(min) ylabel(0 25 50 75 100) xlabel(900 4100) /*
*/ norank smooth cband optimal display cont bwidth(.25)
(234 observations deleted)
                                                              n = 76
-------------------------------------------------------------------
  Given reference test IUGR, for variable GR_N, curves cross at
  score 2694.55 and values (95% CI):
     Sensitivity (obs) = 70.83% (48.91% - 87.39%)
     Specificity (obs) = 71.15% (56.92% - 82.87%)
-------------------------------------------------------------------
Note: score value and results dependent on interval size (85.03),
      which entails  38 cutoff points for variable GR_N.

. roctg IUGR GR_N if GA_R==0, ab(min) ylab(0 25 50 75 100) xlab(900 4100)  /*
*/       nor sm cb op d cont bw(.25)
(123 observations deleted)
                                                              n = 151
-------------------------------------------------------------------
  Given reference test IUGR, for variable GR_N, curves cross at
  score 2799.60 and values (95% CI):
     Sensitivity (obs) = 64.00% (42.52% - 82.03%)
     Specificity (obs) = 64.29% (55.26% - 72.62%)
-------------------------------------------------------------------
Note: score value and results dependent on interval size (58.20),
      which entails  50 cutoff points for variable GR_N.
```

Note that the number of cutoff points for stratum *GA_R=1* has been constrained to 38 due to the small sample size. The graphical output can be seen in Figure 3. Smoothing has been increased in order to enhance the curves, y-labels customized in order to avoid cluttering, and x-labels set to fixed extreme values in order to make both graphs visually comparable.

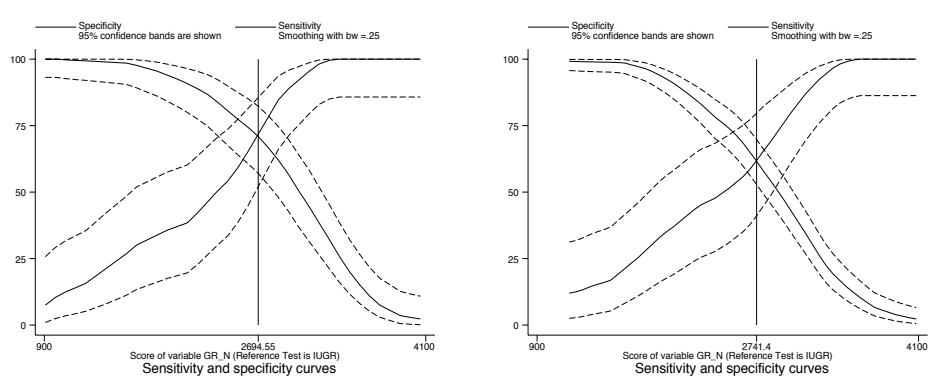

Figure 3: Graph provided by roctg according to sub-groups.

At first sight, *NT* seems to perform slightly better among the *GA_R+*. Also, the "optimal" cutoff point is 50 g lower in this stratum. Nevertheless, minding the lack of precision testified by the large confidence intervals/bands shown in the outputs and graphs, one can conclude that not much is gained from stratification.

5 Saved results

`roctg` saves in the global `S_#` macros:

`S_1`	reference test variable
`S_2`	new test variable

`roctg` saves in `r()`:

Scalars

`r(score)`	"optimal" score	`r(spec_sm)`	specificity (smoothed) *
`r(sens)`	sensitivity (unsmoothed)	`r(cutoff)`	number of cutoff points used for calculations
`r(spec)`	specificity (unsmoothed)		
`r(sens_sm)`	sensitivity (smoothed) *	`r(int)`	interval used for specifying cutoff points

* returned only if option `smooth` is requested

6 References

Cleves, M. A. 1999. sg120: Receiver Operating Characteristic (ROC) analysis. *Stata Technical Bulletin* 52: 19–33. In *Stata Technical Bulletin Reprints*, vol. 9, 212–229. College Station, TX: Stata Press.

Hanley, J. A. and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.

Seed, P. T. and A. Tobias. 2001. sbe36_1: Summary statistics for diagnostic tests. *Stata Technical Bulletin* 59: 9–12. In *Stata Technical Bulletin Reprints*, vol. 10, 90–93. College Station, TX: Stata Press.

Tanner, W. P. J. and J. A. Swets. 1954. A decision making theory of visual detection. *Psychological Review* 61: 401–409.

About the Authors

Michael E. Reichenheim (michael@ims.uerj.br), Departamento de Epidemiologia, Programa de Investigação Epidemiológica em Violência Familiar (PIEVF) / Núcleo de Pesquisa das Violências (NUPEVI), Instituto de Medicina Social (www.ims.uerj.br), Universidade do Estado do Rio de Janeiro, Brasil