Title

roc — Receiver-Operating-Characteristic (ROC) analysis

Syntax

```
roctab refvar classvar [weight] [if exp] [in range] [, bamber hanley detail
    lorenz table binomial level(#) norefline graph specificity summary
    plot(plot) rlopts(line_options) connected_options twoway_options ]

rocfit refvar classvar [weight] [if exp] [in range] [, continuous(#)
    generate(newvar) level(#) nolog maximize_options ]

rocplot [, confband level(#) norefline ciopts(rarea_options)
    plot(plot) rlopts(line_options) scatter_options twoway_options ]

roccomp refvar classvar [classvars] [weight] [if exp] [in range] [, by(varname)
    binormal level(#) test(matname) norefline separate graph summary
    rlopts(line_options) connected_options twoway_options ]

rocgold refvar goldvar classvar [classvars] [weight] [if exp] [in range] [,
    by(varname) binormal level(#) test(matname) graph summary sidak
    rlopts(line_options) connected_options twoway_options ]
```

`fweight`s are allowed; see [U] **14.1.6 weight**.

Description

The above commands are used to perform Receiver Operating Characteristic (ROC) analyses with rating and discrete classification data.

The two variables *refvar* and *classvar* must be numeric. The reference variable indicates the true state of the observation such as diseased and nondiseased or normal and abnormal, and must be coded as 0 and 1. The rating or outcome of the diagnostic test or test modality is recorded in *classvar*, which must be at least ordinal, with higher values indicating higher risk.

`roctab` is used to perform nonparametric ROC analyses. By default, `roctab` calculates the area under the ROC curve. Optionally, `roctab` can plot the ROC curve, display the data in tabular form, and produce Lorenz-like plots.

`rocfit` fits maximum-likelihood ROC models assuming a binormal distribution of the latent variable.

`rocplot` may be used after `rocfit` to plot the fitted ROC curve and simultaneous confidence bands.

`roccomp` tests the equality of two or more ROC areas obtained from applying two or more test modalities to the same sample or to independent samples. `roccomp` expects the data to be in wide form when comparing areas estimated from the same sample, and in long form for areas estimated from independent samples.

`rocgold` independently tests the equality of the ROC area of each of several test modalities, specified by *classvar*, against a "gold" standard ROC curve, *goldvar*. For each comparison, `rocgold` reports the raw and the Bonferroni adjusted significance probability. Optionally, Šidák's adjustment for multiple comparisons can be obtained.

Options

Options unique to roctab

`bamber` specifies that the standard error for the area under the ROC curve be calculated using the method suggested by Bamber (1975). Otherwise, standard errors are obtained as suggested by DeLong, DeLong, and Clarke-Pearson (1988).

`hanley` specifies that the standard error for the area under the ROC curve be calculated using the method suggested by Hanley and McNeil (1982). Otherwise, standard errors are obtained as suggested by DeLong, DeLong, and Clarke-Pearson (1988).

`detail` outputs a table displaying the sensitivity, specificity, percent of subjects correctly classified, and two likelihood ratios for each possible cut-point of *classvar*.

`lorenz` specifies that Gini and Pietra indices be reported. Optionally, `graph` will plot the Lorenz-like curve.

`table` outputs a $2 \times k$ contingency table displaying the raw data.

`binomial` specifies that exact binomial confidence intervals be calculated.

`specificity` produces a graph of sensitivity versus specificity, instead of sensitivity versus (1 − specificity). `specificity` implies `graph`.

Options unique to rocfit

`continuous(#)` specifies that the continuous *classvar* should be divided into # groups of approximately equal length. This option is required when *classvar* takes on more than 20 distinct values.

`continuous(.)` may be specified to indicate that *classvar* is to be used as it is, even though it could have more than 20 distinct values.

`generate(`*newvar*`)` specifies the new variable that is to contain the values indicating the groups produced by `continuous(#)`. `generate()` may only be specified with `continuous()`.

`nolog` prevents `rocfit` from showing the iteration log.

maximize_options control the maximization process; see [R] **maximize**. You should never have to specify any of these options.

Options unique to rocplot

`confband` specifies that simultaneous confidence bands be plotted around the ROC curve.

`ciopts(`*rarea_options*`)` affect the rendition of the confidence bands; see [G] **graph twoway rarea**.

scatter_options affect the rendition of the plotted points; see [G] **graph twoway scatter**.

Options unique to roccomp and rocgold

`by(`*varname*`)` is required when comparing independent ROC areas. The `by()` variable identifies the groups to be compared.

`binormal` specifies that the areas under the ROC curves to be compared should be estimated using the binormal distribution assumption. By default, areas to be compared are computed using the trapezoidal rule.

`test(`*matname*`)` specifies the contrast matrix to be used when comparing ROC areas. By default, the null hypothesis that all areas are equal is tested.

`separate` is meaningful only with `roccomp`; it says that each ROC curve should be placed on its own graph rather than one curve on top of the other.

`sidak` (`rocgold` only) requests that the significance probability be adjusted for the effect of multiple comparisons using Šidák's method. Bonferroni's adjustment is reported by default.

Options common to several commands

`level(`*#*`)` specifies the confidence level, in percent, for the confidence intervals. For `rocplot`, it specifies the confidence level for the confidence bands. The default is `level(95)` or as set by `set level`; see [R] **level**.

`graph` produces graphical output of the ROC curve. For `roctab`, if `lorenz` is specified, graphical output of a Lorenz-like curve will be produced.

`summary` reports the area under the ROC curve, its standard error, and its confidence interval. If `lorenz` is specified with `roctab`, Lorenz indices are reported. This option is only needed when also specifying `graph`.

`norefline` suppresses the plotting of the 45-degree reference line from the graphical output of the ROC curve.

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

`rlopts(`*line_options*`)` affect the rendition of the reference line; see [G] **graph twoway line**.

connected_options affect the rendition of the plotted points connected by lines; see [G] **graph twoway connected**.

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

Remarks

Remarks are presented under the headings

Introduction
Nonparametric ROC *curves*
Parametric ROC *curves*
Lorenz-like curves
Comparing areas under the ROC *curve*

Introduction

Receiver Operating Characteristic (ROC) analysis is used to quantify the accuracy of diagnostic tests or other evaluation modality used to discriminate between two states or conditions. For ease of presentation, we will refer to these two states as normal and abnormal, and to the discriminatory test as a diagnostic test. The discriminatory accuracy of a diagnostic test is measured by its ability to correctly classify known normal and abnormal subjects. The analysis uses the ROC curve, a graph of the sensitivity versus 1 − specificity of the diagnostic test. The sensitivity is the fraction of positive cases that are correctly classified by the diagnostic test, while the specificity is the fraction of negative cases that are correctly classified. Thus, the sensitivity is the true-positive rate, and the specificity the true-negative rate.

The global performance of a diagnostic test is commonly summarized by the area under the ROC curve. This area can be interpreted as the probability that the result of a diagnostic test of a randomly selected abnormal subject will be greater than the result of the same diagnostic test from a randomly selected normal subject. The greater the area under the ROC curve, the better the global performance of the diagnostic test.

Both nonparametric methods and parametric (semi-parametric) methods have been suggested for generating the ROC curve and for calculating its area. In the following sections we present these approaches, and in the last section we present tests for comparing areas under ROC curves.

Nonparametric ROC curves

The points on the nonparametric ROC curve are generated by using each possible outcome of the diagnostic test as a classification cut-point and computing the corresponding sensitivity and 1 − specificity. These points are then connected by straight lines, and the area under the resulting ROC curve is computed using the trapezoidal rule.

▷ Example

Hanley and McNeil (1982) presented data from a study in which a reviewer was asked to classify, using a nine point scale, a random sample of 109 tomographic images from patients with neurological problems. The rating scale was as follows: 1–definitely normal, 2–probably normal, 3–questionable, 4–probably abnormal, and 5–definitely abnormal. The true disease status was normal for 58 of the patients and abnormal for the remaining 51 patients.

Here, we list 9 of the 109 observations.

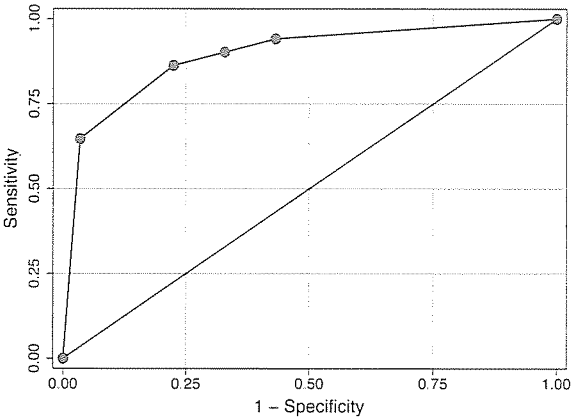
(*Continued on next page*)

```
. use http://www.stata-press.com/data/r8/hanley

. list disease rating in 1/9

     +------------------+
     | disease   rating |
     |------------------|
  1. |       1        5 |
  2. |       0        1 |
  3. |       1        5 |
  4. |       0        4 |
  5. |       0        1 |
     |------------------|
  6. |       0        3 |
  7. |       1        5 |
  8. |       0        5 |
  9. |       0        1 |
     +------------------+
```

For each observation, `disease` identifies the true disease status of the subject (0 = normal, 1 = abnormal), and `rating` contains the classification value assigned by the reviewer.

We can use `roctab` to calculate and plot the nonparametric ROC curve by specifying both the `summary` and `graph` options. By also specifying the `table` option, we obtain a contingency table summarizing our dataset.

```
. roctab disease rating, table graph summary
```



Area under ROC curve = 0.8932

```
           |                        rating
   disease |         1          2          3          4          5 |     Total
-----------+-------------------------------------------------------+----------
         0 |        33          6          6         11          2 |        58
         1 |         3          2          2         11         33 |        51
-----------+-------------------------------------------------------+----------
     Total |        36          8          8         22         35 |       109

                      ROC                    -Asymptotic Normal--
           Obs       Area     Std. Err.      [95% Conf. Interval]
         ------------------------------------------------------------
           109     0.8932       0.0307        0.83295     0.95339
```

By default, `roctab` reports the area under the curve, its standard error, and its confidence interval. The `graph` option can be used to plot the ROC curve.

The ROC curve is plotted by computing the sensitivity and specificity using each value of the rating variable as a possible cut-point. A point is plotted on the graph for each of the cut-points. These plotted points are joined by straight lines to form the ROC curve, and the area under the ROC curve is computed using the trapezoidal rule.

We can tabulate the computed sensitivities and specificities for each of the possible cut-points by specifying `detail`.

```
. roctab disease rating, detail
Detailed report of Sensitivity and Specificity
------------------------------------------------------------------------------
                                            Correctly
Cut point      Sensitivity   Specificity   Classified          LR+          LR-
------------------------------------------------------------------------------
( >= 1 )           100.00%         0.00%       46.79%       1.0000
( >= 2 )            94.12%        56.90%       74.31%       2.1835       0.1034
( >= 3 )            90.20%        67.24%       77.98%       2.7534       0.1458
( >= 4 )            86.27%        77.59%       81.65%       3.8492       0.1769
( >= 5 )            64.71%        96.55%       81.65%      18.7647       0.3655
( >  5 )             0.00%       100.00%       53.21%                    1.0000
------------------------------------------------------------------------------

                     ROC                      -Asymptotic Normal--
           Obs       Area     Std. Err.      [95% Conf. Interval]
         ------------------------------------------------------------
           109     0.8932       0.0307        0.83295       0.95339
```

Each cut-point in the table indicates the ratings used to classify tomographs as being from an abnormal subject. For example, the first cut-point, (>= 1), indicates that all tomographs rated as 1 or greater are classified as coming from abnormal subjects. Because all tomographs have a rating of 1 or greater, all are considered abnormal. Consequently, all abnormal cases are correctly classified (sensitivity = 100%), but none of the normal patients are classified correctly (specificity = 0%). For the second cut-point (>= 2), tomographs with ratings of 1 are classified as normal and those with ratings of 2 or greater are classified as abnormal. The resulting sensitivity and specificity are 94.12% and 56.90%, respectively. Using this cut-point, we correctly classified 74.31% of the 109 tomographs. Similar interpretations can be used on the remaining cut-points. As mentioned, each cut-point corresponds to a point on the nonparametric ROC curve. The first cut-point, (>= 1), corresponds to the point at (1,1) and the last cut-point, (> 5), to the point at (0,0).

`detail` also reports two likelihood ratios suggested by Choi (1998): the likelihood ratio for a positive test result (LR+) and the likelihood ratio for a negative test result (LR–). The likelihood ratio for a positive test result is the ratio of the probability of a positive test among the truly positive subjects to the probability of a positive test among the truly negative subjects. The likelihood ratio for a negative test result (LR–) is the ratio of the probability of a negative test among the truly positive subjects to the probability of a negative test among the truly negative subjects. Choi points out that LR+ corresponds to the slope of the line from the origin to the point on the ROC curve determined by the cut-point. Similarly, LR– corresponds to the slope from the point (1,1) to the point on the ROC curve determined by the cut-point.

By default, `roctab` calculates the standard error for the area under the curve using an algorithm suggested by DeLong, DeLong, and Clarke-Pearson (1988) and asymptotic normal confidence intervals. Optionally, standard errors based on methods suggested by Bamber (1975) or Hanley and McNeil (1982) can be computed by specifying `bamber` or `hanley` respectively, and an exact binomial confidence interval can be obtained by specifying `binomial`.

```
. roctab disease rating, bamber
                      ROC        Bamber       -Asymptotic Normal—
           Obs        Area      Std. Err.      [95% Conf. Interval]
          ------------------------------------------------------------
           109      0.8932       0.0306       0.83317      0.95317
. roctab disease rating, hanley binomial
                      ROC        Hanley       — Binomial Exact —
           Obs        Area      Std. Err.      [95% Conf. Interval]
          ------------------------------------------------------------
           109      0.8932       0.0320       0.81559      0.94180
```

◁

Parametric ROC curves

Dorfman and Alf (1969) developed a generalized approach for obtaining maximum likelihood estimates of the parameters for a smooth fitting ROC curve. The most commonly used method, and the one implemented here, is based upon the binormal model.

The model assumes the existence of an unobserved continuous latent variable that is normally distributed (perhaps after a monotonic transformation) in both the normal and abnormal populations with means μ_n and μ_a, and variances σ_n^2 and σ_a^2, respectively. The model further assumes that the K categories of the rating variable result from partitioning the unobserved latent variable by $K-1$ fixed boundaries. The method fits a straight line to the empirical ROC points plotted using normal probability scales on both axes. Maximum likelihood estimates of the line's slope and intercept and the $K-1$ boundaries are obtained simultaneously. See *Methods and Formulas* for details.

The intercept from the fitted line is a measurement of $(\mu_a - \mu_n)/\sigma_a$, and the slope measures σ_n/σ_a.

Thus, the intercept is the standardized difference between the two latent population means, and the slope is the ratio of the two standard deviations. The null hypothesis of no difference between the two population means is evaluated by testing if the intercept $=0$, and the null hypothesis that the variances in the two populations are equal is evaluated by testing if the slope $=1$.

▷ Example

We use Hanley and McNeil's (1982) dataset, described in the previous example, to fit a smooth ROC curve assuming a binormal model.

(*Continued on next page*)

```
. rocfit disease rating
Fitting binormal model:
Iteration 0:   log likelihood = -123.68069
Iteration 1:   log likelihood = -123.64867
Iteration 2:   log likelihood = -123.64855
Iteration 3:   log likelihood = -123.64855
Binormal model of disease on rating                 Number of obs    =        109
Goodness-of-fit chi2(2) =       0.21
Prob > chi2             =     0.9006
Log likelihood          =  -123.64855

------------------------------------------------------------------------------
             |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   intercept |   1.656782   0.310456     5.34    0.000     1.048300    2.265265
   slope (*) |   0.713002   0.215882    -1.33    0.092     0.289881    1.136123
-------------+----------------------------------------------------------------
       _cut1 |   0.169768   0.165307     1.03    0.152    -0.154227    0.493764
       _cut2 |   0.463215   0.167235     2.77    0.003     0.135441    0.790990
       _cut3 |   0.766860   0.174808     4.39    0.000     0.424243    1.109477
       _cut4 |   1.797938   0.299581     6.00    0.000     1.210770    2.385106
------------------------------------------------------------------------------

------------------------------------------------------------------------------
             |                  Indices from binormal fit
       Index |   Estimate    Std. Err.                   [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ROC area |   0.911331   0.029506                    0.853501    0.969161
    delta(m) |   2.323671   0.502370                    1.339044    3.308298
        d(e) |   1.934361   0.257187                    1.430284    2.438438
        d(a) |   1.907771   0.259822                    1.398530    2.417012
------------------------------------------------------------------------------
(*) z test for slope==1
```

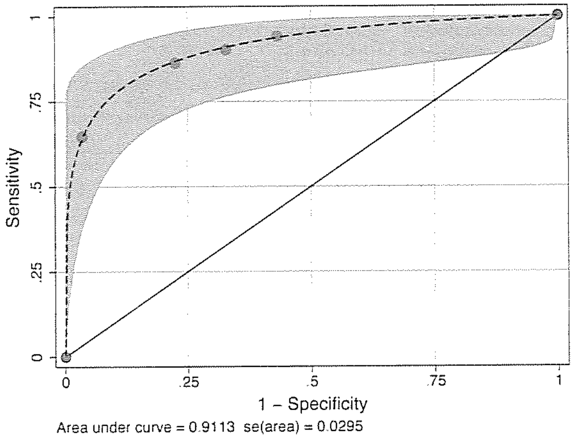
`rocfit` outputs the MLE for the intercept and slope of the fitted regression line along with, in this case, 4 boundaries (because there are 5 ratings) labeled `_cut1` through `_cut4`. In addition, `rocfit` also computes and reports 4 indices based on the fitted ROC curve: the area under the curve (labeled `ROC area`), $\delta(m)$ (labeled `delta(m)`), d_e (labeled `d(e)`), and d_a (labeled `d(a)`). More information about these indices can be found in the *Methods and Formulas* section and in Erdreich and Lee (1981).

Note that in the output table we are testing whether or not the variances of the two latent populations are equal by testing if the slope $= 1$.

We plot the fitted ROC curve.

(*Continued on next page*)

```
. rocplot, confband
```



◁

Lorenz-like curves

For applications where it is known that the risk status increases or decreases monotonically with increasing values of the diagnostic test, the ROC curve and associated indices are useful in assessing the overall performance of a diagnostic test. When the risk status does not vary monotonically with increasing values of the diagnostic test, however, the resulting ROC curve can be nonconvex and its indices unreliable. For these situations, Lee (1999) proposed an alternative to the ROC analysis based on Lorenz-like curves and the associated Pietra and Gini indices.

Lee (1999) mentions at least three specific situations where results from Lorenz curves are superior to those obtained from ROC curves: (1) a diagnostic test with similar means but very different standard deviations in the abnormal and normal populations, (2) a diagnostic test with bimodal distributions in either the normal or abnormal population, and (3) a diagnostic test distributed symmetrically in the normal population and skewed in the abnormal.

When the risk status increases or decreases monotonically with increasing values of the diagnostic test, the ROC and Lorenz curves yield interchangeable results.

▷ Example

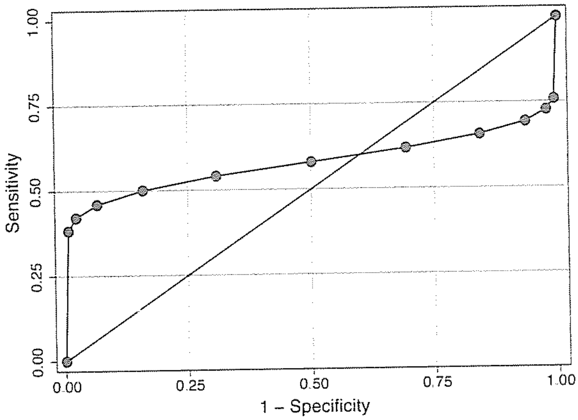
To illustrate the use of the `lorenz` option we constructed a fictitious dataset that yields results similar to those presented in Table III of Lee (1999). The data assume that a 12 point rating scale was used to classify 442 diseased and 442 healthy subjects. We list a few of the observations.

(*Continued on next page*)

```
. use http://stata-press.com/data/r8/lorenz
. list in 1/7, noobs sep(0)

  ┌───────────────────────┐
  │ disease   class   pop │
  ├───────────────────────┤
  │       0       5    66 │
  │       1      11    17 │
  │       0       6    85 │
  │       0       3    19 │
  │       0      10    19 │
  │       0       2     7 │
  │       1       4    16 │
  └───────────────────────┘
```

The data consist of 24 observations, 12 observations from diseased individuals and 12 from nondiseased individuals. Each observation corresponds to one of the 12 classification values of the rating scale variable, `class`. The number of subjects represented by each observation is given by the pop variable, making this a frequency-weighted dataset. The data were generated assuming a binormal distribution of the latent variable with similar means for the normal and abnormal populations, but with the standard deviation for the abnormal population 5 times greater than that of the normal population.

```
. roctab disease class [fweight=pop], graph summary
```



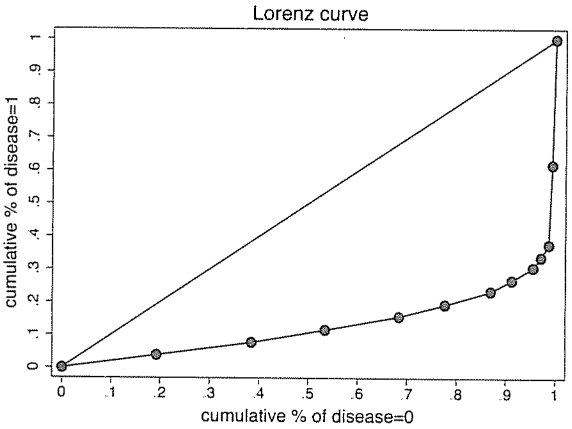
Area under ROC curve = 0.5774

```
                      ROC                    -Asymptotic Normal--
           Obs       Area     Std. Err.      [95% Conf. Interval]
         ------------------------------------------------------------
           884     0.5774       0.0215        0.53517     0.61959
```

The resulting ROC curve is nonconvex or, as termed by Lee, "wiggly". Lee argues that for this and similar situations, the Lorenz curve and indices are preferred.

(*Continued on next page*)

```
. roctab disease class [fweight=pop], lorenz graph summary
```



```
Lorenz curve
─────────────────────────────
  Pietra index =    0.6493
  Gini index   =    0.7441
```

Like ROC curves, a more bowed Lorenz curve suggests a better diagnostic test. This "bowedness" is quantified by the Pietra index, which is geometrically equivalent to twice the largest triangle that can be inscribed in the area between the curve and the diagonal line, and the Gini index, which is equivalent to twice the area between the Lorenz curve and the diagonal. Lee (1999) provides several additional interpretations for the Pietra and Gini indices. If interested, consult the reference for more information.

◁

Comparing areas under the ROC curve

The area under multiple ROC curves can be compared using `roccomp`. The command syntax is slightly different if the ROC curves are correlated (i.e., different diagnostic tests applied to the same sample) or independent (i.e., diagnostic tests applied to different samples).

Correlated data

▷ Example

Hanley and McNeil (1983) presented data from an evaluation of two computer algorithms designed to reconstruct CT images from phantoms. We will call these two algorithms' modalities 1 and 2. A sample of 112 phantoms was selected; 58 phantoms were considered normal and the remaining 54 were abnormal. Each of the two modalities was applied to each phantom and the resulting images rated by a reviewer using a six point scale: 1–definitely normal, 2–probably normal, 3–possibly normal, 4–possibly abnormal, 5–probably abnormal, and 6–definitely abnormal. Because each modality was applied to the same sample of phantoms, the two sets of outcomes are correlated.

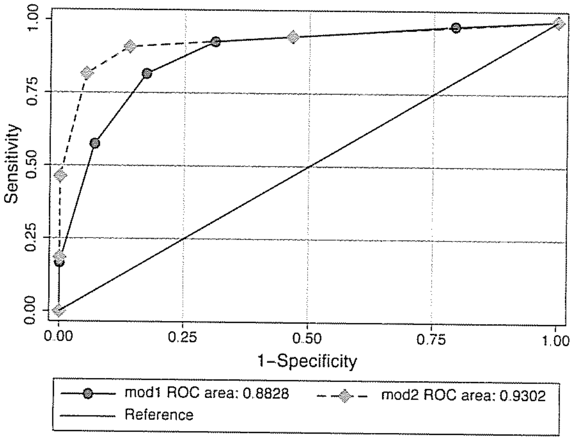
We list the first seven observations:

```
. use http://www.stata-press.com/data/r8/ct
. list in 1/7, sep(0)

     ┌──────────────────────┐
     │ mod1   mod2   status │
     ├──────────────────────┤
  1. │    2      1        0 │
  2. │    5      5        1 │
  3. │    2      1        0 │
  4. │    2      3        0 │
  5. │    5      6        1 │
  6. │    2      2        0 │
  7. │    3      2        0 │
     └──────────────────────┘
```

Note that the data are in wide form. This is required when dealing with correlated data. Each observation corresponds to one phantom. The variable `mod1` identifies the rating assigned for the first modality, and `mod2` identifies the rating assigned for the second modality. The true status of the phantoms is given by `status=0` if normal and `status=1` if abnormal. The observations with at least one missing rating were dropped from the analysis.

We plot the two ROC curves and compare their areas.

```
. roccomp status mod1 mod2, graph summary
```



```
                              ROC                    -Asymptotic Normal--
                 Obs         Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
mod1             112       0.8828       0.0317        0.82067     0.94498
mod2             112       0.9302       0.0256        0.88005     0.98042
-------------------------------------------------------------------------
Ho: area(mod1) = area(mod2)
    chi2(1) =       2.31       Prob>chi2 =    0.1282
```

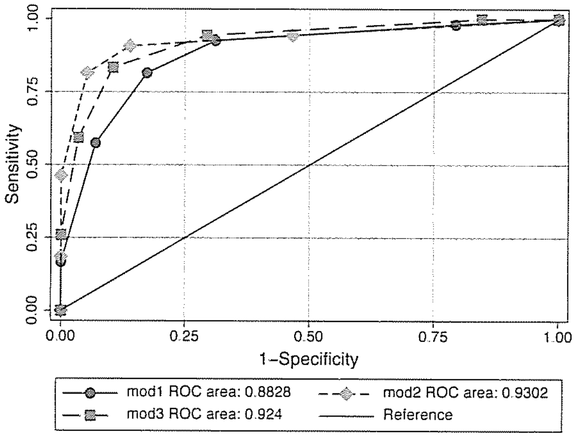
By default, `roccomp`, with the `graph` option specified, plots the ROC curves on the same graph. Optionally, the curves can be plotted side by side, each on its own graph, by also specifying `separate`.

For each curve, `roccomp` reports summary statistics and provides a test for the equality of the area under the curves using an algorithm suggested by DeLong, DeLong, and Clarke-Pearson (1988).

Although the area under the ROC curve for modality 2 is larger than that of modality 1, the chi-squared test yielded a significance probability of 0.1282, suggesting that there is no significant difference between these two areas.

The `roccomp` command can also be used to compare more than two ROC areas. To illustrate this, we modified the previous dataset by including a fictitious third modality.

```
. use http://www.stata-press.com/data/r8/ct2
. roccomp status mod1 mod2 mod3, graph summary
```



```
                             ROC                    -Asymptotic Normal--
                 Obs       Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
mod1             112     0.8828       0.0317        0.82067     0.94498
mod2             112     0.9302       0.0256        0.88005     0.98042
mod3             112     0.9240       0.0241        0.87670     0.97132
-------------------------------------------------------------------------
Ho: area(mod1) = area(mod2) = area(mod3)
    chi2(2) =      6.54       Prob>chi2 =    0.0381
```

By default, `roccomp` tests whether the areas under the ROC curves are all equal. Other comparisons can be tested by creating a contrast matrix and specifying `test(`*matname*`)`, where *matname* is the name of the contrast matrix.

For example, assume that we are interested in testing whether the area under the ROC for `mod1` is equal to that of `mod3`. To do this, we can first create an appropriate contrast matrix and then specify its name with the `test()` option.

Of course, this is a trivial example because we could have just specified

```
. roccomp status mod1 mod3
```

without including `mod2` to obtain the same test results. However, for illustration we will continue with this example.

The contrast matrix must have its number of columns equal to the number of *classvars* (i.e., the total number of ROC curves), a number of rows less than or equal to the number of *classvars*, and the elements of each row must add to zero.

(*Continued on next page*)

```
. matrix C=(1,0,-1)

. roccomp status mod1 mod2 mod3, test(C)
                              ROC                  -Asymptotic Normal--
                 Obs         Area     Std. Err.    [95% Conf. Interval]
-------------------------------------------------------------------------
mod1             112       0.8828       0.0317      0.82067     0.94498
mod2             112       0.9302       0.0256      0.88005     0.98042
mod3             112       0.9240       0.0241      0.87670     0.97132
-------------------------------------------------------------------------
Ho: Comparison as defined by contrast matrix: C
    chi2(1) =      5.25       Prob>chi2 =    0.0220
```

Note that although all three areas are reported, the comparison is made using the specified contrast matrix.

Perhaps more interesting would be a comparison of the area from `mod1` and the average area of `mod2` and `mod3`.

```
. matrix C=(1,-.5,-.5)

. roccomp status mod1 mod2 mod3, test(C)
                              ROC                  -Asymptotic Normal--
                 Obs         Area     Std. Err.    [95% Conf. Interval]
-------------------------------------------------------------------------
mod1             112       0.8828       0.0317      0.82067     0.94498
mod2             112       0.9302       0.0256      0.88005     0.98042
mod3             112       0.9240       0.0241      0.87670     0.97132
-------------------------------------------------------------------------
Ho: Comparison as defined by contrast matrix: C
    chi2(1) =      3.43       Prob>chi2 =    0.0642
```

Other contrasts could be made. For example, we could test if `mod3` is different from at least one of the other two by first creating the following contrast matrix:

```
. matrix C=(-1, 0, 1 \ 0, -1, 1)

. matrix list C

C[2,3]
    c1  c2  c3
r1  -1   0   1
r2   0  -1   1
```

◁

Independent data

▷ Example

In the previous example, we noted that because each test modality was applied to the same sample of phantoms, the classification outcomes were correlated. Now assume that we have collected the same data as presented by Hanley and McNeil (1983), except that we applied the first test modality to one sample of phantoms and the second test modality to a different sample of phantoms. The resulting measurements are now considered independent.

Here are a few of the observations.

```
. use http://www.stata-press.com/data/r8/ct3
. list in 1/7, sep(0)

     +------------------------------+
     | pop   status   rating   mod  |
     |------------------------------|
  1. |  12        0        1     1  |
  2. |  31        0        1     2  |
  3. |   1        1        1     1  |
  4. |   3        1        1     2  |
  5. |  28        0        2     1  |
  6. |  19        0        2     2  |
  7. |   3        1        2     1  |
     +------------------------------+
```

Note that the data are in long form. This is required when dealing with independent data. The data consist of 24 observations, 6 observations corresponding to abnormal phantoms and 6 to normal phantoms evaluated using the first modality, and similarly 6 observations corresponding to abnormal phantoms and 6 to normal phantoms evaluated using the second modality. The number of phantoms corresponding to each observation is given by the pop variable. Once again we have frequency-weighted data. The variable `mod` identifies the modality and `rating` is the assigned classification.

We can better view our data by using the `table` command.

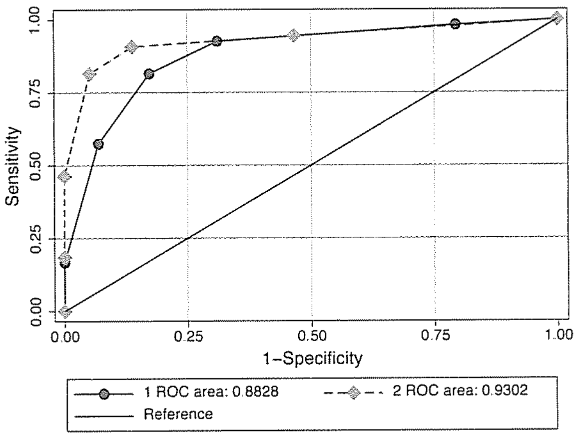
```
. table status rating [fw=pop], by(mod) row col

----------------------------------------------------------
mod and   |                    rating
status    |     1      2      3      4      5      6  Total
----------+-----------------------------------------------
1         |
        0 |    12     28      8      6      4            58
        1 |     1      3      6     13     22      9     54
          |
    Total |    13     31     14     19     26      9    112
----------+-----------------------------------------------
2         |
        0 |    31     19      5      3                   58
        1 |     3      2      5     19     15     10     54
          |
    Total |    34     21     10     22     15     10    112
----------------------------------------------------------
```

The `status` variable indicates the true status of the phantoms, `status=0` if normal and `status=1` if abnormal.

We now compare the areas under the two ROC curves.

(*Continued on next page*)

```
. roccomp status rating [fw=pop], by(mod) graph summary
```



```
                               ROC                      -Asymptotic Normal--
mod                 Obs        Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
1                   112      0.8828       0.0317        0.82067     0.94498
2                   112      0.9302       0.0256        0.88005     0.98042
-------------------------------------------------------------------------
Ho: area(1) = area(2)
    chi2(1) =     1.35       Prob>chi2 =    0.2447
```

◁

Comparing areas with a gold standard

The area under multiple ROC curves can be compared with a gold standard using `rocgold`. The command syntax is similar to that of `roccomp`. The tests are corrected for the effect of multiple comparisons.

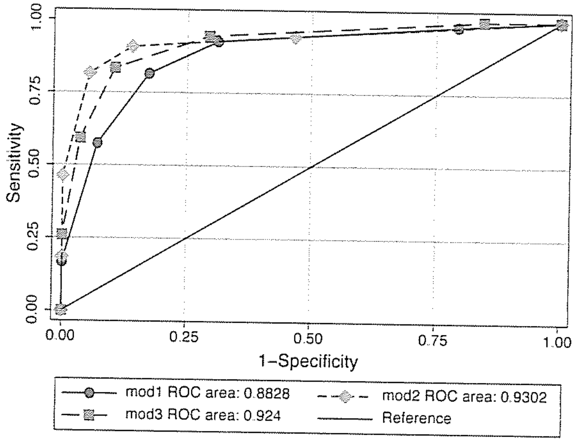
▷ Example

We will use the same data (presented by Hanley and McNeil (1983)) as in the `roccomp` examples. Let's assume the first modality is considered to be the standard with which both the second and third modalities are compared.

We want to plot and compare both the areas of the ROC curves of `mod2` and `mod3` with `mod1`. Since `mod1` is what we are considering to be the gold standard, it is listed first after the reference variable in the `rocgold` command line.

(Continued on next page)

```
. use http://www.stata-press.com/data/r8/ct2
. rocgold status mod1 mod2 mod3, graph summary
```



```
-------------------------------------------------------------------------------
                         ROC                                        Bonferroni
                         Area     Std. Err.      chi2     df    Pr>chi2    Pr>chi2
-------------------------------------------------------------------------------
mod1 (standard)        0.8828      0.0317
mod2                   0.9302      0.0256       2.3146     1     0.1282     0.2563
mod3                   0.9240      0.0241       5.2480     1     0.0220     0.0439
-------------------------------------------------------------------------------
```

Equivalently, we could have done this in two steps by using the `roccomp` command.

```
. roccomp status mod1 mod2, graph summary
. roccomp status mod1 mod3, graph summary msymbol(O S)
```

◁

Saved Results

`roctab` saves in `r()`:

Scalars

`r(N)`	number of observations
`r(se)`	standard error for the area under the ROC curve
`r(lb)`	lower bound of CI for the area under the ROC curve
`r(ub)`	upper bound of CI for the area under the ROC curve
`r(area)`	area under the ROC curve
`r(pietra)`	Pietra index
`r(gini)`	Gini index

(Continued on next page)

`rocfit` saves in `e()`:

Scalars

`e(N)`	number of observations
`e(k)`	number of parameters
`e(k_eq)`	number of equations
`e(k_dv)`	number of dependent variables
`e(df_m)`	model degrees of freedom
`e(ll)`	log likelihood
`e(rc)`	return code
`e(ic)`	number of iterations
`e(rank)`	rank of `e(V)`
`e(chi2_gf)`	goodness-of-fit χ^2
`e(df_gf)`	goodness-of-fit degrees of freedom
`e(p_gf)`	χ^2 goodness-of-fit significance probability
`e(area)`	area under the ROC curve
`e(se_area)`	standard error for the area under the ROC curve
`e(deltam)`	$\delta(m)$
`e(se_delm)`	standard area for $\delta(m)$
`e(de)`	d_e index
`e(se_de)`	standard error for d_e index
`e(da)`	d_a index
`e(se_da)`	standard error for d_a index

Macros

`e(cmd)`	`rocfit`
`e(depvar)`	names of dependent variables
`e(title)`	title in estimation output
`e(wtype)`	weight type
`e(wexp)`	weight expression
`e(user)`	name of likelihood-evaluator program
`e(opt)`	type of optimization
`e(chi2type)`	`GOF`; type of model χ^2 test

Matrices

`e(b)`	coefficient vector
`e(ilog)`	iteration log (up to 20 iterations)
`e(V)`	variance–covariance matrix of the estimators

Functions

`e(sample)`	marks estimation sample

`roccomp` saves in `r()`:

Scalars

`r(N_g)`	number of groups
`r(p)`	significance probability
`r(df)`	χ^2 degrees of freedom
`r(chi2)`	χ^2

Matrices

`r(V)`	variance–covariance matrix

`rocgold` saves in `r()`:

Scalars

`r(N_g)`	number of groups

Matrices

`r(V)`	variance–covariance matrix
`r(chi2)`	χ^2 vector
`r(df)`	χ^2 degrees-of-freedom vector
`r(p)`	significance probability vector
`r(p_adj)`	adjusted significance probability vector

Methods and Formulas

`roctab`, `rocfit`, `rocplot`, `roccomp`, and `rocgold` are implemented as ado-files.

Assume that we applied a diagnostic test to each of N_n normal and N_a abnormal subjects. Further assume that the higher the outcome value of the diagnostic test, the higher the risk of the subject being abnormal. Let $\widehat{\theta}$ be the estimated area under the curve, and let $X_i, i = 1, 2, \ldots, N_a$ and $Y_j, j = 1, 2, \ldots, N_n$ be the values of the diagnostic test for the abnormal and normal subjects, respectively.

Nonparametric ROC

The points on the nonparametric ROC curve are generated by using each possible outcome of the diagnostic test as a classification cut-point and computing the corresponding sensitivity and 1 − specificity. These points are then connected by straight lines, and the area under the resulting ROC curve is computed using the trapezoidal rule.

The default standard error for the area under the ROC curve is computed using the algorithm described by DeLong, DeLong, and Clarke-Pearson (1988). For each abnormal subject, i, define

$$V_{10}(X_i) = \frac{1}{N_n} \sum_{j=1}^{N_n} \psi(X_i, Y_j)$$

and for each normal subject, j, define

$$V_{01}(Y_j) = \frac{1}{N_a} \sum_{i=1}^{N_a} \psi(X_i, Y_j)$$

where

$$\psi(X, Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

Define

$$S_{10} = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} \{V_{10}(X_i) - \widehat{\theta}\}^2$$

and

$$S_{01} = \frac{1}{N_n - 1} \sum_{j=1}^{N_n} \{V_{01}(Y_j) - \widehat{\theta}\}^2$$

The variance of the estimated area under the ROC curve is given by

$$\mathrm{var}(\widehat{\theta}) = \frac{1}{N_a} S_{10} + \frac{1}{N_n} S_{01}$$

The `hanley` standard error for the area under the ROC curve is computed using the algorithm described by Hanley and McNeil (1982). It requires the calculation of two quantities, Q_1 and Q_2, where Q_1 is Pr(two randomly selected abnormal subjects will both have a higher score than a randomly selected normal subject), and Q_2 is Pr(one randomly selected abnormal subject will have a higher score than any two randomly selected normal subjects). The Hanley and McNeil variance of the estimated area under the ROC curve is

$$\mathrm{var}(\widehat{\theta}) = \frac{\widehat{\theta}(1 - \widehat{\theta}) + (N_a - 1)(Q_1 - \widehat{\theta}^2) + (N_n - 1)(Q_2 - \widehat{\theta}^2)}{N_a N_n}$$

The `bamber` standard error for the area under the ROC curve is computed using the algorithm described by Bamber (1975). For any two Y values, Y_j and Y_k, and any X_i value, define

$$b_{yyx} = p(Y_j, Y_k < X_i) + p(X_i < Y_j, Y_k) - 2p(Y_j < X_i < Y_k)$$

and similarly, for any two X values, X_i and X_l, and any Y_j value, define

$$b_{xxy} = p(X_i, X_l < Y_j) + p(Y_j < X_i, X_l) - 2p(X_i < Y_j < X_l)$$

Then Bamber's unbiased estimate of the variance for the area under the ROC curve is

$$\text{var}(\widehat{\theta}) = \frac{1}{4}(N_a-1)(N_n-1)\{p(X \neq Y)+(N_a-1)b_{xxy}+(N_n-1)b_{yyx}-4(N_a+N_n-1)(\widehat{\theta}-0.5)^2\}$$

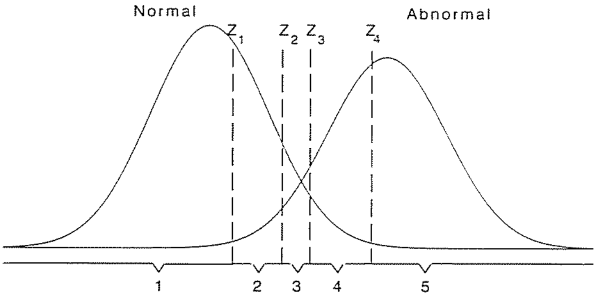
Asymptotic confidence intervals are constructed and reported by default, assuming a normal distribution for the area under the ROC curve.

Exact binomial confidence intervals are calculated as described in [R] **ci**, with p equal to the area under the ROC curve.

Parametric ROC curves

Dorfman and Alf (1969) developed a general procedure for obtaining maximum likelihood estimates of the parameters of a smooth fitting ROC curve. The most common method, and the one implemented in Stata, is based upon the binormal model.

The model assumes that there is an unobserved continuous latent variable that is normally distributed in both the normal and abnormal populations. The idea is better explained with the following illustration:



It is assumed that the latent variable is normally distributed for both the normal and abnormal subjects, perhaps after a monotonic transformation, with means μ_n and μ_a, and variances σ_n^2 and σ_a^2, respectively.

This latent variable is assumed to be partitioned into the k categories of the rating variable by $k-1$ fixed boundaries. In the above figure, the $k=5$ categories of the rating variable identified on the bottom result from the partition of the four boundaries Z_1 through Z_4.

Let R_j for $j = 1, 2, \ldots, k$ indicate the categories of the rating variable, let $i = 1$ if the subject belongs to the normal group, and let $i = 2$ if the subject belongs to the abnormal group.

Then

$$p(R_j | i = 1) = F(Z_j) - F(Z_{j-1})$$

where $Z_k = (x_k - \mu_n)/\sigma_n$. F is the cumulative normal distribution. $F(Z_0) = 0$ and $F(Z_k) = 1$. Also.

$$p(R_j|i = 2) = F(bZ_j - a) - F(bZ_{j-1} - a)$$

where $b = \sigma_n/\sigma_a$ and $a = (\mu_a - \mu_n)/\sigma_a$.

The parameters a, b and the $k-1$ fixed boundaries Z_j are simultaneously estimated by maximizing the log likelihood function

$$\log L = \sum_{i=1}^{2}\sum_{j=1}^{k} r_{ij}\log\{p(R_j|i)\}$$

where r_{ij} is the number of R_j's in group i.

The area under the fitted ROC curve is computed as

$$\Phi\left(\frac{a}{\sqrt{1+b^2}}\right)$$

where Φ is the standard normal cumulative distribution function.

Point estimates for the ROC curve indices are as follows:

$$\delta(m) = \frac{a}{b} \qquad d_e = \frac{2a}{b+1} \qquad d_a = \frac{a\sqrt{2}}{\sqrt{1+b^2}}$$

Variances for these indices are computed using the delta method.

The $\delta(m)$ estimates $(\mu_a - \mu_n)/\sigma_n$, d_e estimates $2(\mu_a - \mu_n)/(\sigma_a - \sigma_n)$, and d_a estimates $\sqrt{2}(\mu_a - \mu_n)/(\sigma_a^2 - \sigma_n^2)^2$.

Simultaneous confidence bands for the entire curve are obtained as suggested by Ma and Hall (1993) by first obtaining Working–Hotelling (1929) confidence bands for the fitted straight line in normal probability coordinates, and then transforming them back to ROC coordinates.

Comparing areas under the ROC curve

Areas under ROC curves are compared using an algorithm suggested by DeLong, DeLong, and Clarke-Pearson (1988). Let $\widehat{\theta} = (\widehat{\theta}^1, \widehat{\theta}^2, \dots, \widehat{\theta}^k)$ be a vector representing the areas under k ROC curves. For the rth area, define

$$V_{10}^r(X_i) = \frac{1}{N_n}\sum_{j=1}^{N_n}\psi(X_i^r, Y_j^r)$$

and for each normal subject, j, define

$$V_{01}^r(Y_j) = \frac{1}{N_a}\sum_{i=1}^{N_a}\psi(X_i^r, Y_j^r)$$

where

$$\psi(X^r, Y^r) = \begin{cases} 1 & Y^r < X^r \\ \frac{1}{2} & Y^r = X^r \\ 0 & Y^r > X^r \end{cases}$$

Define the $k \times k$ matrix $\mathbf{S}_{10}$ such that the (r,s)th element is

$$S_{10}^{r,s} = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} \{V_{10}^r(X_i) - \widehat{\theta}^r\}\{V_{10}^s(X_i) - \widehat{\theta}^s\}$$

and $\mathbf{S}_{01}$ such that the (r,s)th element is

$$S_{01}^{r,s} = \frac{1}{N_n - 1} \sum_{j=1}^{N_n} \{V_{01}^r(Y_i) - \widehat{\theta}^r\}\{V_{01}^s(Y_i) - \widehat{\theta}^s\}$$

Then the covariance matrix is

$$S = \frac{1}{N_a} S_{10} + \frac{1}{N_n} S_{01}$$

Let $\mathbf{L}$ be a contrast matrix defining the comparison, so that

$$(\widehat{\theta} - \theta)'\mathbf{L}'(\mathbf{LSL}')^{-1}\mathbf{L}(\widehat{\theta} - \theta)$$

has a chi-squared distribution with degrees of freedom equal to the rank of $\mathbf{LSL}'$.

References

Bamber, D. 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 12: 387–415.

Choi, B. C. K. 1998. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. *American Journal of Epidemiology* 148: 1127–1132.

Cleves, M. 1999. sg120: Receiver Operating Characteristic (ROC) analysis. *Stata Technical Bulletin* 52: 19–33. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 212–229.

——. 2000a. sg120.1: Two new options added to rocfit command. *Stata Technical Bulletin* 53: 18–19. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 230–231.

——. 2000b. sg120.2: Correction to roccomp command. *Stata Technical Bulletin* 54: 26. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, p. 231.

——. 2002a. Comparative assessment of three common algorithms for estimating the variance of the area under the nonparametric receiver operating characteristic curve. *The Stata Journal* 2: 280–289.

——. 2002b. From the help desk: Comparing areas under receiver operating characteristic curves from two or more probit or logit models. *The Stata Journal* 2: 301–313.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating curves: A nonparametric approach. *Biometrics* 44: 837–845.

Dorfman, D. D. and E. Alf. 1969. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. *Journal of Mathematical Psychology* 6: 487–496.

Erdreich, L. S. and E. T. Lee. 1981. Use of relative operating characteristic analysis in epidemiology: a method for dealing with subjective judgment. *American Journal of Epidemiology* 114: 649–662.

Hanley, J. A. and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 26–36.

——. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148: 839–843.

Lee, W. C. 1999. Probabilistic analysis of global performances of diagnostic test: Interpreting the Lorenz curve-based summary measures. *Statistics in Medicine* 18: 455–471.

Ma, G. and W. J. Hall. 1993. Confidence bands for the receiver operating characteristic curves. *Medical Decision Making* 13: 191–197.

Reichenheim, M. E. and A. Ponce de Leon. 2002. Estimation of sensitivity and specificity arising from validity studies with incomplete designs. *The Stata Journal* 2: 267–279.

Seed, P. T. and A. Tobias. 2001. sbe36.1: Summary statistics for diagnostic tests. *Stata Technical Bulletin* 59: 25–27. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 90–93.

Tobias, A. 2000. sbe36: Summary statistics report for diagnostic tests. *Stata Technical Bulletin* 56: 16–18. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 87–90.

Working, H. and H. Hotelling. 1929. Application of the theory of error to the interpretation of trends. *Journal of the American Statistical Association* 24: 73–85.

Also See

Related: [R] **logistic**