roccomp — Tests of equality of ROC areas

Description

`roccomp` and `rocgold` 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.

`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 *p*-value. Optionally, Šidák's adjustment for multiple comparisons can be obtained.

See [R] **rocfit** and [R] **rocreg** for commands that fit maximum-likelihood ROC models.

Quick start

Equality of AUCs for rating `v1` of true state `true` between samples defined by `catvar`

```
roccomp true v1, by(catvar)
```

Equality of AUCs for ratings `v1` and `v2` for the same sample

```
roccomp true v1 v2
```

Same as above, but plot ROC curves without reporting summary statistics and test of equality

```
roccomp true v1 v2, graph
```

Same as above, but plot `v1` with a dashed line and `v2` with a solid line

```
roccomp true v1 v2, graph plot1opts(lpattern(dash)) ///
   plot2opts(lpattern(solid))
```

Use contrast matrix `mymat` to compare ROC areas for `v1`, `v2`, `v3`, and `v4`

```
matrix mymat = (1,0,-1,0 \ 0,1,0,-1)
roccomp true v1 v2 v3 v4, test(mymat)
```

Test equality of ROC area for `v1` against a "gold standard" `gold`

```
rocgold true gold v1
```

Menu

roccomp

Statistics > Epidemiology and related > ROC analysis > Test equality of two or more ROC areas

rocgold

Statistics > Epidemiology and related > ROC analysis > Test equality of ROC area against gold standard

Syntax

Test equality of ROC areas

`roccomp` *refvar classvar* [*classvars*] [*if*] [*in*] [*weight*] [, *roccomp_options*]

Test equality of ROC area against a standard ROC curve

`rocgold` *refvar goldvar classvar* [*classvars*] [*if*] [*in*] [*weight*] [, *rocgold_options*]

roccomp_options	Description
Main	
`by(`*varname*`)`	split into groups by variable
`test(`*matname*`)`	use contrast matrix for comparing ROC areas
`graph`	graph the ROC curve
`norefline`	suppress plotting the 45-degree reference line
`separate`	place each ROC curve on its own graph
`summary`	report the area under the ROC curve
`binormal`	estimate areas by using binormal distribution assumption
`line#opts(`*cline_options*`)`	affect rendition of the #th binormal fit line
`level(#)`	set confidence level; default is `level(95)`
Plot	
`plot#opts(`*plot_options*`)`	affect rendition of the #th ROC curve
Reference line	
`rlopts(`*cline_options*`)`	affect rendition of the reference line
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***

rocgold_options	Description
Main	
`sidak`	adjust the p-value by using Šidák's method
`test(`*matname*`)`	use contrast matrix for comparing ROC areas
`graph`	graph the ROC curve
`norefline`	suppress plotting the 45-degree reference line
`separate`	place each ROC curve on its own graph
`summary`	report the area under the ROC curve
`binormal`	estimate areas by using binormal distribution assumption
`line#opts(`*cline_options*`)`	affect rendition of the #th binormal fit line
`level(#)`	set confidence level; default is `level(95)`
Plot	
`plot#opts(`*plot_options*`)`	affect rendition of the #th ROC curve; plot 1 is the "gold standard"
Reference line	
`rlopts(`*cline_options*`)`	affect rendition of the reference line
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than `by()` documented in [G-3] ***twoway_options***

plot_options	Description
marker_options	change look of markers (color, size, etc.)
marker_label_options	add marker labels; change look or position
cline_options	change look of the line

`collect` is allowed with `roccomp` and `rocgold`; see [U] **11.1.10 Prefix commands**.

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

Options

Main

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

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

`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.

`graph` produces graphical output of the ROC curve.

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

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

`summary` reports the area under the ROC curve, its standard error, and its confidence interval. This option is needed only when also specifying `graph`.

`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.

`line#opts(`*cline_options*`)` affect the rendition of the line representing the #th ROC curve drawn using the binormal distribution assumption; see [G-3] ***cline_options***. These lines are drawn only if the `binormal` option is specified.

`level(#)` specifies the confidence level, as a percentage, for the confidence intervals. The default is `level(95)` or as set by `set level`; see [R] **level**.

Plot

`plot#opts(`*plot_options*`)` affect the rendition of the #th ROC curve—the curve's plotted points connected by lines. The *plot_options* can affect the size and color of markers, whether and how the markers are labeled, and whether and how the points are connected; see [G-3] ***marker_options***, [G-3] ***marker_label_options***, and [G-3] ***cline_options***.

For `rocgold`, `plot1opts()` are applied to the ROC for the gold standard.

Reference line

`rlopts(`*cline_options*`)` affects the rendition of the reference line; see [G-3] ***cline_options***.

Y axis, X axis, Titles, Legend, Overall

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

Remarks and examples

Remarks are presented under the following headings:

Introduction
Comparing areas under the ROC curve
Correlated data
Independent data
Comparing areas with a gold standard

Introduction

`roccomp` provides comparison of the ROC curves of multiple classifiers. `rocgold` compares the ROC curves of multiple classifiers with a single "gold standard" classifier. Adjustment of inference for multiple comparisons is also provided by `rocgold`.

See Pepe (2003) for a discussion of ROC analysis. Pepe has posted Stata datasets and programs used to reproduce results presented in the book (https://www.stata.com/bookstore/pepe.html).

Comparing areas under the ROC curve

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

Correlated data

▷ Example 1

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 were 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.

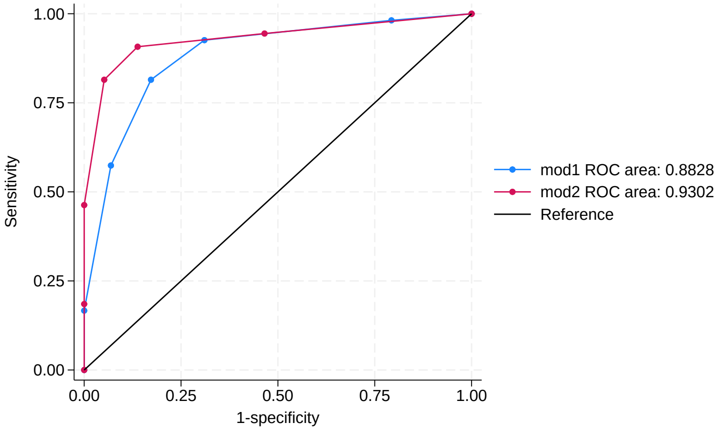
We list the first 7 observations:

```
. use https://www.stata-press.com/data/r19/ct
(Reconstruction of CT images)

. 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  |
     +-----------------------+
```

The data are in wide form, which 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 they are normal and `status=1` if they are 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
--------------------------------------------------------------------------------
H0: area(mod1) = area(mod2)
    chi2(1) =      2.31       Prob>chi2 =   0.1282
```



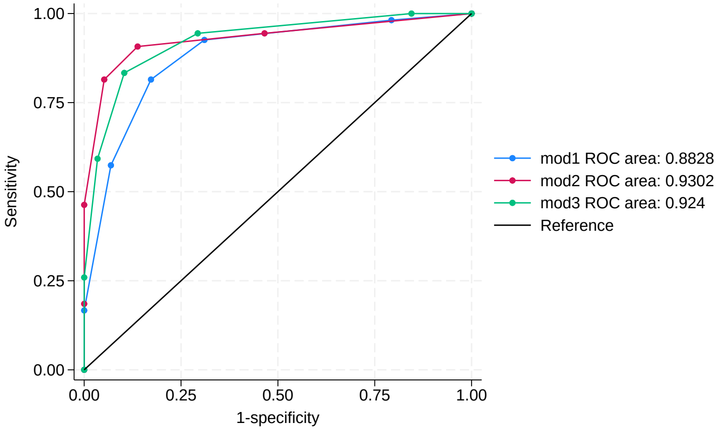
By default, `roccomp`, with the `graph` option specified, plots the ROC curves on the same graph. Optionally, the curves can be plotted side by side, each on its own graph, by also specifying `separate`.

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 χ^2 test yielded a p-value 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 https://www.stata-press.com/data/r19/ct2
(Reconstruction of CT images)

. 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
─────────────────────────────────────────────────────────────────────────────

H0: 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* (that is, the total number of ROC curves) and a number of rows less than or equal to the number of *classvars*, and the elements of each row must add to zero.

```
. 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
------------------------------------------------------------------------
H0: Comparison as defined by contrast matrix: C
    chi2(1) =     5.25       Prob>chi2 =   0.0220
```

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
------------------------------------------------------------------------
H0: 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)

. mat list C

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

◁

Independent data

▷ Example 2

In example 1, 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 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 https://www.stata-press.com/data/r19/ct3
(Reconstruction of CT images)

. 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    |
     +--------------------------------+
```

The data are in long form, which 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.

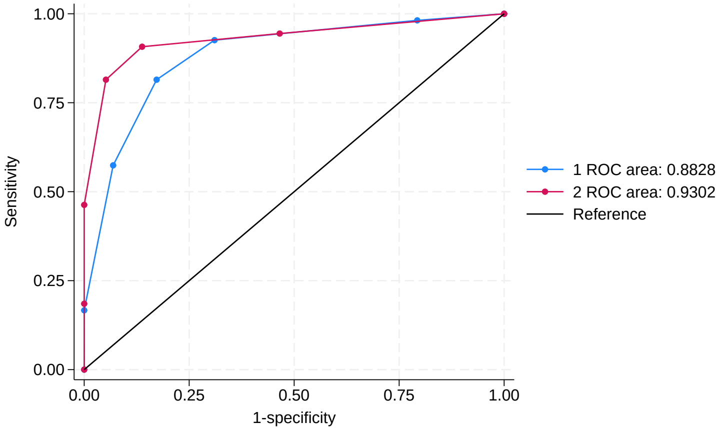
We can better view our data by using the `table` command.

```
. table (mod status) (rating) [fw=pop], totals(mod mod#status mod#rating)

--------------------------------------------------
              |               Rating
              |   1    2    3    4    5    6  Total
--------------+-----------------------------------
Modality      |
  1           |
    Status    |
      0       |  12   28    8    6    4          58
      1       |   1    3    6   13   22    9     54
      Total   |  13   31   14   19   26    9    112
  2           |
    Status    |
      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 they are normal and `status` $=1$ if they are abnormal.

We now compare the areas under the two ROC curves.

```
. 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
─────────────────────────────────────────────────────────────────────────

H0: area(1) = area(2)
    chi2(1) =       1.35       Prob>chi2 =   0.2447
```



◁

Comparing areas with a gold standard

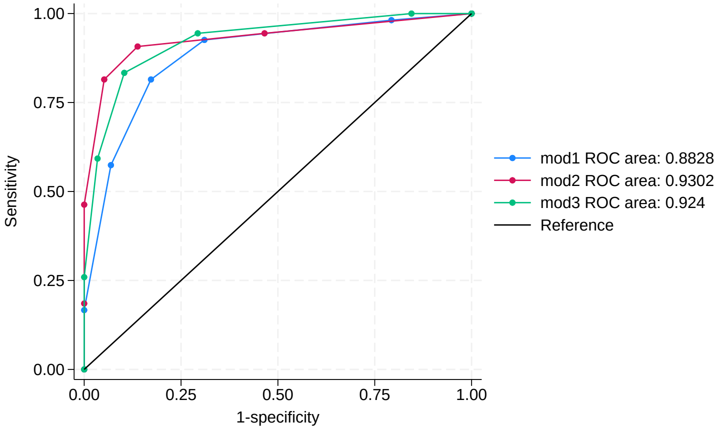
The area under multiple ROC curves can be compared with a gold standard using `rocgold`. The command syntax is similar to that of `roccomp`. The tests are corrected for the effect of multiple comparisons.

▷ Example 3

We will use the same data (presented by Hanley and McNeil [1983]) as in the `roccomp` examples. Let's assume that the first modality is considered to be the standard against 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`. Because we consider `mod1` to be the gold standard, it is listed first after the reference variable in the `rocgold` command line.

```
. use https://www.stata-press.com/data/r19/ct2
(Reconstruction of CT images)

. 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
```

◁

Stored results

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

Scalars

`r(N_g)`	number of groups	`r(df)`	χ^2 degrees of freedom
`r(p)`	p-value for χ^2 test	`r(chi2)`	χ^2

Matrices

`r(V)`	variance–covariance matrix

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

Scalars

`r(N_g)`	number of groups

Matrices

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

Methods and formulas

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 $X_i, i = 1, 2, \dots, N_a$ and $Y_j, j = 1, 2, \dots, N_n$ be the values of the diagnostic test for the abnormal and normal subjects, respectively.

Areas under ROC curves are compared using an algorithm suggested by DeLong, DeLong, and Clarke-Pearson (1988). Let $\widehat{\boldsymbol{\theta}} = (\widehat{\theta^1}, \widehat{\theta^2}, \dots, \widehat{\theta^k})$ be a vector representing the areas under k ROC curves. See *Methods and formulas* in [R] **roctab** for the definition of these area estimates.

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

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

has a χ^2 distribution with degrees of freedom equal to the rank of $\mathbf{LSL}'$.

References

Cleves, M. A. 2002a. Comparative assessment of three common algorithms for estimating the variance of the area under the nonparametric receiver operating characteristic curve. *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. *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 characteristic curves: A nonparametric approach. *Biometrics* 44: 837–845. https://doi.org/10.2307/2531595.

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. https://doi.org/10.1093/oxfordjournals.aje.a113236.

Hanley, J. A., and B. J. McNeil. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148: 839–843. https://doi.org/10.1148/radiology.148.3.6878708.

Harbord, R. M., and P. Whiting. 2009. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal* 9: 211–229.

Juul, S., and M. Frydenberg. 2021. *An Introduction to Stata for Health Researchers*. 5th ed. College Station, TX: Stata Press.

Ma, G., and W. J. Hall. 1993. Confidence bands for the receiver operating characteristic curves. *Medical Decision Making* 13: 191–197. https://doi.org/10.1177/0272989X9301300304.

Pepe, M. S. 2003. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press.

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

Working, H., and H. Hotelling. 1929. Application of the theory of error to the interpretation of trends. *Journal of the American Statistical Association* 24 (Suppl.): 73–85. https://doi.org/10.2307/2277011.

Also see

[R] **logistic postestimation** — Postestimation tools for logistic

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

[R] **rocfit** — Parametric ROC models

[R] **rocreg** — Parametric and nonparametric ROC regression

[R] **roctab** — Nonparametric ROC analysis

For suggested citations, see the FAQ on citing Stata documentation.