

roctab — Nonparametric ROC analysis

Description	Quick start	Menu	Syntax
Options	Remarks and examples	Stored results	Methods and formulas
References	Also see		

Description

The above command is 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` performs 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.

See [\[R\] rocfits](#) for a command that fits maximum-likelihood ROC models.

Quick start

AUC, standard error, and CI for ROC curve of binary true state, `true`, as a function of classification variable `v`

```
roctab true v
```

Also plot the ROC curve

```
roctab true v, graph summary
```

As above, but plot sensitivity versus specificity instead of the ROC curve

```
roctab true v, specificity summary
```

Exact binomial CI instead of asymptotic normal CI for AUC estimate

```
roctab true v, binomial
```

Gini and Pietra indices

```
roctab true v, lorenz
```

Menu

Statistics > Epidemiology and related > ROC analysis > Nonparametric ROC analysis without covariates

Syntax

```
roctab refvar classvar [if] [in] [weight] [, options]
```

<i>options</i>	Description
Main	
<u>l</u> orenz	report Gini and Pietra indices
<u>b</u> inomial	calculate exact binomial confidence intervals
<u>n</u> olabel	display numeric codes rather than value labels
<u>d</u> etail	show details on sensitivity/specificity for each cutpoint
<u>t</u> able	display the raw data in a $2 \times k$ contingency table
<u>b</u> amber	calculate standard errors by using the Bamber method
<u>h</u> anley	calculate standard errors by using the Hanley method
<u>g</u> raph	graph the ROC curve
<u>n</u> oreflne	suppress plotting the 45-degree reference line
<u>s</u> ummary	report the area under the ROC curve
<u>s</u> pecificity	graph sensitivity versus specificity
<u>l</u> evel(#)	set confidence level; default is level(95)
Plot	
<u>p</u> lotopts(<i>plot_options</i>)	affect rendition of the ROC curve
Reference line	
<u>r</u> lopts(<i>cline_options</i>)	affect rendition of the reference line
Add plots	
<u>a</u> ddplot(<i>plot</i>)	add other plots to the generated graph
Y axis, X axis, Titles, Legend, Overall	
<i>twoway_options</i>	any options other than by() documented in [G-3] <i>twoway_options</i>
collect is allowed; see [U] 11.1.10 Prefix commands.	
fweights are allowed; see [U] 11.1.6 weight.	

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

Options

Main

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

binomial specifies that exact binomial confidence intervals be calculated.

no label specifies that numeric codes be displayed rather than value labels.

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

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

`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 \(1988\)](#).

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

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

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

`specificity` produces a graph of sensitivity versus specificity instead of sensitivity versus $(1 - \text{specificity})$. `specificity` implies `graph`.

`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

`plotopts(plot_options)` affects the rendition of the plotted 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](#).

Reference line

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

Add plots

`addplot(plot)` provides a way to add other plots to the generated graph; see [\[G-3\] addplot_option](#).

Y axis, X axis, Titles, Legend, Overall

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

Remarks and examples

stata.com

Remarks are presented under the following headings:

[Introduction](#)
[Nonparametric ROC curves](#)
[Lorenz-like curves](#)

Introduction

The `roctab` command provides nonparametric estimation of the ROC for a given classifier and true-status reference variable. The Lorenz curve functionality of `roctab`, which provides an alternative to standard ROC analysis, is discussed in *Lorenz-like curves*.

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>).

Nonparametric ROC curves

The points on the nonparametric ROC curve are generated using each possible outcome of the diagnostic test as a classification cutpoint 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 1

[Hanley and McNeil \(1982\)](#) presented data from a study in which a reviewer was asked to classify, using a five-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:

```
. use https://www.stata-press.com/data/r17/hanley
(Tomographic images)
. list disease rating in 1/9
```

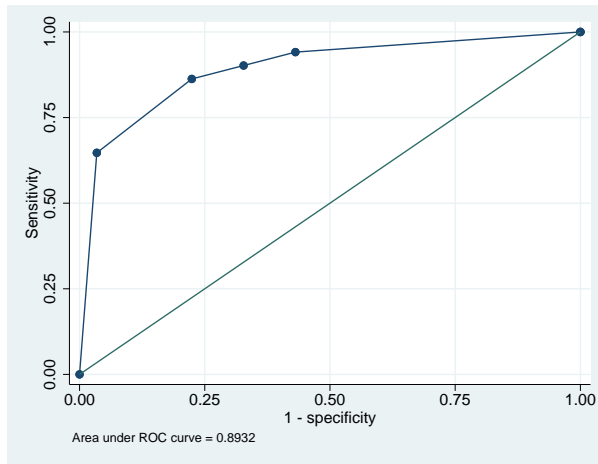
	disease	rating
1.	1	5
2.	0	1
3.	1	5
4.	0	4
5.	0	1
6.	0	3
7.	1	5
8.	0	5
9.	0	1

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
```

True disease status of subject	rating					Total
	1	2	3	4	5	
0	33	6	6	11	2	58
1	3	2	2	11	33	51
Total	36	8	8	22	35	109
Obs	ROC area	Std. err.	Asymptotic normal [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 cutpoint. A point is plotted on the graph for each of the cutpoints. 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 cutpoints by specifying `detail`.

```
. roctab disease rating, detail
Detailed report of sensitivity and specificity
```

Cutpoint	Sensitivity	Specificity	Correctly 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

Obs	ROC area	Std. err.	Asymptotic normal [95% conf. interval]	
109	0.8932	0.0307	0.83295	0.95339

Each cutpoint in the table indicates the ratings used to classify tomographs as being from an abnormal subject. For example, the first cutpoint (≥ 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 is classified correctly (specificity = 0%). For the second cutpoint (≥ 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 cutpoint, we correctly classified 74.31% of the 109 tomographs. Similar interpretations can be used on the remaining cutpoints. As mentioned, each cutpoint corresponds to a point on the nonparametric ROC curve. The first cutpoint (≥ 1) corresponds to the point at (1,1), and the last cutpoint (> 5) corresponds 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 LR+ 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 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 cutpoint. Similarly, LR- corresponds to the slope from the point (1,1) to the point on the ROC curve determined by the cutpoint.

By default, `roctab` calculates the standard error for the area under the curve by 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
      Obs      ROC      Bamber      Asymptotic normal
      area      std. err.      [95% conf. interval]
-----
      109      0.8932      0.0306      0.83317      0.95317
. roctab disease rating, hanley binomial
      Obs      ROC      Hanley      Binomial exact
      area      std. err.      [95% conf. interval]
-----
      109      0.8932      0.0320      0.81559      0.94180

```

◀

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 can be 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 2

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.

```

. use https://www.stata-press.com/data/r17/lorenz, clear
. list in 1/7, noobs sep(0)

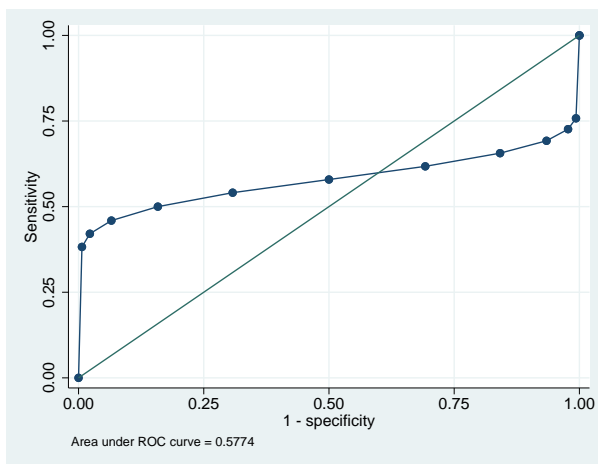
```

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 five times greater than that of the normal population.

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

Obs	ROC area	Std. err.	Asymptotic normal [95% conf. interval]	
884	0.5774	0.0215	0.53517	0.61959

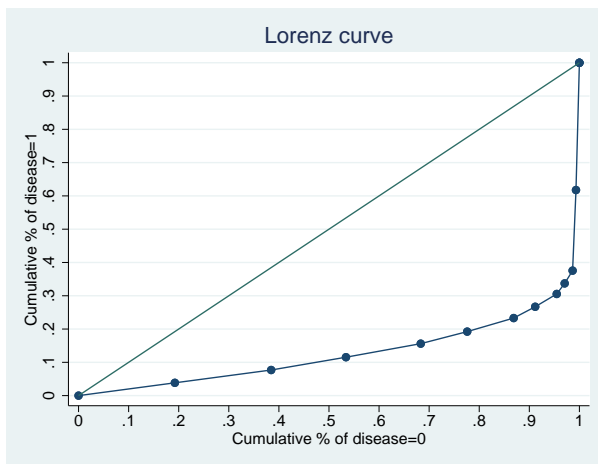


The resulting ROC curve is nonconvex or, as termed by Lee, “wiggly”. Lee argues that for this and similar situations, the Lorenz curve and indices are preferred.

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

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.



Stored results

roctab stores the following in `r()`:

Scalars

<code>r(N)</code>	number of observations
<code>r(se)</code>	standard error for the area under the ROC curve
<code>r(lb)</code>	lower bound of CI for the area under the ROC curve
<code>r(ub)</code>	upper bound of CI for the area under the ROC curve
<code>r(level)</code>	confidence level
<code>r(area)</code>	area under the ROC curve
<code>r(pietra)</code>	Pietra index
<code>r(gini)</code>	Gini index

Macros

<code>r(cutpoints)</code>	description of cutpoints (<code>detail</code> only)
---------------------------	--

Matrices

<code>r(detail)</code>	matrix with details on sensitivity and specificity for each cutpoint (<code>detail</code> only)
------------------------	--

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 $\hat{\theta}$ 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.

The points on the nonparametric ROC curve are generated using each possible outcome of the diagnostic test as a classification cutpoint 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

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

and

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

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

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

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 is $\Pr(\text{two randomly selected abnormal subjects will both have a higher score than a randomly selected normal subject})$, and Q_2 is $\Pr(\text{one randomly selected abnormal subject will have a higher score than any two randomly selected normal subjects})$. The Hanley and McNeil variance of the estimated area under the ROC curve is

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

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

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

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

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

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.

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. [https://doi.org/10.1016/0022-2496\(75\)90001-2](https://doi.org/10.1016/0022-2496(75)90001-2).
- 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. <https://doi.org/10.1093/oxfordjournals.aje.a009592>.

- 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. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>.
- Harbord, R. M., and P. Whiting. 2009. *metandi*: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal* 9: 211–229.
- Hong, L., G. Alfani, C. Gigliarano, and M. Bonneti. 2018. *giniinc*: A Stata package for measuring inequality from incomplete income and survival data. *Stata Journal* 18: 692–715.
- Juul, S., and M. Frydenberg. 2014. *An Introduction to Stata for Health Researchers*. 4th ed. College Station, TX: Stata Press.
- Lee, W. C. 1999. Probabilistic analysis of global performances of diagnostic tests: Interpreting the Lorenz curve-based summary measures. *Statistics in Medicine* 18: 455–471. [https://doi.org/10.1002/\(sici\)1097-0258\(19990228\)18:4<455::aid-sim44>3.0.co;2-a](https://doi.org/10.1002/(sici)1097-0258(19990228)18:4<455::aid-sim44>3.0.co;2-a).
- 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>.
- Pacifico, D., and F. Poege. 2017. Estimating measures of multidimensional poverty with Stata. *Stata Journal* 17: 687–703.
- 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] **roccomp** — Tests of equality of ROC areas
- [R] **rocfitt** — Parametric ROC models
- [R] **rocreg** — Receiver operating characteristic (ROC) regression