Chapter 547

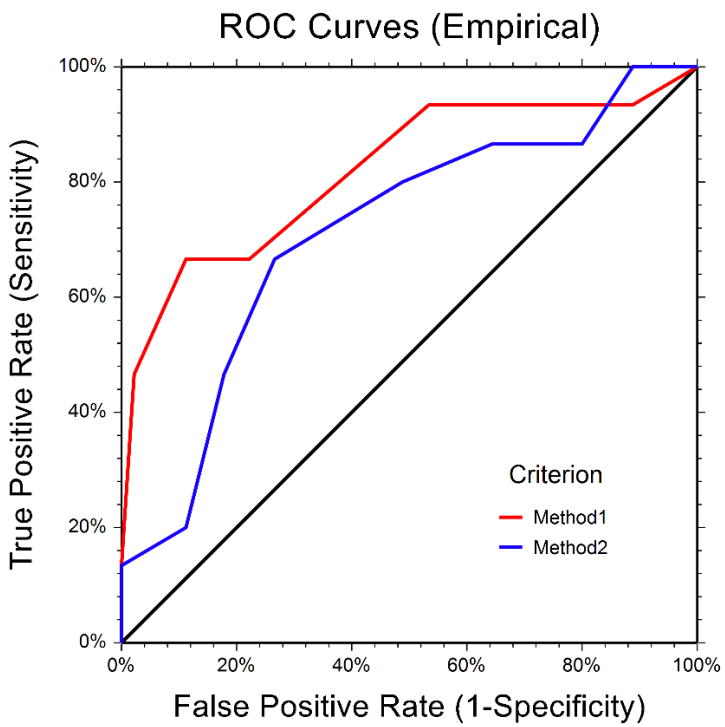
Comparing Two ROC Curves – Paired Design

Introduction

This procedure is used to compare two ROC curves for the paired sample case wherein each subject has a known condition value and test values (or scores) from two diagnostic tests. The test values are paired because they are measured on the same subject.

In addition to producing a wide range of cutoff value summary rates for each criterion, this procedure produces difference tests, equivalence tests, non-inferiority tests, and confidence intervals for the difference in the area under the ROC curve.

This procedure includes analyses for both empirical (nonparametric) and Binormal ROC curve estimation.



© NCSS, LLC. All Rights Reserved.

Discussion and Technical Details

Although ROC curve analysis can be used for a variety of applications across a number of research fields, we will examine ROC curves through the lens of diagnostic testing. In a typical diagnostic test, each unit (e.g., individual or patient) is measured on some scale or given a score with the intent that the measurement or score will be useful in classifying the unit into one of two conditions (e.g., Positive / Negative, Yes / No, Diseased / Non-diseased). Based on a (hopefully large) number of individuals for which the score and condition is known, researchers may use ROC curve analysis to determine the ability of the score to classify or predict the condition. When two diagnostic tests are administered (measured) for each subject, the resulting information can be used to compare the ability of the two diagnostic tests to classify the condition.

ROC Curve and Cutoff Analysis for each Diagnostic Test

The details of the many summary measures and rates for each cutoff value are discussed in the chapter One ROC Curve and Cutoff Analysis. We invite the reader to go to that chapter for details on classification tables, as well as true positive rate (sensitivity), true negative rate (specificity), false negative rate (miss rate), false positive rate (fall-out), positive predictive value (precision), negative predictive value, false omission rate, false discovery rate, prevalence, proportion correctly classified (accuracy), proportion incorrectly classified, Youden index, sensitivity plus specificity, distance to corner, positive likelihood ratio, negative likelihood ratio, diagnostic odds ratio, and cost analysis for each cutoff value.

The One ROC Curve and Cutoff Analysis chapter also contains details about finding the optimal cutoff value, as well as hypothesis tests and confidence intervals for individual areas under the ROC curve.

ROC Curves

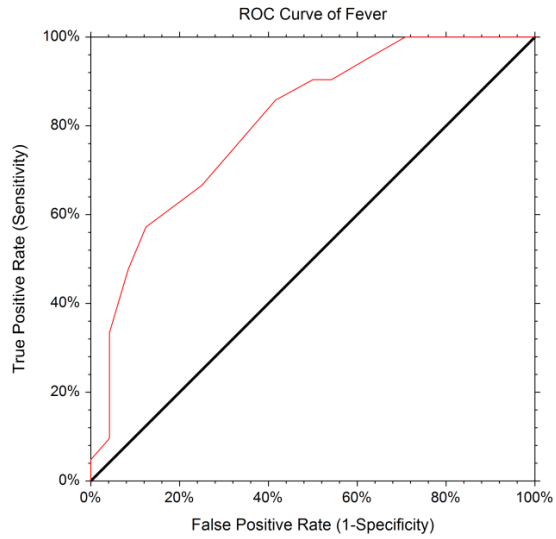
A receiver operating characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1 – specificity) for *all possible* cutoff values. General discussions of ROC curves can be found in Altman (1991), Swets (1996), Zhou et al. (2002), and Krzanowski and Hand (2009). Gehlbach (1988) provides an example of its use.

Two types of ROC curves can be generated in **NCSS**: the empirical ROC curve and the binormal ROC curve.

© NCSS, LLC. All Rights Reserved.

Empirical ROC Curve

The empirical ROC curve is the more common version of the ROC curve. The empirical ROC curve is a plot of the true positive rate versus the false positive rate for all possible cut-off values.

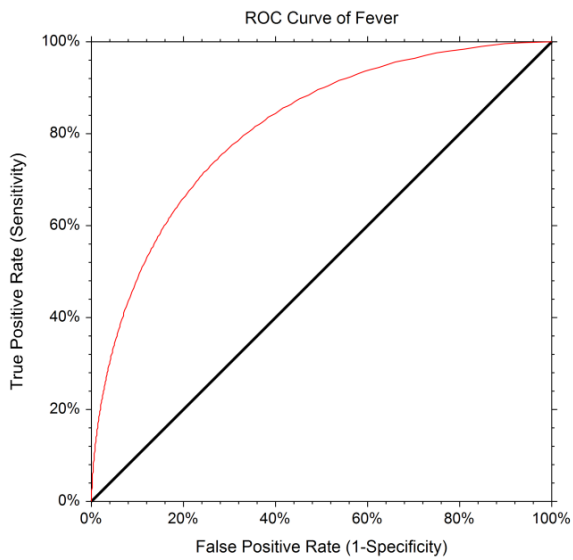


That is, each point on the ROC curve represents a different cutoff value. The points are connected to form the curve. Cutoff values that result in low false-positive rates tend to result in low true-positive rates as well. As the true-positive rate increases, the false positive rate increases. The better the diagnostic test, the more quickly the true positive rate nears 1 (or 100%). A near-perfect diagnostic test would have an ROC curve that is almost vertical from (0,0) to (0,1) and then horizontal to (1,1). The diagonal line serves as a reference line since it is the ROC curve of a diagnostic test that randomly classifies the condition.

© NCSS, LLC. All Rights Reserved.

Binormal ROC Curve

The Binormal ROC curve is based on the assumption that the diagnostic test scores corresponding to the positive condition and the scores corresponding to the negative condition can each be represented by a Normal distribution. To estimate the Binormal ROC curve, the sample mean and sample standard deviation are estimated from the known positive group, and again for the known negative group. These sample means and sample standard deviations are used to specify two Normal distributions. The Binormal ROC curve is then generated from the two Normal distributions. When the two Normal distributions closely overlap, the Binormal ROC curve is closer to the 45-degree diagonal line. When the two Normal distributions overlap only in the tails, the Binormal ROC curve has a much greater distance from the 45-degree diagonal line.



It is recommended that researchers identify whether the scores for the positive and negative groups need to be transformed to more closely follow the Normal distribution before using the Binormal ROC Curve methods.

Area under the ROC Curve (AUC)

The area under an ROC curve (AUC) is a popular measure of the accuracy of a diagnostic test. In general, higher AUC values indicate better test performance. The possible values of AUC range from 0.5 (no diagnostic ability) to 1.0 (perfect diagnostic ability).

The AUC has a physical interpretation. The AUC is the probability that the criterion value of an individual drawn at random from the population of those with a positive condition is larger than the criterion value of another individual drawn at random from the population of those where the condition is negative.

Another interpretation of AUC is the average true positive rate (average sensitivity) across all possible false positive rates.

Two methods are commonly used to estimate the AUC. One method is the empirical (nonparametric) method by DeLong et al. (1988). This method has become popular because it does not make the strong normality assumptions that the Binormal method makes. The other method is the Binormal method presented by Metz (1978) and McClish (1989). This method results in a smooth ROC curve from which the complete (and partial) AUC may be calculated.

© NCSS, LLC. All Rights Reserved.

AUC of an Empirical ROC Curve

The empirical (nonparametric) method by DeLong et al. (1988) is a popular method for computing the AUC. This method has become popular because it does not make the strong Normality assumptions that the Binormal method makes.

The value of AUC using the empirical method is calculated by summing the area of the trapezoids that are formed below the connected points making up the ROC curve. From DeLong et al. (1988), define the T_1 component of the ith subject, $V(T_{1i})$ as

$$V(T_{1i}) = \frac{1}{n_0}\sum_{j=1}^{n_0}\Psi\left(T_{1i}, T_{0j}\right)$$

and define the T_0 component of the jth subject, $V(T_{0j})$ as

$$V\left(T_{0j}\right) = \frac{1}{n_1}\sum_{i=1}^{n_1}\Psi\left(T_{1i}, T_{0j}\right)$$

where

$\Psi(X,Y) = 0$ if $Y > X$,

$\Psi(X,Y) = 1/2$ if $Y = X$,

$\Psi(X,Y) = 1$ if $Y < X$

The empirical AUC is estimated as

$$A_{Emp} = \sum_{i=1}^{n_1} V(T_{1i})/n_1 = \sum_{j=1}^{n_0} V\left(T_{0j}\right)/n_0$$

The variance of the estimated AUC is estimated as

$$V\left(A_{Emp}\right) = \frac{1}{n_1}S_{T_1}^2 + \frac{1}{n_0}S_{T_0}^2$$

where $S_{T_1}^2$ and $S_{T_0}^2$ are the variances

$$S_{T_i}^2 = \frac{1}{n_i - 1}\sum_{i=1}^{n_i}\left[V(T_{1i}) - A_{Emp}\right]^2, \qquad i = 0,1$$

© NCSS, LLC. All Rights Reserved.

AUC of a Binormal ROC Curve

The formulas that we use here come from McClish (1989). Suppose there are two populations, one made up of individuals with the condition being positive and the other made up of individuals with the negative condition. Further, suppose that the value of a criterion variable is available for all individuals. Let *X* refer to the value of the criterion variable in the negative population and *Y* refer to the value of the criterion variable in the positive population. The binormal model assumes that both *X* and *Y* are normally distributed with different means and variances. That is,

$$X \sim N(\mu_x, \sigma_x^2),\ Y \sim N\left(\mu_y, \sigma_y^2\right)$$

The ROC curve is traced out by the function

$$\{FP(c), TP(c)\} = \left\{\Phi\left(\frac{\mu_x - c}{\sigma_x}\right), \Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\right\}, \quad -\infty < c < \infty$$

where $\Phi(z)$ is the cumulative normal distribution function.

The area under the whole ROC curve is

$$\begin{aligned} A &= \int_{-\infty}^{\infty} TP(c)FP'(c)\, dc \\ &= \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\Phi\left(\frac{\mu_x - c}{\sigma_x}\right)\right] dc \\ &= \Phi\left[\frac{a}{\sqrt{1 + b^2}}\right] \end{aligned}$$

where

$$a = \frac{\mu_y - \mu_x}{\sigma_y} = \frac{\Delta}{\sigma_y},\ b = \frac{\sigma_x}{\sigma_y},\ \Delta = \mu_y - \mu_x$$

The area under a portion of the AUC curve is given by

$$\begin{aligned} A &= \int_{c_1}^{c_2} TP(c)FP'(c)\, dc \\ &= \frac{1}{\sigma_x}\int_{c_2}^{c_1} \left[\Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\Phi\left(\frac{\mu_x - c}{\sigma_x}\right)\right] dc \end{aligned}$$

© NCSS, LLC. All Rights Reserved.

The partial area under an ROC curve is usually defined in terms of a range of false-positive rates rather than the criterion limits c_1 and c_2. However, the one-to-one relationship between these two quantities, given by

$$c_i = \mu_x + \sigma_x \Phi^{-1}(FP_i)$$

allows the criterion limits to be calculated from desired false-positive rates.

The MLE of *A* is found by substituting the MLE's of the means and variances into the above expression and using numerical integration. When the area under the whole curve is desired, these formulas reduce to

$$\hat{A} = \Phi\left[\frac{\hat{a}}{\sqrt{1+\hat{b}^2}}\right]$$

Note that for ease of reading we will often omit the use of the *hat* to indicate an MLE in the following.

The variance of $\hat{A}$ is derived using the method of differentials as

$$V(\hat{A}) = \left(\frac{\partial A}{\partial \Delta}\right)^2 V(\hat{\Delta}) + \left(\frac{\partial A}{\partial \sigma_x^2}\right)^2 V(s_x^2) + \left(\frac{\partial A}{\partial \sigma_y^2}\right)^2 V(s_y^2)$$

where

$$\frac{\partial A}{\partial \Delta} = \frac{E}{\sqrt{2\pi(1+b^2)\sigma_y^2}}[\Phi(\tilde{c}_1) - \Phi(\tilde{c}_0)]$$

$$\frac{\partial A}{\partial \sigma_x^2} = \frac{E}{4\pi(1+b^2)\sigma_x\sigma_y}[e^{-k_0} - e^{-k_1}] - \frac{abE}{2\sigma_x\sigma_y\sqrt{2\pi}(1+b^2)^{3/2}}[\Phi(\tilde{c}_1) - \Phi(\tilde{c}_0)]$$

$$E = \exp\left(-\frac{a^2}{2(1+b^2)}\right)$$

$$\frac{\partial A}{\partial \sigma_y^2} = -\frac{a}{2\sigma_y}\left(\frac{\partial A}{\partial \Delta}\right) - b^2\left(\frac{\partial A}{\partial \sigma_x^2}\right)$$

$$\tilde{c}_i = \left[\Phi^{-1}(FP_i) + \frac{ab}{(1+b^2)}\right]\sqrt{(1+b^2)}$$

$$k_i = \frac{\tilde{c}_i^2}{2}$$

$$\mathrm{V}(\hat{\Delta}) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

© NCSS, LLC. All Rights Reserved.

$$\mathrm{V}(s_x^2) = \frac{2\sigma_x^4}{n_x - 1}$$

$$\mathrm{V}(s_y^2) = \frac{2\sigma_y^4}{n_y - 1}$$

Comparing the AUC of Paired Data ROC Curves

Comparing ROC curves may be done using either the empirical (nonparametric) methods described by DeLong (1988) or the Binormal model methods as described in McClish (1989).

Comparing Paired Data AUCs based on Empirical ROC Curve Estimation

Following Zhou et al. (2002) page 185, a *z*-test may be used for comparing the AUC of two diagnostic tests in a paired design

$$z = \frac{A_1 - A_2}{\sqrt{V(A_1 - A_2)}}$$

where

$$\mathrm{V}(A_1 - A_2) = \mathrm{V}(A_1) + \mathrm{V}(A_2) - 2\mathrm{Cov}(A_1, A_2)$$

Each Variance is defined as

$$V(A_k) = \frac{S_{T_{k1}}}{n_{k1}} + \frac{S_{T_{k0}}}{n_{k0}}$$

where

$$S_{T_{ki}} = \frac{1}{n_{ki} - 1}\sum_{j=1}^{n_{ki}}\left[V(T_{kij}) - A_k\right]^2, \quad k = 1,2 \quad i = 0,1$$

$$\mathrm{V}(T_{k1i}) = \frac{1}{n_{k0} - 1}\sum_{j=1}^{n_{k0}}\Psi(T_{k1i}, T_{k0j}), \quad k = 1,2$$

$$V(T_{k0j}) = \frac{1}{n_{k1} - 1}\sum_{i=1}^{n_{k1}}\Psi(T_{k1i}, T_{k0j}), \quad k = 1,2$$

© NCSS, LLC. All Rights Reserved.

$$A_k = \frac{\sum_{i=1}^{n_{k1}} \mathrm{V}(T_{k1i})}{n_{k1}} = \frac{\sum_{j=1}^{n_{k0}} \mathrm{V}(T_{k0j})}{n_{k0}}, \quad k = 1,2$$

$$\Psi(X,Y) = \begin{cases} 0 & \text{if } Y > X \\ 1/2 & \text{if } Y = X \\ 1 & \text{if } Y < X \end{cases}$$

Here T_{k0_j} represents the observed diagnostic test result for the jth subject in group k without the condition and T_{k1_j} represents the observed diagnostic test result for the jth subject in group k with the condition.

$$\mathrm{Cov}(A_1, A_2) = \frac{S_{T_{11}T_{21}}}{n_1} + \frac{S_{T_{10}T_{20}}}{n_0}$$

$$S_{T_{11}T_{21}} = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} [\mathrm{V}(T_{11j}) - A_1][\mathrm{V}(T_{21j}) - A_2]$$

$$S_{T_{10}T_{20}} = \frac{1}{n_0 - 1} \sum_{j=1}^{n_1} [\mathrm{V}(T_{10j}) - A_1][\mathrm{V}(T_{20j}) - A_2]$$

Comparing Paired Data AUCs based on Binormal ROC Curve Estimation

When the binormal assumption is viable, the hypothesis that the areas under the two ROC curves are equal may be tested using

$$z = \frac{A_1 - A_2}{\sqrt{\mathrm{V}(A_1 - A_2)}}$$

where

$$\mathrm{V}(A_1 - A_2) = \mathrm{V}(A_1) + \mathrm{V}(A_2) - 2\mathrm{Cov}(A_1, A_2)$$

where $\mathrm{V}(A_1)$ and $\mathrm{V}(A_2)$ are calculated using the formula for $\mathrm{V}(A)$ given above in the section on a single Binormal ROC curve.

Since the data are paired, a covariance term must also be calculated. This is done using the differential method as follows

$$\mathrm{Cov}(A_1, A_2) = \left(\frac{\partial A_1}{\partial \Delta_1}\right)\left(\frac{\partial A_2}{\partial \Delta_2}\right)\mathrm{Cov}(\hat{\Delta}_1, \hat{\Delta}_2) + \left(\frac{\partial A_1}{\partial \sigma_{x_1}^2}\right)\left(\frac{\partial A_2}{\partial \sigma_{x_2}^2}\right)\mathrm{Cov}(s_{x_1}^2, s_{x_2}^2) + \left(\frac{\partial A_1}{\partial \sigma_{y_1}^2}\right)\left(\frac{\partial A_2}{\partial \sigma_{y_2}^2}\right)\mathrm{Cov}(s_{y_1}^2, s_{y_2}^2)$$

where

$$\mathrm{Cov}(\hat{\Delta}_1, \hat{\Delta}_2) = \frac{\rho_x \sigma_{x_1} \sigma_{x_2}}{n_x} + \frac{\rho_y \sigma_{y_1} \sigma_{y_2}}{n_y}$$

$$\mathrm{Cov}(s_{x_1}^2, s_{x_2}^2) = \frac{2\rho_x \sigma_{x_1}^2 \sigma_{x_2}^2}{n_x - 1}$$

© NCSS, LLC. All Rights Reserved.

$$\mathrm{Cov}\left(s_{y_1}^2, s_{y_2}^2\right) = \frac{2\rho_y \sigma_{y_1}^2 \sigma_{y_2}^2}{n_y - 1}$$

and $\rho_y(\rho_x)$ is the correlation between the two sets of criterion values in the diseased (non-diseased) population.

McClish (1989) ran simulations to study the accuracy of the normality approximation of the above *z* statistic for various portions of the AUC curve. She found that a logistic-type transformation resulted in a *z* statistic that was closer to normality. This transformation is

$$\theta(A) = \ln\left(\frac{FP_2 - FP_1 + A}{FP_2 - FP_1 - A}\right)$$

which has the inverse version

$$A = (FP_2 - FP_1)\frac{e^\theta - 1}{e^\theta + 1}$$

The variance of this quantity is given by

$$V(\theta) = \left(\frac{2(FP_2 - FP_1)}{(FP_2 - FP_1)^2 - A^2}\right)^2 V(A)$$

and the covariance is given by

$$\mathrm{Cov}(\theta_1, \theta_2) = \frac{4(FP_2 - FP_1)^2}{[(FP_2 - FP_1)^2 - A_1^2][(FP_2 - FP_1)^2 - A_2^2]}\mathrm{Cov}(A_1, A_2)$$

The adjusted *z* statistic is

$$\begin{aligned} z &= \frac{\theta_1 - \theta_2}{\sqrt{V(\theta_1 - \theta_2)}} \\ &= \frac{\theta_1 - \theta_2}{\sqrt{V(\theta_1) + V(\theta_2) - 2\mathrm{Cov}(\theta_1, \theta_2)}} \end{aligned}$$

© NCSS, LLC. All Rights Reserved.

Data Structure

The data are entered in three columns. One column specifies the true condition of the individual. The two other columns contain the criterion values for the tests being examined.

Paired Criteria Dataset (Subset)

Condition	Method1	Method2
Present	8	7
Absent	6	3
Absent	3	3
Absent	7	6
Absent	1	2
Present	5	9
Absent	3	3
Present	4	2
.	.	.
.	.	.
.	.	.

© NCSS, LLC. All Rights Reserved.

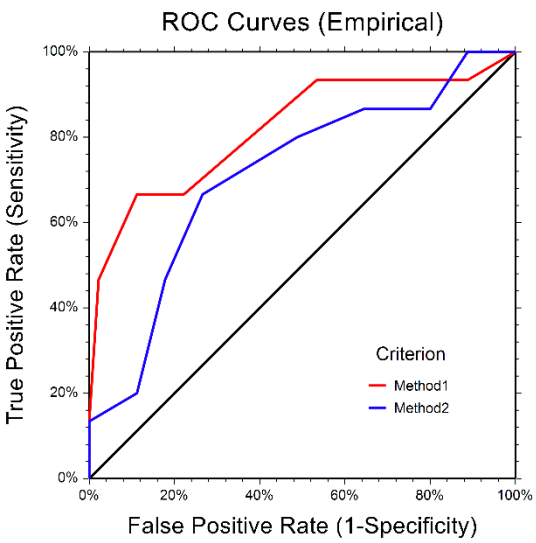
ROC Plot Format Window Options

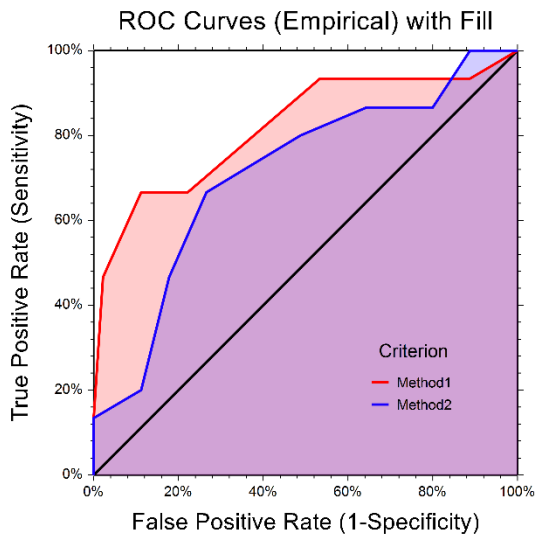
This section describes some of the options available on the ROC Plot Format window, which is displayed when the ROC Plot Chart Format button is clicked. Common options, such as axes, labels, legends, and titles are documented in the Graphics Components chapter.

ROC Plot Tab

Empirical ROC Line Section

You can specify the format of the empirical ROC curve lines using the options in this section.

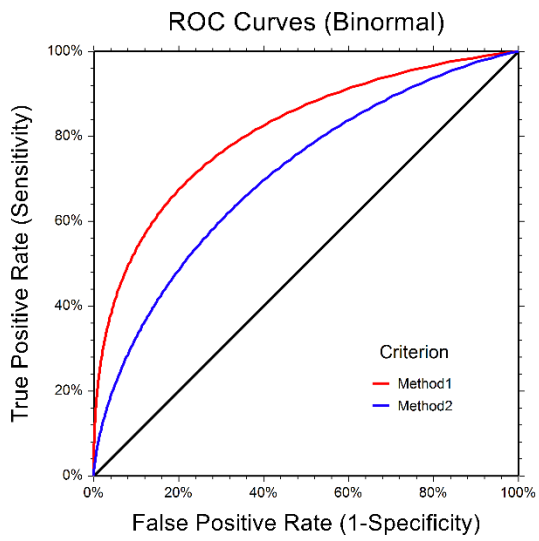


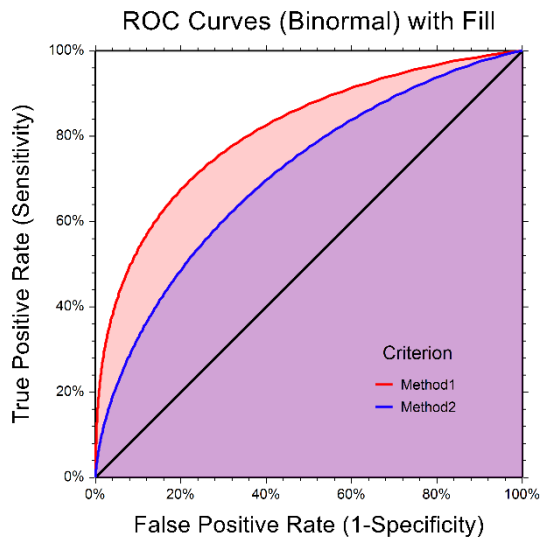


© NCSS, LLC. All Rights Reserved.

Binormal ROC Line Section

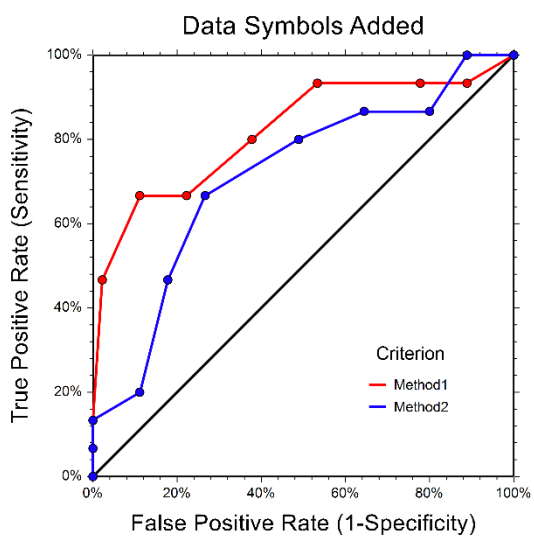
You can specify the format of the Binormal ROC curves lines using the options in this section.





Symbols Section

You can modify the attributes of the symbols using the options in this section.



Reference Line Section

You can modify the attributes of the 45° reference line using the options in this section.

Titles, Legend, Numeric Axis, Group Axis, Grid Lines, and Background Tabs

Details on setting the options in these tabs are given in the Graphics Components chapter.

© NCSS, LLC. All Rights Reserved.

Example 1 – Comparing Two Paired ROC Curves

This section presents an example of a producing a statistical comparison of areas under the ROC curves using a *Z*-test. The paired classification data are given in the Paired Criteria dataset. It is anticipated that higher Score values are associated with the condition being present.

Setup

To run this example, complete the following steps:

1 Open the Paired Criteria example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Paired Criteria** and click **OK**.

2 Specify the Comparing Two ROC Curves – Paired Design procedure options

- Find and open the **Comparing Two ROC Curves – Paired Design** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Condition Variable .....**Condition**
Positive Condition Value.....**Present**
Criterion Variables **Method1, Method2**
Criterion Direction.....**Higher values indicate a Positive Condition**

Cutoff Tab

Cutoff Value List .....**Data**
Counts, TPR (Sensitivity),.....**Checked**
TNR (Specificity), PPV, Accuracy,
TPR + TNR, Prevalence
All Other Reports .....**Unchecked**

AUC Reports Tab

Area Under Curve (AUC) Analysis.....**Checked**
(Empirical Estimation)
Test Comparing Two AUCs (Empirical .....**Checked**
Estimation)
Confidence Intervals for Comparing Two.....**Checked**
AUCs (Empirical Estimation)

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

© NCSS, LLC. All Rights Reserved.

Common Rates and Indices for Each Cutoff Value

Common Rates and Indices for Each Cutoff Value

Criterion Variable: Method1

Estimated Prevalence = 15 / 60 = 0.25
Estimated Prevalence is the proportion of the sample with a positive condition of PRESENT, or (A + C) / (A + B + C + D) for all cutoff values. The estimated prevalence should only be used as a valid estimate of the population prevalence when the entire sample is a random sample of the population.

	Table Counts								
Cutoff Value	**TPs A**	**FPs B**	**FNs C**	**TNs D**	**TPR (Sensitivity)**	**TNR (Specificity)**	**PPV**	**Accuracy**	**TPR + TNR**
≥ 1	15	45	0	0	1.0000	0.0000	0.2500	0.2500	1.0000
≥ 2	14	40	1	5	0.9333	0.1111	0.2593	0.3167	1.0444
≥ 3	14	35	1	10	0.9333	0.2222	0.2857	0.4000	1.1556
≥ 4	14	24	1	21	0.9333	0.4667	0.3684	0.5833	1.4000
≥ 5	12	17	3	28	0.8000	0.6222	0.4138	0.6667	1.4222
≥ 6	10	10	5	35	0.6667	0.7778	0.5000	0.7500	1.4444
≥ 7	10	5	5	40	0.6667	0.8889	0.6667	0.8333	1.5556
≥ 8	7	1	8	44	0.4667	0.9778	0.8750	0.8500	1.4444
≥ 9	2	0	13	45	0.1333	1.0000	1.0000	0.7833	1.1333

Criterion Variable: Method2

Estimated Prevalence = 15 / 60 = 0.25
Estimated Prevalence is the proportion of the sample with a positive condition of PRESENT, or (A + C) / (A + B + C + D) for all cutoff values. The estimated prevalence should only be used as a valid estimate of the population prevalence when the entire sample is a random sample of the population.

	Table Counts								
Cutoff Value	**TPs A**	**FPs B**	**FNs C**	**TNs D**	**TPR (Sensitivity)**	**TNR (Specificity)**	**PPV**	**Accuracy**	**TPR + TNR**
≥ 1	15	45	0	0	1.0000	0.0000	0.2500	0.2500	1.0000
≥ 2	15	40	0	5	1.0000	0.1111	0.2727	0.3333	1.1111
≥ 3	13	36	2	9	0.8667	0.2000	0.2653	0.3667	1.0667
≥ 4	13	29	2	16	0.8667	0.3556	0.3095	0.4833	1.2222
≥ 5	12	22	3	23	0.8000	0.5111	0.3529	0.5833	1.3111
≥ 6	10	12	5	33	0.6667	0.7333	0.4545	0.7167	1.4000
≥ 7	7	8	8	37	0.4667	0.8222	0.4667	0.7333	1.2889
≥ 8	3	5	12	40	0.2000	0.8889	0.3750	0.7167	1.0889
≥ 9	2	0	13	45	0.1333	1.0000	1.0000	0.7833	1.1333
≥ 10	1	0	14	45	0.0667	1.0000	1.0000	0.7667	1.0667

Definitions:

Cutoff Value	Indicates the criterion value range that predicts a positive condition.
A	The number of True Positives.
B	The number of False Positives.
C	The number of False Negatives.
D	The number of True Negatives.
TPR	The True Positive Rate or Sensitivity = A / (A + C).
TNR	The True Negative Rate or Specificity = D / (B + D).
PPV	The Positive Predictive Value or Precision = A / (A + B).
Accuracy	The Proportion Correctly Classified = (A + D) / (A + B + C + D).
TPR + TNR	The Sensitivity + Specificity.

The report displays, for each criterion, some of the more commonly used rates for each cutoff value.

© NCSS, LLC. All Rights Reserved.

Area Under Curve Analysis (Empirical Estimation)

Area Under Curve Analysis (Empirical Estimation)

Estimated Prevalence = 15 / 60 = 0.25
Estimated Prevalence is the proportion of the sample with a positive condition of PRESENT. The estimated prevalence should only be used as a valid estimate of the population prevalence when the entire sample is a random sample of the population.

				Upper One-Sided Test of H0: AUC ≤ 0.5 vs. H1: AUC > 0.5		95% Confidence Interval Limits	
Criterion	Count	AUC	Standard Error	Z-Value	P-Value	Lower	Upper
Method1	60	0.8193	0.0730	4.372	0.0000	0.6165	0.9201
Method2	60	0.7126	0.0797	2.667	0.0038	0.5190	0.8366

Definitions:

Criterion	The Criterion Variable containing the scores of the individuals.
Count	The number of the individuals used in the analysis.
AUC	The area under the ROC curve using the empirical (trapezoidal) approach.
Standard Error	The standard error of the AUC estimate.
Z-Value	The Z-score for testing the designated hypothesis test.
P-Value	The p-value associated with the Z-Value.
Confidence Interval Limits	Form the confidence interval for AUC.

This report gives statistical tests comparing the area under the curve to the value 0.5. The small *P*-values indicate a significant difference from 0.5 for both criteria. The report also gives the 95% confidence interval for each estimated AUC.

Test Comparing Two AUCs (Empirical Estimation)

Test Comparing Two AUCs (Empirical Estimation)

H0: AUC1 = AUC2
H1: AUC1 ≠ AUC2
Sample Size: 60

Paired Criterion Variables				Difference				
Criterion 1	Criterion 2	AUC1	AUC2	AUC1 - AUC2	Standard Error	Percent	Z-Value	P-Value
Method1	Method2	0.8193	0.7126	0.1067	0.0742	-13.02	1.438	0.1504

Definitions:

Criterion 1	The first specified Criterion Variable.
Criterion 2	The second specified Criterion Variable.
AUC1	The calculated area under the ROC curve for Criterion 1.
AUC2	The calculated area under the ROC curve for Criterion 2.
Difference (AUC1 - AUC2)	The simple difference AUC1 minus AUC2.
Difference Standard Error	The standard error of the AUC difference.
Difference Percent	The difference (AUC1 - AUC2) expressed as a percent difference from AUC1.
Z-Value	The calculated Z-statistic for testing H0: AUC1 = AUC2.
P-Value	The probability that the true AUC1 equals AUC2, given the sample data.

This report gives a two-sided statistical test comparing the area under the curve of Method1 to the area under the curve of Method2. The *P*-value does not indicate a significant difference between the AUCs.

© NCSS, LLC. All Rights Reserved.

Confidence Intervals for Comparing Two AUCs (Empirical Estimation)

Confidence Intervals for Comparing Two AUCs (Empirical Estimation)

Sample Size: 60

Paired Criterion Variables				Difference			
						95% Confidence Interval Limits	
Criterion 1	Criterion 2	AUC1	AUC2	AUC1 - AUC2	Standard Error	Lower	Upper
Method1	Method2	0.8193	0.7126	0.1067	0.0742	-0.0387	0.2521

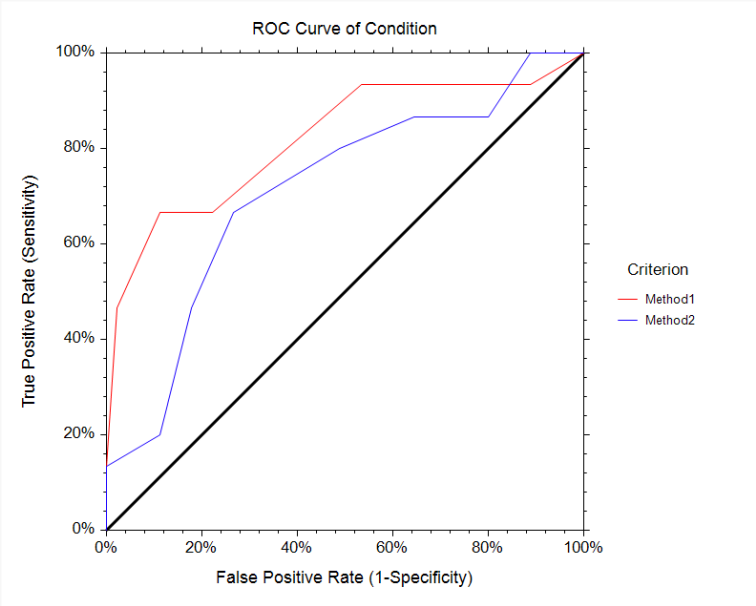
Definitions:

Criterion 1	The first specified Criterion Variable.
Criterion 2	The second specified Criterion Variable.
AUC1	The calculated area under the ROC curve for Criterion 1.
AUC2	The calculated area under the ROC curve for Criterion 2.
Difference (AUC1 - AUC2)	The simple difference AUC1 minus AUC2.
Difference Standard Error	The standard error of the AUC difference.
Confidence Interval Limits	Form the confidence interval for the difference between the AUCs.

This report provides the confidence interval for the difference of the area under the curve of Method1 and the area under the curve of Method2.

ROC Plot

ROC Plot



The coordinates of the points of the ROC curves are the TPR and FPR for each of the unique Method1 and Method2 values. The diagonal (45 degree) line is an ROC curve of random classification and serves as a baseline. Each ROC curve shows the overall ability of using the score to classify the condition. Although the Method1 curve seems to show better diagnostic capabilities, the significance test did not indicate a significant difference in the two areas under the ROC curves.

© NCSS, LLC. All Rights Reserved.

Example 2 – Comparing Two ROC Curves using Binormal Estimation

This section presents an example of a producing a statistical comparison of two (paired) ROC curves using Binormal estimation methods. The dataset used is the Paired Criteria dataset.

Setup

To run this example, complete the following steps:

1 Open the Paired Criteria example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Paired Criteria** and click **OK**.

2 Specify the Comparing Two ROC Curves – Paired Design procedure options

- Find and open the **Comparing Two ROC Curves – Paired Design** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Condition Variable ..... **Condition**
Positive Condition Value ..... **Present**
Criterion Variables **Method1, Method2**
Criterion Direction ..... **Higher values indicate a Positive Condition**

AUC Reports Tab

Area Under Curve (AUC) Analysis (Binormal Estimation) ..... **Checked**
Test Comparing Two AUCs (Binormal Estimation) ..... **Checked**
Confidence Intervals for Comparing Two AUCs (Binormal Estimation) ..... **Checked**

Plots Tab

ROC Plot ..... **Checked**

ROC Plot Format (*Click the Button*)

Empirical ROC Line ..... **Checked**
Binormal ROC Line ..... **Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

© NCSS, LLC. All Rights Reserved.

Area Under Curve Analysis (Binormal Estimation)

Area Under Curve Analysis (Binormal Estimation)

Estimated Prevalence = 15 / 60 = 0.25
Estimated Prevalence is the proportion of the sample with a positive condition of PRESENT. The estimated prevalence should only be used as a valid estimate of the population prevalence when the entire sample is a random sample of the population.

				Upper One-Sided Test of H0: AUC ≤ 0.5 vs. H1: AUC > 0.5		95% Confidence Interval Limits	
Criterion	Count	AUC	Standard Error	Z-Value	P-Value	Lower	Upper
Method1	60	0.8118	0.0660	4.723	0.0000	0.6368	0.9072
Method2	60	0.7082	0.0766	2.720	0.0033	0.5245	0.8290

Definitions:

Criterion	The Criterion Variable containing the scores of the individuals.
Count	The number of the individuals used in the analysis.
AUC	The area under the ROC curve using the Binormal estimation approach.
Standard Error	The standard error of the AUC estimate.
Z-Value	The Z-score for testing the designated hypothesis test.
P-Value	The p-value associated with the Z-Value.
Confidence Interval Limits	Form the confidence interval for AUC.

This report gives a statistical test comparing the area under the curve to the value 0.5 for each criterion. The small *P*-values indicate a significant difference from 0.5 for both criteria. The report also gives the 95% confidence interval for each estimated AUC.

Test Comparing Two AUCs (Binormal Estimation)

Test Comparing Two AUCs (Binormal Estimation)

H0: AUC1 = AUC2
H1: AUC1 ≠ AUC2
Sample Size: 60

Paired Criterion Variables				Difference				
Criterion 1	Criterion 2	AUC1	AUC2	AUC1 - AUC2	Standard Error	Percent	Z-Value	P-Value
Method1	Method2	0.8118	0.7082	0.1035	0.0786	-12.753	1.287	0.1981

Definitions:

Criterion 1	The first specified Criterion Variable.
Criterion 2	The second specified Criterion Variable.
AUC1	The calculated area under the ROC curve for Criterion 1.
AUC2	The calculated area under the ROC curve for Criterion 2.
Difference (AUC1 - AUC2)	The simple difference AUC1 minus AUC2.
Difference Standard Error	The standard error of the AUC difference.
Difference Percent	The difference (AUC1 - AUC2) expressed as a percent difference from AUC1.
Z-Value	The calculated Z-statistic for testing H0: AUC1 = AUC2. A logistic-type transformation is used in the calculation of the Z-Value (see documentation).
P-Value	The probability that the true AUC1 equals AUC2, given the sample data.

This report gives a two-sided statistical test comparing the area under the curve of Group 1 to the area under the curve of Group 2. The *P*-value does not indicate a significant difference between the AUCs.

© NCSS, LLC. All Rights Reserved.

Confidence Intervals for Comparing Two AUCs (Binormal Estimation)

Confidence Intervals for Comparing Two AUCs (Binormal Estimation)

Sample Size: 60

Paired Criterion Variables				Difference			
						95% Confidence Interval Limits	
Criterion 1	Criterion 2	AUC1	AUC2	AUC1 - AUC2	Standard Error	Lower	Upper
Method1	Method2	0.8118	0.7082	0.1035	0.0786	-0.0506	0.2576

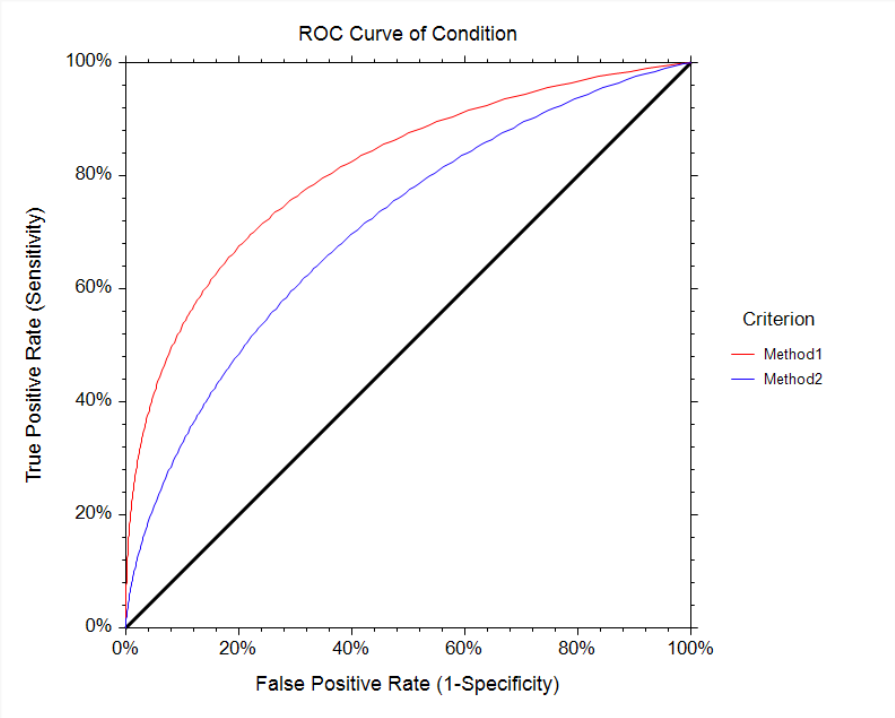
Definitions:

Criterion 1	The first specified Criterion Variable.
Criterion 2	The second specified Criterion Variable.
AUC1	The calculated area under the ROC curve for Criterion 1.
AUC2	The calculated area under the ROC curve for Criterion 2.
Difference (AUC1 - AUC2)	The simple difference AUC1 minus AUC2.
Difference Standard Error	The standard error of the AUC difference.
Confidence Interval Limits	Form the confidence interval for the difference between the AUCs.

This report provides the confidence interval for the difference of the area under the curve of Method1 and the area under the curve of Method2.

ROC Plot

ROC Plot

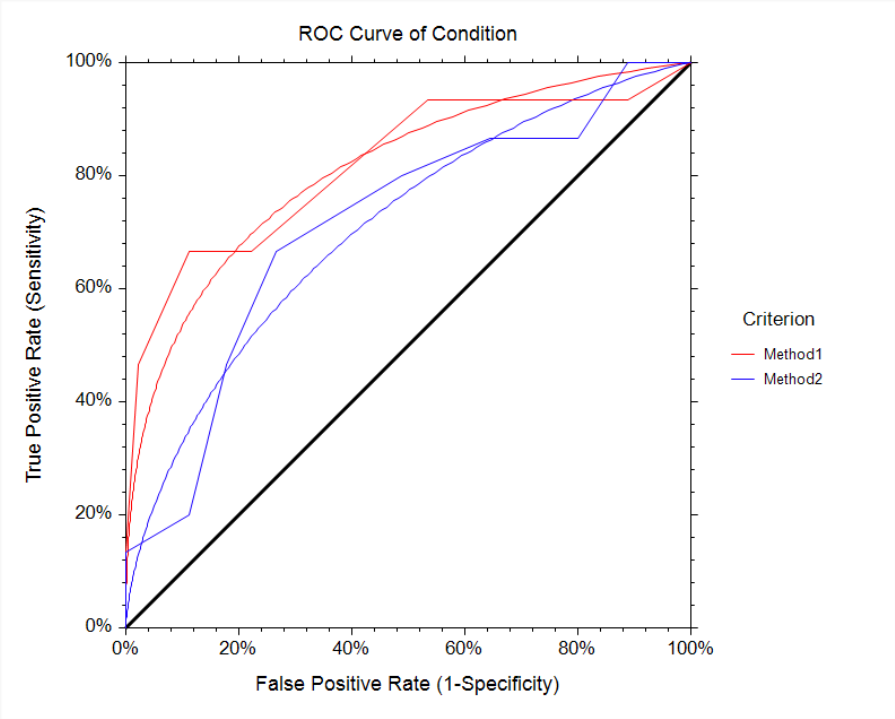


The Binormal estimation ROC plot is a smooth curve estimation of the true ROC curves. The diagonal (45 degree) line is an ROC curve of random classification and serves as a baseline.

© NCSS, LLC. All Rights Reserved.

The Binormal estimation ROC plot and the empirical estimation ROC plot can be superimposed in one plot using the plot format button:

ROC Plot



© NCSS, LLC. All Rights Reserved.

Example 3 – Non-Inferiority Test for Two Paired AUCs

This section presents an example of testing the non-inferiority of one area under the ROC curve to another. Suppose researchers wish to show that a new, less expensive classification method works at least as well as the current method. The non-inferiority margin is set at 0.1. The dataset used is the Disease Classification dataset.

Setup

To run this example, complete the following steps:

1 Open the Disease Classification example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select Disease Classification and click OK.

2 Specify the Comparing Two ROC Curves – Paired Design procedure options

- Find and open the **Comparing Two ROC Curves – Paired Design** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Condition Variable ..... **Disease**
Positive Condition Value ..... **Yes**
Criterion Variables ..... **New, Current**
Criterion Direction ..... **Higher values indicate a Positive Condition**

AUC Reports Tab

Area Under Curve (AUC) Analysis..... **Checked**
(Empirical Estimation)
Non-Inferiority Test for Two AUCs ..... **Checked**
(Empirical Estimation)
Non-Inferiority Margin **0.1**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

© NCSS, LLC. All Rights Reserved.

Area Under Curve Analysis (Empirical Estimation)

Area Under Curve Analysis (Empirical Estimation)

Estimated Prevalence = 19 / 70 = 0.2714
Estimated Prevalence is the proportion of the sample with a positive condition of YES. The estimated prevalence should only be used as a valid estimate of the population prevalence when the entire sample is a random sample of the population.

				Upper One-Sided Test of H0: AUC ≤ 0.5 vs. H1: AUC > 0.5		95% Confidence Interval Limits	
Criterion	Count	AUC	Standard Error	Z-Value	P-Value	Lower	Upper
New	70	0.8607	0.0451	8.005	0.0000	0.7422	0.9270
Current	70	0.8003	0.0586	5.122	0.0000	0.6526	0.8894

Definitions:

Criterion	The Criterion Variable containing the scores of the individuals.
Count	The number of the individuals used in the analysis.
AUC	The area under the ROC curve using the empirical (trapezoidal) approach.
Standard Error	The standard error of the AUC estimate.
Z-Value	The Z-score for testing the designated hypothesis test.
P-Value	The p-value associated with the Z-Value.
Confidence Interval Limits	Form the confidence interval for AUC.

This report gives a statistical test comparing the area under the curve to the value 0.5 for each group. The small P-values indicate a significant difference from 0.5 for both groups. The report also gives the 95% confidence interval for each estimated AUC.

Non-Inferiority Test for Two AUCs (Empirical Estimation)

Non-Inferiority Test for Two AUCs (Empirical Estimation)

Non-Inferiority Margin: 0.1
H0: AUC1 - AUC2 ≤ -0.1
H1: AUC1 - AUC2 > -0.1
Sample Size: 70

Paired Criterion Variables					
Criterion 1	Criterion 2	Difference AUC1 - AUC2	Non-Inferiority P-Value	95% One-Sided Lower Limit	Conclusion at α = 0.05
New	Current	0.0604	0.0149	-0.0611	Reject H0

Definitions:

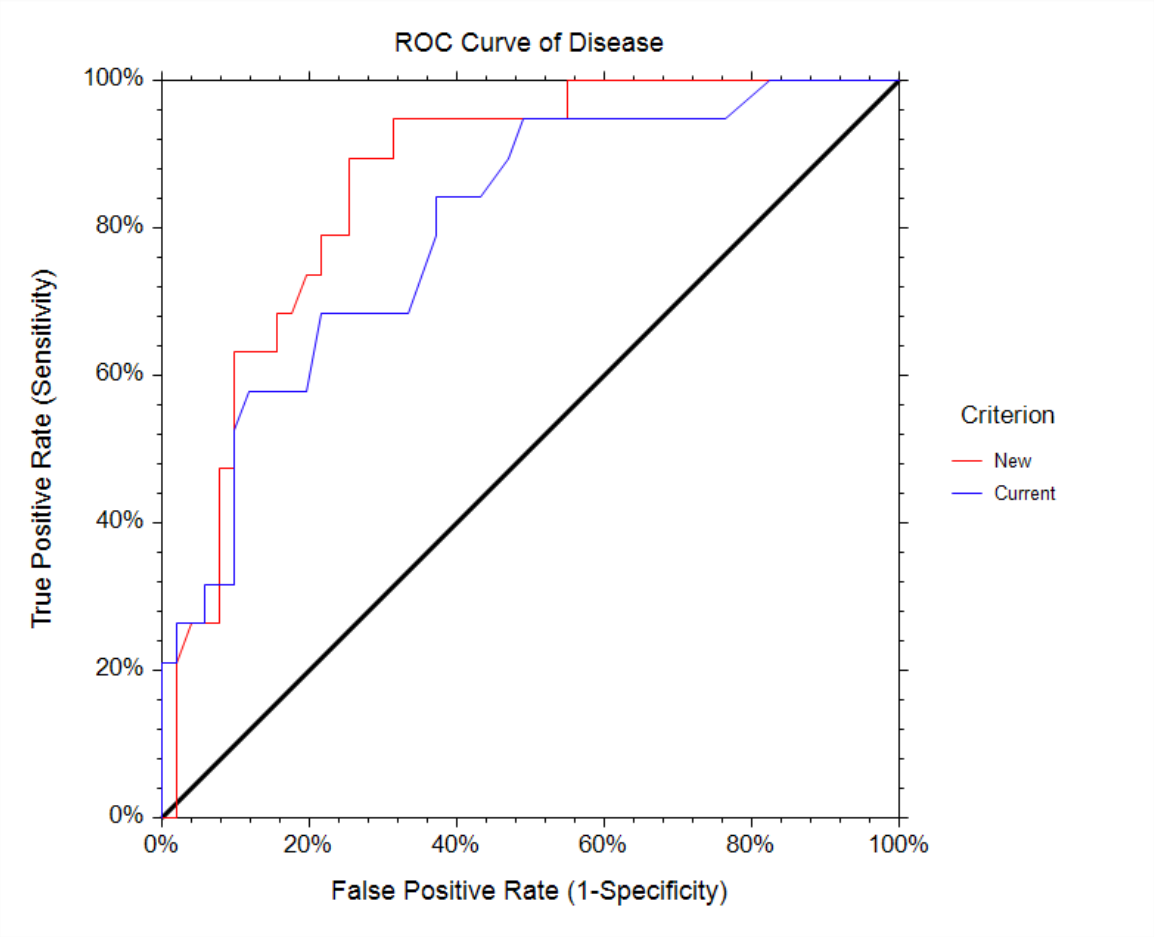
Criterion 1	The first specified Criterion Variable.
Criterion 2	The second specified Criterion Variable.
AUC1	The calculated area under the ROC curve for Criterion 1.
AUC2	The calculated area under the ROC curve for Criterion 2.
Difference (AUC1 - AUC2)	The simple difference AUC1 minus AUC2.
Non-Inferiority P-Value	The P-value for testing whether AUC1 is non-inferior to AUC2. If the Non-Inferiority P-Value is less than α, H0 is rejected, and non-inferiority may be concluded.
One-Sided Lower Limit of the Difference	Gives the cutoff (relative to the Non-Inferiority Margin) at which non-inferiority is concluded. If the One-Sided Lower Limit is greater than the negative of the Non-Inferiority Margin, H0 is rejected, and non-inferiority may be concluded.
Conclusion	The determination concerning H0, based on the Non-Inferiority P-Value (or the One-Sided Lower Limit.)

© NCSS, LLC. All Rights Reserved.

The Non-Inferiority *P*-value indicates evidence that the new criterion is non-inferior to the current criterion. Also, the one-sided lower 95% confidence limit for the difference is greater than the negative of the non-inferiority margin (-0.1).

ROC Plot

ROC Plot



The plot can be made to contain the empirical ROC curve, the Binormal ROC curve, or both, by making the proper selection after clicking the ROC Plot Format button.

The ROC plot gives the visual comparison of the two areas under the ROC curve.

© NCSS, LLC. All Rights Reserved.