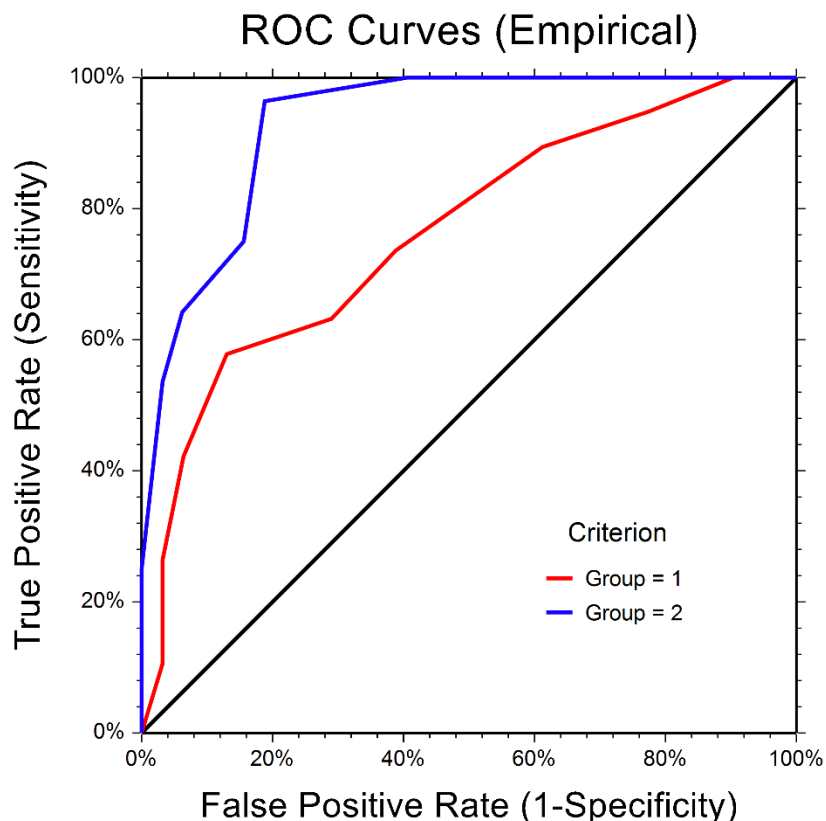


## Chapter 548

# Comparing Two ROC Curves – Independent Groups Design

## Introduction

This procedure is used to compare two ROC curves generated from data from two independent groups. In addition to producing a wide range of cutoff value summary rates for each group, 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.



## 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.

### 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.

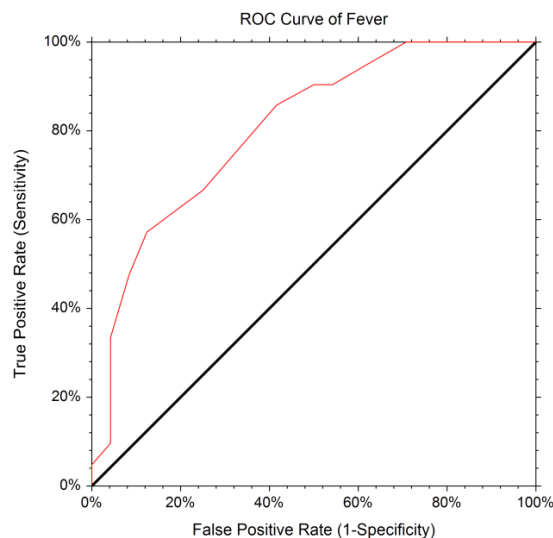
### ROC Curves

A receiver operating characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate ( $1 - \text{specificity}$ ) 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.

#### 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.

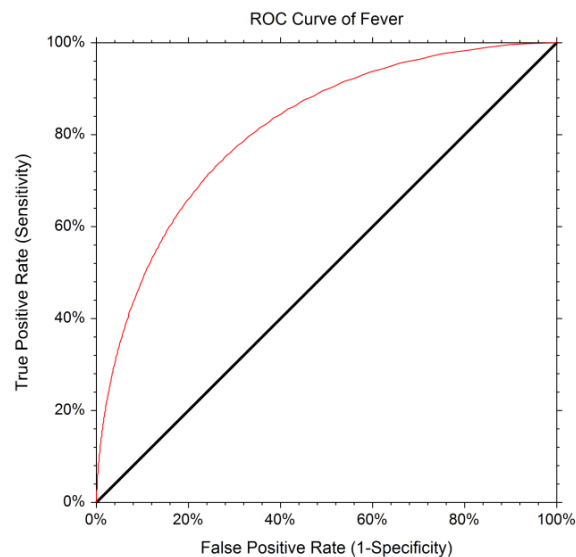


## Comparing Two ROC Curves – Independent Groups Design

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 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.

### 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

## Comparing Two ROC Curves – Independent Groups Design

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.

### 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  $i^{\text{th}}$  subject,  $V(T_{1i})$  as

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

and define the  $T_0$  component of the  $j^{\text{th}}$  subject,  $V(T_{0j})$  as

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

where

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

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

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

The empirical AUC is estimated as

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

The variance of the estimated AUC is estimated as

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

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

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

## Comparing Two ROC Curves – Independent Groups Design

### 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,

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

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

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

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

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

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

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

## Comparing Two ROC Curves – Independent Groups Design

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

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

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

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

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

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

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

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

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

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

## Comparing the AUC of Independent Sample 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 Independent Sample AUCs based on Empirical ROC Curve Estimation

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

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

where

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

Each Variance is defined as

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

where

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

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

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

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

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

Here  $T_{k0j}$  represents the observed diagnostic test result for the  $j$ th subject in group  $k$  without the condition and  $T_{k1j}$  represents the observed diagnostic test result for the  $j$ th subject in group  $k$  with the condition.

## Comparing Two ROC Curves – Independent Groups Design

### Comparing Independent Sample 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

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

where

$$V(A_1 - A_2) = V(A_1) + V(A_2)$$

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

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

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

The adjusted  $z$  statistic is

$$z = \frac{\theta_1 - \theta_2}{\sqrt{V(\theta_1 - \theta_2)}} = \frac{\theta_1 - \theta_2}{\sqrt{V(\theta_1) + V(\theta_2)}}$$

## Data Structure

The data are entered in three columns. One column specifies the true condition of the individual. Another column contains the criterion values. The third column defines the groups.

### Criterion Groups dataset

Condition	Score	Group
0	1	1
0	3	1
0	4	1
1	7	1
0	4	1
0	5	1
1	9	1
.	.	.
.	.	.
.	.	.



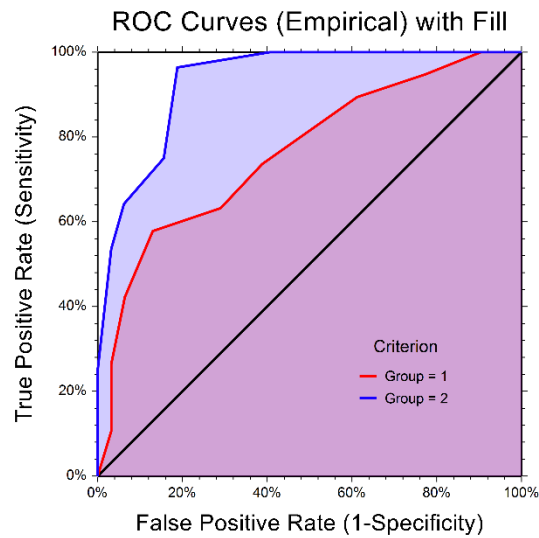
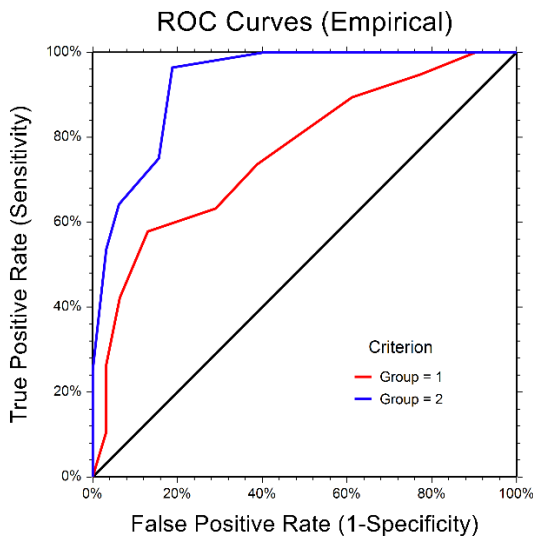
## ROC Plot Format Window Options

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

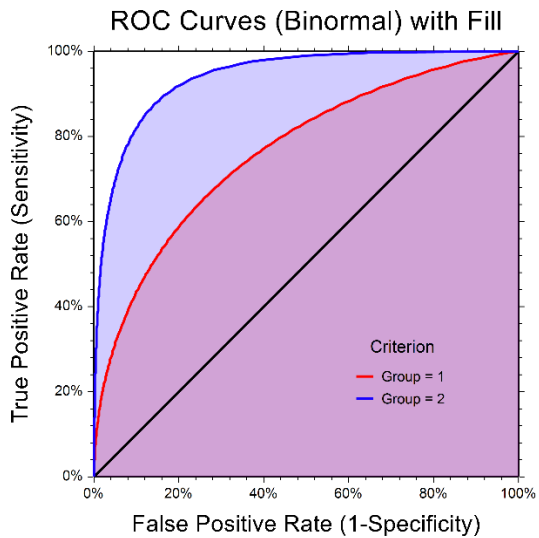
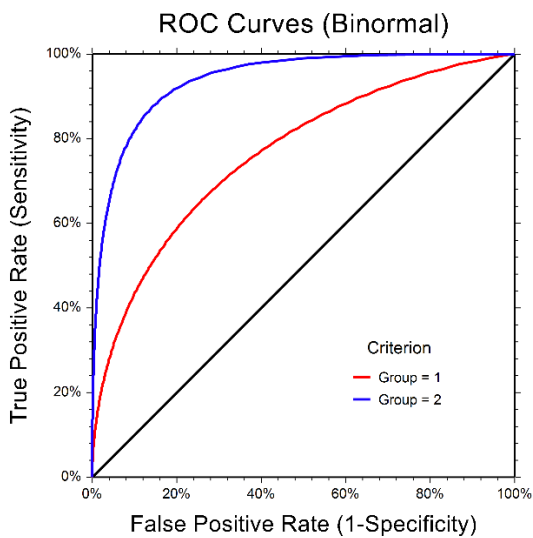
#### Empirical ROC Line Section

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



#### Binormal ROC Line Section

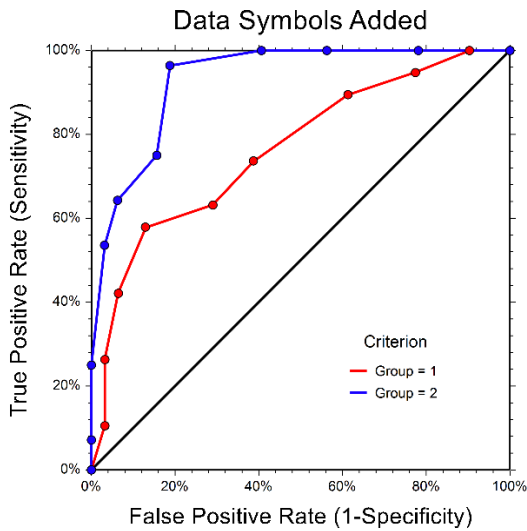
You can specify the format of the Binormal ROC curves lines using the options in this section.



## Comparing Two ROC Curves – Independent Groups Design

### 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.

## Example 1 – Comparing Two ROC Curves

This section presents an example of a producing a statistical comparison of two ROC curves using a Z-test. In the Criterion Groups dataset, a 1 for Condition indicates the condition is present, while a 0 indicates the condition is absent. 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 Criterion Groups example dataset**

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

**2 Specify the Comparing Two ROC Curves – Independent Groups Design procedure options**

- Find and open the **Comparing Two ROC Curves – Independent Groups 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 template. To load this template, click **Open Example Template** in the Help Center or File menu.

<u>Option</u>	<u>Value</u>
<b>Variables Tab</b>	
Condition Variable .....	<b>Condition</b>
Positive Condition Value.....	<b>1</b>
Criterion Variable .....	<b>Score</b>
Criterion Direction .....	<b>Higher values indicate a Positive Condition</b>
Group Variable.....	<b>Group</b>
<b>Cutoff Reports Tab</b>	
Cutoff Value List .....	<b>Data</b>
Counts, TPR (Sensitivity), .....	<b>Checked</b>
TNR (Specificity), PPV, Accuracy, TPR + TNR, Prevalence	
All Other Reports .....	<b>Unchecked</b>
<b>AUC Reports Tab</b>	
Area Under Curve (AUC) Analysis .....	<b>Checked</b>
(Empirical Estimation)	
Test Comparing Two AUCs.....	<b>Checked</b>
(Empirical Estimation)	
Confidence Intervals for Comparing .....	<b>Checked</b>
Two AUCs (Empirical Estimation)	
All Other Reports .....	<b>Unchecked</b>

**3 Run the procedure**

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

## Comparing Two ROC Curves – Independent Groups Design

## Common Rates and Indices for each Cutoff Value

## Common Rates and Indices for each Cutoff Value

Criterion Variable: Score when Group = 1

Estimated Prevalence =  $19 / 50 = 0.3800$ 

Estimated Prevalence is the proportion of the sample with a positive condition of 1, 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.

Cutoff Value	Table Counts				TPR (Sens.)	TNR (Spec.)	PPV	Accuracy	TPR + TNR
	TPs A	FPs B	FNs C	TNs D					
≥ 1.00	19	31	0	0	1.0000	0.0000	0.3800	0.3800	1.0000
≥ 2.00	19	28	0	3	1.0000	0.0968	0.4043	0.4400	1.0968
≥ 3.00	18	24	1	7	0.9474	0.2258	0.4286	0.5000	1.1732
≥ 4.00	17	19	2	12	0.8947	0.3871	0.4722	0.5800	1.2818
≥ 5.00	14	12	5	19	0.7368	0.6129	0.5385	0.6600	1.3497
≥ 6.00	12	9	7	22	0.6316	0.7097	0.5714	0.6800	1.3413
≥ 7.00	11	4	8	27	0.5789	0.8710	0.7333	0.7600	1.4499
≥ 8.00	8	2	11	29	0.4211	0.9355	0.8000	0.7400	1.3565
≥ 9.00	5	1	14	30	0.2632	0.9677	0.8333	0.7000	1.2309
≥ 10.00	2	1	17	30	0.1053	0.9677	0.6667	0.6400	1.0730

## Common Rates and Indices for each Cutoff Value

Criterion Variable: Score when Group = 2

Estimated Prevalence =  $28 / 60 = 0.4667$ 

Estimated Prevalence is the proportion of the sample with a positive condition of 1, 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.

Cutoff Value	Table Counts				TPR (Sens.)	TNR (Spec.)	PPV	Accuracy	TPR + TNR
	TPs A	FPs B	FNs C	TNs D					
≥ 1.00	28	32	0	0	1.0000	0.0000	0.4667	0.4667	1.0000
≥ 2.00	28	25	0	7	1.0000	0.2188	0.5283	0.5833	1.2188
≥ 3.00	28	18	0	14	1.0000	0.4375	0.6087	0.7000	1.4375
≥ 4.00	28	13	0	19	1.0000	0.5938	0.6829	0.7833	1.5938
≥ 5.00	27	6	1	26	0.9643	0.8125	0.8182	0.8833	1.7768
≥ 6.00	21	5	7	27	0.7500	0.8438	0.8077	0.8000	1.5938
≥ 7.00	18	2	10	30	0.6429	0.9375	0.9000	0.8000	1.5804
≥ 8.00	15	1	13	31	0.5357	0.9688	0.9375	0.7667	1.5045
≥ 9.00	7	0	21	32	0.2500	1.0000	1.0000	0.6500	1.2500
≥ 10.00	2	0	26	32	0.0714	1.0000	1.0000	0.5667	1.0714

Definitions:

Cutoff Value indicates the criterion value range that predicts a positive condition.

A is the number of True Positives.

B is the number of False Positives.

C is the number of False Negatives.

D is the number of True Negatives.

TPR is the True Positive Rate or Sensitivity =  $A / (A + C)$ .TNR is the True Negative Rate or Specificity =  $D / (B + D)$ .PPV is the Positive Predictive Value or Precision =  $A / (A + B)$ .Accuracy is the Proportion Correctly Classified =  $(A + D) / (A + B + C + D)$ .

TPR + TNR is the Sensitivity + Specificity.

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

## Comparing Two ROC Curves – Independent Groups Design

## Area Under Curve Analysis (Empirical Estimation)

## Area Under Curve Analysis (Empirical Estimation)

Estimated Prevalence (1) =  $19 / 50 = 0.3800$

Estimated Prevalence (2) =  $28 / 60 = 0.4667$

Estimated Prevalence is the proportion of the sample with a positive condition of 1. 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.

Group	Count	AUC	Standard Error	Z-Value to Test AUC > 0.5	Upper 1-Sided P-Value	95% Confidence Limits	
						Lower	Upper
1	50	0.7640	0.0710	3.720	0.0001	0.5860	0.8717
2	60	0.9314	0.0304	14.172	0.0000	0.8392	0.9715

## Definitions:

Group is the Criterion group label.

Count is the number of the individuals used in the analysis.

AUC is the area under the ROC curve using the empirical (trapezoidal) approach.

Standard Error is the standard error of the AUC estimate.

Z-Value is the Z-score for testing the designated hypothesis test.

P-Value is the probability level associated with the Z-Value.

The Lower and Upper Confidence 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 groups. 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

Total Sample Size: 110

## Group Variable: Group

Group 1	Group 2	AUC1	AUC2	Difference AUC1 - AUC2	Difference Std Error	Difference Percent	Z-Value	P-Value
1	2	0.7640	0.9314	-0.1674	0.0772	21.905	-2.167	0.0302

## Definitions:

Group 1 is the category of the Group Variable assigned to Group 1.

Group 2 is the category of the Group Variable assigned to Group 2.

AUC1 is the calculated area under the ROC curve for Group 1.

AUC2 is the calculated area under the ROC curve for Group 2.

Difference (AUC1 - AUC2) is the simple difference AUC1 minus AUC2.

Difference Std Error is the standard error of the AUC difference.

Difference Percent is the Difference (AUC1 - AUC2) expressed as a percent difference from AUC1.

Z-Value is the calculated Z-statistic for testing H0: AUC1 = AUC2.

P-Value is 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 small *P*-value indicates a significant difference between the AUCs.

Comparing Two ROC Curves – Independent Groups Design

**Confidence Intervals for Comparing Two AUCs (Empirical Estimation)**

**Confidence Intervals for Comparing Two AUCs (Empirical Estimation)**

Total Sample Size: 110

Group Variable: Group

Group 1	Group 2	AUC1	AUC2	Difference AUC1 - AUC2	Difference Std Error	95% Confidence Limits Lower	Upper
1	2	0.7640	0.9314	-0.1674	0.0772	-0.3187	-0.0160

Definitions:

Group 1 is the category of the Group Variable assigned to Group 1.

Group 2 is the category of the Group Variable assigned to Group 2.

AUC1 is the calculated area under the ROC curve for Group 1.

AUC2 is the calculated area under the ROC curve for Group 2.

Difference (AUC1 - AUC2) is the simple difference AUC1 minus AUC2.

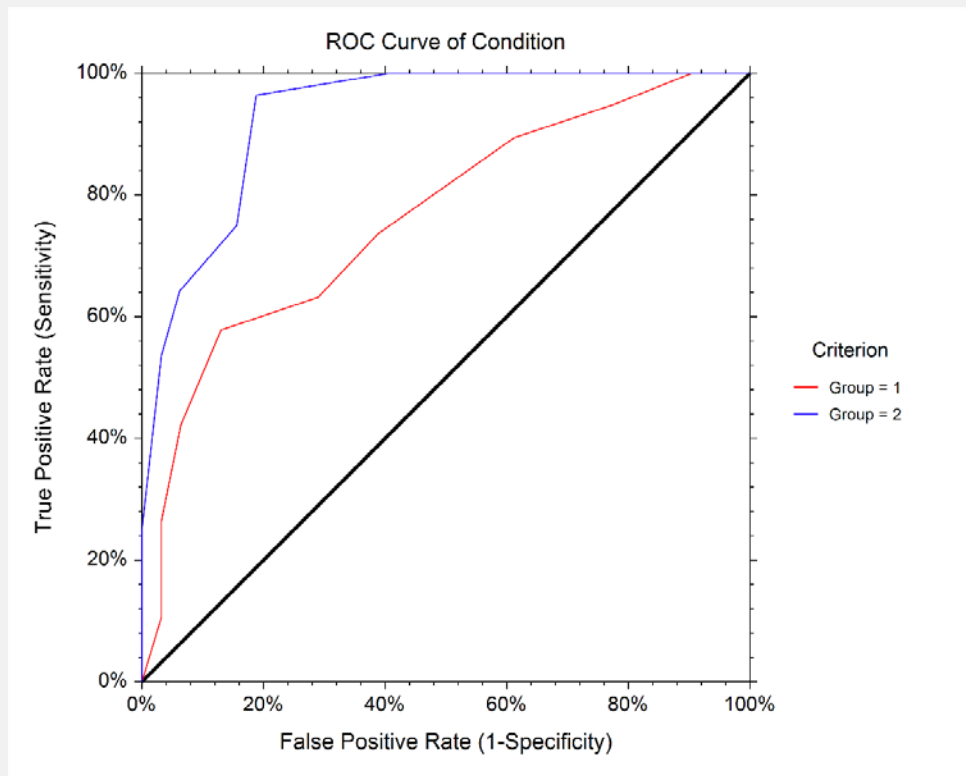
Difference Std Error is the standard error of the AUC difference.

The Lower and Upper Confidence Limits form the confidence interval for the difference between the AUCs.

This report provide the confidence interval for the difference of the area under the curve of Group 1 and the area under the curve of Group 2.

**ROC Plot Section**

**ROC Plot Section**



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 coordinates of the points of the ROC curves are the TPR and FPR for each of the unique Score 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. The Group 2 curve appears to show better classification ability and the Group 1 curve.

## Example 2 – Comparing Two ROC Curves using Binormal Estimation

This section presents an example of producing a statistical comparison of two ROC curves using Binormal estimation methods. The dataset used is the Criterion Groups dataset.

### Setup

To run this example, complete the following steps:

- 1 **Open the Criterion Groups example dataset**
  - From the File menu of the NCSS Data window, select **Open Example Data**.
  - Select **Criterion Groups** and click **OK**.
  
- 2 **Specify the Comparing Two ROC Curves – Independent Groups Design procedure options**
  - Find and open the **Comparing Two ROC Curves – Independent Groups 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 template. To load this template, click **Open Example Template** in the Help Center or File menu.

<u>Option</u>	<u>Value</u>
<b>Variables Tab</b>	
Condition Variable .....	<b>Condition</b>
Positive Condition Value.....	<b>1</b>
Criterion Variable.....	<b>Score</b>
Criterion Direction.....	<b>Higher values indicate a Positive Condition</b>
Group Variable.....	<b>Group</b>
<b>AUC Reports Tab</b>	
Area Under Curve (AUC).....	<b>Checked</b>
Analysis (Binormal Estimation)	
Test Comparing Two AUCs.....	<b>Checked</b>
(Binormal Estimation)	
Confidence Intervals for Comparing.....	<b>Checked</b>
Two AUCs (Binormal Estimation)	
<b>Plots Tab</b>	
ROC Plot.....	<b>Checked</b>
ROC Plot Format ( <i>Click the Button</i> )	
Empirical ROC Line.....	<b>Checked</b>
Binormal ROC Line .....	<b>Checked</b>

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

## Comparing Two ROC Curves – Independent Groups Design

## Area Under Curve Analysis (Binormal Estimation)

## Area Under Curve Analysis (Binormal Estimation)

Estimated Prevalence (1) =  $19 / 50 = 0.3800$

Estimated Prevalence (2) =  $28 / 60 = 0.4667$

Estimated Prevalence is the proportion of the sample with a positive condition of 1. 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.

Group	Count	AUC	Standard Error	Z-Value to Test AUC > 0.5	Upper 1-Sided P-Value	95% Confidence Limits	
						Lower	Upper
1	50	0.7654	0.0686	3.868	0.0001	0.5944	0.8702
2	60	0.9411	0.0274	16.106	0.0000	0.8560	0.9765

## Definitions:

Group is the Criterion group label.

Count is the number of the individuals used in the analysis.

AUC is the area under the ROC curve using the Binormal estimation approach.

Standard Error is the standard error of the AUC estimate.

Z-Value is the Z-score for testing the designated hypothesis test.

P-Value is the probability level associated with the Z-Value.

The Lower and Upper Confidence 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.

## Test Comparing Two AUCs (Binormal Estimation)

## Test Comparing Two AUCs (Binormal Estimation)

H0: AUC1 = AUC2

H1: AUC1 ≠ AUC2

Total Sample Size: 110

## Group Variable: Group

Group 1	Group 2	AUC1	AUC2	Difference AUC1 - AUC2	Difference Std Error	Difference Percent	Z-Value	P-Value
1	2	0.7654	0.9411	-0.1757	0.0739	22.953	-2.536	0.0112

## Definitions:

Group 1 is the category of the Group Variable assigned to Group 1.

Group 2 is the category of the Group Variable assigned to Group 2.

AUC1 is the calculated area under the ROC curve for Group 1.

AUC2 is the calculated area under the ROC curve for Group 2.

Difference (AUC1 - AUC2) is the simple difference AUC1 minus AUC2.

Difference Std Error is the standard error of the AUC difference.

Difference Percent is the Difference (AUC1 - AUC2) expressed as a percent difference from AUC1.

Z-Value is 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 is 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 small P-value indicates a significant difference between the AUCs.



## Comparing Two ROC Curves – Independent Groups Design

## Confidence Intervals for Comparing Two AUCs (Binormal Estimation)

## Confidence Intervals for Comparing Two AUCs (Binormal Estimation)

Total Sample Size: 110

Group Variable: Group

Group 1	Group 2	AUC1	AUC2	Difference AUC1 - AUC2	Difference Std Error	95% Confidence Limits	
						Lower	Upper
1	2	0.7654	0.9411	-0.1757	0.0739	-0.3205	-0.0309

## Definitions:

Group 1 is the category of the Group Variable assigned to Group 1.

Group 2 is the category of the Group Variable assigned to Group 2.

AUC1 is the calculated area under the ROC curve for Group 1.

AUC2 is the calculated area under the ROC curve for Group 2.

Difference (AUC1 - AUC2) is the simple difference AUC1 minus AUC2.

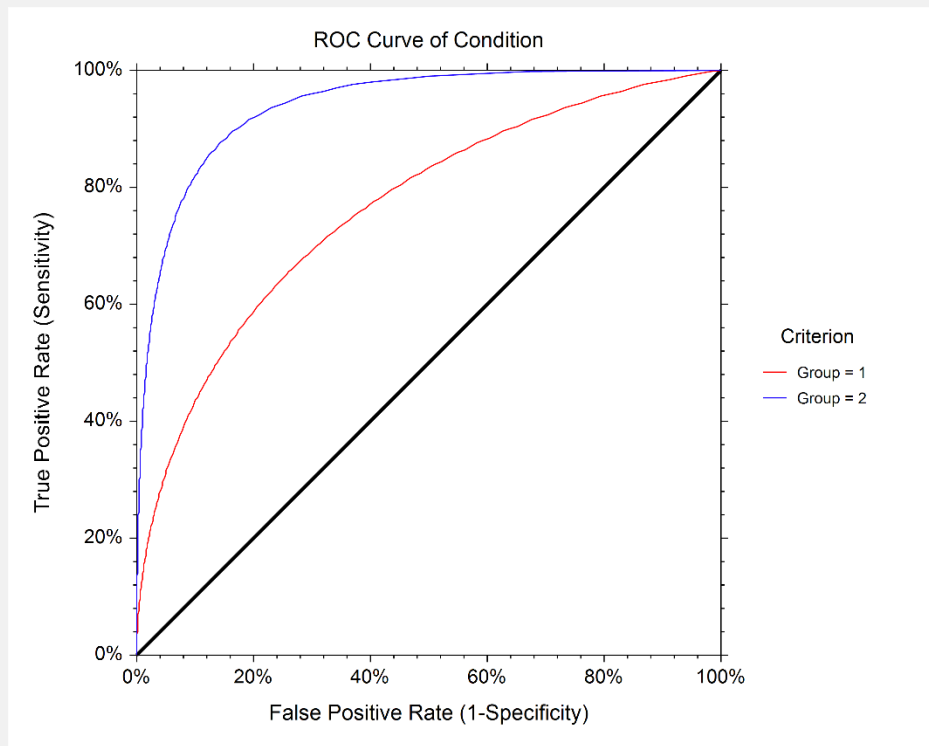
Difference Std Error is the standard error of the AUC difference.

The Lower and Upper Confidence 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 Group 1 and the area under the curve of Group 2.

## ROC Plot Section

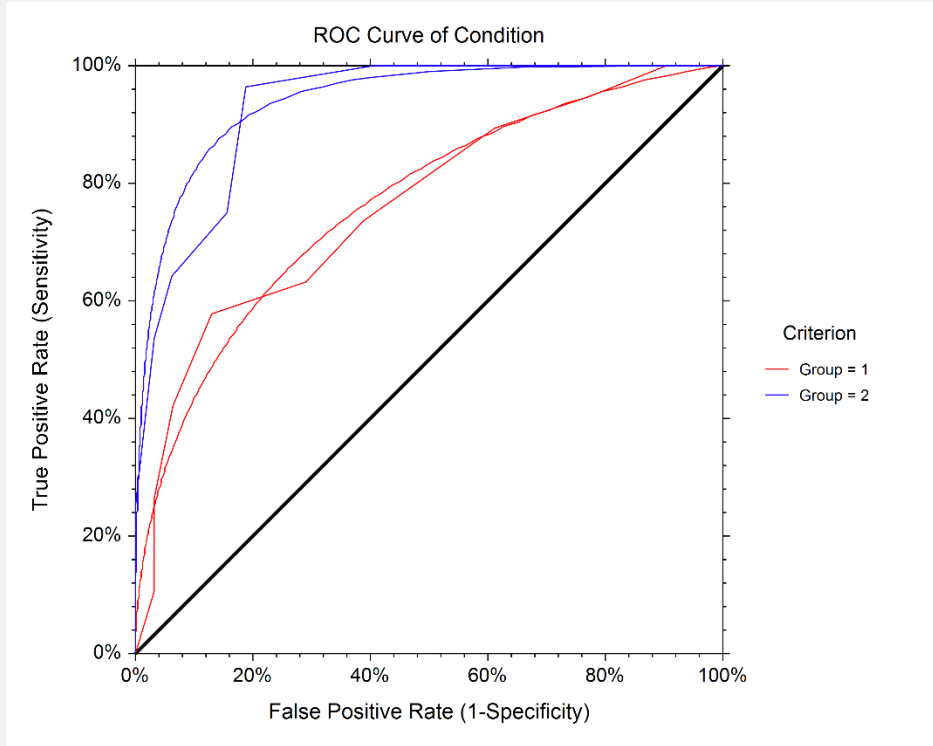
## ROC Plot Section



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

### Comparing Two ROC Curves – Independent Groups Design

#### ROC Plot Section



## Example 3 – Equivalence Test for Two AUCs

This section presents an example of testing the equivalence of two areas under the ROC curve. Suppose researchers wish to show that a new, less expensive classification method works equally well to that of the current method. The equivalence margin is set at 0.15. The dataset used is the Disease Diagnosis dataset.

### Setup

To run this example, complete the following steps:

#### 1 Open the Disease Diagnosis example dataset

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

#### 2 Specify the Comparing Two ROC Curves – Independent Groups Design procedure options

- Find and open the **Comparing Two ROC Curves – Independent Groups 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 template. To load this template, click **Open Example Template** in the Help Center or File menu.

<u>Option</u>	<u>Value</u>
<b>Variables Tab</b>	
Condition Variable .....	<b>Disease</b>
Positive Condition Value.....	<b>Yes</b>
Criterion Variable .....	<b>Score</b>
Criterion Direction .....	<b>Higher values indicate a Positive Condition</b>
Group Variable.....	<b>Method</b>
<b>AUC Reports Tab</b>	
Area Under Curve (AUC) Analysis .....	<b>Checked</b>
(Empirical Estimation)	
Equivalence Test for Two AUCs.....	<b>Checked</b>
(Empirical Estimation)	
Lower Equivalence Bound.....	<b>-0.15</b>
Upper Equivalence Bound.....	<b>0.15</b>

#### 3 Run the procedure

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

## Comparing Two ROC Curves – Independent Groups Design

## Area Under Curve Analysis (Empirical Estimation)

## Area Under Curve Analysis (Empirical Estimation)

Estimated Prevalence (Current) =  $10 / 40 = 0.2500$

Estimated Prevalence (New) =  $9 / 40 = 0.2250$

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.

Method	Count	AUC	Standard Error	Z-Value	Upper	95% Confidence Limits	
				to Test AUC > 0.5	1-Sided P-Value	Lower	Upper
Current	40	0.8883	0.0534	7.268	0.0000	0.7246	0.9571
New	40	0.8710	0.0578	6.416	0.0000	0.7002	0.9475

## Definitions:

Group is the Criterion group label.

Count is the number of the individuals used in the analysis.

AUC is the area under the ROC curve using the empirical (trapezoidal) approach.

Standard Error is the standard error of the AUC estimate.

Z-Value is the Z-score for testing the designated hypothesis test.

P-Value is the probability level associated with the Z-Value.

The Lower and Upper Confidence 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.

## Equivalence Test for Two AUCs (Empirical Estimation)

## Equivalence Test for Two AUCs (Empirical Estimation)

Lower Equivalence Bound (LEB): -0.1500

Upper Equivalence Bound (UEB): 0.1500

H0:  $AUC1 - AUC2 \leq -0.1500$  or  $AUC1 - AUC2 \geq 0.1500$

H1:  $-0.1500 < AUC1 - AUC2 < 0.1500$

Total Sample Size: 80

Group Variable: Method

Group 1	Group 2	Difference AUC1 - AUC2	Two One-Sided Tests			90% C. I.		Conclusion ( $\alpha = 0.05$ )
			Lower P-Value	Upper P-Value	Equiv. P-Value	Lower	Upper	
Current	New	0.0174	0.0168	0.0460	0.0460	-0.1121	0.1469	Reject H0

## Definitions:

Group 1 is the category of the Group Variable assigned to Group 1.

Group 2 is the category of the Group Variable assigned to Group 2.

AUC1 is the calculated area under the ROC curve for Group 1.

AUC2 is the calculated area under the ROC curve for Group 2.

Difference (AUC1 - AUC2) is the simple difference AUC1 minus AUC2.

Lower P-Value is the P-value for testing H0:  $AUC1 - AUC2 \leq LEB$  vs. H1:  $AUC1 - AUC2 > LEB$ .

Upper P-Value is the P-value for testing H0:  $AUC1 - AUC2 \geq UEB$  vs. H1:  $AUC1 - AUC2 < UEB$ .

Equivalence P-Value is the P-Value for testing overall equivalence. It is the larger of the Lower and Upper P-Values.

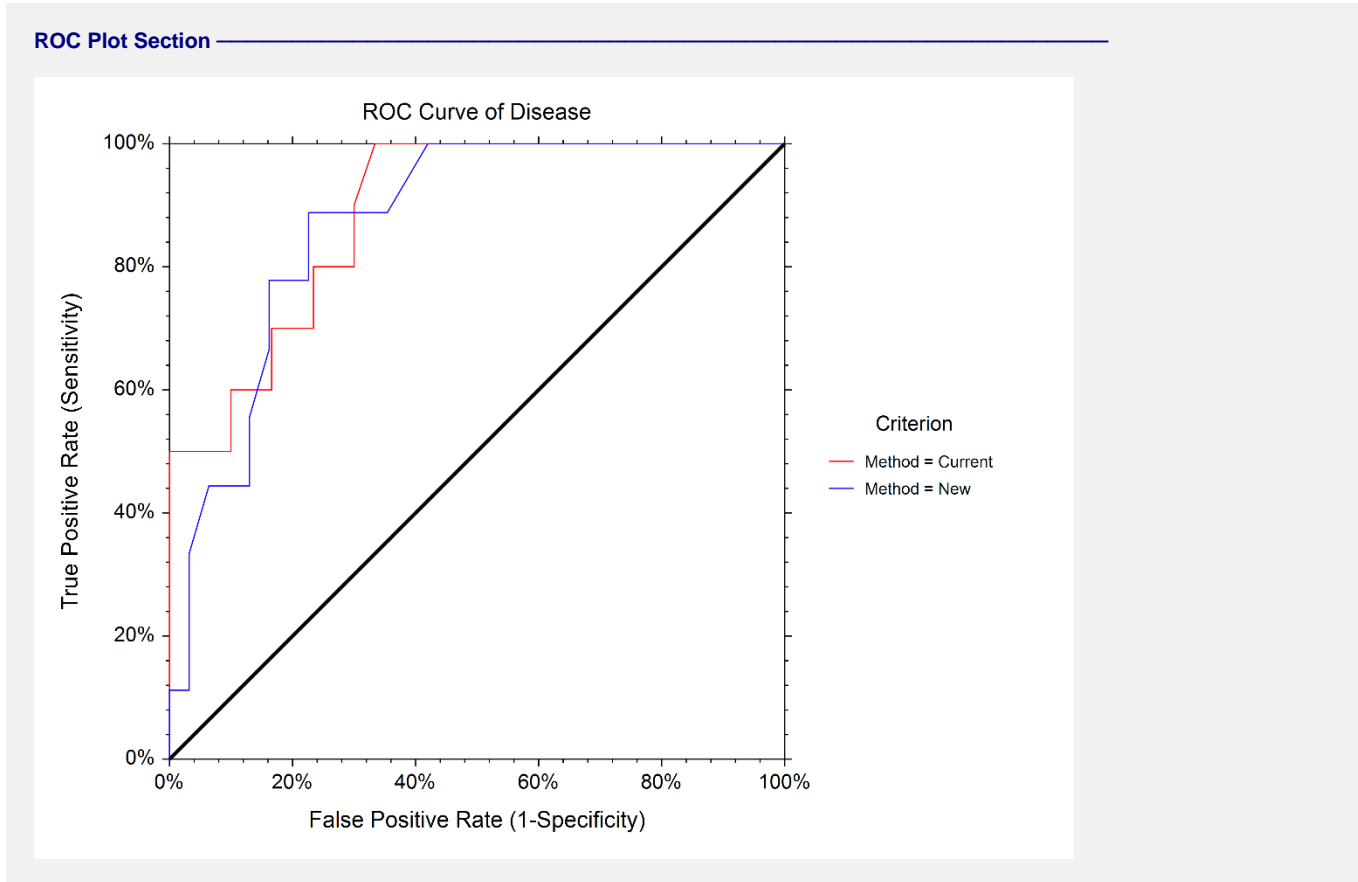
If the Equivalence P-Value is less than  $\alpha$ , H0 is rejected and equivalence may be concluded.

The Lower and Upper Confidence Limits form the equivalence confidence interval. The confidence level ( $100 * (1 - 2\alpha)$ ) corresponds to a test based on  $\alpha$ . If the equivalence confidence interval is inside the equivalence bounds, H0 is rejected and equivalence may be concluded.

Conclusion is the determination concerning H0, based on the Equivalence P-Value (or the equivalence confidence interval).

The Equivalence P-value indicates evidence that the two areas under the curve are equal. Also, the 90% confidence interval is contained by the equivalence bounds.

## ROC Plot Section



The ROC plot shows the similarity of the two areas under the ROC curve.