Chapter 545

# ROC Curves (Old Version)

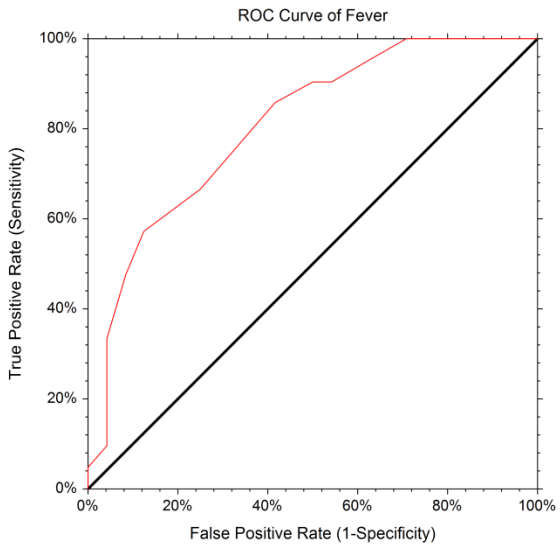
## Introduction

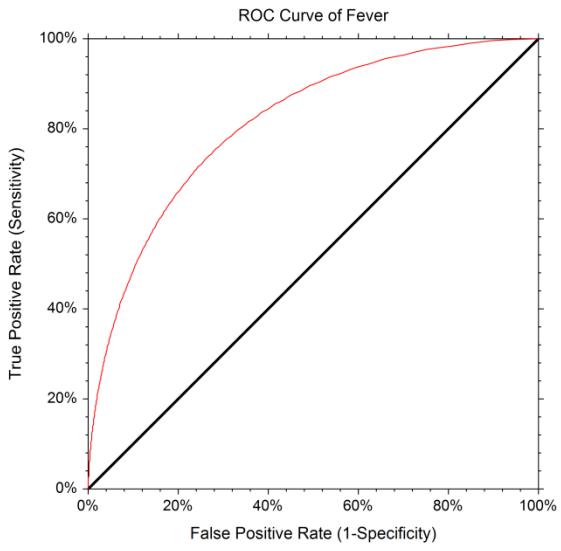
This procedure generates both binormal and empirical (nonparametric) ROC curves. It computes comparative measures such as the whole, and partial, area under the ROC curve. It provides statistical tests comparing the AUCs and partial AUCs for paired and independent sample designs.

### Discussion

A diagnostic test yields a measurement (*criterion* value) that is used to diagnose some condition of interest such as a disease. (In the sequel, we will often call the 'condition of interest' the 'disease.') The measurement might be a rating along a discrete scale or a value along a continuous scale. A positive or negative diagnosis is made by comparing the measurement to a cutoff value. If the measurement is less (or greater as the case may be) than the cutoff, the test is negative. Otherwise, the test is positive. Thus, the cutoff value helps determine the rates of false positives and false negatives.

A receiver operating characteristic (ROC) curve shows the characteristics of a diagnostic test by graphing the false-positive rate (1-*specificity*) on the horizontal axis and the true-positive rate (*sensitivity*) on the vertical axis for various cutoff values. Examples of an empirical ROC curve and a binormal ROC curve are shown below.




Each point on the ROC curve represents a different cutoff value. Cutoff values that result in low false-positive rates tend to result low true-positive rates as well. As the true-positive rate increases, so does the false positive rate. Obviously, a useful diagnostic test should have a cutoff value at which the true-positive rate is high and the false-positive rate is low. In fact, 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 is useless in determining the disease.

Complete discussions about ROC curves can be found in Altman (1991), Swets (1996), and Zhou et al (2002). Gehlbach (1988) provides an example of its use.

# Methods for Creating ROC Curves

Several methods have been proposed to generate ROC curves. These include the binormal and the empirical (nonparametric) methods.

## Binormal

The most commonly used method to generate smooth ROC curves is the binormal method popularized by a group of researchers including Metz (1978) (who developed the popular ROCFIT software). This method considers two populations: those with, and those without, the disease. It assumes that the criterion variable (or a scale-transformation of it) follows a normal distribution in each population. Using this normality assumption, a smooth ROC curve can be drawn using the sample means and variances of the two populations. Researchers have shown through various simulation studies that this *binormal* assumption is not as limiting as at first thought since non-normal data can often be transformed to a near-normal scale. **So, if you want to use this method, you should make sure that your data has been transformed so that it is nearly normal.**

## Empirical or Nonparametric

An empirical (nonparametric) approach that does not depend on the normality assumptions was developed by DeLong, DeLong, and Clarke-Pearson (1988). These ROC curves are especially useful when the diagnostic test results in a continuous criterion variable.

# Types of ROC Experimental Designs

Either of two experimental designs are usually employed when comparing ROC curves. These designs are *paired* or *independent samples*. Separate methods of analysis are needed to compare ROC curves depending upon which experimental design was used.

## Independent Sample (Non-correlated) Designs

In this design, individuals with, and without, the disease are randomly assigned into two (or more) groups. The first group receives diagnostic test A and the second group receives diagnostic test B. Each individual receives only one diagnostic test.

## Paired (Correlated) Designs

In this design, individuals with, and without, the disease each receive both diagnostic tests. This allows each subject to 'serve as their own control.'

## An Example Using a Paired Design

ROC curves are explained with an example paraphrased from Gehlbach (1988). Forty-five patients with fever, headache, and a history of tick bite were classified into two groups: those with Rocky Mountain Spotted Fever (RMSF) and those without it. The serum-sodium level of each patient is measured using two techniques. We want to determine if serum-sodium level is useful in detecting RMSF, which technique is most accurate in diagnosing RMSF, and what the diagnostic cutoff value of the selected test should be. The data are presented next.

| RMSF = Yes | | | | RMSF = No | | | |
|---|---|---|---|---|---|---|---|
| **ID** | **Method1** | **Method2** | **Diagnosis** | **ID** | **Method1** | **Method2** | **Diagnosis** |
| 1 | 124 | 122 | 1 | 22 | 129 | 124 | 0 |
| 2 | 125 | 124 | 1 | 23 | 131 | 128 | 0 |
| 3 | 126 | 125 | 1 | 24 | 131 | 130 | 0 |
| 4 | 126 | 125 | 1 | 25 | 134 | 133 | 0 |
| 5 | 127 | 126 | 1 | 26 | 134 | 133 | 0 |
| 6 | 128 | 126 | 1 | 27 | 135 | 133 | 0 |
| 7 | 128 | 127 | 1 | 28 | 136 | 134 | 0 |
| 8 | 128 | 128 | 1 | 29 | 136 | 134 | 0 |
| 9 | 128 | 128 | 1 | 30 | 136 | 134 | 0 |
| 10 | 129 | 128 | 1 | 31 | 137 | 134 | 0 |
| 11 | 129 | 130 | 1 | 32 | 137 | 136 | 0 |
| 12 | 131 | 130 | 1 | 33 | 138 | 136 | 0 |
| 13 | 132 | 133 | 1 | 34 | 138 | 137 | 0 |
| 14 | 133 | 133 | 1 | 35 | 139 | 138 | 0 |
| 15 | 133 | 134 | 1 | 36 | 139 | 138 | 0 |
| 16 | 135 | 134 | 1 | 37 | 139 | 140 | 0 |
| 17 | 135 | 134 | 1 | 38 | 139 | 140 | 0 |
| 18 | 135 | 134 | 1 | 39 | 140 | 141 | 0 |
| 19 | 136 | 136 | 1 | 40 | 140 | 141 | 0 |
| 20 | 138 | 138 | 1 | 41 | 141 | 142 | 0 |
| 21 | 139 | 140 | 1 | 42 | 142 | 142 | 0 |
| | | | | 43 | 142 | 142 | 0 |
| | | | | 44 | 142 | 142 | 0 |
| | | | | 45 | 143 | 144 | 0 |

The first step in analyzing these data is to create a two-by-two table showing the diagnostic accuracy of each method at given cutoff value. If the cutoff is set at X, the table would appear as follows:

**Generic Table for Cutoff=X**

| Method<br>Cutoff=X | RMSF = Yes | RMSF = No |
|---|---|---|
| **Sodium <= X, Positive** | **A** | **B** |
| **Sodium > X, Negative** | **C** | **D** |

The letters A, B, C, and D represent counts of the number of individuals in each of the four possible categories.

For example, if a cutoff of 130 is used to diagnose those with the disease, the tables for each sodium measurement method would be:

**Table for Method 1, Cutoff = 130**

| **Method 1 Cutoff=130** | **RMSF=Yes** | **RMSF=No** |
|---|---|---|
| **Sodium<=130, Positive** | **11** | **1** |
| **Sodium>130, Negative** | **10** | **23** |

**Table for Method 2, Cutoff = 130**

| Method 2 Cutoff=130 | RMSF=Yes | RMSF=No |
|---|---|---|
| Sodium<=130, **Positive** | 12 | 3 |
| Sodium>130, **Negative** | 9 | 21 |

If a cutoff of 137 is used to diagnose those with the disease, the tables for each sodium measurement method would be:

**Table for Method 1, Cutoff = 137**

| Method 1 Cutoff=137 | RMSF=Yes | RMSF=No |
|---|---|---|
| Sodium<=137, **Positive** | 19 | 11 |
| Sodium>137, **Negative** | 2 | 13 |

**Table for Method 2, Cutoff = 137**

| Method 2 Cutoff=137 | RMSF=Yes | RMSF=No |
|---|---|---|
| Sodium<=137, **Positive** | 19 | 13 |
| Sodium>137, **Negative** | 12 | 11 |

As you study these tables, you can see changing the cutoff value changes the table counts. An ROC curve is constructed by creating many of these tables and plotting the sensitivity versus one minus the specificity.

# Definition of Terms

We will now define the indices that are used to create ROC curves.

## Sensitivity

Sensitivity is the proportion of those with the disease that are correctly identified as having the disease by the test. In terms of our two-by-two tables, sensitivity = *A/(A+C).*

## Specificity

Specificity is the proportion of those without the disease that are correctly identified as not having the disease by the test. In terms of our two-by-two tables, specificity = *D/(B+D).*

## Prevalence

Prevalence is the overall proportion of individuals with the disease. In terms of our two-by-two tables, prevalence = *(A+C)/(A+B+C+D).* Notice that since the prevalence is defined in terms of the marginal values, it does not depend on the cutoff value.

## Positive Predictive Value (PPV)

PPV is the proportion of individuals with positive test results who have the disease. In terms of our two-by-two tables, PPV = *(A)/(A+B).*

## Negative Predictive Value (NPV)

NPV is the proportion of individuals with negative test results who do not have the disease. In terms of our two-by-two tables, NPV = *(D)/(C+D).*

## Discussion about PPV and NPV

A problem with sensitivity and specificity is that they do not assess the probability of making a correct diagnosis. To overcome this, practitioners have developed two other indices: PPV and NPV. Unfortunately, these indices have the disadvantage that they are directly impacted by the prevalence of the disease in the population. For example, if your sampling procedure is constructed to obtain more individuals with the disease than is the case in the whole population of interest, the PPV and NPV need to be adjusted.

Using Bayes theorem, adjusted values of PPV and NPV are calculated based on new prevalence values as follows:

$$PPV = \frac{sensitivity \times prevalence}{sensitivity \times prevalence + (1 - specificity) \times (1 - prevalence)}$$

$$NPV = \frac{specificity \times (1 - prevalence)}{(1 - sensitivity) \times prevalence + specificity \times (1 - prevalence)}$$

Another way of interpreting these terms is as follows. The prevalence of a disease is the prior probability that a subject has the disease before the diagnostic test is run. The values of PPV and 1-NPV are the posterior probabilities of a subject having the disease after the diagnostic test is conducted.

## Likelihood Ratio

The likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis. It is calculated as follows:

$$LR = \frac{\Pr(positive\ test|disease)}{\Pr(positive\ test|no\ disease)} = \frac{sensitivity}{1 - specificity}$$

# Finding the Optimal Criterion Value

The optimal criterion value is that value that minimizes the average cost. The approach we use was given by Metz (1978) and Zhou et al. (2002). This approach is based on an analysis of the costs (and benefits) of the four possible outcomes of a diagnostic test: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The cost of each of these outcomes must be determined. This is no small task. In fact, a whole field of study has arisen to determine these costs. Once these costs are found, the average overall cost *C* of performing a test is given by

$$C = C_0 + C_{TP}P(TP) + C_{TN}P(TN) + C_{FP}P(FP) + C_{FN}P(FN)$$

Here, $C_0$ is the fixed cost of performing the test, $C_{TP}$ is the cost associated with a true positive, P(TP) is the proportion of TP's in the population, and so on. Note that P(TP) is equal to

$$P(TP) = sensitivity[P(Condition = True)]$$

Metz (1978) showed that the point along the ROC curve where the average cost is minimum is the point where $sensitivity - m(1 - specificity)$ is maximized, where

$$m = \frac{P(Condition = False)}{P(Condition = True)}\left(\frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}\right)$$

P(*Condition* = True) is called the *prevalence* of the disease. Depending on the method used to obtain the sample, it may or may not be estimated from the sample. Note that the costs enter this equation as the ratio of the net cost for a test of an individual without the disease to the net cost for a test of an individual with the disease.

Using the above result, the cut optimum cutoff may be found by scanning a report that shows *C*(*Cutoff*) for every value of the cutoff variable.

# Area Under the ROC Curve (AUC)

## The AUC of a Single ROC Curve

The area under an ROC curve (AUC) is a popular measure of the accuracy of a diagnostic test. Other things being equal, the larger the AUC, the better the test is at predicting the existence of the disease. 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 with the disease is larger than the criterion value of another individual drawn at random from the population without the disease.

A statistical test of usefulness of a diagnostic test is to compare it to the value 0.5. Such a statistical test can be made if we are willing to assume that the sample is large enough so that the estimated AUC follows the normal distribution. The statistical test is

$$z = \frac{\tilde{A} - 0.5}{\sqrt{V(\tilde{A})}}$$

where $\tilde{A}$ is the estimated AUC and $V(\tilde{A})$ is the estimated variance of $\tilde{A}$.

Two methods are commonly used to estimate the AUC. The first is the *binormal* method presented by Metz (1978) and McClish (1989). This method results in a smooth ROC curve from which both the complete and partial AUC may be calculated. The second 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 above *z* test may be used for both methods, as long as an appropriate estimate of $V(\tilde{A})$ is used.

## The AUC of a Single 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 disease and the other made up of individuals without the disease. 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 non-diseased population and *Y* refer to the value of the criterion variable in the diseased 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

$$A = \int_{-\infty}^{\infty} TP(c)FP'(c)\,\mathrm{dc}$$

$$= \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\Phi\left(\frac{\mu_x - c}{\sigma_x}\right)\right]\mathrm{d}c$$

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

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

$$A = \int_{c_1}^{c_2} TP(c)FP'(c)\mathrm{d}c$$

$$= \frac{1}{\sigma_x}\int_{c_2}^{c_1} \left[\Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\Phi\left(\frac{\mu_x - c}{\sigma_x}\right)\right]\mathrm{d}c$$

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 sequel.

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

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

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

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

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

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

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

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

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

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

Once estimates of $\hat{A}$ and $V(\hat{A})$ are calculated, hypothesis tests and confidence intervals can be calculated using standard methods. However, following the advice of Zhou et al. (2002) page 125, we use the following transformation which results in statistics that are closer to normality and ensures confidence limits that are inside the zero-one range. The transformation is

$$\hat{\Psi} = \ln\left(\frac{1+\hat{A}}{1-\hat{A}}\right)$$

The variance of $\widehat{\Psi}$ is estimated using

$$\mathrm{V}(\widehat{\Psi}) = \frac{4}{\left(1-\hat{A}^2\right)^2} V(\hat{A})$$

A $100(1-\alpha)\%$ confidence interval for $\Psi$ may then be constructed as

$$L, U = \widehat{\Psi} \pm z_{1-\alpha/2}\sqrt{V(\widehat{\Psi})}$$

Using the inverse transformation, the confidence interval for *A* is given by the two limits

$$\frac{1-e^{-L}}{1+e^{-L}} \quad \text{and} \quad \frac{1-e^{-U}}{1+e^{-U}}$$

## The AUC of a Single 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 formula for computing this estimate of the AUC and its variance are given later in the section on comparing two empirical ROC curves.

## Comparing the AUC of Two ROC Curves

Occasionally, it is of interest to compare the areas under the ROC curve (AUC) of two diagnostic tests using a hypothesis test. This may be done using either the binormal model results shown in McClish (1989) or the empirical (nonparametric) results of DeLong (1988).

## Comparing the AUC of Two Empirical ROC Curves

A statistical test may be constructed that uses empirical estimates of the AUCs, their variances, and covariance. The variance and covariance formulas used depend on whether the design is paired or independent samples. Following Zhou et al. (2002) page 185, the formula to compare two AUCs is the following *z* test (which asymptotically follows the standard normal distribution) is given by

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

where

$$V(A_1 - A_2) = V(A_1) + V(A_2) - 2Cov(A_1, A_2)$$

## Independent Samples

For independent samples in which each subject receives only one of the two diagnostic tests, the covariance is zero and the two variances are

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

where

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

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

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

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

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

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

### Paired Samples

For paired samples in which each subject receives both of the two diagnostic tests, the variances are given as above and the covariance is given by

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

where

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

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

## Comparing Two Binormal AUCs

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

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

### Independent Samples Design

When an independent sample design is used, the variance of the difference in AUC's is the sum of the variances since the covariance is zero. That is,

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

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

### Paired Design

When a paired design is used, the variance of the difference in AUC's is

$$V(A_1 - A_2) = V(A_1) + V(A_2) - 2Cov(A_1, A_2)$$

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

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

where

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

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

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

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

## Transformation to Achieve Normality

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

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

The adjusted *z* statistic is

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

$$= \frac{\theta_1 - \theta_2}{\sqrt{V(\theta_1) + \mathrm{V}(\theta_2) - 2Cov(\theta_1, \theta_2)}}$$

# Cost Benefit Ratios

This is the ratio of the net cost when the condition is absent to the net cost when the condition is present. In symbols, this is

$$\frac{Cost(FP) - Benefit(TN)}{Cost(FN) - Benefit(TP)}$$

This value is used to compute the optimum criterion value. Since it is difficult to calculate this value exactly, you can enter a set of up to 4 values. These values will be used in the Cost-Benefit Report.

# Data Structure

The data are entered in two or more variables. One variable specifies the true condition of the individual. The other variable(s) contain the criterion value(s) for the tests being compared.

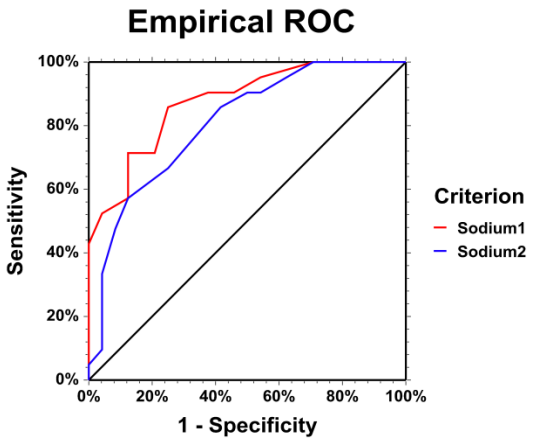
# ROC Plot Format Window Options

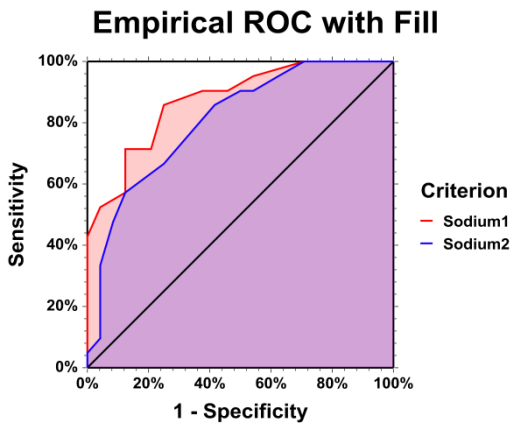
This section describes the specific options available on the ROC Plot Format window, which is displayed when the ROC Plot Chart Format button is clicked. Common options, such as axes, labels, legends, and titles are documented in the Graphics Components chapter.

## ROC Plot Tab

### Empirical ROC Line Section

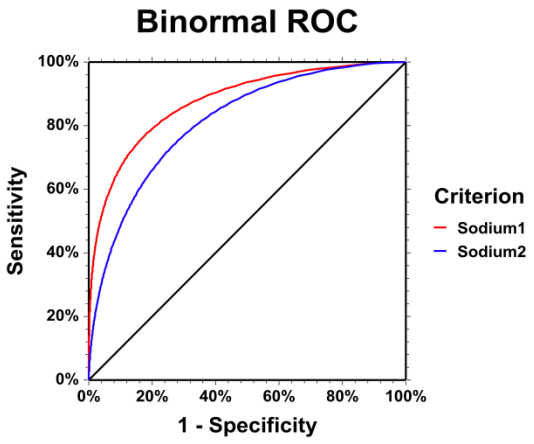
You can specify the format of the empirical ROC curve lines using the options in this section.

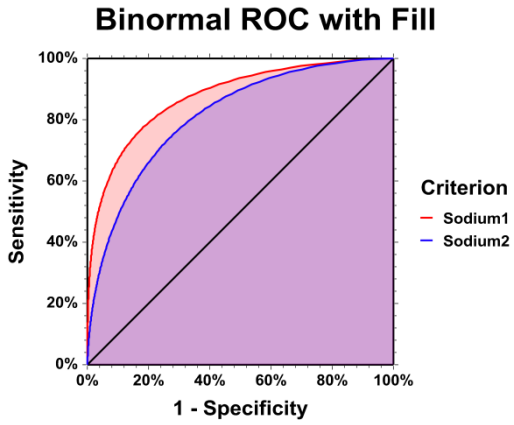



### Binormal ROC Line Section

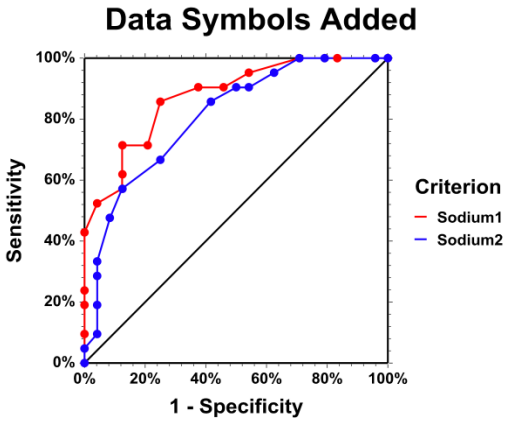
You can specify the format of the binormal ROC curves lines using the options in this section.



## 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 – ROC Curve for a Paired Design

This section presents an example of how to generate an ROC curve for the RMSF data contained in the ROC dataset. This is an example of data from a paired design.

## Setup

To run this example, complete the following steps:

**1 Open the ROC example dataset**

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

**2 Specify the ROC Curves (Old Version) procedure options**

- Find and open the **ROC Curves (Old Version)** 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

| | |
|---|---|
| Actual Condition Variable | **Fever** |
| Positive Condition Value | **1** |
| Actual Condition Prevalence | **0.1** |
| Cost Benefit Ratio(s) | **1.1 1.3 1.5 1.7** |
| Criterion Variable(s) | **Sodium1, Sodium2** |
| Test Direction | **Low X = Positive** |
| Criterion List | **120 to 140 by 5** |
| Max Equivalence Difference | **0.05** |

Reports Tab

| | |
|---|---|
| ROC Data Report (Empirical) | **Checked** |
| ROC Data Report (Binormal) | **Checked** |
| Cost-Benefit Report (Empirical) | **Checked** |
| Cost-Benefit Report (Binormal) | **Checked** |
| Predictive Value Report (Empirical) | **Checked** |
| Predictive Value Report (Binormal) | **Checked** |
| Area Under Curve (AUC) Analysis: One Curve | **Checked** |
| Area Under Curve (AUC) Analysis: Compare Two Curves | **Checked** |
| Area Under Curve (AUC) Analysis: Two Curve Equivalence | **Checked** |
| Skip Line After | **20** |

Report Options (*in the Toolbar*)

| | |
|---|---|
| Variable Labels | **Column Names** |

### 3 Run the procedure

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

## ROC Data using the Empirical ROC Curve

**ROC Data for Condition = Fever using the Empirical ROC Curve**

**Sodium1**

| Cutoff Value | Count +\|P A | Count +\|A B | Count -\|P C | Count -\|A D | Sensitivity A/(A+C) | C/(A+C) | False+ B/(B+D) | Specificity D/(B+D) |
|---|---|---|---|---|---|---|---|---|
| 120.00 | 0 | 0 | 21 | 24 | 0.00000 | 1.00000 | 0.00000 | 1.00000 |
| 125.00 | 2 | 0 | 19 | 24 | 0.09524 | 0.90476 | 0.00000 | 1.00000 |
| 130.00 | 11 | 1 | 10 | 23 | 0.52381 | 0.47619 | 0.04167 | 0.95833 |
| 135.00 | 18 | 6 | 3 | 18 | 0.85714 | 0.14286 | 0.25000 | 0.75000 |
| 140.00 | 21 | 19 | 0 | 5 | 1.00000 | 0.00000 | 0.79167 | 0.20833 |

**Sodium2**

| Cutoff Value | Count +\|P A | Count +\|A B | Count -\|P C | Count -\|A D | Sensitivity A/(A+C) | C/(A+C) | False+ B/(B+D) | Specificity D/(B+D) |
|---|---|---|---|---|---|---|---|---|
| 120.00 | 0 | 0 | 21 | 24 | 0.00000 | 1.00000 | 0.00000 | 1.00000 |
| 125.00 | 4 | 1 | 17 | 23 | 0.19048 | 0.80952 | 0.04167 | 0.95833 |
| 130.00 | 12 | 3 | 9 | 21 | 0.57143 | 0.42857 | 0.12500 | 0.87500 |
| 135.00 | 18 | 10 | 3 | 14 | 0.85714 | 0.14286 | 0.41667 | 0.58333 |
| 140.00 | 21 | 17 | 0 | 7 | 1.00000 | 0.00000 | 0.70833 | 0.29167 |

Notes:

| | |
|---|---|
| A | The number of subjects with a POSITIVE test when the condition was PRESENT. |
| B | The number of subjects with a POSITIVE test when the condition was ABSENT. |
| C | The number of subjects with a NEGATIVE test when the condition was PRESENT. |
| D | The number of subjects with a NEGATIVE test when the condition was ABSENT. |
| Sensitivity | The Pr(Positive Test\|Condition Present). |
| False+ | The Pr(Positive Test\|Condition Absent). |
| Specificity | The Pr(Negative Test\|Condition Absent). |

The report displays the numeric information used to generate the empirical ROC curve.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option of the Reports panel.

### A B C D

These four columns give the counts of the two-by-two tables that are formed at each of the corresponding cutoff points.

### Sensitivity A/(A+C)

This is the proportion of those that had the disease that were correctly diagnosed by the test.

### C/(A+C)

This is the proportion of those that had the disease that were incorrectly diagnosed.

### False + B/(B+D)

The proportion of those who did not have the disease who were incorrectly diagnosed by the test as having it.

### Specificity D/(B+D)

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

## ROC Data using the Binormal ROC Curve

**ROC Data for Condition = Fever using the Binormal ROC Curve**

**Sodium1**

| Cutoff Value | Count +\|P A | Count +\|A B | Count -\|P C | Count -\|A D | Sensitivity A/(A+C) | C/(A+C) | False+ B/(B+D) | Specificity D/(B+D) |
|---|---|---|---|---|---|---|---|---|
| 120.00 | 0 | 0 | 21 | 24 | 0.00751 | 1.00000 | 0.00000 | 1.00000 |
| 125.00 | 2 | 0 | 19 | 24 | 0.09734 | 0.90476 | 0.00000 | 0.99957 |
| 130.00 | 11 | 1 | 10 | 23 | 0.43561 | 0.47619 | 0.04167 | 0.97664 |
| 135.00 | 18 | 6 | 3 | 18 | 0.83464 | 0.14286 | 0.25000 | 0.74153 |
| 140.00 | 21 | 19 | 0 | 5 | 0.98246 | 0.00000 | 0.79167 | 0.24423 |

**Sodium2**

| Cutoff Value | Count +\|P A | Count +\|A B | Count -\|P C | Count -\|A D | Sensitivity A/(A+C) | C/(A+C) | False+ B/(B+D) | Specificity D/(B+D) |
|---|---|---|---|---|---|---|---|---|
| 120.00 | 0 | 0 | 21 | 24 | 0.01869 | 1.00000 | 0.00000 | 0.99949 |
| 125.00 | 4 | 1 | 17 | 23 | 0.14344 | 0.80952 | 0.04167 | 0.98899 |
| 130.00 | 12 | 3 | 9 | 21 | 0.48070 | 0.42857 | 0.12500 | 0.90223 |
| 135.00 | 18 | 10 | 3 | 14 | 0.83352 | 0.14286 | 0.41667 | 0.61742 |
| 140.00 | 21 | 17 | 0 | 7 | 0.97642 | 0.00000 | 0.70833 | 0.24291 |

Notes:

| | |
|---|---|
| A | The number of subjects with a POSITIVE test when the condition was PRESENT. |
| B | The number of subjects with a POSITIVE test when the condition was ABSENT. |
| C | The number of subjects with a NEGATIVE test when the condition was PRESENT. |
| D | The number of subjects with a NEGATIVE test when the condition was ABSENT. |
| Sensitivity | The Pr(Positive Test\|Condition Present). |
| False+ | The Pr(Positive Test\|Condition Absent). |
| Specificity | The Pr(Negative Test\|Condition Absent). |

The report displays the numeric information used to generate the binormal ROC curve.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option of the Reports panel.

### A B C D

These four columns give the counts of the two-by-two tables that are formed at each of the corresponding cutoff points.

### Sensitivity A/(A+C)

This is the proportion of those that had the disease that were correctly diagnosed by the test. Note that these values are based on the binormal model.

### C/(A+C)

This is the proportion of those that had the disease that were incorrectly diagnosed. Note that these values are based on the binormal model.

### False + B/(B+D)

The proportion of those who did not have the disease who were incorrectly diagnosed by the test as having it. Note that these values are based on the binormal model.

### Specificity D/(B+D)

This is the proportion of those who did not have the disease who were correctly diagnosed as such. Note that these values are based on the binormal model.

## Cost-Benefit Analysis – Empirical Curve

**Cost - Benefit Analysis for Condition = Fever with Prevalence = 0.1 using Empirical Curve**

**Sodium1**

| Cutoff Value | Sensitivity | Specificity | Cost - Benefit When Ratio = 1.1000 | Cost - Benefit When Ratio = 1.3000 | Cost - Benefit When Ratio = 1.5000 | Cost - Benefit When Ratio = 1.7000 |
|---|---|---|---|---|---|---|
| 120.00 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 125.00 | 0.0952 | 1.0000 | 0.0952 | 0.0952 | 0.0952 | 0.0952 |
| 130.00 | 0.5238 | 0.9583 | 0.1113 | 0.0363 | -0.0387 | -0.1137 |
| 135.00 | 0.8571 | 0.7500 | -1.6179 | -2.0679 | -2.5179 | -2.9679 |
| 140.00 | 1.0000 | 0.2083 | -6.8375 | -8.2625 | -9.6875 | -11.1125 |

**Sodium2**

| Cutoff Value | Sensitivity | Specificity | Cost - Benefit When Ratio = 1.1000 | Cost - Benefit When Ratio = 1.3000 | Cost - Benefit When Ratio = 1.5000 | Cost - Benefit When Ratio = 1.7000 |
|---|---|---|---|---|---|---|
| 120.00 | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 125.00 | 0.1905 | 0.9583 | -0.2220 | -0.2970 | -0.3720 | -0.4470 |
| 130.00 | 0.5714 | 0.8750 | -0.6661 | -0.8911 | -1.1161 | -1.3411 |
| 135.00 | 0.8571 | 0.5833 | -3.2679 | -4.0179 | -4.7679 | -5.5179 |
| 140.00 | 1.0000 | 0.2917 | -6.0125 | -7.2875 | -8.5625 | -9.8375 |

Notes:
The cost-benefit ratio is the ratio of the net cost when the condition is absent to the net cost when it is present.
Select the cutoff value for which the computed cost value is maximized (or minimized).
Prevalence is the actual probability of the condition in the population.

The report displays the numeric information used to generate the ROC curve.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option on the Reports tab.

### Sensitivity

This is the proportion of those that had the disease that were correctly diagnosed by the test.

### Specificity

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

### Cost-Benefit When Ratio = 1.1

The cost-benefit ratio is the ratio of the net cost when the condition is absent to the net cost when it is present. The optimum cutoff value is that one at which the computed cost value is maximized (or minimized).

## Cost-Benefit Analysis – Binormal Curve

**Cost - Benefit Analysis for Condition = Fever with Prevalence = 0.1 using Binormal Curve**

**Sodium1**

| Cutoff Value | Sensitivity | Specificity | Cost - Benefit When Ratio = 1.1000 | Cost - Benefit When Ratio = 1.3000 | Cost - Benefit When Ratio = 1.5000 | Cost - Benefit When Ratio = 1.7000 |
|---|---|---|---|---|---|---|
| 120.00 | 0.0075 | 1.0000 | 0.0075 | 0.0075 | 0.0075 | 0.0075 |
| 125.00 | 0.0973 | 0.9996 | 0.0930 | 0.0922 | 0.0915 | 0.0907 |
| 130.00 | 0.4356 | 0.9766 | 0.2044 | 0.1623 | 0.1203 | 0.0783 |
| 135.00 | 0.8346 | 0.7415 | -1.7242 | -2.1895 | -2.6547 | -3.1200 |
| 140.00 | 0.9825 | 0.2442 | -6.4997 | -7.8600 | -9.2204 | -10.5808 |

**Sodium2**

| Cutoff Value | Sensitivity | Specificity | Cost - Benefit When Ratio = 1.1000 | Cost - Benefit When Ratio = 1.3000 | Cost - Benefit When Ratio = 1.5000 | Cost - Benefit When Ratio = 1.7000 |
|---|---|---|---|---|---|---|
| 120.00 | 0.0187 | 0.9995 | 0.0137 | 0.0127 | 0.0118 | 0.0109 |
| 125.00 | 0.1434 | 0.9890 | 0.0345 | 0.0146 | -0.0052 | -0.0250 |
| 130.00 | 0.4807 | 0.9022 | -0.4872 | -0.6632 | -0.8392 | -1.0151 |
| 135.00 | 0.8335 | 0.6174 | -2.9540 | -3.6427 | -4.3313 | -5.0200 |
| 140.00 | 0.9764 | 0.2429 | -6.5188 | -7.8816 | -9.2443 | -10.6071 |

Notes:
The cost-benefit ratio is the ratio of the net cost when the condition is absent to the net cost when it is present.
Select the cutoff value for which the computed cost value is maximized (or minimized).
Prevalence is the actual probability of the condition in the population.

The report displays the numeric information used to generate the ROC curve.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option on the Reports tab.

### Sensitivity

This is the proportion of those that had the disease that were correctly diagnosed by the test.

### Specificity

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

### Cost-Benefit When Ratio = 1.1

The cost-benefit ratio is the ratio of the net cost when the condition is absent to the net cost when it is present. The optimum cutoff value is that one at which the computed cost value is maximized (or minimized).

## Predicted Value Section – Empirical Method

**Predictive Value Section for Fever using the Empirical ROC Curve**

**Sodium1**

| Cutoff Value | Sensitivity | Specificity | Likelihood Ratio | Prev. PPV | = 0.47 NPV | Prev. PPV | = 0.10 NPV |
|---|---|---|---|---|---|---|---|
| 120.00 | 0.00000 | 1.00000 | | 0.00000 | 0.53333 | 0.00000 | 0.90000 |
| 125.00 | 0.09524 | 1.00000 | | 1.00000 | 0.55814 | 1.00000 | 0.90865 |
| 130.00 | 0.52381 | 0.95833 | 12.57143 | 0.91667 | 0.69697 | 0.58278 | 0.94768 |
| 135.00 | 0.85714 | 0.75000 | 3.42857 | 0.75000 | 0.85714 | 0.27586 | 0.97927 |
| 140.00 | 1.00000 | 0.20833 | 1.26316 | 0.52500 | 1.00000 | 0.12308 | 1.00000 |

**Sodium2**

| Cutoff Value | Sensitivity | Specificity | Likelihood Ratio | Prev. PPV | = 0.47 NPV | Prev. PPV | = 0.10 NPV |
|---|---|---|---|---|---|---|---|
| 120.00 | 0.00000 | 1.00000 | | 0.00000 | 0.53333 | 0.00000 | 0.90000 |
| 125.00 | 0.19048 | 0.95833 | 4.57143 | 0.80000 | 0.57500 | 0.33684 | 0.91420 |
| 130.00 | 0.57143 | 0.87500 | 4.57143 | 0.80000 | 0.70000 | 0.33684 | 0.94839 |
| 135.00 | 0.85714 | 0.58333 | 2.05714 | 0.64286 | 0.82353 | 0.18605 | 0.97351 |
| 140.00 | 1.00000 | 0.29167 | 1.41176 | 0.55263 | 1.00000 | 0.13559 | 1.00000 |

Notes:

| | |
|---|---|
| Sensitivity | The Pr(Positive Test\|Condition Present). |
| Specificity | The Pr(Negative Test\|Condition Absent). |
| Likelihood Ratio | The ratio Pr(Positive Test\|Condition Present)/Pr(Positive Test\|Condition Absent). |
| Prev | The prevalence of the disease. The first value is from the data. The other was input. |
| PPV | Positive Predictive Value. The Pr(Condition Present\|Positive Test). |
| NPV | Negative Predictive Value. The Pr(Condition Absent\|Negative Test). |

The report displays the information to assess the predicted value of the diagnostic test.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option on the Reports tab.

### Sensitivity

This is the proportion of those that had the disease that were correctly diagnosed by the test.

### Specificity

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

### Likelihood Ratio

The likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis. It is calculated as follows:

$$LR = \frac{\Pr(positive\ test|disease)}{\Pr(positive\ test|no\ disease)} = \frac{sensitivity}{1 - specificity}$$

### Prev = x.xxxx PPV

The values of PPV for the two prevalence values. The first prevalence value is the one that was calculated from the data. The second prevalence value was set by the user.

### Prev = x.xxxx NPV

The values of NPV for the two prevalence values. The first prevalence value is the one that was calculated from the data. The second prevalence value was set by the user.

## Predicted Value Section – Binormal Method

**Predictive Value Section for Fever using the Binormal ROC Curve**

**Sodium1**

| Cutoff Value | Sensitivity | Specificity | Likelihood Ratio | Prev. PPV | = 0.47 NPV | Prev. PPV | = 0.10 NPV |
|---|---|---|---|---|---|---|---|
| 120.00 | 0.00751 | 1.00000 | 5003.14160 | 0.99977 | 0.53521 | 0.99820 | 0.90068 |
| 125.00 | 0.09734 | 0.99957 | 223.92737 | 0.99492 | 0.55860 | 0.96136 | 0.90881 |
| 130.00 | 0.43561 | 0.97664 | 18.65033 | 0.94226 | 0.66416 | 0.67451 | 0.93966 |
| 135.00 | 0.83464 | 0.74153 | 3.22914 | 0.73860 | 0.83673 | 0.26405 | 0.97582 |
| 140.00 | 0.98246 | 0.24423 | 1.29995 | 0.53215 | 0.94088 | 0.12621 | 0.99208 |

**Sodium2**

| Cutoff Value | Sensitivity | Specificity | Likelihood Ratio | Prev. PPV | = 0.47 NPV | Prev. PPV | = 0.10 NPV |
|---|---|---|---|---|---|---|---|
| 120.00 | 0.01869 | 0.99949 | 36.74964 | 0.96984 | 0.53790 | 0.80328 | 0.90164 |
| 125.00 | 0.14344 | 0.98899 | 13.02933 | 0.91936 | 0.56888 | 0.59145 | 0.91221 |
| 130.00 | 0.48070 | 0.90223 | 4.91677 | 0.81140 | 0.66506 | 0.35330 | 0.93989 |
| 135.00 | 0.83352 | 0.61742 | 2.17868 | 0.65592 | 0.80911 | 0.19490 | 0.97091 |
| 140.00 | 0.97642 | 0.24291 | 1.28969 | 0.53018 | 0.92170 | 0.12534 | 0.98933 |

Notes:

| | |
|---|---|
| Sensitivity | The Pr(Positive Test\|Condition Present). |
| Specificity | The Pr(Negative Test\|Condition Absent). |
| Likelihood Ratio | The ratio Pr(Positive Test\|Condition Present)/Pr(Positive Test\|Condition Absent). |
| Prev | The prevalence of the disease. The first value is from the data. The other was input. |
| PPV | Positive Predictive Value. The Pr(Condition Present\|Positive Test). |
| NPV | Negative Predictive Value. The Pr(Condition Absent\|Negative Test). |

The report displays the information to assess the predicted value of the diagnostic test.

### Cutoff

The cutoff values of the criterion variable as set in the Criterion List option on the Reports tab.

### Sensitivity

This is the proportion of those that had the disease that were correctly diagnosed by the test.

### Specificity

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

### Likelihood Ratio

The likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis. It is calculated as follows:

$$LR = \frac{\Pr(positive\ test|disease)}{\Pr(positive\ test|no\ disease)} = \frac{sensitivity}{1 - specificity}$$

### Prev = x.xxxx PPV

The values of PPV for the two prevalence values. The first prevalence value is the one that was calculated from the data. The second prevalence value was set by the user.

### Prev = x.xxxx NPV

The values of NPV for the two prevalence values. The first prevalence value is the one that was calculated from the data. The second prevalence value was set by the user.

## Area Under Curve Hypothesis Tests

**Empirical Area Under Curve Analysis for Condition = Fever**

| Criterion | Empirical Estimate of AUC | AUC's Standard Error | Z-Value to Test AUC > 0.5 | 1-Sided Prob Level | 2-Sided Prob Level | Prevalence of Fever | Count |
|---|---|---|---|---|---|---|---|
| Sodium1 | 0.87500 | 0.05052 | 7.42 | 0.0000 | 0.0000 | 0.46667 | 45 |
| Sodium2 | 0.80754 | 0.06431 | 4.78 | 0.0000 | 0.0000 | 0.46667 | 45 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and hypothesis tests above use the empirical approach for correlated (paired) samples developed by DeLong, DeLong, and Clarke-Pearson.
3. The Z-Value compares the AUC to 0.5, since the AUC of a 'useless' criterion is 0.5. The one-sided test is usually used here since your only interest is that the criterion is better than 'useless'.
4. The Z test used here is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

**Binormal Area Under Curve Analysis for Condition = Fever**

| Criterion | Binormal Estimate of AUC | AUC's Standard Error | Z-Value to Test AUC > 0.5 | 1-Sided Prob Level | 2-Sided Prob Level | Prevalence of Fever | Count |
|---|---|---|---|---|---|---|---|
| Sodium1 | 0.87720 | 0.03995 | 4.70 | 0.0000 | 0.0000 | 0.46667 | 45 |
| Sodium2 | 0.81350 | 0.06379 | 3.12 | 0.0009 | 0.0018 | 0.46667 | 45 |

Notes:
1. The AUCs, SEs, and hypothesis tests above use the binormal approach given by McClish (1989).
2. The Z-Value compares the AUC to 0.5, since the AUC of a 'useless' criterion is 0.5. The one-sided test is usually used here since your only interest is that the criterion is better than 'useless'.
3. The Z tests used here are only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.
4. The Z tests use a logistic-type transformation to achieve better normality.

These reports display areas under the ROC curve and associated standard errors and hypotheses tests for each of the criterion variables. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method.

The one-sided and two-sided hypothesis tests test the hypothesis that the diagnostic test is better than flipping a coin to make the diagnosis. The actual formulas used where presented earlier in this chapter.

## Area Under Curve Confidence Intervals

**Empirical Confidence Interval of AUC for Condition = Fever**

| Criterion | Empirical Estimate of AUC | AUC's Standard Error | Lower 95.0% Confidence Limit | Upper 95.0% Confidence Limit | Prevalence of Fever | Count |
|---|---|---|---|---|---|---|
| Sodium1 | 0.87500 | 0.05052 | 0.73131 | 0.94432 | 0.46667 | 45 |
| Sodium2 | 0.80754 | 0.06431 | 0.63966 | 0.90188 | 0.46667 | 45 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and hypothesis tests above use the nonparametric approach developed by DeLong, DeLong, and Clarke-Pearson.
3. The confidence interval is based on the transformed AUC as given by Zhou et al (2002).
4. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

**Binormal Confidence Interval of AUC for Condition = Fever**

| Criterion | Binormal Estimate of AUC | AUC's Standard Error | Lower 95.0% Confidence Limit | Upper 95.0% Confidence Limit | Prevalence of Fever | Count |
|---|---|---|---|---|---|---|
| Sodium1 | 0.87720 | 0.03995 | 0.77143 | 0.93580 | 0.46667 | 45 |
| Sodium2 | 0.81350 | 0.06379 | 0.64553 | 0.90640 | 0.46667 | 45 |

Notes:
1. The AUCs, SEs, and hypothesis tests above use the binormal approach given by McClish (1989).
2. The confidence intervals given here are only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.
3. The confidence interval use a logistic-type transformation to achieve better normality.

These reports display areas under the ROC curve and associated standard errors and confidence intervals for each of the criterion variables. The first report is based on the empirical ROC method and the second

report is based on the binormal ROC method. The actual formulas used where presented earlier in this chapter.

## Tests of (AUC1-AUC2) = 0

**Empirical Test of (AUC1 - AUC2) = 0 for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | Difference Value | Difference Std Error | Difference Percent | Z-Value | Prob Level |
|---|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87500 | 0.80754 | 0.06746 | 0.02130 | -7.71 | 3.17 | 0.0015 |
| Sodium2, Sodium1 | 0.80754 | 0.87500 | -0.06746 | 0.02130 | 8.35 | -3.17 | 0.0015 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and hypothesis tests above use the nonparametric approach developed by DeLong, DeLong, and Clarke-Pearson.
3. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

**Binormal Test of (AUC1 - AUC2) = 0 for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | Difference Value | Difference Std Error | Difference Percent | Z-Value | Prob Level |
|---|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87720 | 0.81350 | 0.06371 | 0.02717 | -7.26 | 4.61 | 0.0000 |
| Sodium2, Sodium1 | 0.81350 | 0.87720 | -0.06371 | 0.02717 | 7.83 | -4.61 | 0.0000 |

Notes:
1. The AUCs, SEs, and hypothesis tests above use the binormal approach.
2. The z-test is based on a logistic-type transformation of the areas.

These reports display the results of hypothesis tests concerning the equality of two AUCs. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method. The actual formulas used where presented earlier in this chapter.

## Variances and Covariances of (AUC1-AUC2) = 0

**Empirical Test Variances and Covariances for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | AUC1 Variance | AUC2 Variance | AUC1,AUC2 Covariance | Difference Variance |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87500 | 0.80754 | 0.00255 | 0.00414 | 0.00312 | 0.00045 |
| Sodium2, Sodium1 | 0.80754 | 0.87500 | 0.00414 | 0.00255 | 0.00312 | 0.00045 |

Notes:
1. These AUCs, variances, and covariances use the nonparametric approach developed by DeLong, DeLong, and Clarke-Pearson.
2. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

**Binormal Test Variances and Covariances for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | AUC1 Variance | AUC2 Variance | AUC1,AUC2 Covariance | Difference Variance |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87720 | 0.81350 | 0.00160 | 0.00407 | 0.00246 | 0.00074 |
| Sodium2, Sodium1 | 0.81350 | 0.87720 | 0.00407 | 0.00160 | 0.00246 | 0.00074 |

Notes:
1. The AUCs, SEs, and hypothesis tests above use the binormal approach.
2. The z-test is based on a logistic-type transformation of the areas.

These reports display the variances and covariances associated with the hypothesis tests given in the last set of reports. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method. The actual formulas used where presented earlier in this chapter.

## Confidence Intervals of (AUC1-AUC2) = 0

**Empirical Confidence Intervals of Differences in AUCs for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | Difference Value | Difference Std Error | Lower 95.0% Confidence Limit | Upper 95.0% Confidence Limit |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87500 | 0.80754 | 0.06746 | 0.02130 | 0.02571 | 0.10921 |
| Sodium2, Sodium1 | 0.80754 | 0.87500 | -0.06746 | 0.02130 | -0.10921 | -0.02571 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and confidence limits above use the nonparametric approach developed by DeLong, DeLong, and Clarke-Pearson.
3. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

**Binormal Confidence Interval of Difference in AUCs for Condition = Fever**

| Criterions 1,2 | AUC1 | AUC2 | Difference Value | Difference Std Error | Lower 95.0% Confidence Limit | Upper 95.0% Confidence Limit |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.87720 | 0.81350 | 0.06371 | 0.02717 | 0.01046 | 0.11696 |
| Sodium2, Sodium1 | 0.81350 | 0.87720 | -0.06371 | 0.02717 | -0.11696 | -0.01046 |

Notes:
1. The AUCs, SEs, and hypothesis tests above use the binormal approach.
2. The z-test is based on a logistic-type transformation of the areas.

These reports display the confidence intervals for difference between a pair of AUCs. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method. The actual formulas used where presented earlier in this chapter.

## Equivalence Tests Comparing AUC1 and AUC2

**Empirical Equivalence Test of the AUCs for Condition = Fever**

| Criterion Variable1, Variable2 | Prob Level | Lower 90.0% Conf. Limit | Upper 90.0% Conf. Limit | Lower Equiv. Bound | Upper Equiv. Bound | Reject H0 and Conclude Equivalence at the 5.0% Significance Level |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.7938 | 0.03242 | 0.10250 | -0.05000 | 0.05000 | Cannot reject H0 |
| Sodium2, Sodium1 | 0.7938 | -0.10250 | -0.03242 | -0.05000 | 0.05000 | Cannot reject H0 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and confidence limits above use the empirical approach developed by DeLong, DeLong, and Clarke-Pearson.
3. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.
4. Equivalence means that Test2 does not differ from Test1 except for a small, negligible amount which we call the 'Equivalence Bound'.

**Binormal Equivalence Test of the AUCs for Condition = Fever**

| Criterion Variable1, Variable2 | Prob Level | Lower 90.0% Conf. Limit | Upper 90.0% Conf. Limit | Lower Equiv. Bound | Upper Equiv. Bound | Reject H0 and Conclude Equivalence at the 5.0% Significance Level |
|---|---|---|---|---|---|---|
| Sodium1, Sodium2 | 0.6930 | 0.01902 | 0.10839 | -0.05000 | 0.05000 | Cannot reject H0 |
| Sodium2, Sodium1 | 0.6930 | -0.10839 | -0.01902 | -0.05000 | 0.05000 | Cannot reject H0 |

Notes:
1. The AUCs, SEs, and confidence limits above use the binormal approach.
2. Equivalence means that Test2 does not differ from Test1 except for a small, negligible amount which we call the 'Equivalence Bound'.
3. The logistic-type transformation is not used in these calculations.

These reports display the results of an equivalence test. This hypothesis test tests whether the two diagnostics are equivalence in the sense that their AUC's are no more different than the maximum amount specified. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method.

Often, you want to show that one diagnostic test is equivalent to another. In this case, you would use a test of equivalence.

## Noninferiority Tests Comparing AUC1 and AUC2

**Empirical Noninferiority Test of the AUCs for Condition = Fever**

| Criterion Variable1, Variable2 | Prob Level | 1-Sided 95.0% Conf. Limit | Noninferiority Bound | Reject H0 and Conclude Noninferiority at the 5.0% Significance Level |
|---|---|---|---|---|
| Sodium1, Sodium2 | 0.0000 | 0.03242 | 0.05000 | Yes: (AUC1-AUC2)<0.05 |
| Sodium2, Sodium1 | 0.7938 | -0.10250 | 0.05000 | Cannot reject H0 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and confidence limits above use the empirical approach developed by DeLong, DeLong, and Clarke-Pearson.
3. This method is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.
4. Noninferiority means that Test2 is no worse than Test1 except for a small, negligible amount which we call the 'Noninferiority Bound'.

**Binormal Noninferiority Test of the AUCs for Condition = Fever**

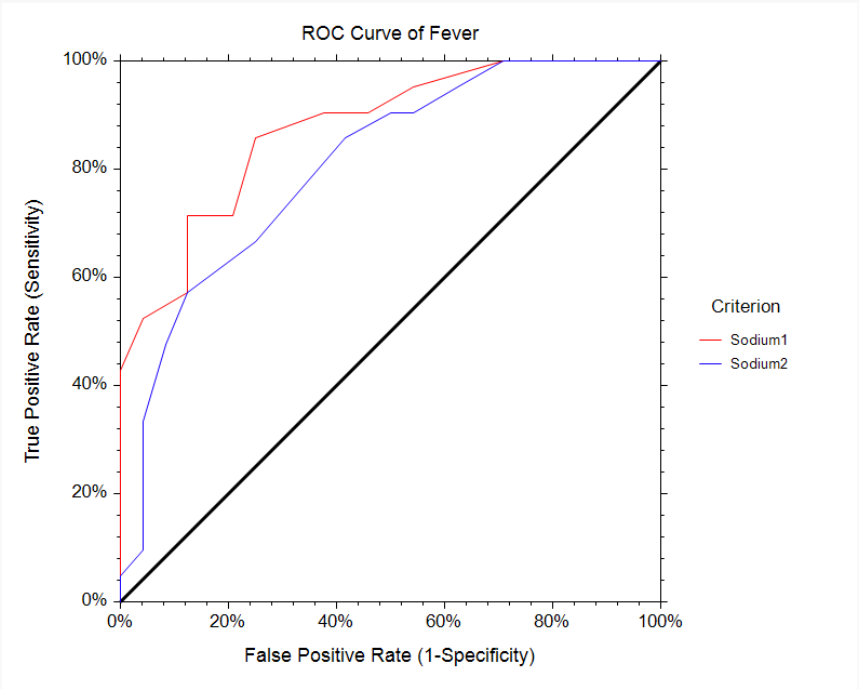
| Criterion Variable1, Variable2 | Prob Level | 1-Sided 95.0% Conf. Limit | Noninferiority Bound | Reject H0 and Conclude Noninferiority at the 5.0% Significance Level |
|---|---|---|---|---|
| Sodium1, Sodium2 | 0.0000 | 0.01902 | 0.05000 | Yes: (AUC1-AUC2)<0.05 |
| Sodium2, Sodium1 | 0.6930 | -0.10839 | 0.05000 | Cannot reject H0 |

Notes:
1. The AUCs, SEs, and confidence limits above use the binormal approach.
2. Noninferiority means that Test2 is no worse than Test1 except for a small, negligible amount which we call the 'Noninferiority Bound'.
3. The logistic-type transformation is not used in these calculations.

These reports display the results of a noninferiority test. This hypothesis test tests whether diagnostic test 2 is no worse than diagnostic test 1 in the sense that AUC2 is not less than AUC1 by more than a small amount. The first report is based on the empirical ROC method and the second report is based on the binormal ROC method.

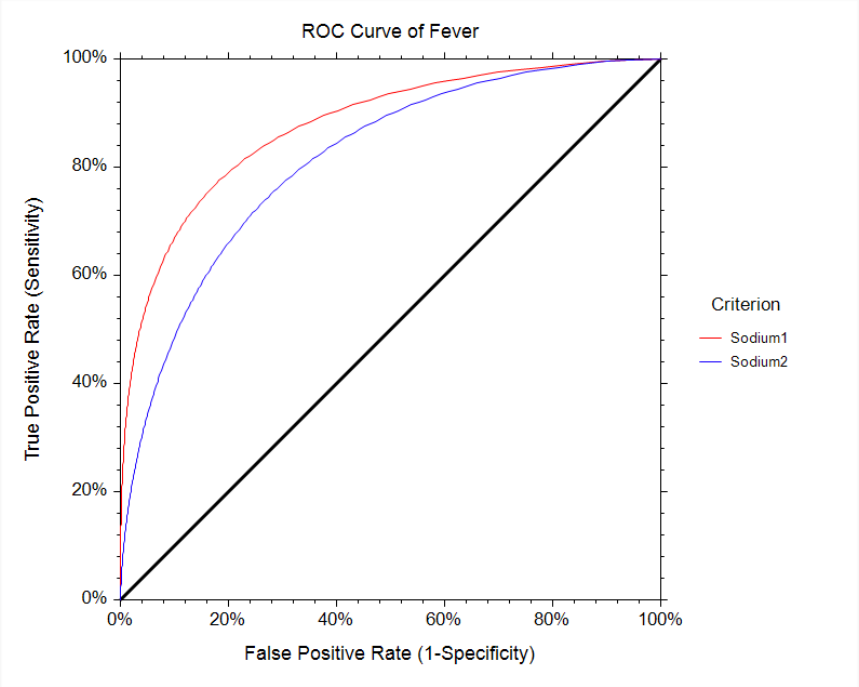
# ROC Plot Section

**ROC Plot Section**




A 2nd run is required to generate the plot below after clicking on the ROC Plot format button and changing the check box to Binormal ROC Line.

**ROC Plot Section**



Both the empirical and binormal ROC curves are displayed here (a second run is needed to get the second plot after clicking on the ROC Plot Format button and changing the check box to Binormal ROC Line). The empirical curve is shown first. It always has a 'zig-zag' pattern. The smooth, binormal ROC curve is shown second.

The ROC curves plot the proportion of those who actually had the disease who were correctly diagnosed on the vertical axis versus the proportion of those who did not have the disease who were falsely diagnosed as having it on the horizontal axis. Hence, an optimum test procedure is one whose ROC curve proceeds from the lower-left corner vertically until it reaches the top and then horizontally across the top to the right side. The 45 degree line represents what you would expect from a chance (flip of the coin) classification procedure.

When you are comparing two curves as in this example, you would generally take the outside curve (the one furthest from the middle line). However, it is possible for the curves to cross so that one test is optimum in a certain range but not in another.

# Example 2 – Validation Using Zhou et al. (2002)

Zhou et al. (2002) page 175 presents an example comparing the results of two mammography tests: plain film and digitized film. In this example, both tests were administered to 58 people, yielding a paired design. The results given in Zhou (2002) contain a few typos. We have obtained the corrected results from the authors, which are as follows:

| Test | AUC | SE |
|---|---|---|
| Plain | 0.83504 | 0.06581 |
| Digitized | 0.84701 | 0.05987 |

## Setup

To run this example, complete the following steps:

**1 Open the Zhou175 example dataset**

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Zhou175** and click **OK**.

**2 Specify the ROC Curves (Old Version) procedure options**

- Find and open the **ROC Curves (Old Version)** 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

| | |
|---|---|
| Actual Condition Variable | **Cancer** |
| Positive Condition Value | **1** |
| Actual Condition Prevalence | **0.1** |
| Cost Benefit Ratio(s) | **1** |
| Frequency (Count) Variable | **Count** |
| Criterion Variable(s) | **PlainFilm-DigiFilm** |
| Test Direction | **High X = Positive** |
| Criterion List | **1 2 3 4** |
| Frequency (Count) Variable | **Count** |
| Max Equivalence Difference | **0.05** |

Reports Tab

| | |
|---|---|
| Area Under Curve (AUC) Analysis: One Curve | **Checked** |
| Skip Line After | **20** |

**3 Run the procedure**

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

## Output

**Empirical Area Under Curve Analysis for Condition = Cancer**

| Criterion | Empirical Estimate of AUC | AUC's Standard Error | Z-Value to Test AUC > 0.5 | 1-Sided Prob Level | 2-Sided Prob Level | Prevalence of Cancer | Count |
|---|---|---|---|---|---|---|---|
| PlainFilm | 0.83504 | 0.06581 | 5.09 | 0.0000 | 0.0000 | 0.22414 | 58 |
| DigiFilm | 0.84701 | 0.05987 | 5.80 | 0.0000 | 0.0000 | 0.22414 | 58 |

Notes:
1. This approach underestimates AUC when there are only a few (3 to 7) unique criterion values.
2. The AUCs, SEs, and hypothesis tests above use the empirical approach for correlated (paired) samples developed by DeLong, DeLong, and Clarke-Pearson.
3. The Z-Value compares the AUC to 0.5, since the AUC of a 'useless' criterion is 0.5. The one-sided test is usually used here since your only interest is that the criterion is better than 'useless'.
4. The Z test used here is only accurate for larger samples with at least 30 subjects with, and another 30 subjects without, the condition of interest.

Note that the estimated AUC's and standard errors match those given by Zhou (2002) exactly.