

Chapter 265

Tests for Two ROC Curves

Introduction

Receiver operating characteristic (ROC) curves are used to summarize the accuracy of diagnostic tests. The technique is used when a criterion variable is available which is used to make a yes or no decision. The area under the ROC curve (AUC) is a popular summary index of an ROC curve.

This module computes power and sample size for comparing the AUC's of two diagnostic tests obtained from the same patients. The methodology of Obuchowski and McClish (1997) is used when the criterion variable yields a discrete value. The methodology of Hanley and McNeil (1983) is used when the criterion variable yields a continuous value.

Technical Details

In the following, we suppose that we have two groups of patients, those with a condition of interest (the disease) and those without it. A patient's classification may be known from extensive diagnosis or based on the value of another diagnostic test. The diagnostic tests of interest are performed on each patient and the resulting test values are recorded. At each specified cutoff value of the criterion variable, the true positive rate (TPR) and the false positive rate (FPR) are calculated. An ROC curve is generated by plotting TPR versus FPR. The plot allows the consequences of using various cutoff values to be evaluated. The area under the ROC curve, either for the whole or partial range, is often used as a summary measure of the accuracy of the test.

It should be noted that TPR is similar to the statistical power of the diagnostic test at a particular cutoff value of the criterion variable. Similarly, FPR is an estimate of the probability that the diagnostic test results in a type I (alpha) error. Thus the ROC curve may be interpreted as a plot of the diagnostic test's power versus its significance level at various possible criterion cutoff values.

Users of ROC curves have developed special names for TPR and FPR. They call TPR the *sensitivity* of the test and $1 - \text{FPR}$ the *specificity* of the test. Statisticians will be more familiar with using the word *power* instead of sensitivity and the phrase ' $1 - \alpha$ ' instead of specificity.

An ROC curve may be summarized by the area under it (AUC). This area has an additional interpretation. Suppose that a rater is asked to study two subjects, one that is actually disease positive and one that is disease negative. The AUC is equal to the probability that the rater will give the disease positive subject a higher score than the disease negative subject. That is, the AUC is the probability that the rater will correctly order the two subjects as to which is more likely to have the disease.

Several methods of computing the AUC have been proposed. One method uses the trapezoidal rule to calculate the AUC directly. Another method, called the *binormal model*, computes the area by fitting two normal distributions to the data.

Tests for Two ROC Curves

The Binormal Model

Let X denote the distribution of the criterion variable for normal (non-diseased) patients and Y denote the distribution of the criterion variable for abnormal (diseased) patients. It is assumed that

$$X \sim N(\mu_-, \sigma_-^2)$$

and

$$Y \sim N(\mu_+, \sigma_+^2)$$

The partial area under the ROC curve, AUC , is defined as

$$\theta_i = \int_{c_1}^{c_2} \Phi(A_i + B_i v) \phi(v) dv$$

where $\Phi(z)$ is the cumulative normal distribution, $c_j = \Phi^{-1}(FPR_j)$, and

$$A_i = \frac{\mu_{i+} - \mu_{i-}}{\sigma_{i+}}$$

$$B_i = \frac{\sigma_{i-}}{\sigma_{i+}}$$

Note that for the full range area under the curve, $c_1 = -\infty$ and $c_2 = \infty$.

Maximum likelihood estimates of A and B can be computed. The variances and covariance of these MLE's can be estimated from Fisher's information matrix.

Define $\Delta = \theta_1 - \theta_2$ to be the difference in the accuracies (AUC's) of the two tests. A test of whether the two AUC's are different amounts to testing whether $\Delta = 0$. The test statistic for this test is

$$Z = \frac{\hat{\Delta} - 0}{\sqrt{\text{var}_0(\hat{\Delta})}}$$

where $\text{var}_0(\hat{\Delta})$ is the variance of $\hat{\Delta}$ under the null hypothesis of equality. The above test statistic results in the following formula for computing sample size

$$N_+ = \frac{\left(z_\alpha \sqrt{V_0(\hat{\Delta})} + z_\beta \sqrt{V_{Alt}(\hat{\Delta})} \right)^2}{\Delta^2}$$

Rating Data

When the criterion values are discrete rating values, Obuchowski and McClish (1997) showed that the variances could be calculated using

$$V_0(\hat{\Delta}) = V(\hat{\theta}_1) + V(\hat{\theta}_1) - 2C(\hat{\theta}_1, \hat{\theta}_1)$$

$$V_{Alt}(\hat{\Delta}) = V(\hat{\theta}_1) + V(\hat{\theta}_2) - 2C(\hat{\theta}_1, \hat{\theta}_2)$$

Tests for Two ROC Curves

where

$$V(\hat{\theta}_i) = f_i^2 \left(1 + \frac{B_i^2}{R} + \frac{A_i^2}{2} \right) + g_i^2 \left(B_i^2 \left(\frac{1+R}{2R} \right) \right)$$

$$C(\hat{\theta}_1, \hat{\theta}_2) = f_1 f_2 \left(r_+ + r_- \frac{B_1 B_2}{R} + r_+^2 \frac{A_1 A_2}{2} \right) + \frac{g_1 g_2 B_1 B_2 (r_-^2 + R r_+^2)}{2R} + \frac{f_1 g_2 A_1 B_2 r_+^2}{2} + \frac{f_2 g_1 A_2 B_1 r_+^2}{2}$$

$$f_i = \frac{E_{1i} E_{3i}}{\sqrt{2\pi E_{2i}}}$$

$$g_i = \frac{E_{1i} E_{4i}}{\sqrt{2\pi E_{2i}}} - \frac{A_i B_i E_{1i} E_{3i}}{\sqrt{2\pi E_{2i}^3}}$$

$$E_{1i} = \exp\left(-\frac{A_i^2}{2 + 2B_i^2}\right)$$

$$E_{2i} = 1 + B_i^2$$

$$E_{3i} = \Phi(c_2) - \Phi(c_1)$$

$$E_{4i} = \exp\left(-\frac{c_1^2}{2}\right) - \exp\left(-\frac{c_2^2}{2}\right)$$

$$c_j = \frac{\Phi^{-1}(FPR_j) + \frac{A_j B_j}{1 + B_j^2}}{\sqrt{1 + B_j^2}}$$

$$R = \frac{N_-}{N_+}$$

$$A_i = B_i \Phi^{-1}(TNR_i) - \Phi^{-1}(FPR_i)$$

r_- and r_+ are the correlations between the results of the two diagnostics tests for normal and abnormal patients, respectively. For the most conservative results, set $B_i = 1$.

Continuous Data

When the criterion values are continuous, Obuchowski (1998) suggests that the following formulas of Hanley and McNeil (1983) are more appropriate. Note that these formulas cannot be used for evaluating the AUC for a partial range.

$$V(\hat{\Delta}) = V(\hat{\theta}_1) + V(\hat{\theta}_2) - 2C(\hat{\theta}_1, \hat{\theta}_2)$$

Tests for Two ROC Curves

where

$$V(\hat{\theta}_i) = \frac{\theta_i}{R(2 - \theta_i)} + \frac{2\theta_i^2}{1 + \theta_i} - \theta_i^2 \left(\frac{1 + R}{R} \right)$$

$$C(\hat{\theta}_1, \hat{\theta}_2) = 2r\sqrt{V(\theta_1)V(\theta_2)}$$

and r is derived from a special table provided by Hanley and McNeil (1983).

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* or *Sample Size (N+)*.

Select *Sample Size (N+)* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment that has already been run.

Test

Alternative Hypothesis

Specify whether the test is *one-sided* or *two-sided*. When a two-sided test is selected, the value of alpha is divided by two.

Note that most researchers assume that, unless stated otherwise, all statistical tests are two-sided. If you use a one-sided test, you should clearly state and justify this in all reports.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Tests for Two ROC Curves

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on N_+ or N_- or both.

The options are

- **Equal ($N_+ = N_-$)**
This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.
- **Enter N_+ , solve for N_-**
Select this option when you wish to fix N_+ at some value (or values), and then solve only for N_- . Please note that for some values of N_+ , there may not be a value of N_- that is large enough to obtain the desired power.
- **Enter N_- , solve for N_+**
Select this option when you wish to fix N_- at some value (or values), and then solve only for N_+ . Please note that for some values of N_- , there may not be a value of N_+ that is large enough to obtain the desired power.
- **Enter $R = N_-/N_+$, solve for N_+ and N_-**
For this choice, you set a value for the ratio of N_- to N_+ , and then PASS determines the needed N_+ and N_- , with this ratio, to obtain the desired power. An equivalent representation of the ratio, R , is

$$N_- = R * N_+.$$
- **Enter percentage in Group 1, solve for N_+ and N_-**
For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed N_+ and N_- with this percentage to obtain the desired power.

N_+ (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter N_+ , solve for N_- "

N_+ is the number of items or individuals sampled from the Group 1 population.

N_+ must be ≥ 2 . You can enter a single value or a series of values.

N_- (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N_- , solve for N_+ "

N_- is the number of items or individuals sampled from the Group 2 population.

N_- must be ≥ 2 . You can enter a single value or a series of values.

Tests for Two ROC Curves

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N-/N+, solve for N+ and N-."

R is the ratio of N- to N+. That is,

$$R = N- / N+.$$

Use this value to fix the ratio of N- to N+ while solving for N+ and N-. Only sample size combinations with this ratio are considered.

N- is related to N+ by the formula:

$$N- = [R \times N+],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1 (e.g., $N+ = 10$ and $N- = 20$, or $N+ = 50$ and $N- = 100$).

R must be greater than 0. If $R < 1$, then N- will be less than N+; if $R > 1$, then N- will be greater than N+. You can enter a single or a series of values.

Percent in Group 1 (+)

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N+ and N-."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for N+ and N-. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When Not Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal (N+ = N-)**
This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.
- **Enter N+ and N- individually**
This choice permits you to enter different values for N+ and N-.
- **Enter N+ and R, where N- = R * N+**
Choose this option to specify a value (or values) for N+, and obtain N- as a ratio (multiple) of N+.
- **Enter total sample size and percentage in Group 1**
Choose this option to specify a value (or values) for the total sample size (N), obtain N+ as a percentage of N, and then N- as N - N+.

Tests for Two ROC Curves

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal ($N_+ = N_-$)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for N_+ , and also the value for N_- .

The Sample Size Per Group must be ≥ 2 . You can enter a single value or a series of values.

N_+ (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter N_+ and N_- individually" or "Enter N_+ and R , where $N_- = R * N_+$."*

N_+ is the number of items or individuals sampled from the Group 1 population.

N_+ must be ≥ 2 . You can enter a single value or a series of values.

N_- (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N_+ and N_- individually."

N_- is the number of items or individuals sampled from the Group 2 population.

N_- must be ≥ 2 . You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N_+ and R , where $N_- = R * N_+$."*

R is the ratio of N_- to N_+ . That is,

$$R = N_-/N_+$$

Use this value to obtain N_- as a multiple (or proportion) of N_+ .

N_- is calculated from N_+ using the formula:

$$N_- = [R \times N_+],$$

where the value $[Y]$ is the next integer $\geq Y$.

For example, setting $R = 2.0$ results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If $R < 1$, then N_- will be less than N_+ ; if $R > 1$, then N_- will be greater than N_+ . You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines N_+ and N_- .

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

You can enter a single value or a series of values.

Percent in Group 1 (+)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Tests for Two ROC Curves

Effect Size – Area Under the Curve

AUC1 (Area Under Curve 1)

Specify one or more values of the AUC for diagnostic test 1. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis).

Since the AUC may include a portion of the ROC curve that is not of interest because the FPR values are unrealistic, you may be interested in only a portion of the area. In this case, you can specify a range of FPR values for which the area is to be calculated. Unfortunately, the definition of the area becomes more difficult. When analyzing the whole ROC curve, the area is known to be between 0.50 and 1.0. Following the suggestion of Obuchowski and McClish (1997), the following transformation is applied so that the values of AUC remain between 0.5 and 1.0.

$$AUC' = \frac{1}{2} \left(1 + \frac{AUC - \min}{\max - \min} \right)$$

where

$$\max = FPR2 - FPR1$$

$$\min = \frac{\max}{2} (FPR2 + FPR1)$$

Thus, when a partial range is entered for FPR1 and FPR2, the values entered here are assumed to be AUC' and are translated to AUC using the above formulas.

AUC2 (Area Under Curve 2)

Specify one or more values of the AUC for diagnostic test 2. The range of values is from 0.5 (indicative of a test useless in diagnosis) to 1.0 (indicative of a test that is perfect in diagnosis). Note that, as discussed above, this is the value of AUC' when a partial area is being analyzed.

Effect Size – False Positive Rate Limits

Lower FPR

This option specifies the lower (left) limit of the false positive rate (FPR) for which the area is to be computed. If the area under the whole ROC curve is wanted, set this value to 0.0. If the partial area is wanted, set this value to the desired left limit.

Note that the range of possible values is $0.0 \leq \text{Lower FPR} < \text{Upper FPR} \leq 1.0$

Upper FPR

This option specifies the upper (right) limit of the false positive rate (FPR). If the area under the whole ROC curve is wanted, set this value to 1.0. If the partial area is wanted, set this value to the desired right limit.

Note that the range of possible values is $0.0 \leq \text{Lower FPR} < \text{Upper FPR} \leq 1.0$

Effect Size – Correlations

Correlation+

This is the correlation between the two diagnostic-test scores for the positive group. Although correlations can range between -1 and 1, typical values are from 0.3 to 0.6.

Note that if you want to analyze a design in which a separate set of patients receive each diagnostic test, this may be done by setting this correlation value to 0.

Tests for Two ROC Curves

Correlation-

This is the correlation between the two diagnostic-test scores for the negative group. Although correlations can range between -1 and 1, typical values are from 0.3 to 0.6.

Note that if you want to analyze a design in which a separate set of patients receive each diagnostic test, this may be done by setting this correlation value to 0.

Effect Size – Type of Data

Type of Data

Specify the type of data that will be collected from the tests. The formulas for the variance are determined by this option. Possible types are:

- **Continuous**

The test results are from a continuum of possible values. The Hanley and McNeil (1983) variance formulas are used. Note that this option does not allow a partial range of FPR values to be analyzed.

- **Discrete (Ratings)**

The test results are from a small set of rating values such as 1, 2, 3, 4, 5. The Obuchowski & McClish (1997) variance formulas are used.

B1 (SD Ratio)

$B1$ is the ratio of the standard deviation of the negative group to the positive group (SD_-/SD_+) for diagnostic test 1. That is, assuming the binormal model

$$B1 = \frac{\sigma_{1-}}{\sigma_{1+}}$$

Note that this parameter is ignored for continuous data.

Although $B1$ can be any positive number, typical values are between 0.3 and 3.0. Obuchowski suggests that if the value of $B1$ is not known, a value of 1.0 is used since this will result in a conservative (extra large) sample size. She reports that in her experience, typical values are much less than 1.0, often near 0.3.

B2 (SD Ratio)

$B2$ is the ratio of the standard deviation of the negative group to the positive group (SD_-/SP_+) for diagnostic test 2. That is, assuming the binormal model

$$B2 = \frac{\sigma_{2-}}{\sigma_{2+}}$$

Note that this parameter is ignored for continuous data.

Although $B2$ can be any positive number, typical values are between 0.3 and 3.0. Obuchowski suggests that if the value of $B2$ is not known, a value of 1.0 is used since this will result in a conservative (extra large) sample size. She reports that in her experience, typical values are much less than 1.0, often near 0.3.

Tests for Two ROC Curves

Example 1 – Calculating Power

An investigator wants to compare the accuracy of two diagnostic tests which yield measurements on a rating scale from 1 to 5. Historically, such tests have had an AUC of 0.80. The investigator wants to investigate three alternative AUC values: 0.825, 0.850, and 0.900. A two-sided test is planned with a significance level of 0.05. Historically, both the positive and negative correlations between the responses on two such tests have been close to 0.60. Since no other information is available, B1 and B2 are both set to 1.0. The investigator would like to achieve a power of 90% in the study. Patients without the disease under study are about twice as frequent as patients with the disease. The investigator wants to see results for a sample size of up to 6000 patients.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two ROC Curves** procedure window by clicking on **ROC**, and then clicking on **Tests for Two ROC Curves**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Power
Alternative Hypothesis	Two-Sided Test
Alpha	0.05
Group Allocation	Enter N+ and R, where N- = R x N+
N+ (Size of Positive Group)	20 50 100 250 500 1000 2000
R (Sample Allocation Ratio)	2
AUC1 (Area Under Curve 1)	0.80
AUC2 (Area Under Curve 2)	0.825 0.85 0.9
Lower FPR	0.00
Upper FPR	1.00
Correlation+	0.6
Correlation-	0.6
Type of Data	Discrete (Ratings)
B1 (SD Ratio)	1
B2 (SD Ratio)	1

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Report

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data

Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Power	N+	N-	N	Target R	Actual R	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.0501	20	40	60	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.0733	50	100	150	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.1084	100	200	300	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.2104	250	500	750	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.3744	500	1000	1500	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.6426	1000	2000	3000	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.9090	2000	4000	6000	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050

(report continues)

Tests for Two ROC Curves

Report Definitions

Power is the probability of rejecting a false null hypothesis.
 N+ and N- are the number of items sampled from each population.
 N is the total sample size, N+ + N-.
 Target R is the desired ratio (or ratios) of R entered in the procedure. R is the ratio of N- to N+, so that
 $N- = R \times N+$.
 Actual R is the value for R obtained in this scenario. Because N+ and N- are discrete, this value is sometimes slightly different than the target R.
 AUC1' and AUC2' are the adjusted areas under the ROC curve for diagnostic tests 1 and 2, respectively.
 Diff' is AUC2 - AUC1. This is the adjusted difference to be detected.
 AUC1 and AUC2 are the actual areas under the ROC curve for diagnostic tests 1 and 2, respectively.
 Diff is AUC2 - AUC1. This is the difference to be detected.
 Alpha is the probability of rejecting a true null hypothesis.
 FPR1, FPR2 are the lower and upper bounds on the false positive rates.
 B1 and B2 are the ratios of the standard deviations of the negative and positive groups for each test.

Summary Statements

A sample of 20 from the positive group and 40 from the negative group achieve 5% power to detect a difference of 0.0250 between a diagnostic test with an area under the ROC curve (AUC) of 0.8000 and another diagnostic test with an AUC of 0.8250 using a two-sided z-test at a significance level of 0.0500. The data are discrete (rating scale) responses. The AUC is computed between false positive rates of 0.000 and 1.000. The ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group for diagnostic test 1 is 1.000 and for diagnostic test 2 is 1.000. The correlation between the two diagnostic tests is assumed to be 0.600 for the positive group and 0.600 for the negative group.

This report shows the power for each of the sample sizes. Most of the definitions are standard. However, a special explanation must be given for AUC and AUC'.

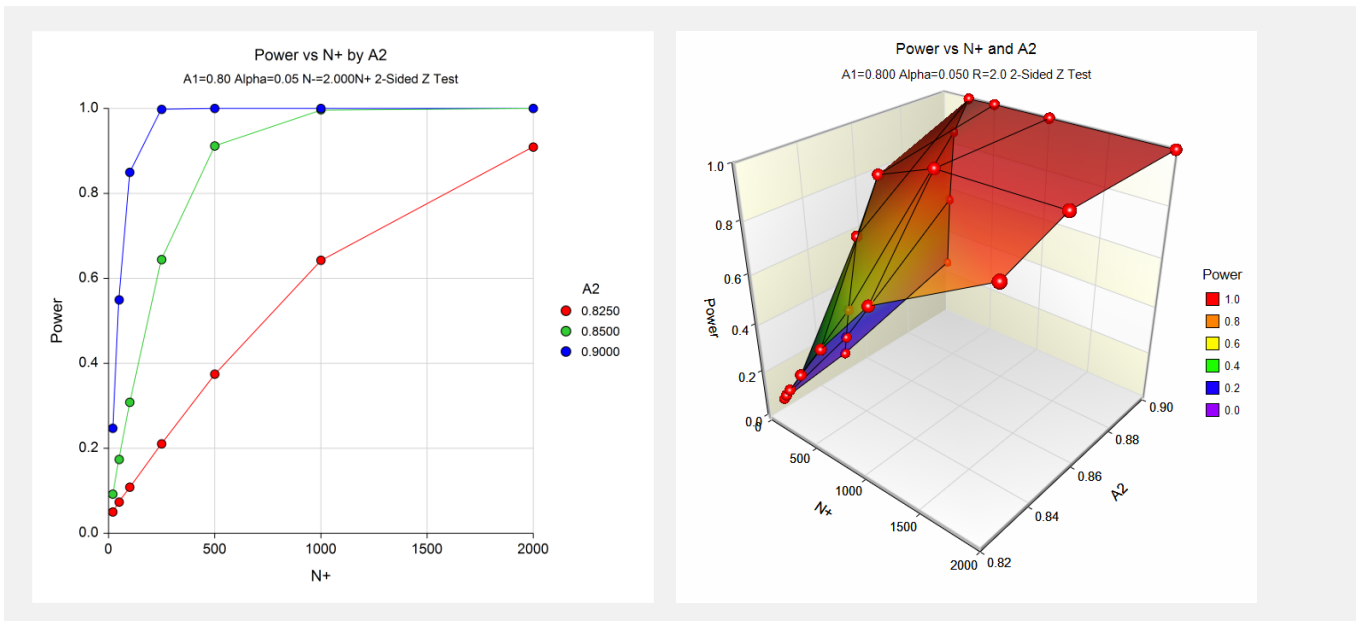
AUC'

This is the adjusted area under the curve. A rescaling, discussed earlier, has been applied so that the minimum area is 0.5 and the maximum area is 1.0.

AUC

This is the actual area under the curve. This value will equal the adjusted area when the FPR range is set from 0.0 to 1.0. Otherwise, these values will be different.

Plots Section



These plots show the power versus the sample size for the three values of AUC1.

Tests for Two ROC Curves

Example 2 – Calculating Sample Size

Continuing Example 1, the investigator wants to know the exact sample size needed for each of the three values of AUC2. The investigator wants to look at the Numeric Report.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two ROC Curves** procedure window by clicking on **ROC**, and then clicking on **Tests for Two ROC Curves**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided Test
Power	0.90
Alpha	0.05
Group Allocation	Enter R = N-/N+, solve for N+ and N-
R (Sample Allocation Ratio)	2
AUC1 (Area Under Curve 1)	0.80
AUC2 (Area Under Curve 2)	0.825 0.85 0.9
Lower FPR	0.00
Upper FPR	1.00
Correlation+	0.6
Correlation-	0.6
Type of Data	Discrete (Ratings)
B1 (SD Ratio)	1
B2 (SD Ratio)	1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data

Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Target Power	Actual Power	N+	N-	N	Target R	Actual R	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.90	0.9001	1937	3874	5811	2.0	2.0	0.8000	0.8250	0.0250	0.8000	0.8250	0.0250	0.050
0.90	0.9002	480	960	1440	2.0	2.0	0.8000	0.8500	0.0500	0.8000	0.8500	0.0500	0.050
0.90	0.9012	117	234	351	2.0	2.0	0.8000	0.9000	0.1000	0.8000	0.9000	0.1000	0.050

This report shows the sample size needed to achieve 90% power for each value of AUC2.

Tests for Two ROC Curves

Example 3 – Partial Area Under Curve

Continuing Example 2, the investigator knows that FPR values between 0.0 and 0.20 are the only values of interest. Hence, he wants to investigate the sample size needed when the FPR range is confined to this range.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two ROC Curves** procedure window by clicking on **ROC**, and then clicking on **Tests for Two ROC Curves**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided Test
Power.....	0.90
Alpha.....	0.05
Group Allocation	Enter R = N-/N+, solve for N+ and N-
R (Sample Allocation Ratio)	2
AUC1 (Area Under Curve 1).....	0.80
AUC2 (Area Under Curve 2).....	0.825 0.85 0.9
Lower FPR.....	0.00
Upper FPR.....	0.20
Correlation+	0.6
Correlation-	0.6
Type of Data	Discrete (Ratings)
B1 (SD Ratio).....	1
B2 (SD Ratio).....	1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data

Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Target Power	Actual Power	N+	N-	N	Target R	Actual R	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.90	0.9000	4095	8190	12285	2.0	2.0	0.8000	0.8250	0.0250	0.1280	0.1370	0.0090	0.050
0.90	0.9002	1012	2024	3036	2.0	2.0	0.8000	0.8500	0.0500	0.1280	0.1460	0.0180	0.050
0.90	0.9001	242	484	726	2.0	2.0	0.8000	0.9000	0.1000	0.1280	0.1640	0.0360	0.050

Note that the necessary sample size has more than doubled.

Tests for Two ROC Curves

Example 4 – Validation using Obuchowski

The formulas used in this module were given in Obuchowski and McClish (1997). On pages 1538 - 1540, they provide an example which will be duplicated here. The study compared an automated classification system with an expert mammographer in their ability to find malignant breast lesions. The measure of diagnostic accuracy is the AUC from an FPR of 0.0 to an FPR of 0.2. The allocation ratio is 2. $B_1 = B_2 = 1.0$. Correlation+ = Correlation- = 0.6. The values of A1 and A2 are found to be 2.6 and 1.9. These translate to adjusted AUC's of 0.922222 and 0.819444. A two-tailed test is envisioned in which alpha is 0.05. A power of 80% is desired. In their article, they found $N_+ = 109$ and $N_- = 218$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two ROC Curves** procedure window by clicking on **ROC**, and then clicking on **Tests for Two ROC Curves**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	Two-Sided Test
Power	0.80
Alpha	0.05
Group Allocation	Enter R = N-/N+, solve for N+ and N-
R (Sample Allocation Ratio)	2
AUC1 (Area Under Curve 1)	0.922222
AUC2 (Area Under Curve 2)	0.819444
Lower FPR	0.00
Upper FPR	0.20
Correlation+	0.6
Correlation-	0.6
Type of Data	Discrete (Ratings)
B1 (SD Ratio)	1
B2 (SD Ratio)	1

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results**Numeric Results for Testing AUC1 = AUC2 with Discrete (Rating) Data**

Test Type = Two-Sided. FPR1 = 0.0. FPR2 = 0.200. B1 = 1.000. B2 = 1.000. Allocation Ratio = 2.000.

Target Power	Actual Power	N+	N-	N	Target R	Actual R	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.80	0.8027	109	218	327	2.0	2.0	0.9222	0.8194	-0.1028	0.1720	0.1350	-0.0370	0.050

Note that the sample sizes of 109 and 218 match exactly with the results of Obuchowski.

Tests for Two ROC Curves

Example 5 – Validation using Hanley

The formulas for continuous data were given in Hanley and McNeil (1982). On page 34 of their article they provide a table of sample sizes calculated using their formulas. We will duplicate their results for $AUC1 = 0.70$ and $AUC2 = 0.75$. Using a one-sided test of significance with $\alpha = 0.05$ and a sample allocation ratio of 1.0, they found the number of subjects for both the positive and negative groups to be 652, 897, and 1131 for statistical powers of 80%, 90%, and 95%, respectively.

When using Hanley and McNeil's formulation, the values of $B1$, $B2$, $FPR1$, and $FPR2$ are ignored. Also, in this case, the correlations are set to 0.0.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two ROC Curves** procedure window by clicking on **ROC**, and then clicking on **Tests for Two ROC Curves**. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Alternative Hypothesis	One-Sided Test
Power	0.8 0.9 0.95
Alpha	0.05
Group Allocation	Equal (N+ = N-)
AUC1 (Area Under Curve 1)	0.7
AUC2 (Area Under Curve 2)	0.75
Lower FPR	0.00
Upper FPR	1.00
Correlation+	0.0
Correlation-	0.0
Type of Data	Continuous

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results 1

Numeric Results for Testing $AUC1 = AUC2$ with Continuous Data

Test Type = One-Sided. $FPR1 = 0.0$. $FPR2 = 1.0$. $B1 = 1.000$. $B2 = 1.000$. Allocation Ratio = 1.000.

Target Power	Actual Power	N+	N-	N	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.80	0.8003	652	652	1304	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.90	0.9001	897	897	1794	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.95	0.9501	1129	1129	2258	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050

Note that the sample sizes of 897 and 652 match exactly with the results of Hanley and McNeil. The 1129 is two less than their 1131. This difference may be due to refinements in computing the normal probability distribution used in PASS. You can compare these sample sizes by calculating the power.

Tests for Two ROC Curves

Numeric Results 2

Numeric Results for Testing $AUC1 = AUC2$ with Continuous Data

Test Type = One-Sided. FPR1 = 0.0. FPR2 = 1.0. B1 = 1.000. B2 = 1.000. Allocation Ratio = 1.000.

Power	N+	N-	N	AUC1'	AUC2'	Diff'	AUC1	AUC2	Diff	Alpha
0.9499	1128	1128	2256	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.9501	1129	1129	2258	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.9502	1130	1130	2260	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.9504	1131	1131	2262	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050
0.9505	1132	1132	2264	0.7000	0.7500	0.0500	0.7000	0.7500	0.0500	0.050

Note that the power for 1129 is 0.9501 while the power for 1131 is 0.9505. This is only a slight difference and explains why this value showed up in their table.