

## 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.

## 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  $\Delta = \theta_2 - \theta_1$  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

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

## Discrete (Rating) Data

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

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

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

where

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

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

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

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

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

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

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

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

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

## Tests for Two ROC Curves

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

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

$r_-$  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.

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

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).

## 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

#### Design Tab

Solve For .....	<b>Power</b>
Alternative Hypothesis .....	<b>Two-Sided Test</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Enter N+ and R, where N- = R × N+</b>
N+ .....	<b>20 50 100 250 500 1000 2000</b>
R .....	<b>2</b>
AUC1 (Area Under Curve 1) .....	<b>0.80</b>
AUC2 (Area Under Curve 2) .....	<b>0.825 0.85 0.9</b>
Type of Data .....	<b>Discrete (Ratings)</b>
B1 (SD Ratio).....	<b>1.0</b>
B2 (SD Ratio).....	<b>1.0</b>
Lower FPR.....	<b>0.00</b>
Upper FPR.....	<b>1.00</b>
Correlation+ .....	<b>0.6</b>
Correlation- .....	<b>0.6</b>

# Output

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

## Numeric Reports

### Numeric Results

Solve For: [Power](#)  
 Alternative Hypothesis: Two-Sided (H1: AUC1 ≠ AUC2)  
 Type of Data: Discrete (Ratings)  
 B1 (SD Ratio): 1  
 B2 (SD Ratio): 1  
 FPR Limits: Lower = 0, Upper = 1  
 Correlation+: 0.6  
 Correlation-: 0.6

Power	Sample Size			Area Under the ROC Curve								
	N+	N-	N	R (N- / N+)		Actual			Adjusted for FPR Limits			Alpha
				Target	Actual	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'	
0.0501	20	40	60	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.0733	50	100	150	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.1084	100	200	300	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.2104	250	500	750	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.3744	500	1000	1500	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.6426	1000	2000	3000	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.9090	2000	4000	6000	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.0920	20	40	60	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.1737	50	100	150	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.3083	100	200	300	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.6442	250	500	750	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.9116	500	1000	1500	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.9969	1000	2000	3000	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
1.0000	2000	4000	6000	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.2470	20	40	60	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
0.5494	50	100	150	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
0.8496	100	200	300	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
0.9978	250	500	750	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
1.0000	500	1000	1500	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
1.0000	1000	2000	3000	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05
1.0000	2000	4000	6000	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05

Power: The probability of rejecting a false null hypothesis when the alternative hypothesis is true.  
 N+ and N-: The number of items sampled from each population.  
 N: The total sample size.  $N = (N+) + (N-)$ .  
 Target R: 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: 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: The actual areas under the ROC curve for diagnostic tests 1 and 2, respectively.  
 Difference: The difference to be detected.  $Difference = AUC2 - AUC1$ .  
 AUC1' and AUC2': The adjusted (rescaled to account for FPR limits) areas under the ROC curve for diagnostic tests 1 and 2, respectively.  
 Difference': The adjusted (rescaled) difference to be detected.  $Difference' = AUC2' - AUC1'$ .  
 Alpha: The probability of rejecting a true null hypothesis.  
 B1, B2: The ratios of the standard deviations of the negative and positive groups for each test.  
 FPR1, FPR2: The lower and upper bounds on the false positive rates.  
 Correlation+, Correlation-: The correlation between scores on the two diagnostic tests for the subjects that are actually positive and for the subjects that are actually negative, respectively.

Tests for Two ROC Curves

**Summary Statements**

A two-group (positive/negative or with condition/without) design with discrete (rating scale) response data will be used to compare the area under the ROC curve (AUC) of two diagnostic tests. Each subject will receive both diagnostic tests. The comparison will be made using a two-sided Z-test with a Type I error rate ( $\alpha$ ) of 0.05. The area under the curve will be computed between the X-axis (false positive rate) values of 0 and 1. For the first diagnostic test, the ratio of the standard deviation of the responses in the negative group to the standard deviation of the responses in the positive group (SD1- / SD1+) is assumed to be 1. For the second diagnostic test, the ratio (SD2- / SD2+) is assumed to be 1. The correlation between scores of the two diagnostic tests is assumed to be 0.6 for the positive (with condition) group and 0.6 for the negative (without condition) group. To detect an area under the curve of 0.8 for the first diagnostic test, and 0.825 for the second (or a difference of 0.025), with sample sizes of 20 for the positive (with condition) group and 40 for the negative (without condition) group, the power is 0.0501.

**Dropout-Inflated Sample Size**

Dropout Rate	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
	N+	N-	N	N+'	N-'	N'	D+	D-	D
20%	20	40	60	25	50	75	5	10	15
20%	50	100	150	63	125	188	13	25	38
20%	100	200	300	125	250	375	25	50	75
20%	250	500	750	313	625	938	63	125	188
20%	500	1000	1500	625	1250	1875	125	250	375
20%	1000	2000	3000	1250	2500	3750	250	500	750
20%	2000	4000	6000	2500	5000	7500	500	1000	1500

- Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
- N+, N-, and N The evaluable sample sizes at which power is computed (as entered by the user). If N+ and N- subjects are evaluated out of the N+' and N-' subjects that are enrolled in the study, the design will achieve the stated power.
- N+', N-', and N' The number of subjects that should be enrolled in the study in order to obtain N+, N-, and N evaluable subjects, based on the assumed dropout rate. N+' and N-' are calculated by inflating N+ and N- using the formulas  $N+' = N+ / (1 - DR)$  and  $N-' = N- / (1 - DR)$ , with N+' and N-' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lohknygina, Y. (2018) pages 32-33.)
- D+, D-, and D The expected number of dropouts.  $D+ = N+' - N+$ ,  $D- = N-' - N-$ , and  $D = D+ + D-$ .

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 25 subjects should be enrolled in Group 1, and 50 in Group 2, to obtain final group sample sizes of 20 and 40, respectively.

**References**

Hanley, J. A. and McNeil, B. J. 1983. 'A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases.' Radiology, 148, 839-843. September, 1983.  
 Obuchowski, N. and McClish, D. 1997. 'Sample Size Determination for Diagnostic Accuracy Studies Involving Binormal ROC Curve Indices.' Statistics in Medicine, 16, pages 1529-1542.

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'.

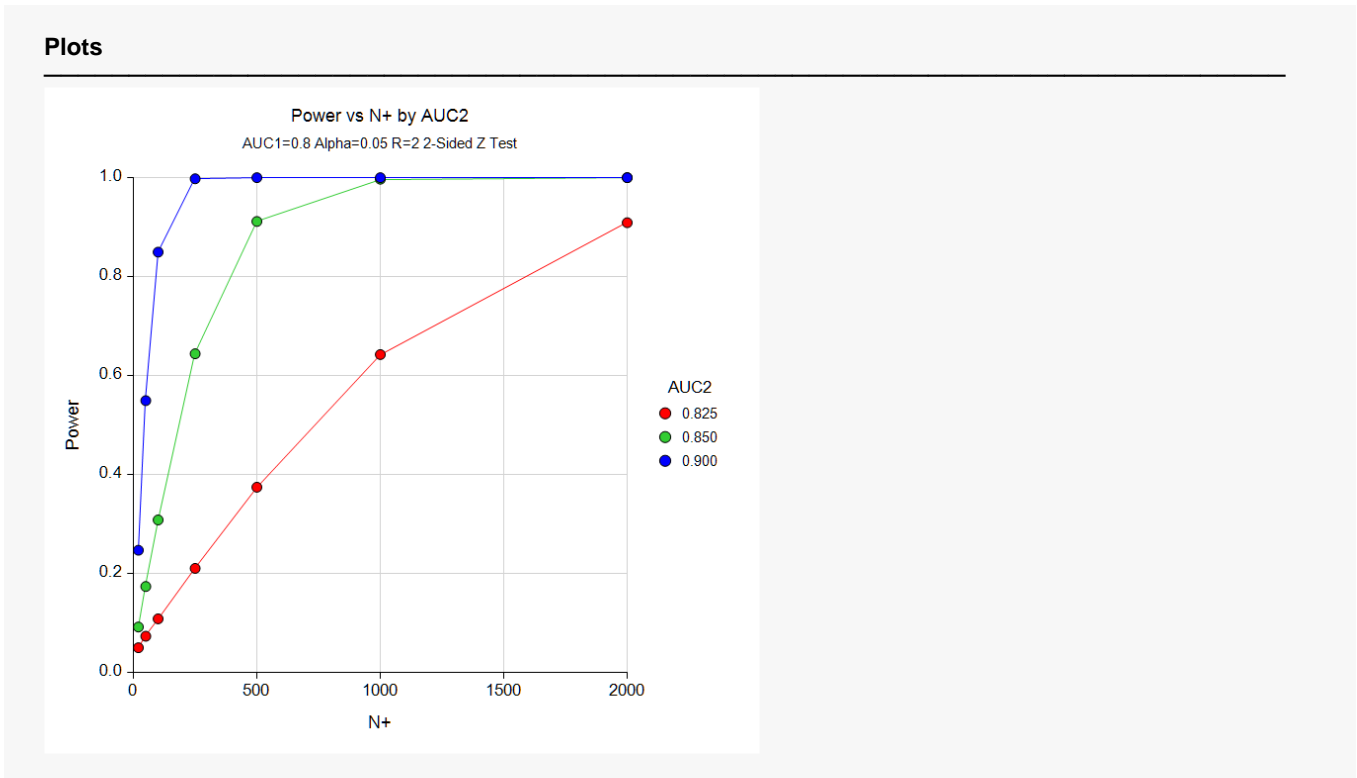
## AUC

This is the 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.

## AUC'

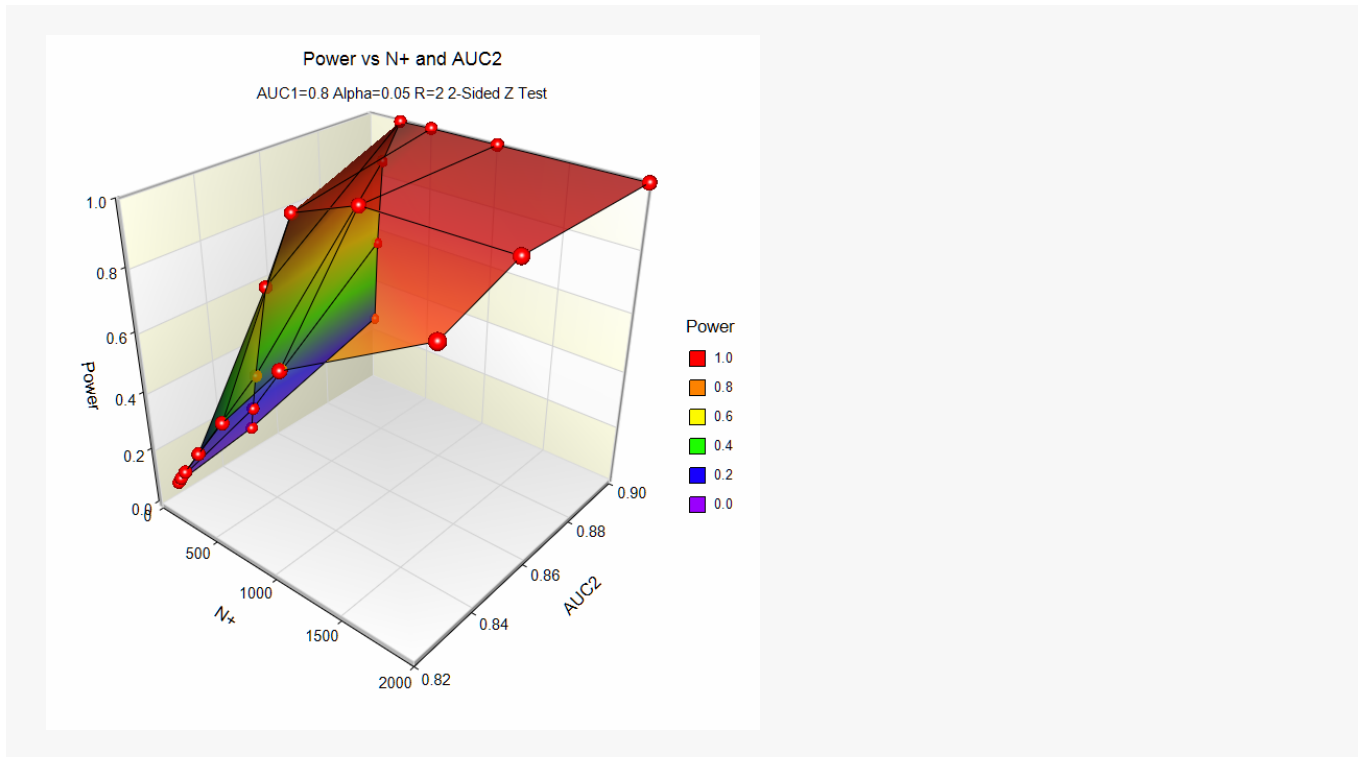
This adjusted value is only useful when the FPR range is not 0.0 to 1.0. An adjustment is applied so that the minimum area is 0.5 and the maximum area is 1.0. This will yield the values of AUC that resulted when the FPR range was 0.0 to 1.0.

## Plots Section





Tests for Two ROC Curves



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

## 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

#### Design Tab

Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided Test</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Enter R = N-/N+, solve for N+ and N-</b>
R .....	<b>2</b>
AUC1 (Area Under Curve 1) .....	<b>0.80</b>
AUC2 (Area Under Curve 2) .....	<b>0.825 0.85 0.9</b>
Type of Data .....	<b>Discrete (Ratings)</b>
B1 (SD Ratio).....	<b>1.0</b>
B2 (SD Ratio).....	<b>1.0</b>
Lower FPR.....	<b>0.00</b>
Upper FPR.....	<b>1.00</b>
Correlation+ .....	<b>0.6</b>
Correlation- .....	<b>0.6</b>

## Tests for Two ROC Curves

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: AUC_1 \neq AUC_2$ )  
 Type of Data: Discrete (Ratings)  
 B1 (SD Ratio): 1  
 B2 (SD Ratio): 1  
 FPR Limits: Lower = 0, Upper = 1  
 Correlation+: 0.6  
 Correlation-: 0.6

Power		Sample Size			R (N- / N+)		Area Under the ROC Curve			Adjusted for FPR Limits			Alpha
Target	Actual	N+	N-	N	Target	Actual	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'	
0.9	0.9001	1937	3874	5811	2	2	0.8	0.825	0.025	0.8	0.825	0.025	0.05
0.9	0.9002	480	960	1440	2	2	0.8	0.850	0.050	0.8	0.850	0.050	0.05
0.9	0.9012	117	234	351	2	2	0.8	0.900	0.100	0.8	0.900	0.100	0.05

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

## 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided Test</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Enter R = N-/N+, solve for N+ and N-</b>
R .....	<b>2</b>
AUC1 (Area Under Curve 1) .....	<b>0.80</b>
AUC2 (Area Under Curve 2) .....	<b>0.825 0.85 0.9</b>
Type of Data .....	<b>Discrete (Ratings)</b>
B1 (SD Ratio).....	<b>1.0</b>
B2 (SD Ratio).....	<b>1.0</b>
Lower FPR.....	<b>0.00</b>
Upper FPR.....	<b>0.20</b>
Correlation+ .....	<b>0.6</b>
Correlation- .....	<b>0.6</b>

## Tests for Two ROC Curves

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: AUC1 \neq AUC2$ )  
 Type of Data: Discrete (Ratings)  
 B1 (SD Ratio): 1  
 B2 (SD Ratio): 1  
 FPR Limits: Lower = 0, Upper = 0.2  
 Correlation+: 0.6  
 Correlation-: 0.6

Power		Sample Size			Area Under the ROC Curve								
		N+	N-	N	R (N- / N+)		Actual			Adjusted for FPR Limits			Alpha
Target	Actual				Target	Actual	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'	
0.9	0.9000	4095	8190	12285	2	2	0.128	0.137	0.009	0.8	0.825	0.025	0.05
0.9	0.9002	1012	2024	3036	2	2	0.128	0.146	0.018	0.8	0.850	0.050	0.05
0.9	0.9001	242	484	726	2	2	0.128	0.164	0.036	0.8	0.900	0.100	0.05

Note that the necessary sample size has more than doubled.

## Example 4 – Validation using Obuchowski and McClish (1997)

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.  $B1 = B2 = 1.0$ .  $\text{Correlation}^+ = \text{Correlation}^- = 0.6$ . 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 4** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>Two-Sided Test</b>
Power.....	<b>0.80</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Enter R = N-/N+, solve for N+ and N-</b>
R .....	<b>2</b>
AUC1 (Area Under Curve 1) .....	<b>0.922222</b>
AUC2 (Area Under Curve 2) .....	<b>0.819444</b>
Type of Data .....	<b>Discrete (Ratings)</b>
B1 (SD Ratio).....	<b>1.0</b>
B2 (SD Ratio).....	<b>1.0</b>
Lower FPR.....	<b>0.00</b>
Upper FPR.....	<b>0.20</b>
Correlation+ .....	<b>0.6</b>
Correlation- .....	<b>0.6</b>

## Tests for Two ROC Curves

## Output

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

### Numeric Results

Solve For: [Sample Size](#)  
 Alternative Hypothesis: Two-Sided ( $H_1: AUC_1 \neq AUC_2$ )  
 Type of Data: Discrete (Ratings)  
 B1 (SD Ratio): 1  
 B2 (SD Ratio): 1  
 FPR Limits: Lower = 0, Upper = 0.2  
 Correlation+: 0.6  
 Correlation-: 0.6

Power		Sample Size			R (N- / N+)		Area Under the ROC Curve			Adjusted for FPR Limits			Alpha
Target	Actual	N+	N-	N	Target	Actual	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'	
0.8	0.8027	109	218	327	2	2	0.172	0.135	-0.037	0.9222	0.8194	-0.1028	0.05

Note that the sample sizes of 109 and 218 match exactly with the results of Obuchowski and McClish (1997).

## Example 5 – Validation using Hanley and McNeil (1982)

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  $AUC_1 = 0.70$  and  $AUC_2 = 0.75$ . 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  $B_1$ ,  $B_2$ ,  $FPR_1$ , and  $FPR_2$  are ignored. Also, in this case, the correlations are set to 0.0.

### Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 5a** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

#### Design Tab

Solve For .....	<b>Sample Size</b>
Alternative Hypothesis .....	<b>One-Sided Test</b>
Power.....	<b>0.8 0.9 0.95</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N+ = N-)</b>
AUC1 (Area Under Curve 1) .....	<b>0.7</b>
AUC2 (Area Under Curve 2) .....	<b>0.75</b>
Type of Data .....	<b>Continuous</b>
Correlation+ .....	<b>0.0</b>
Correlation- .....	<b>0.0</b>



## Output

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

### Numeric Results 1

Numeric Results												
Solve For:		Sample Size										
Alternative Hypothesis:		One-Sided (H1: AUC1 > AUC2 or H1: AUC1 < AUC2)										
Type of Data:		Continuous										
FPR Limits:		Lower = 0, Upper = 1										
Correlation+:		0										
Correlation-:		0										
Area Under the ROC Curve												
Power		Sample Size			Actual			Adjusted for FPR Limits			Alpha	
Target	Actual	N+	N-	N	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'		
0.80	0.8003	652	652	1304	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.90	0.9001	897	897	1794	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.95	0.9501	1129	1129	2258	0.7	0.75	0.05	0.7	0.75	0.05	0.05	

Note that the sample sizes of 897 and 652 match exactly with the results of Hanley and McNeil (1982). 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 for a range of sample sizes between 1128 and 1132 (**Example 5b**).

### Numeric Results 2

Numeric Results											
Solve For:		Power									
Alternative Hypothesis:		One-Sided (H1: AUC1 > AUC2 or H1: AUC1 < AUC2)									
Type of Data:		Continuous									
FPR Limits:		Lower = 0, Upper = 1									
Correlation+:		0									
Correlation-:		0									
Area Under the ROC Curve											
Power	Sample Size			Actual			Adjusted for FPR Limits			Alpha	
	N+	N-	N	AUC1	AUC2	Difference	AUC1'	AUC2'	Difference'		
0.9499	1128	1128	2256	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.9501	1129	1129	2258	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.9502	1130	1130	2260	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.9504	1131	1131	2262	0.7	0.75	0.05	0.7	0.75	0.05	0.05	
0.9505	1132	1132	2264	0.7	0.75	0.05	0.7	0.75	0.05	0.05	

Note that the power for 1129 is 0.9501 while the power for 1131 is 0.9504. This is only a slight difference and explains why this value showed up in their table.