Chapter 261

Confidence Intervals for the Area Under an ROC Curve

Introduction

Receiver operating characteristic (ROC) curves are used to assess the accuracy of a diagnostic test. The technique is used when you have a criterion variable which will be used to make a yes or no decision based on the value of this variable. The area under the ROC curve (AUC) is a popular summary index of an ROC curve.

This module computes the sample size necessary to achieve a specified width of a confidence interval. We use the approach of Hanley and McNeil (1982) in which the criterion variable is continuous.

Technical Details

In the following, we suppose that we have two groups of patients, those with a condition of interest (the positive group or group 1) and those without it (the negative group or group 2). This classification may be known from extensive diagnosis or based on the value of another diagnostic test. The diagnostic test of interest is performed on each patient and the resulting test value is recorded. At each specified cutoff value of the criterion variable, the true positive rate (TPR) and the false positive rate (FPR) are calculated. A plot of the TPR versus the FPR allows the study of the consequences of using various cutoff values. This plot is called the *ROC curve*.

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 is a plot of the diagnostic test's power versus it's 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 - 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 determine which of the two subjects 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 negative (normal) patients and *Y* denote the distribution of the criterion variable for positive (diseased) patients. It is assumed that

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

and

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

For a particular cutoff value of the criterion variable, *c*, the true positive rate is given by

$$TPR(c) = P(Y > c)$$

$$= 1 - \Phi\left(\frac{c - \mu_+}{\sigma_+}\right)$$

$$= \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right)$$

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

Similarly, the false positive rate is given by

$$FPR(c) = P(X > c)$$

$$= 1 - \Phi\left(\frac{c - \mu_-}{\sigma_-}\right)$$

$$= \Phi\left(\frac{\mu_- - c}{\sigma_-}\right)$$

The ROC curve is thus the curve traced out by the functions

$$[FPR(c), TPR(c)] = \left[\Phi\left(\frac{\mu_- - c}{\sigma_-}\right), \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right)\right]$$

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

$$\theta = \int_{-\infty}^{\infty} TPR(c) FPR'(c) dc$$

$$= \int_{-\infty}^{\infty} \Phi\left(\frac{\mu_+ - c}{\sigma_+}\right) \phi\left(\frac{\mu_- - c}{\sigma_-}\right)\left(-\frac{1}{\sigma_-}\right) dc$$

$$= \int_{-\infty}^{\infty} \Phi(A + Bv)\phi(v) dv$$

$$= \Phi\left(\frac{A}{\sqrt{1 + B^2}}\right)$$

where

$$c = \mu_- - v\sigma_-$$

$$A = \frac{|\mu_+ - \mu_-|}{\sigma_+}$$

$$B = \frac{\sigma_-}{\sigma_+}$$

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

Confidence Interval for AUC

Let *AUC* denote the sample AUC value. For large samples, the distribution of AUC is approximately normal. Hence, a100(1 – α)% confidence interval for AUC may be computed using the standard normal distribution as follows

$$AUC \pm z_{\alpha/2} SE(AUC)$$

The width of the confidence interval is $2z_{\alpha/2}SE(AUC)$. One-sided limits may be obtained by replacing $\alpha/2$ by α.

The formula for SE(AUC) was given by Hanley and McNeil (1982) is

$$SE(AUC) = \sqrt{\frac{AUC(1 - AUC) + (N_1 - 1)(Q_1 - AUC^2) + (N_2 - 1)(Q_2 - AUC^2)}{N_1 N_2}}$$

where

$$Q_1 = \frac{AUC}{2 - AUC}$$

$$Q_2 = \frac{2AUC^2}{1 + AUC}$$

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

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

Confidence

Confidence Level (1 – Alpha)

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n1 and n2 items are drawn from populations using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95* or *0.90 to 0.99 by 0.01*.

Sample Size (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on *N1* or *N2* or both.

The options are

- **Equal (N1 = N2)**

  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 N1, solve for N2**

  Select this option when you wish to fix *N1* at some value (or values), and then solve only for *N2*. Please note that for some values of *N1*, there may not be a value of *N2* that is large enough to obtain the desired power.

- **Enter N2, solve for N1**

  Select this option when you wish to fix *N2* at some value (or values), and then solve only for *N1*. Please note that for some values of *N2*, there may not be a value of *N1* that is large enough to obtain the desired power.

- **Enter R = N2/N1, solve for N1 and N2**

  For this choice, you set a value for the ratio of *N2* to *N1*, and then PASS determines the needed *N1* and *N2*, with this ratio, to obtain the desired power. An equivalent representation of the ratio, *R*, is

  $$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for N1 and N2**

  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 *N1* and *N2* with this percentage to obtain the desired power.

N1 (Sample Size, Group 1)

This option is displayed if Group Allocation = "Enter N1, solve for N2"

N1 is the number of items or individuals sampled from the Group 1 population.

N1 must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N2, solve for N1"

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be $\geq$ 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 R = N2/N1, solve for N1 and N2."

R is the ratio of *N2* to *N1*. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of *N2* to *N1* while solving for *N1* and *N2*. Only sample size combinations with this ratio are considered.

N2 is related to *N1* by the formula:

$$N2 = [R \times N1],$$

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., *N1* = 10 and *N2* = 20, or *N1* = 50 and *N2* = 100).

R must be greater than 0. If *R* < 1, then *N2* will be less than *N1*; if *R* > 1, then *N2* will be greater than *N1*. 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 N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for *N1* and *N2*. 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 (N1 = N2)**

  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 N1 and N2 individually**

  This choice permits you to enter different values for *N1* and *N2*.

- **Enter N1 and R, where N2 = R * N1**

  Choose this option to specify a value (or values) for *N1*, and obtain *N2* as a ratio (multiple) of *N1*.

- **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 *N1* as a percentage of *N*, and then *N2* as *N - N1*.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal (N1 = N2)."

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 *N1*, and also the value for *N2*.

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

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter N1 and N2 individually" or "Enter N1 and R, where N2 = R * N1."*

N1 is the number of items or individuals sampled from the Group 1 population.

N1 must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

N2 is the number of items or individuals sampled from the Group 2 population.

N2 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 N1 and R, where N2 = R * N1."*

R is the ratio of *N2* to *N1*. That is,

$$R = N2/N1$$

Use this value to obtain *N2* as a multiple (or proportion) of *N1*.

N2 is calculated from *N1* using the formula:

$$N2=[R \ x \ N1],$$

where the value *[Y]* is the next integer ≥ *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 *N2* will be less than N1; if *R* > 1, then *N2* will be greater than *N1*. 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 *N1* and *N2*.

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.

Precision

Width of Confidence Interval

This is the distance between the lower confidence limit (*LCL*) and the upper confidence limit (*UCL*). Its calculation is *UCL - LCL*. It is a measure of the precision of the confidence interval. This width usually ranges between 0 and 1.

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

Distance from *AUC* to Limit

This is the distance from *AUC* to the lower confidence limit (*LCL*) or the upper confidence limit (*UCL*). It is calculated using | *AUC - LCL*| or |*UCL - AUC* |. The range is between 0 and 1.

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

Area Under ROC Curve

AUC (Area Under Curve)

Specify one or more values of the AUC for the diagnostic test. 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).

Example 1 – Calculating Sample Size

A study is planned in which a researcher wishes to construct a two-sided 95% confidence interval for AUC. The confidence level is set to 0.95. The researcher would like to try AUC values 0.6, 0.7, 0.8, and 0.9. The research would like to see the sample necessary for confidence intervals between 0.05 and 0.1 in width.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Area Under an ROC Curve** procedure window by clicking on **ROC**, and then clicking on **Confidence Intervals for the Area Under an ROC Curve**. 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**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Interval Type	**Two-Sided**
Confidence Level	**0.95**
Group Allocation	**Equal (N1 = N2)**
Confidence Interval Width	**0.05 0.10**
AUC	**0.6 0.7 0.8 0.9**

Annotated Output

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

Numeric Report

Numeric Results for Two-Sided Confidence Interval for ROC Curve's AUC

Confidence Level	Total Subjects N	Ratio N2/N1 R	Number Positive N1	Number Negative N2	Sample AUC	C.I. Width UCL-LCL	Lower Conf Limit LCL	Upper Conf Limit UCL
0.950	1952	1.000	976	976	0.600	0.050	0.575	0.625
0.950	1660	1.000	830	830	0.700	0.050	0.675	0.725
0.950	1204	1.000	602	602	0.800	0.050	0.775	0.825
0.950	628	1.000	314	314	0.900	0.050	0.875	0.925
0.950	490	1.000	245	245	0.600	0.100	0.550	0.650
0.950	416	1.000	208	208	0.700	0.100	0.650	0.750
0.950	302	1.000	151	151	0.800	0.100	0.750	0.850
0.950	158	1.000	79	79	0.900	0.100	0.850	0.950

Report Definitions
Confidence Level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true coefficient alpha.
N is the total number of subjects sampled.
R is N2 / N1, so that N2 = R x N1.
N1 is the number of subjects sampled from the 'positive' group.
N2 is the number of subjects sampled from the 'negative' group.
Sample AUC is the anticipated value of the sample area under the ROC curve.
C.I. Width (UCL-LCL) is the width of the confidence interval. It is the distance from the lower limit to the upper limit.
Lower and Upper Confidence Limits are the actual limits that would result from a dataset with these statistics. They may not be exactly equal to the specified values because of the discrete nature of the N1 and N2.

References
Hanley, J.A. and NcNeil, B.J. 1982. 'The Meaning and Use of the Area under a Receiver Operating Characteristic
   (ROC) Curve.' Radiology, Vol 148, 29-36.
Kryzanowski, W.J. and Hand, D.J. 2009. 'ROC Curves for Continuous Data.' Chapman & Hall/CRC Press.

Summary Statements
A random sample of 976 subjects from the positive population and 976 subjects from the negative
population produce a two-sided 95.0% confidence interval with a width of 0.050 when the sample
AUC is 0.600.

This report shows the sample size necessary to meet the confidence interval width requirements of the study.

Plot Section
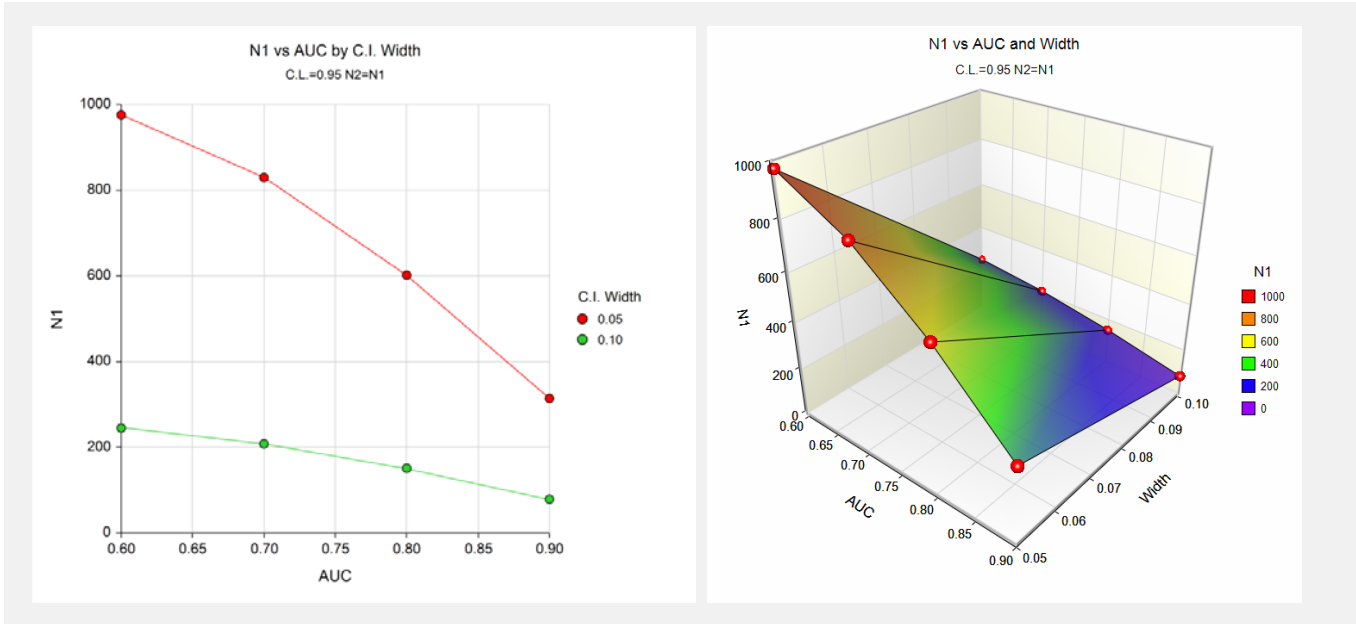
These plots show the sample size versus AUC for the two confidence interval widths.

Example 2 – Validation using Mathews

Mathews 2010 gives a sample size calculation on page 160. In this example, a two-sided, 95% confidence interval is to have a width of 0.10 when AUC is 0.90. The sample size is determined to be 77 per group. Note that this result is approximate because various simplifications in the SE(AUC) formula are made.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for the Area Under an ROC Curve** procedure window by clicking on **ROC**, and then clicking on **Confidence Intervals for the Area Under an ROC Curve**. 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**.

Option   **Value**

Design Tab

Solve For .... **Sample Size**
Interval Type .... **Two-Sided**
Confidence Level **0.95**
Group Allocation .... **Equal (N1 = N2)**
Confidence Interval Width **0.1**
AUC **0.9**

Annotated Output

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

Numeric Report

Numeric Results for Two-Sided Confidence Interval for ROC Curve's AUC

Confidence Level	Total Subjects N	Ratio N2/N1 R	Number Positive N1	Number Negative N2	Sample AUC	C.I. Width UCL-LCL	Lower Conf Limit LCL	Upper Conf Limit UCL
0.950	158	1.000	79	79	0.900	0.100	0.850	0.950

PASS has calculated N1 = N2 = 79 which is close to Mathews value of 77. Remember that 77 was based on some simplifying approximations the PASS does not use.