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

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

and

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

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

$$TPR(c) = \mathrm{P}(Y > c)$$

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

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

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

Similarly, the false positive rate is given by

$$FPR(c) = \mathrm{P}(X > c)$$

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

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

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

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

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

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

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

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 *α*/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

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

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

# 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

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 ........................................................ **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**

## Output

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

### Numeric Reports

**Numeric Results**

Solve For: Sample Size
Interval Type: Two-Sided

| | Sample Size | | | | | | Confidence Interval Limits | |
|---|---|---|---|---|---|---|---|---|
| Confidence Level | Number Positive N1 | Number Negative N2 | Total Subjects N | Ratio N2/N1 R | Area Under Curve AUC | Confidence Interval Width | Lower | Upper |
| 0.95 | 976 | 976 | 1952 | 1 | 0.6 | 0.05 | 0.575 | 0.625 |
| 0.95 | 830 | 830 | 1660 | 1 | 0.7 | 0.05 | 0.675 | 0.725 |
| 0.95 | 602 | 602 | 1204 | 1 | 0.8 | 0.05 | 0.775 | 0.825 |
| 0.95 | 314 | 314 | 628 | 1 | 0.9 | 0.05 | 0.875 | 0.925 |
| 0.95 | 245 | 245 | 490 | 1 | 0.6 | 0.10 | 0.550 | 0.650 |
| 0.95 | 208 | 208 | 416 | 1 | 0.7 | 0.10 | 0.650 | 0.750 |
| 0.95 | 151 | 151 | 302 | 1 | 0.8 | 0.10 | 0.750 | 0.850 |
| 0.95 | 79 | 79 | 158 | 1 | 0.9 | 0.10 | 0.850 | 0.950 |

Confidence Level The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true coefficient alpha.
N1 The number of subjects sampled from the "positive" group.
N2 The number of subjects sampled from the "negative" group.
N The total number of subjects sampled.
R Ratio equal to N2/N1, so that N2 = R x N1.

| | |
|---|---|
| AUC | The anticipated value of the sample area under the ROC curve. |
| Confidence Interval Width | The width of the confidence interval. It is the distance from the lower limit to the upper limit. |
| Confidence Interval Limits | 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. |

**Summary Statements**

A two-group (positive/negative or with condition/without) design will be used to obtain a two-sided 95% confidence interval for the area under an ROC curve (AUC). The sample AUC value is assumed to be 0.6. The standard error calculation method given by Hanley and McNeil (1982) will be used to compute the confidence interval limits. To produce a confidence interval width of 0.05, the number of subjects needed will be 976 in the positive (with condition) group and 976 in the negative (without condition) group.

**Dropout-Inflated Sample Size**

| | Sample Size | | | Dropout-Inflated Enrollment Sample Size | | | Expected Number of Dropouts | | |
|---|---|---|---|---|---|---|---|---|---|
| Dropout Rate | N1 | N2 | N | N1' | N2' | N' | D1 | D2 | D |
| 20% | 976 | 976 | 1952 | 1220 | 1220 | 2440 | 244 | 244 | 488 |
| 20% | 830 | 830 | 1660 | 1038 | 1038 | 2076 | 208 | 208 | 416 |
| 20% | 602 | 602 | 1204 | 753 | 753 | 1506 | 151 | 151 | 302 |
| 20% | 314 | 314 | 628 | 393 | 393 | 786 | 79 | 79 | 158 |
| 20% | 245 | 245 | 490 | 307 | 307 | 614 | 62 | 62 | 124 |
| 20% | 208 | 208 | 416 | 260 | 260 | 520 | 52 | 52 | 104 |
| 20% | 151 | 151 | 302 | 189 | 189 | 378 | 38 | 38 | 76 |
| 20% | 79 | 79 | 158 | 99 | 99 | 198 | 20 | 20 | 40 |

| | |
|---|---|
| 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. |
| N1, N2, and N | The evaluable sample sizes at which the confidence interval is computed. If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated confidence interval. |
| N1', N2', and N' | The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. After solving for N1 and N2, N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.) |
| D1, D2, and D | The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2. |

**Dropout Summary Statements**

Anticipating a 20% dropout rate, 1220 subjects should be enrolled in Group 1, and 1220 in Group 2, to obtain final group sample sizes of 976 and 976, respectively.

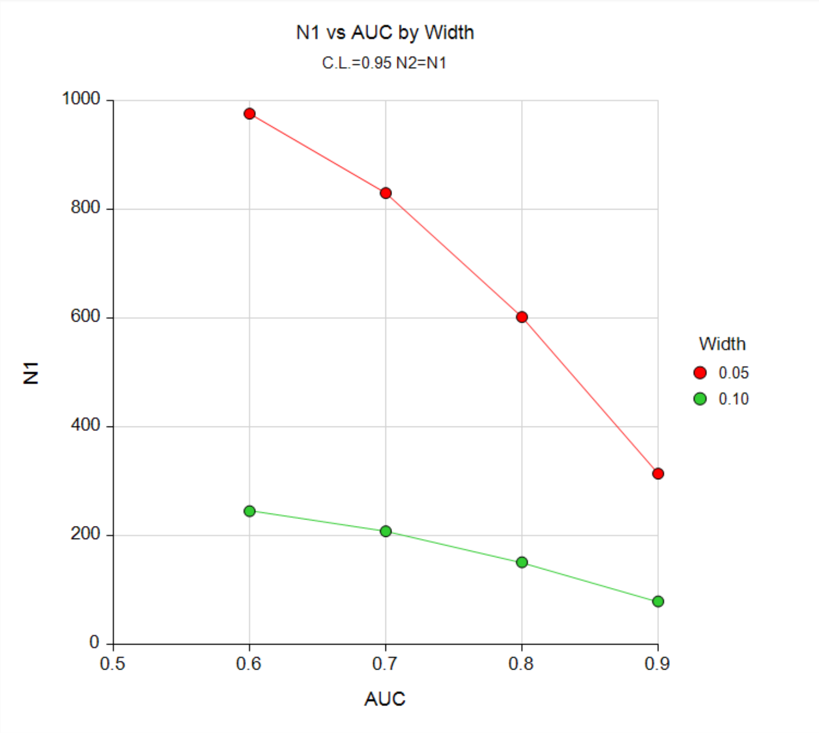
**References**

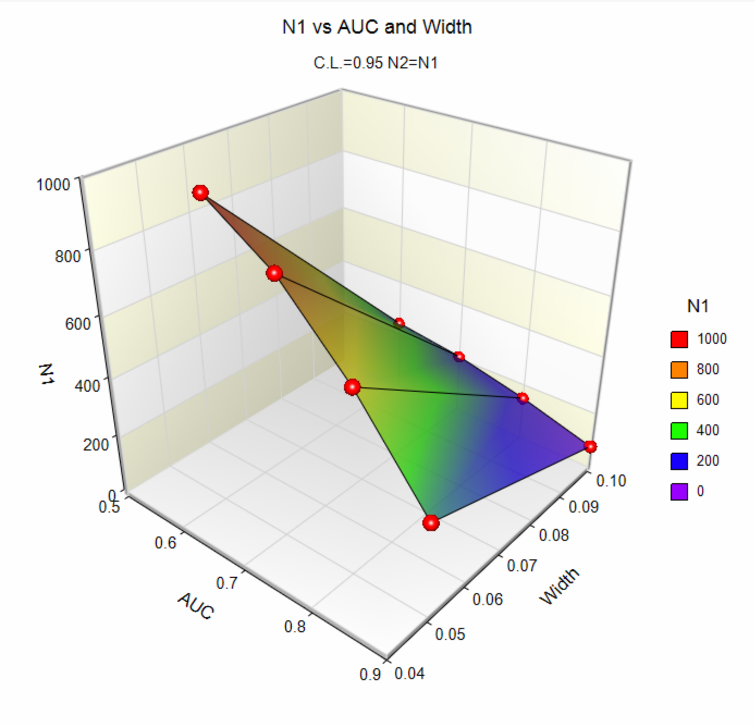
Hanley, J.A. and NcNeil, B.J. 1982. 'The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.' Radiology, Vol 148, 29-36.
Kryzanowski, W.J. and Hand, D.J. 2009. 'ROC Curves for Continuous Data.' Chapman & Hall/CRC Press.

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

## Plots Section

### Plots




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

# Example 2 – Validation using Mathews (2010)

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

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

## Output

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

**Numeric Results**

Solve For: Sample Size
Interval Type: Two-Sided

| | Sample Size | | | | | | Confidence Interval Limits | |
|---|---|---|---|---|---|---|---|---|
| Confidence Level | Number Positive N1 | Number Negative N2 | Total Subjects N | Ratio N2/N1 R | Area Under Curve AUC | Confidence Interval Width | Lower | Upper |
| 0.95 | 79 | 79 | 158 | 1 | 0.9 | 0.1 | 0.85 | 0.95 |

**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.