

Chapter 535

Binary Diagnostic Tests – Single Sample

Introduction

This procedure generates a number of measures of the accuracy of a diagnostic test. Some of these measures include sensitivity, specificity, proportion correctly specified, positive predictive value, and likelihood ratios. Confidence intervals are also available for many of the reported statistics.

Discussion and Technical Details

In a typical binary diagnostic test, a positive or negative diagnosis is made for each individual (e.g., patient, subject, or unit). When the diagnosis is compared to the true (known) condition, there are four possible outcomes: true positive, true negative, false positive, false negative.

Classification Table

		Test Result	
		Positive	Negative
True Condition	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Binary Diagnostic Tests – Single Sample

When all of the individuals are assigned to the four outcomes, a count for each outcome is produced. The four counts are labeled A, B, C, and D in the table below.

Classification Table (Counts)

		Test Result		Total
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	A + C
	Negative	False Positive (B)	True Negative (D)	B + D
Total		A + B	C + D	A + B + C + D

Various rates (proportions) can be used to describe a classification table. Some rates are based on the true condition, some rates are based on the predicted condition, and some rates are based on the whole table. These rates will be described in the following sections.

Rates Assuming a True Condition

The following rates assume one of the two true conditions.

True Positive Rate (TPR) or Sensitivity = $A / (A + C)$

The true positive rate is the proportion of the individuals with a known positive condition for which the test result is positive.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	$TPR = A / (A + C)$ (Y Axis on ROC Curve)
	Negative	False Positive (B)	True Negative (D)	

True Negative Rate (TNR) or Specificity = $D / (B + D)$

The true negative rate is the proportion of the individuals with a known negative condition for which the test result is negative. This rate is often called the specificity.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	$TNR = D / (B + D)$
	Negative	False Positive (B)	True Negative (D)	

False Negative Rate (FNR) or Miss Rate = $C / (A + C)$

The false negative rate is the proportion of the individuals with a known positive condition for which the test result is negative. This rate is sometimes called the miss rate.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	$FNR = C / (A + C)$
	Negative	False Positive (B)	True Negative (D)	

False Positive Rate (FPR) or Fall-out = $B / (B + D)$

The false positive rate is the proportion of the individuals with a known negative condition for which the test result is positive. This rate is sometimes called the fall-out.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	$FPR = B / (B + D)$ (X Axis on ROC Curve)
	Negative	False Positive (B)	True Negative (D)	

Rates Assuming a Predicted Condition

The following rates assume one of the two predicted conditions.

Positive Predictive Value (PPV) or Precision = $A / (A + B)$

The positive predictive value is the proportion of the individuals with a positive test result for which the true condition is positive. This rate is sometimes called the precision.

		Test Result	
		Positive	Negative
True Condition	Positive	True Positive (A)	False Negative (C)
	Negative	False Positive (B)	True Negative (D)

$$PPV = A / (A + B)$$

Positive Predictive Value Adjusted for Known Prevalence

When the prevalence (or pre-test probability of a positive condition) is known, an adjusted formula for positive predictive value, based on the known prevalence value, can be used.

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

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

Negative Predictive Value (NPV) = $D / (C + D)$

The negative predictive value is the proportion of the individuals with a negative test result for which the true condition is negative.

		Test Result	
		Positive	Negative
True Condition	Positive	True Positive (A)	False Negative (C)
	Negative	False Positive (B)	True Negative (D)

$$NPV = D / (C + D)$$

Negative Predictive Value Adjusted for Known Prevalence

When the prevalence (or pre-test probability of a positive condition) is known, an adjusted formula for negative predictive value, based on the known prevalence value, can be used.

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

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

False Omission Rate (FOR) = C / (C + D)

The false omission rate is the proportion of the individuals with a negative test result for which the true condition is positive.

		Test Result	
		Positive	Negative
True Condition	Positive	True Positive (A)	False Negative (C)
	Negative	False Positive (B)	True Negative (D)

$$FOR = C / (C + D)$$

False Discovery Rate (FDR) = B / (A + B)

The false discovery rate is the proportion of the individuals with a positive test result for which the true condition is negative.

		Test Result	
		Positive	Negative
True Condition	Positive	True Positive (A)	False Negative (C)
	Negative	False Positive (B)	True Negative (D)

$$FDR = B / (A + B)$$

Whole Table Rates

The following rates are proportions based on all the individuals.

Prevalence = $(A + C) / (A + B + C + D)$

The prevalence may be estimated from the table if all the individuals are randomly sampled from the population.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	Prevalence = $(A + C) / (A + B + C + D)$
	Negative	False Positive (B)	True Negative (D)	

Accuracy or Proportion Correctly Classified = $(A + D) / (A + B + C + D)$

The accuracy reflects the total proportion of individuals that are correctly classified.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	Accuracy = $(A + D) / (A + B + C + D)$
	Negative	False Positive (B)	True Negative (D)	

Proportion Incorrectly Classified = $(B + C) / (A + B + C + D)$

The proportion incorrectly classified reflects the total proportion of individuals that are incorrectly classified.

		Test Result		
		Positive	Negative	
True Condition	Positive	True Positive (A)	False Negative (C)	$PIC = (B + C) / (A + B + C + D)$
	Negative	False Positive (B)	True Negative (D)	

Other Diagnostic Accuracy Indices

Over the past several decades, a number of table summary indices have been considered, above those described above. Those available in NCSS are described below.

Sensitivity + Specificity

The addition of the sensitivity and the specificity may be used as an overall accuracy measure.

Positive Likelihood Ratio (LR+) = TPR / FPR

The positive likelihood ratio is the ratio of the true positive rate (sensitivity) to the false positive rate (1 – specificity). This likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis.

$$LR+ = TPR / FPR$$

Negative Likelihood Ratio (LR-) = FNR / TNR

The negative likelihood ratio is the ratio of the false negative rate to the true negative rate (specificity).

$$LR- = FNR / TNR$$

Diagnostic Odds Ratio (DOR) = $LR+ / LR-$

The diagnostic odds ratio is the ratio of the positive likelihood ratio to the negative likelihood ratio. In some calculation methods, ½ is added to all counts before the calculation of DOR, to avoid dividing by 0.

$$DOR = LR+ / LR-$$

Confidence Intervals for Rates (Proportions)

For each of the rates (proportions) described above, confidence limits may be calculated from among the four methods described in the One Proportion chapter of the documentation. Although less commonly described in textbooks, the score method of Wilson (1927) has been shown by Agresti and Coull (1998) to have much better coverage probabilities than either the exact method of inverting the binomial or the simple Wald confidence interval.

Confidence Intervals for Likelihood Ratios and Odds Ratios

The calculations for the likelihood ratios and the odds ratios are described in the Two Proportions chapter of the documentation.

Data Structure

In this procedure the four counts are entered directly. The dataset is not used.

Example 1 – Binary Diagnostic Test of a Single Sample

This section presents an example of how to run a binary diagnostic test analysis. Suppose 600 individuals are randomly sampled from the population of interest. A diagnostic test is given for a potential disease. After one week, it is known whether each of the individuals has contracted the disease. Of those with the disease, 57 tested positive and 23 tested negative on the diagnostic test. Of those without the disease, 134 tested positive and 386 tested negative on the diagnostic test.

Setup

To run this example, complete the following steps:

1 Specify the Binary Diagnostic Tests – Single Sample procedure options

- Find and open the **Binary Diagnostic Tests – Single Sample** 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.

Data Tab

Condition=Positive, Result=Positive**57**
 Condition=Positive, Result=Negative.....**23**
 Condition=Negative, Result=Positive.....**134**
 Condition=Negative, Result=Negative**386**

2 Run the procedure

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

Counts and Proportions Sections

Counts

True Condition	Diagnostic Test Result		Total
	Positive	Negative	
Positive	57 (A)	23 (C)	80
Negative	134 (B)	386 (D)	520
Total	191	409	600 (N)

Binary Diagnostic Tests – Single Sample

Table Proportions

True Condition	Diagnostic Test Result		Total
	Positive	Negative	
Positive	0.0950	0.0383	0.1333
Negative	0.2233	0.6433	0.8667
Total	0.3183	0.6817	1.0000

Row Proportions

True Condition	Diagnostic Test Result		Total
	Positive	Negative	
Positive	0.7125	0.2875	1.0000
Negative	0.2577	0.7423	1.0000
Total	0.3183	0.6817	1.0000

Column Proportions

True Condition	Diagnostic Test Result		Total
	Positive	Negative	
Positive	0.2984	0.0562	0.1333
Negative	0.7016	0.9438	0.8667
Total	1.0000	1.0000	1.0000

These reports give a summary of the input counts and resulting totals and proportions.

Rates and Confidence Intervals Section

Rates and Confidence Intervals

Measure	Value	95% C. I.		Formula	Calculation
		Lower	Upper		
Sensitivity (TPR)	0.7125	0.6005	0.8082	$A / (A + C)$	57 / 80
Specificity (TNR)	0.7423	0.7024	0.7794	$D / (B + D)$	386 / 520
Sensitivity + Specificity	1.4548			$TPR + TNR$	0.7125 + 0.7423
False Negative Rate	0.2875	0.1918	0.3995	$C / (A + C)$	23 / 80
False Positive Rate	0.2577	0.2206	0.2976	$B / (B + D)$	134 / 520
Positive Predictive Value	0.2984	0.2345	0.3687	$A / (A + B)$	57 / 191
Negative Predictive Value	0.9438	0.9168	0.9640	$D / (C + D)$	386 / 409
False Omission Rate	0.0562	0.0360	0.0832	$C / (C + D)$	23 / 409
False Discovery Rate	0.7016	0.6313	0.7655	$B / (A + B)$	134 / 191
Prevalence	0.1333	0.1072	0.1632	$(A + C) / N$	80 / 600
Prop. Correctly Classified	0.7383	0.7012	0.7731	$(A + D) / N$	443 / 600
Prop. Incorrectly Classified	0.2617	0.2269	0.2988	$(B + C) / N$	157 / 600
LR+	2.7649	2.2600	3.3826	TPR / FPR	0.7125 / 0.2577
LR-	0.3873	0.2733	0.5489	FNR / TNR	0.2875 / 0.7423
Likelihood Odds Ratio	7.1389	4.2335	12.0381	$LR+ / LR-$	2.7649 / 0.3873

Proportion Confidence Interval Method: Exact Binomial
 Ratio Confidence Interval Method: Katz Logarithm
 Odds Ratio Confidence Interval Method: Logarithm

This report displays a variety of summary measures with corresponding confidence limits. The value formula and calculation are given in the right two columns. The prevalence estimate here is a reasonable estimate, since the entire sample of 600 individuals was a random sample from the population.