ROC curve: Receiver Operating Characteristic

1) Introduction

The diagnostic performance of a test, or the accuracy of a test to discriminate diseased cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis. ROC curves can also be used to compare the diagnostic performance of two or more raters.

2) Interpretation of ROC curves

First of all, we remind the following definitions:

- Sensitivity = True Positive rate (proportion of patients with the disease Correctly diagnosed).
- Specificity = True Negative rate (proportion of patients without the disease correctly diagnosed).
- 1-Specificity = False Positive rate (proportion of patients without the disease who are incorrectly diagnosed as having the disease).

In a ROC (Receiver Operating Characteristic) curve, the True Positive rate (TP = Sensitivity) is plotted as a function of the False Positive rate (FP = 1 - Specificity).

Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (diseased/normal). The ROC curve is a fundamental tool for diagnostic test evaluation.

When the variable under study cannot distinguish between the two groups, i.e. where there is no difference between the two distributions, the area will be equal to 0.5 (the ROC curve will coincide with the diagonal, see Blue line on Figure 1). When there is a perfect separation of the values of the two groups, i.e. no overlapping of the distributions, the area under the ROC curve equals 1 (the ROC curve will reach the upper left corner of the plot, see Yellow line on Figure 1).

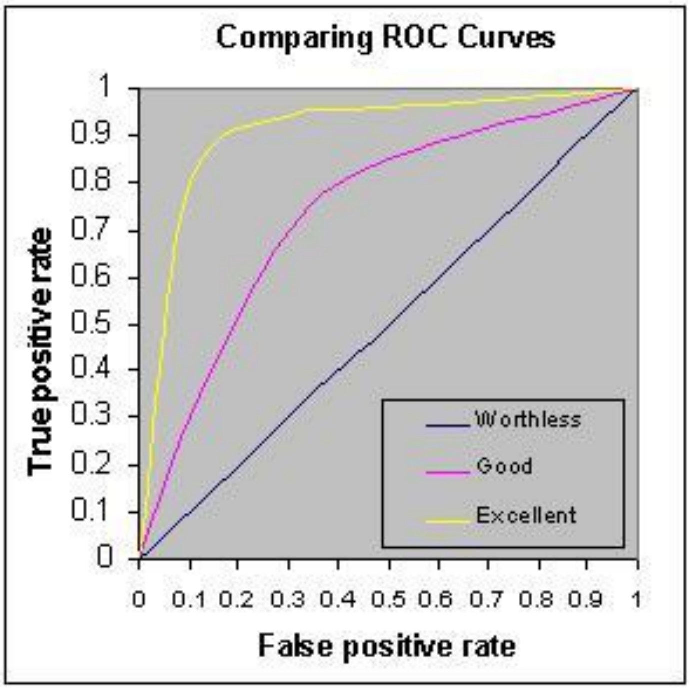


Figure 1. Example of ROC curves.

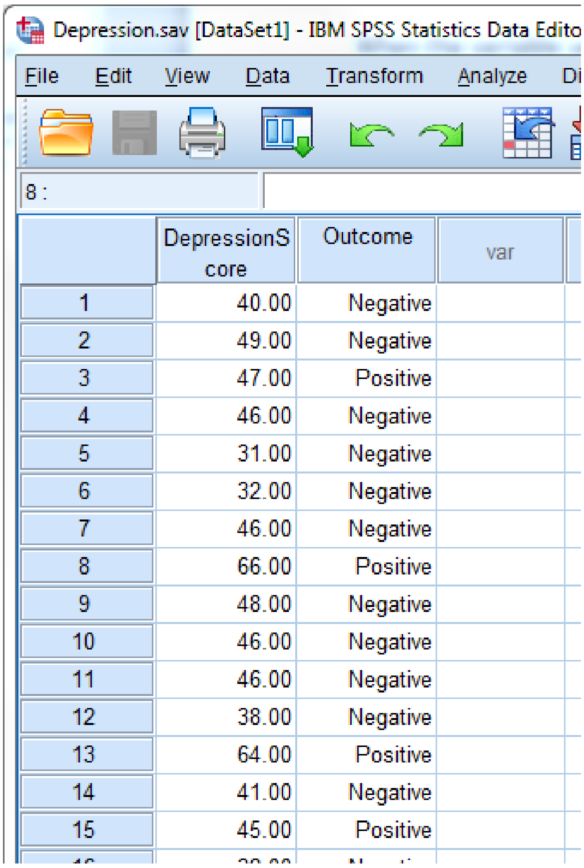
3) Assumptions

In order to perform ROC curve analysis, you should meet the following three assumptions:

- **Assumption 1:** You should usually have a **continuous measurement of interest** (= the parameter you want to study) and **an independent diagnosis**, i.e. a **dichotomous variable** also called **state variable**, which classifies your study subjects into two distinct groups: a diseased and non-diseased group.
- **Assumption 2:** The **state variable** should be **independent** from the **measurement of interest**.
- **Assumption 3:** The cases (e.g., participants) are a random sample from the population of interest.

4) Presentation of the data and Research question

A researcher is interested to see if a measurement of depression can predict if a patient will have a major depressive episode during the six months after this measurement has been taken. The data consists of a cohort of 32 patients who have had their measurement score measured as a scale. The researcher wants to know if a higher depression score tends to lead to a positive outcome, i.e. a major depressive episode. The data is in “Depression.sav” (available on the MASH website practice data set).

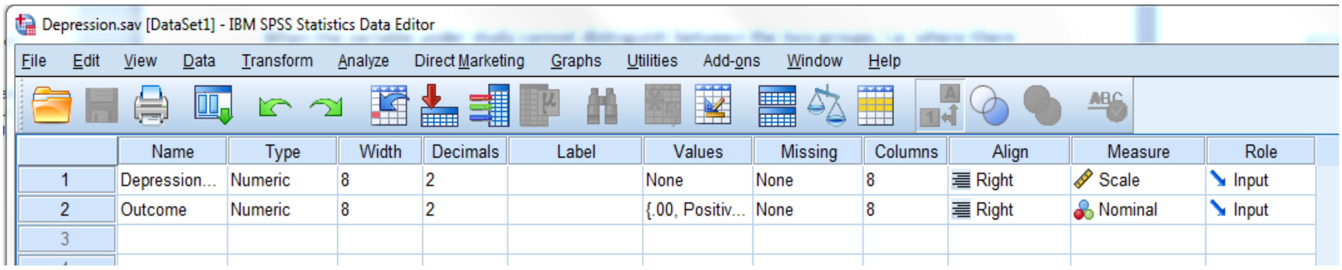


Depression.sav [DataSet1] - IBM SPSS Statistics Data Edito

File Edit View Data Transform Analyze Di

8 :

	DepressionScore	Outcome	var
1	40.00	Negative	
2	49.00	Negative	
3	47.00	Positive	
4	46.00	Negative	
5	31.00	Negative	
6	32.00	Negative	
7	46.00	Negative	
8	66.00	Positive	
9	48.00	Negative	
10	46.00	Negative	
11	46.00	Negative	
12	38.00	Negative	
13	64.00	Positive	
14	41.00	Negative	
15	45.00	Positive	



Depression.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Depression...	Numeric	8	2		None	None	8	Right	Scale	Input
2	Outcome	Numeric	8	2		{.00, Positiv...	None	8	Right	Nominal	Input
3											

Figure 2. Presentation of the data.

The outcome variable ("outcome") is dichotomous and represents whether the participants did or did not have any major depressive episode within 6 months after being administered the instrument that generated their depression score "DepressionScore" (measurement of interest). We believe that a high score of "DepressionScore" tends to be associated with a Positive outcome. The state variable is going to be the outcome and we set 0 as being the positive outcome (i.e. major depressive episode within the six months after measurement) and 1 as being the negative outcome.

5) Procedure on SPSS

Click **Analyze > ROC Curve...** on the main menu (as shown below).

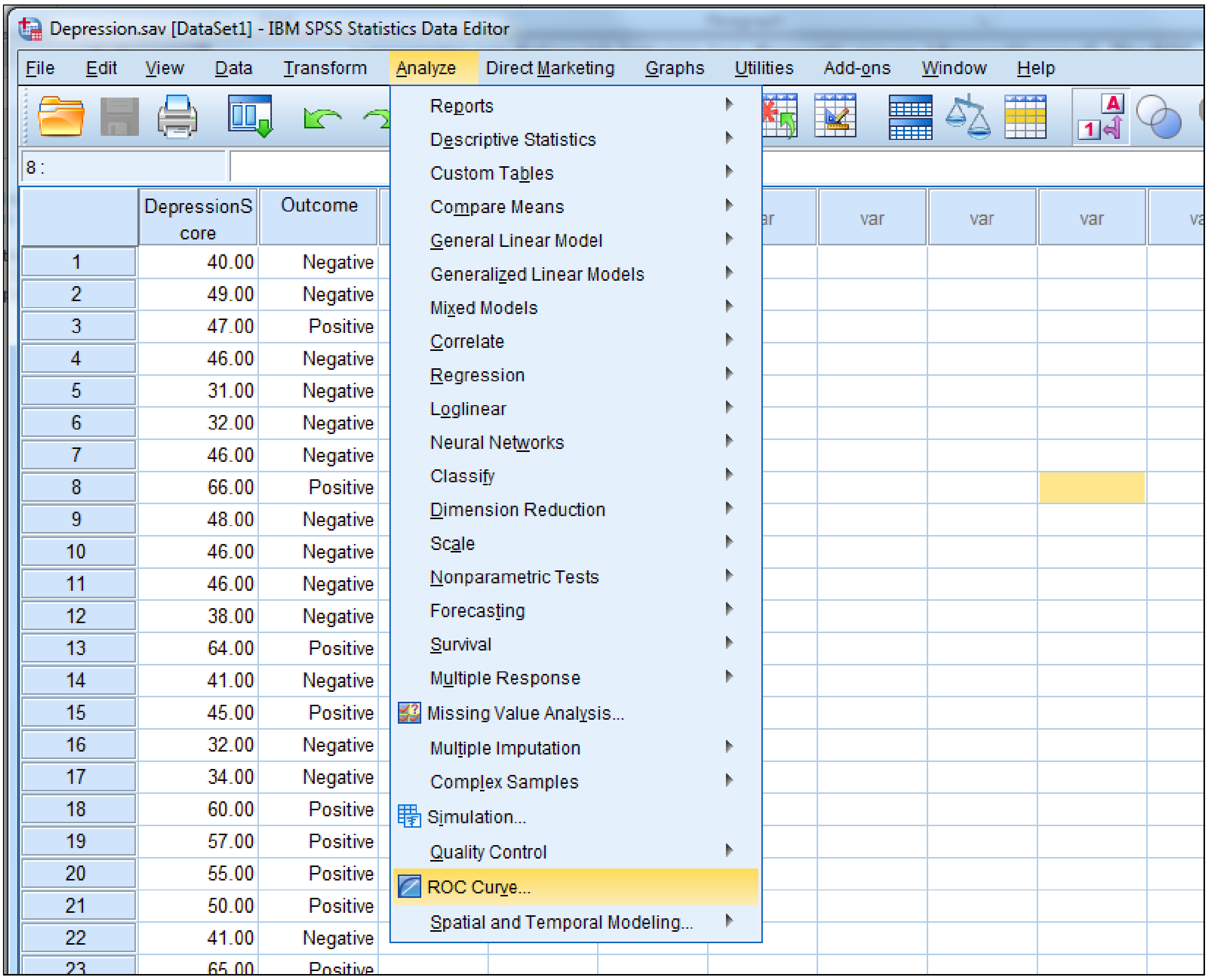


Figure 3. Selecting ROC curve window.

In the ROC Curve dialog box (see Figure 4), put the measurement of interest (DepressionScore) in “Test Variable” and the state variable “outcome” in “State Variable”. You should enter in the Value of State Variable the value for which the event occurs (i.e. the positive outcome: 0). Then it is recommended to tick all the options below, in particular “ROC Curve” and “With diagonal reference line”. In the Options box, make sure that the confidence level in Parameters for Standard Error of Area is 95%.

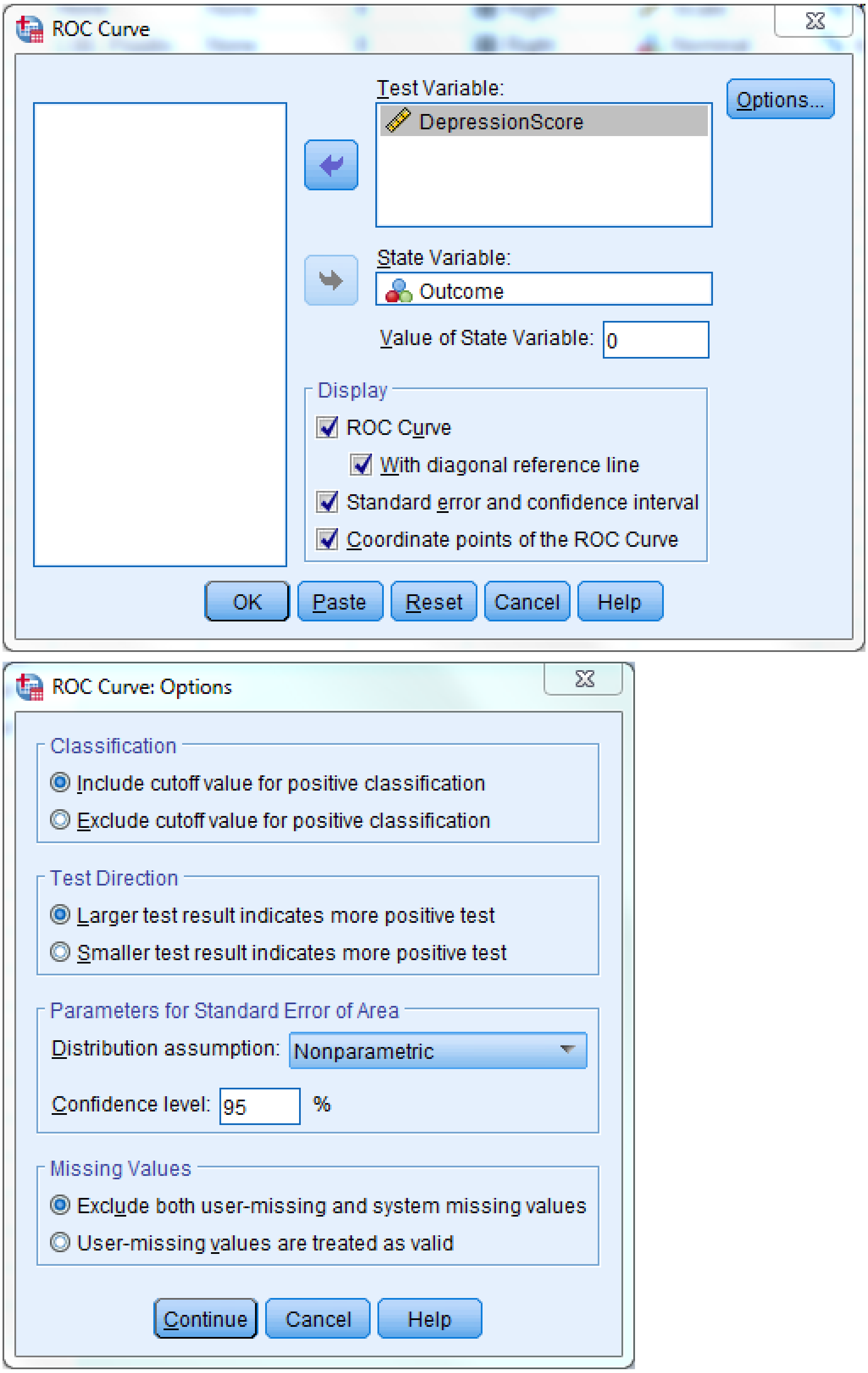


Figure 4. Options to select for the ROC curves analysis.

6) Results

On Figure 5, we can see that the blue line represents the true positive rate against the false positive rate and the bigger the area under the blue curve, the better the diagnosis is.

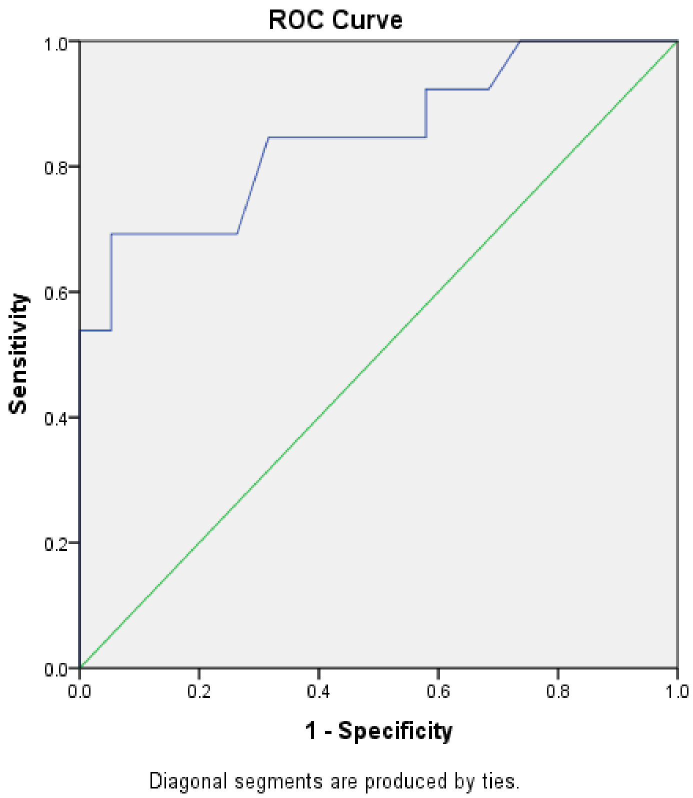


Figure 5. Graph representing the ROC curves.

The Significance level or p-value is the probability that the observed sample Area under the ROC curve is found when in fact, the true (population) Area under the ROC curve is 0.5 (null hypothesis: Area = 0.5). If p is small ($p<0.05$) then it can be concluded that the Area under the ROC curve is significantly different from 0.5 and therefore there is evidence that the laboratory test does have an ability to distinguish between the two groups. The p-value shown on the table on Figure 6 is equal to 0.001 and thus suggests very strong evidence against the null hypothesis. We will then conclude that the depression score can significantly predict the outcome variable.

The 95% Confidence Interval is the interval in which the true (population) Area under the ROC curve lies with 95% confidence.

Area Under the Curve

Test Result Variable(s): DepressionScore

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.848	.073	.001	.704	.992

The test result variable(s): DepressionScore has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

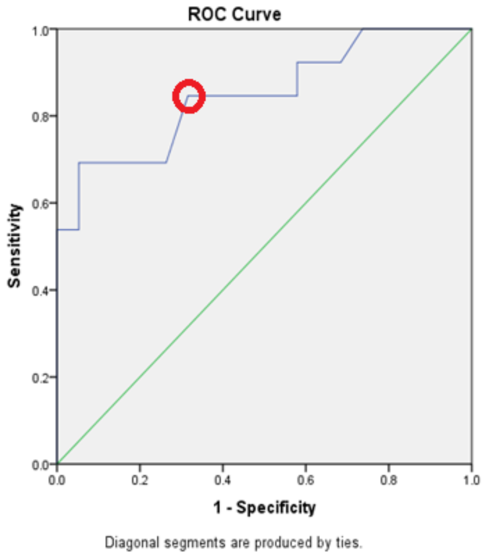
a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 6. Analysis of the area under the curve.

A cut-off score is not necessarily easy to determine. On Figure 7, if we take 46.5000 as our cut off value, we can see that the sensitivity is 84.6% and that the False Positive rate is equivalent to 31.6% (i.e. a Specificity of 68.4%). It means that if we were to choose 46.5000 as our cut-off point, 84.6% of the positive outcome would be correctly predicted by the depression score if the depression score were above or equal to 46.5000. On the other hand, 31.6% of the positive outcome would be incorrectly predicted. Choosing the cut-off score will depend on what percentage of positive outcome are correctly predicted and on what proportion of positive outcome are incorrectly predicted.

The choice of the cut-off value is quite subjective, depending on what rate of errors you would tolerate. We advise you to look at the ROC curve and to pick up the point on the curve which is the furthest away from the line.



The furthest away point corresponds to the coordinates (0.316,0.846), leading to the cutoff value 46.5000 from the table below:

Coordinates of the Curve

Test Result Variable(s): DepressionScore

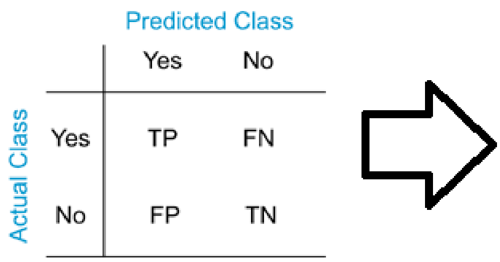
Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
30.0000	1.000	1.000
31.5000	1.000	.947
33.0000	1.000	.842
36.0000	1.000	.789
39.0000	1.000	.737
40.5000	.923	.684
43.0000	.923	.579
45.5000	.846	.579
46.5000	.846	.316
47.5000	.692	.263
48.5000	.692	.211
49.5000	.692	.053
52.5000	.615	.053
55.5000	.538	.053
56.5000	.538	.000
58.5000	.462	.000
61.5000	.308	.000
63.5000	.231	.000
64.5000	.154	.000
65.5000	.077	.000
67.0000	.000	.000

Figure 7. Determining the cut off values of the depression score.

If we choose 46.5000 as cut-off value to determine if a subject is positive (Yes) or negative (No), we build in fact a classifier model. An important feature is to determine how good this classifier is, i.e. how good my model based on the 46.5000 cut-off value is.

For that, we count the number of people who are predicted accurately and inaccurately. There will be some cases with the disease correctly classified as positive (TP = True Positive fraction), but some cases with the disease will be classified negative (FN = False Negative fraction). On the other hand, some cases without the disease will be correctly classified as negative (TN = True Negative fraction), but some cases without the disease will be classified as positive (FP = False Positive fraction). We group all these counts in a table (see below on the left).

In order to assess our classifier, we can compute the Accuracy, which is defined as the number of accurate predictions divided by the total number of individuals.



$$Accuracy = \frac{TP + TN}{P + N}$$

$$Error\ rate = \frac{FP + FN}{P + N}$$

$$Sensitivity = \frac{TP}{P}$$

$$Specificity = \frac{TN}{N}$$

The table on next page is built from the prediction (Positive/Negative) based on the choice of 46.5000 as cut-off value. The accuracy is (11+2)/(11+2+6+13) = 0.75.

	Predicted Positive	Predicted Negative	Total
Actual Positive	11 (84.6%)	2 (15.4%)	13 (100%)
Actual Negative	6 (31.6%)	13 (68.4%)	19 (100%)