CHAPTER 89

ROC Curves

Introduction

Receiver operating characteristic (ROC) curves are used to compare the performance of two competing diagnostic tests and to determine the appropriate cutoff values of those tests. The technique is used when you have a continuous criterion (predictor or independent) variable which will be used to make a yes or no decision. ROC curves are most often employed in the medical fields. It may be used with any classification procedure. For example, in may be used with two-group discriminant analysis to help determine the appropriate cutoff value of the discriminant score for classifying a individuals.

The most complete discussion we found was in Altman (1991). Gehlbach (1988) provides an example of its use.

An Example

ROC curves are explained with a fictional example paraphrased from Gehlbach (1988). Forty-five patients with fever, headache, and a history of tick bite were classified into two groups: those with Rocky Mountain Spotted Fever (RMSF) and those without it. The serum sodium level of each patient is measured using both techniques. We want to determine if serum sodium level is useful in detecting RMSF, which technique is most accurate in diagnosing RMSF, and what the diagnostic cutoff value of the selected test should be.

RMSF=Yes					**RMSF=No**		
ID	**Method1**	**Method2**	**Diagnosis**	**ID**	**Method1**	**Method2**	**Diagnosis**
1	124	122	1	22	129	124	0
2	125	124	1	23	131	128	0
3	126	125	1	24	131	130	0
4	126	125	1	25	134	133	0
5	127	126	1	26	134	133	0
6	128	126	1	27	135	133	0
7	128	127	1	28	136	134	0
8	128	128	1	29	136	134	0
9	128	128	1	30	136	134	0
10	129	128	1	31	137	134	0
11	129	130	1	32	137	136	0
12	131	130	1	33	138	136	0
13	132	133	1	34	138	137	0
14	133	133	1	35	139	138	0
15	133	134	1	36	139	138	0
16	135	134	1	37	139	140	0
17	135	134	1	38	139	140	0
18	135	134	1	39	140	141	0
19	136	136	1	40	140	141	0
20	138	138	1	41	141	142	0
21	139	140	1	42	142	142	0
				43	142	142	0
				44	142	142	0
				45	143	144	0

The next step in analyzing these data is to create a two-by-two table showing the diagnostic accuracy of a method at given cutoff value. If the cutoff is set at X, the table would appear as follows:

Generic Table for Cutoff=X

Method Cutoff=X	RMSF = Yes	RMSF = No
Sodium <= X, Positive	A	B
Sodium > X, Negative	C	D

The letters A, B, C, and D represent counts of the number of individuals in each of the four possible categories.

For example, if a cutoff of 130 is used to diagnose those with the disease, the tables for each sodium measurement method would be:

Table for Method 1, Cutoff=130

Method 1 Cutoff=130	RMSF=Yes	RMSF=No
Sodium<=130, Positive	11	1
Sodium>130, Negative	10	23

Table for Method 2, Cutoff=130

Method 2 Cutoff=130	RMSF=Yes	RMSF=No
Sodium<=130, Positive	12	3
Sodium>130, Negative	9	21

If a cutoff of 137 is used to diagnose those with the disease, the tables for each sodium measurement method would be:

Table for Method 1, Cutoff=137

Method 1 Cutoff=137	RMSF=Yes	RMSF=No
Sodium<=137, Positive	19	11
Sodium>137, Negative	2	13

Table for Method 2, Cutoff=137

Method 2 Cutoff=137	RMSF=Yes	RMSF=No
Sodium<=137, Positive	19	13
Sodium>137, Negative	12	11

As you study these tables, you can see that changing the cutoff value drastically changed the methods ability to diagnose those with RMSF. Unfortunately, it is difficult to make distinct comparisons because we are dealing with counts. To aid in interpretation, analysts have developed several proportions that allow direct comparison from table to table, independent of sample size. Two of the most popular ratios are sensitivity and specificity. We will now define these and a few other common terms used in this type of analysis. In these definitions, the terms positive and negative refer to the presence or absence of the condition of interest.

Sensitivity

Sensitivity is the proportion of those with the disease (positives) that are correctly identified as having the disease. In terms of our two-by-two tables, sensitivity = *A/(A+C).*

Specificity

Specificity is the proportion of those without the disease (negatives) that are correctly identified as not having the disease. In terms of our two-by-two tables, specificity = *D/(B+D).*

Prevalence

Prevalence is the overall proportion of individuals with the disease. In terms of our two-by-two tables, prevalence = *(A+C)/(A+B+C+D).*

Positive Predictive Value (PPV)

PPV is the proportion of individuals with positive test results who are correctly diagnosed as having the condition. In terms of our two-by-two tables, PPV = *(A)/(A+B).*

Negative Predictive Value (NPV)

NPV is the proportion of individuals with negative test results who are correctly diagnosed as not having the condition. In terms of our two-by-two tables, NPV = *(D)/(C+D).*

Discussion

The problem with sensitivity and specificity is that they do not assess the probability of making a correct diagnosis. To overcome this, practitioners have developed two other indices: PPV and NPV. Unfortunately, these indices have the disadvantage that they are directly impacted by the prevalence of the disease in the population. For example, if your sampling procedure is constructed to obtain more individuals with the disease than is the case in the whole population of interest, the PPV and NPV need to be adjusted.

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

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

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

Another way of interpreting these terms is as follows. The prevalence of a disease is the *prior probability* that a subject has the disease before the diagnostic test is run. The values of PPV and 1-NPV are the *posterior probabilities* of a subject having the disease after the test is conducted.

Likelihood Ratio

The likelihood ratio statistic measures the value of the test for increasing certainty about a positive diagnosis. It is calculated as follows:

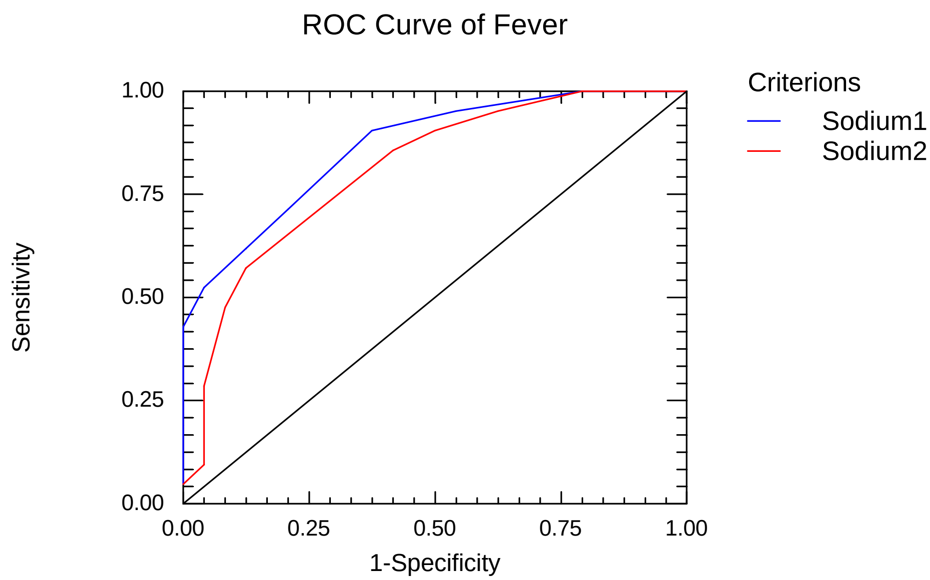
$$LR = \frac{\Pr(\textit{positive test}|\textit{disease})}{\Pr(\textit{positive test}|\textit{no disease})} = \frac{\textit{sensitivity}}{1 - \textit{specificity}}$$

ROC Curve

When a continuous variable (such as serum sodium) is used in a diagnostic test, the choice of a cutoff value is not simple. One approach to this problem is to plot the sensitivity on the vertical axis versus 1-specificity on the horizontal for several possible cutoff values and join the points with a line. The resulting curve is called the *receiver operating characteristic* (ROC) curve because the technique was developed in studies of signal detection by radar operators.

Although popular, ROC curves do not take into account disease prevalence. This is a liability of the procedure.

Following is a sample ROC curve.



ROC curves may be compared in several ways. The most popular seems to be comparing the areas under the curves. Notice that the area of a test the predicts a certain state (disease) perfectly is one. A chance predictor (such as flipping a coin) will produce a 45-degree line with an area of one-half. Hence, the closer the area is to one the better. Also, the larger the area at a given point, the better.

The program calculates the estimated area under each curve using the trapezoidal rule as well as the standard error of the estimated area. The formulas used to calculate the standard errors, based on the negative exponential approximation to the exact formula as proposed by Hanley (1982), are:

$$SE(q) = \sqrt{\frac{q(1-q) + (n_y - 1)(Q_1 - q^2) + (n_n - 1)(Q_2 - q^2)}{n_y n_n}}$$

where q is the area, n_y is the number with the condition, n_n is the number without the condition, and

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

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

Data Structure

The data are entered in two or more variables. One variable specifies the true condition of the individual. The other variable(s) contain the criterion value(s) for the tests being compared.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

This panel specifies which variables are used in the analysis.

Response Variable

A binary response variable which represents whether or not the individual actually has the condition of interest. The value representing a yes is specified in the Positive Test Value box.

Frequency Variable

An optional variable containing a set of frequencies. Normally, each row represents one individual. On occasion, however, a row may represent more than one individual. This variable contains the number of individuals that a row represents.

Criterion Variable(s)

A list of one or more criterion (test, score, discriminant, etc.) variables. If more than one variable is listed, a separate curve is drawn for each.

Test Direction

This option lets indicate whether low or high values of the criterion variable are associated with a positive condition of the response variable. For example, low values of one criterion variable may indicate the presents of a disease, while high values of another criterion variable may associated with the disease.

Low X = Positive

A low value of the criterion variable indicates a positive test result. That is, a low value will indicate a positive test.

High X = Positive

A high value of the criterion variable indicates a positive test result. That is, a high value will indicate a positive test.

Positive Test Value

This is the value of the Response Variable that indicates that the individual has the condition of interest. All other values of the Response Variable are considered as not having the condition of interest. Often, the positive value is set to ‘1’ and the negative value is set to ‘0.’ However, any numbering scheme may be used.

Filter Active

This option indicates whether the currently defined filter (if any) should be used when running this procedure.

Reports Tab

The following options control the reports that are displayed.

Page Title

Specify a page title to be displayed in report headings.

Number Intervals

The range of a criterion variable is divided into this many intervals and a two-by-two table is calculated. Note that each interval ends with a cutoff point.

Prevalence 1-2

Prevalence is the proportion of individuals in the population that have the condition of interest. As a proportion, this number varies between zero and one. These options allow you to specify two additional prevalence values to go along with the one that is computed from the data. Using these values, adjusted values of PPV and NPV are calculated.

Criterion Minimum - Maximum

Normally, the data range of each criterion variable is dividing into a specified number of intervals. This option lets you override the data range and put in your own range. When two or more criterion variables are used, using the minimum and maximum causes a uniform scaling of these variables on all reports.

Criterion Decimals

Specifies the number of decimal places to use when displaying the criterion variable values.

ROC Plot

Check to display this plot or report.

ROC Report

Check to display this plot or report.

Diagnosis Report

Check to display this plot or report.

AUC Report

Check to display this plot or report.

Variable Names

Specify whether to use variable names or (the longer) variable labels in report headings.

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while the double precision will display thirteen-place accuracy. Note that all reports are formatted for single precision only.

ROC Curve Tab

The options on this panel control the appearance of the ROC Curve.

Plot Style File

Designate a scatter plot style file. This file sets all scatter plot options that are not set directly on this panel. Unless you choose otherwise, the ROC style file is used. These style files are created in the Scatter Plot procedure.

Plot Title

This is the text of the title. The characters *{Y}* is replaced by the response variable name. Press the button on the right of the field to specify the font of the text.

Label (Y and X)

This is the text of the label. Press the button on the right of the field to specify the font of the text.

Tick Marks - Ref. Numbers (Y and X)

Pressing these buttons brings up a window that sets the font, rotation, and number of decimal places displayed in the reference numbers along the vertical and horizontal axes.

Major Ticks - Minor Ticks (Y and X)

These options set the number of major and minor tickmarks displayed on each axis.

Grid Lines (Y and X)

These check boxes indicate whether the grid lines should be displayed.

Lines Tab

These options set the color, width, and pattern of the up to fifteen lines representing the criterion variables. Note that the color of the 45 degree line is specified in the group immediately after the criterion variables. For example, if you had three criterion variables, the color of the Group 4 option would be the color of the 45 degree line.

Double-clicking the line, or clicking the button to the right of the symbol, brings up a line specification window. This window lets you specify the characteristics of each line in detail.

Color

The color of the line.

Width

The width of the line.

Pattern

The line pattern (solid, dot, dash, etc.).

Legend Tab

These options control the legend.

Show Legend

Specifies whether to display the legend.

Legend Text

Specifies the title of the legend. Click the button on the right to specify the font size, color, and style of the legend text.

Storage Tab

Various proportions may be stored on the current database for further analysis. This group of options lets you designate which statistics (if any) should be stored and which variables should receive these statistics. The selected statistics are automatically stored to the current database.

Note that the variables you specify must already have been named on the current database.

Note that existing data is replaced. Be careful that you do not specify variables that contain important data.

Criterion - Variable Name

The values of each of these proportions is stored in the indicated variables. When you leave a box blank, nothing is stored.

Template Tab

The options on this panel allow various sets of options to be loaded (File menu: Load Template) or stored (File menu: Save Template). A template file contains all the settings for this procedure.

File Name

Designate the name of the template file either to be loaded or stored.

Template Files

A list of previously stored template files for this procedure.

Template Id's

A list of the Template Id's of the corresponding files. This id value is loaded in the box at the bottom of the panel.

Tutorial

This section presents an example of how to generate an ROC curve. The ROC database contains the data for the RMSF example that was discussed earlier. We will show you how to generate the ROC curve and associated reports to analyze these data.

1 **Open the ROC dataset.**

- From the File menu of the NCSS Data window, select Open.
- Select the Data subdirectory of the NCSS97 directory.
- Click on the file **ROC.s0**.
- Click **Open**.

2 **Open the ROC Curves window.**

- On the menus, select Analysis, then Survival / Reliability, then ROC Curves. The ROC Curves procedure will be displayed.
- On the menus, select File, then New Template. This will fill the procedure with the

default template.

3 Specify the variables.

- On the ROC Curves window, select the **Variables tab**.
- Double-click in the Response Variable text box. This will bring up the variable selection window.
- Select **Fever** from the list of variables and then click **Ok**. “Fever” will appear in the Response Variable box.
- Double-click in the Criterion Variables text box. This will bring up the variable selection window.
- Select **Sodium1, Sodium2** from the list of variables and then click **Ok**. “Sodium1-Sodium2” will appear in the Criterion Variable box.

4 Run the procedure.

- From the Run menu, select Run Procedure. Alternatively, just click the Run button (the left-most button on the button bar at the top).

ROC Section

ROC Section for Fever by Sodium1

Cutoff	YesYes	YesNo	NoYes	NoNo	Sensitivity		False+	Specificity	Cumulative
Sodium1	A	B	C	D	A/(A+C)	C/(A+C)	B/(B+D)	D/(B+D)	Area
124.00	1	0	20	24	0.0476	0.9524	0.0000	1.0000	0.000000
126.11	4	0	17	24	0.1905	0.8095	0.0000	1.0000	0.000000
128.22	9	0	12	24	0.4286	0.5714	0.0000	1.0000	0.000000
130.33	11	1	10	23	0.5238	0.4762	0.0417	0.9583	0.019841
132.44	13	3	8	21	0.6190	0.3810	0.1250	0.8750	0.067460
134.56	15	5	6	19	0.7143	0.2857	0.2083	0.7917	0.123016
136.67	19	9	2	15	0.9048	0.0952	0.3750	0.6250	0.257937
138.78	20	13	1	11	0.9524	0.0476	0.5417	0.4583	0.412698
140.89	21	19	0	5	1.0000	0.0000	0.7917	0.2083	0.656746
143.00	21	24	0	0	1.0000	0.0000	1.0000	0.0000	0.865079

ROC Section for Fever by Sodium2

Cutoff	YesYes	YesNo	NoYes	NoNo	Sensitivity		False+	Specificity	Cumulative
Sodium2	A	B	C	D	A/(A+C)	C/(A+C)	B/(B+D)	D/(B+D)	Area
122.00	1	0	20	24	0.0476	0.9524	0.0000	1.0000	0.000000
124.44	2	1	19	23	0.0952	0.9048	0.0417	0.9583	0.002976
126.89	6	1	15	23	0.2857	0.7143	0.0417	0.9583	0.002976
129.33	10	2	11	22	0.4762	0.5238	0.0833	0.9167	0.018849
131.78	12	3	9	21	0.5714	0.4286	0.1250	0.8750	0.040675
134.22	18	10	3	14	0.8571	0.1429	0.4167	0.5833	0.249008
136.67	19	12	2	12	0.9048	0.0952	0.5000	0.5000	0.322421
139.11	20	15	1	9	0.9524	0.0476	0.6250	0.3750	0.438492
141.56	21	19	0	5	1.0000	0.0000	0.7917	0.2083	0.601190
144.00	21	24	0	0	1.0000	0.0000	1.0000	0.0000	0.809524

The report displays the numeric information used to generate the ROC curve.

Cutoff

The cutoff values of the criterion variable.

A B C D

These four columns give the counts of the two-by-two tables that are formed at each of the corresponding cutoff points.

Sensitivity A/(A+C)

This is the proportion of those that had the disease that were correctly diagnosed by the test.

C/(A+C)

This is the proportion of those that had the disease that were incorrectly diagnosed.

False + B/(B+D)

The proportion of those who did not have the disease who were incorrectly diagnosed by the test as having it.

Specificity D/(B+D)

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

Cumulative Area

This is the cumulative area under the ROC curve. Usually, only the final value is used.

Predicted Value Section

Predicted Value Section for Fever

Cutoff Sodium1	Sensi-tivity	Speci-ficity	Likelihood Ratio	Prev = PPV	0.4667 NPV	Prev = PPV	0.1000 NPV	Prev = PPV	0.9000 NPV
124.00	0.0476	1.0000		1.0000	0.5455	1.0000	0.9043	1.0000	0.1045
126.11	0.1905	1.0000		1.0000	0.5854	1.0000	0.9175	1.0000	0.1207
128.22	0.4286	1.0000		1.0000	0.6667	1.0000	0.9403	1.0000	0.1628
130.33	0.5238	0.9583	12.5714	0.9167	0.6970	0.5828	0.9477	0.9912	0.1827
132.44	0.6190	0.8750	4.9524	0.8125	0.7241	0.3549	0.9539	0.9781	0.2033
134.56	0.7143	0.7917	3.4286	0.7500	0.7600	0.2759	0.9614	0.9686	0.2354
136.67	0.9048	0.6250	2.4127	0.6786	0.8824	0.2114	0.9834	0.9560	0.4217
138.78	0.9524	0.4583	1.7582	0.6061	0.9167	0.1634	0.9886	0.9406	0.5168
140.89	1.0000	0.2083	1.2632	0.5250	1.0000	0.1231	1.0000	0.9191	1.0000
143.00	1.0000	0.0000	1.0000	0.4667	1.0000	0.1000	1.0000	0.9000	1.0000

Cutoff Sodium2	Sensi-tivity	Speci-ficity	Likelihood Ratio	Prev = PPV	0.4667 NPV	Prev = PPV	0.1000 NPV	Prev = PPV	0.9000 NPV
122.00	0.0476	1.0000		1.0000	0.5455	1.0000	0.9043	1.0000	0.1045
124.44	0.0952	0.9583	2.2857	0.6667	0.5476	0.2025	0.9051	0.9536	0.1053
126.89	0.2857	0.9583	6.8571	0.8571	0.6053	0.4324	0.9235	0.9841	0.1297
129.33	0.4762	0.9167	5.7143	0.8333	0.6667	0.3883	0.9403	0.9809	0.1628
131.78	0.5714	0.8750	4.5714	0.8000	0.7000	0.3368	0.9484	0.9763	0.1849
134.22	0.8571	0.5833	2.0571	0.6429	0.8235	0.1860	0.9735	0.9488	0.3121
136.67	0.9048	0.5000	1.8095	0.6129	0.8571	0.1674	0.9793	0.9421	0.3684
139.11	0.9524	0.3750	1.5238	0.5714	0.9000	0.1448	0.9861	0.9320	0.4667
141.56	1.0000	0.2083	1.2632	0.5250	1.0000	0.1231	1.0000	0.9191	1.0000
144.00	1.0000	0.0000	1.0000	0.4667	1.0000	0.1000	1.0000	0.9000	1.0000

The report displays the numeric information used to generate the ROC curve.

Cutoff

The cutoff values of the criterion variable.

Sensitivity

This is the proportion of those that had the disease that were correctly diagnosed by the test.

Specificity

This is the proportion of those who did not have the disease who were correctly diagnosed as such.

Likelihood Ratio

This is the value of the Likelihood Ratio statistic which was discussed earlier.

Prev = x.xxxx PPV

The values of PPV for the three prevalence values. The first prevalence value is the one that was calculated from the data. The final two values were set as options. The default was 0.1 and 0.9, which is what was used here.

Prev = x.xxxx NPV

The values of NPV for the three prevalence values. The first prevalence value is the one that was calculated from the data. The final two values were set as options. The default was 0.1 and 0.9, which is what was used here.

Area Under Curve Section

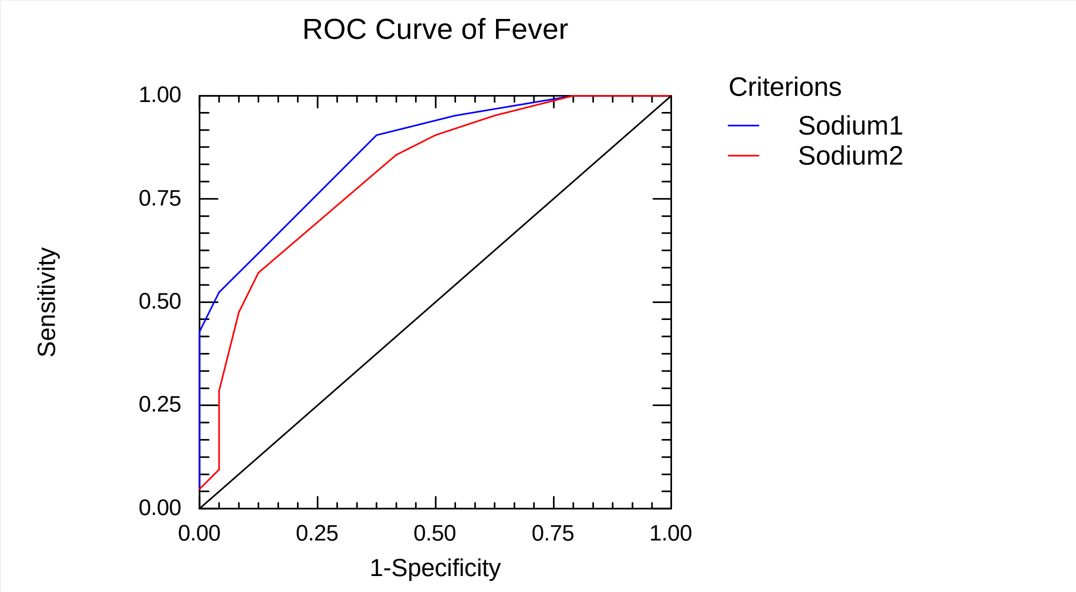
Area Under Curve Section

Criterion Variable	Area Under Curve	Standard Error	Count Yes's	Count No's
Sodium1	0.865079	0.057033	21	24
Sodium2	0.809524	0.066580	21	24

This report displays areas under the ROC curve and associated standard errors for each of the criterion variables. Note that in this example, the first method, Sodium1, has a higher value and so would be chosen as the diagnostic method.

ROC Plot Section

ROC Plot Section



The ROC curve plots the proportion of those who actually had the disease who were correctly diagnosed on the vertical axis versus the proportion of those who did not have the disease who were falsely diagnosed as having it on the horizontal axis. Hence, an optimum test procedure is one whose ROC curve proceeds from the lower-left corner vertically until it reaches the top and then horizontally across the top to the right side. The 45 degree line represents what you would expect from a chance (flip of the coin) classification procedure.

When you are comparing two curves as in this example, you would generally take the outside curve (the one furthest from the middle line). However, it is possible for the curves to cross so that one test is optimum in a certain range but not in another.