

## Worksheet for Illustrative ROC Calculations by Rich Perline

Terminology:

	Actual 1	Actual 0
Predicted 1	TP	FP
Predicted 0	FN	TN

TP=True Positives

FP=False Positives

FN=False Negatives

TN=True Negatives

Sensitivity = True Positive Rate (TPR) =  $TP / (TP+FN)$  Actual Positives

1-Specificity = False Positive Rate (FPR) =  $FP / (FP+TN)$  (Actual Negatives)

Assume six cases with the following actual and predicted values:

Actual	Predicted Prob
1	0.8
1	0.7
0	0.6
1	0.4
0	0.3
0	0.2

Now construct an ROC curve using different cutoff values for the predicted probability.

Table 1.

		Actual 1	Actual 0
cutoff=1	Predicted 1	0	0
	Predicted 0	3	3

For Table 1:  $TPR = 0/3 = 0$  and  $FPR = 0/3 = 0$

Table 2

		Actual 1	Actual 0
cutoff=.7	Predicted 1	2	0
	Predicted 0	1	3

For Table 2:  $TPR = 2/3 = .67$  and  $FPR = 0/3 = 0$

Assume six cases with the following actual and predicted values:

Actual	Predicted Prob
1	0.8
1	0.7
0	0.6
1	0.4
0	0.3
0	0.2

Table 3

		Actual 1	Actual 0
cutoff=.6	Predicted 1	2	1
	Predicted 0	1	2

For Table 3:  $TPR = 2/3 = .67$  and  $FPR = 1/3 = .33$

Table 4

		Actual 1	Actual 0
cutoff=.4	Predicted 1	3	1
	Predicted 0	0	2

For Table 4:  $TPR = 3/3 = 1.0$  and  $FPR = 1/3 = .33$

Table 5

		Actual 1	Actual 0
cutoff=.3	Predicted 1	3	2
	Predicted 0	0	1

For Table 5:  $TPR = 3/3 = 1.0$  and  $FPR = 2/3 = .67$

Table 6

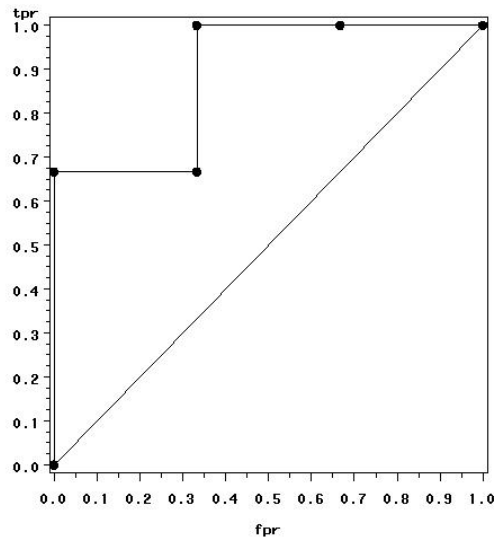
		Actual 1	Actual 0
cutoff=0	Predicted 1	3	3
	Predicted 0	0	0

For Table 6:  $TPR = 3/3 = 1.0$  and  $FPR = 3/3 = 1.0$

In summary, we get the following results from these six tables:

Cutoff	TPR	FPR
1	0	0
0.7	0.67	0
0.6	0.67	0.33
0.4	1	0.33
0.3	1	0.67
0	1	1

Plotting Sensitivity (True Positive Rate) on the vertical axis and 1-Specificity (False Positive Rate) on the horizontal axis for these six pairs of points gives this:



Here is the ROC plot for the sample data. The 45-degree reference line represents a “random” model.

The area under the ROC curve for the sample data can be easily computed as 8/9. Recall that this should be the same as the concordance statistic (“C-statistic”). To check that, we compute the C-statistic:

		Actual 1		
		0.4	0.7	0.8
Actual 0	0.2	Y	Y	Y
	0.3	Y	Y	Y
	0.6	N	Y	Y

This table compares each predicted probability for an Actual 0 case with each predicted probability for an Actual 1 case (i.e.,  $3 \times 3 = 9$  comparisons in all). Whenever the predicted probability for an Actual 0 case is less than the predicted probability for an Actual 1 case, that is indicated as a Y (i.e., concordant). There is only one discordant case indicated as N. Therefore, the C-statistic here is 8/9, which is the same as the area under the ROC curve.

### Lift and Concentration

Lift is based on concentration (also known as *sensitivity*) in the data that have been rank ordered by the posterior probabilities assigned by the model. Concentration is calculated as the number of events in each decile divided by the total number of events in the data. Lift is concentration divided by the decile or depth in the rank-ordered data. Note that concentration is very similar to the ROC. Concentration is Sensitivity, and this can be plotted against depth to derive the concentration curve, shown below.

Decile	Probability from the Model	Truth	Concentration	Lift
0.1	0.9	1	$1/5=0.2$	2.00
0.2	0.83	1	$2/5=0.4$	2.00
0.3	0.7	0	$2/5=0.4$	1.33
0.4	0.62	1	$3/5=0.6$	1.50
0.5	0.6	1	$4/5=0.8$	1.60
0.6	0.5	0	$4/5=0.8$	1.33
0.7	0.44	0	$4/5=0.8$	1.14
0.8	0.21	1	$5/5=1$	1.25
0.9	0.18	0	$5/5=1$	1.11
1	0.11	0	$5/5=1$	1.00

