Tutorial: ROC Curves in Stata

ROC curves illustrate the inherent trade-off of between sensitivity and specificity. We examine ROC curves in the context of risk prediction.

Consider the following scenario: you are responsible for telling a patient that they are at high or low risk for CHD, given some baseline prognostic factors. Using Framingham dataset, you can predict the probability that an individual gets CHD, given their baseline prognostic factors.

Constructing an ROC curve to evaluate a risk prediction model:

1) Using systolic blood pressure, number of cigarettes smoked per day, total cholesterol, sex, and BMI at baseline, predict the probability that each individual in the Framingham dataset had CHD. Call this probability p.

2) As in the diagnostic testing setting, select a cut-off probability c to distinguish high and low risk patients. If $p < c$, the patient is low risk. If $p \geq c$, the patient is high risk.

3) Classify all patients in the dataset as high risk or low risk using the cutoff c.

4) Calculate P(high risk | CHD) = sensitivity. Calculate P(high risk | no CHD) = 1 – specificity.

Steps 1 – 4 are beyond the scope of this module. These values are provided for you in the dataset **`roc.dta`**. **Open the dataset `roc.dta` in Stata.**

5) For the various values of c, plot the false positive rate versus sensitivity. Connect the lines to generate your ROC curve.

Consider the following questions:

- How do the sensitivity and specificity change as the cutoff increases from 0 to 1?

- What value of c would you choose in distinguishing high risk versus low risk patients? Why?

Table: Points on ROC curve for risk prediction model.

Cut-off (c)	Sensitivity	Specificity	False Positive (1-Specificity)
1	0	1.0000	0.0000
0.7901	0.0051	1.0000	0.0000
0.7152	0.0071	0.9988	0.0012
0.6592	0.0111	0.9966	0.0034
0.6055	0.0547	0.9931	0.0069
0.5695	0.0993	0.9875	0.0125
0.5055	0.1682	0.9654	0.0346
0.4595	0.2381	0.9480	0.0520
0.4049	0.3506	0.9221	0.0779
0.3555	0.4205	0.8794	0.1206
0.3029	0.5228	0.7997	0.2003
0.2545	0.6383	0.6963	0.3037
0.2031	0.7285	0.5723	0.4277
0.1559	0.8379	0.4184	0.5816
0.1044	0.9119	0.2408	0.7592
0.0571	0.9899	0.0551	0.9449
0	1.0000	0	1

Plot: ROC curve for risk prediction model