Tutorial: More on ROC curves and complicated graphs in Stata

We construct a new, simpler risk prediction model using only systolic blood pressure, diastolic blood pressure, and age as our prognostic factors. We compare this risk prediction model to the model in the previous tutorial.

- Open the dataset `roc.dta` on the course website.

We construct a plot that includes:

- the ROC curve for the first model from the previous tutorial (with many prognostic factors), called Model 1
- the second model, with only sex and blood pressure, Model 2, and
- a reference line representing arbitrary classification as high or low risk.

Overlaying lines in Stata is relatively easy within the *Twoway graph* window. Using the ROC plot, consider the following questions:

- Model 1 outperforms model 2. How can you tell this from the ROC curve?

- Which model would you recommend?

Later in the course, we learn how to fit the model to obtain the predicted risks. With new biomarkers and genetic risk factors popping up all the time, risk prediction is a hot topic in statistics right now and ROC curves are used frequently!

Table: Points on ROC curve for model 2.

Cut-off (c)	Sensitivity	Specificity	False Positive (1-Specificity)
1	0.0000	1.0000	0.0000
0.7901	0.0273	0.9997	0.0003
0.7152	0.0298	0.9991	0.0009
0.6592	0.0324	0.9975	0.0025
0.6055	0.0434	0.9942	0.0058
0.5695	0.0792	0.9804	0.0196
0.5055	0.1014	0.9699	0.0301
0.4595	0.2002	0.9460	0.0540
0.4049	0.2666	0.9064	0.0936
0.3555	0.4370	0.8224	0.1776
0.3029	0.5860	0.7006	0.2994
0.2545	0.6959	0.6046	0.3954
0.2031	0.7641	0.4736	0.5264
0.1559	0.8739	0.3371	0.6629
0.1044	0.9838	0.0865	0.9135
0.0571	1.0000	0.0000	1.0000
0	1.0000	0.0000	1.0000

Plot: ROC curve for Models 1 and 2, with reference line