Statistical Methods for Diagnostic Tests

Practical Exercise

Question 1

A new test, possibly cheaper than existing options, to detect the presence of *Helicobacter pylori* is being developed. In the sample of individuals tested (drawn from the attenders of a dyspepsia clinic in Glasgow), diagnosis was confirmed by endoscopic biopsy.

The data are in the .csv format file: **Helicobacter.csv**. The two variables are coded as follows:

Variable	**Coding**
Test	0 = Negative 1 = Positive
Biopsy	0 = *H. pylori* absent 1 = *H. pylori* present

a) Calculate appropriate summary measures for test performance and describe the test's performance in this population.

b) What would you estimate the PPV and NPV of the test to be if it was applied in London, if the prevalence of *H. pylori* was about 60% there?

Question 2

AnginaMI.csv contains the creatine kinase (CK) data from 120 patients suffering from either unstable angina or acute myocardial infarction (AMI). The variables are as follows:

Variable	**Coding**
Diagnosis	Character variable – "Angina" or "AMI"
CreatineKinase	Continuous value of CK
DiagnosisCat	0 = Angina 1 = AMI

(a) Draw a ROC curve to show test performance across the range of CK values. What overall performance can be attained?

(b) If you wished to create (instead) a good rule-out test, how would you modify the decision rule? (A rule-out test would be highly effective in saying that a patient would be unlikely to go on to develop AMI, but not necessarily very accurate in predicting AMI in a positive sense.)