Assignment 3

Keith G. Williams - 800690755

Thursday, June 04, 2015

Description

600 cases from a defined population have been subjected to some diagnostic Test A, Test B and Test C. Suppose that 100 actually positive cases and 500 actually negative cases were ultimately found in the population studied.

- In Test A, 120 people test result was positive and 480 people test result was negative. The diagnostic test to be evaluated yielded 70 true positive (TP) decisions and 50 false positive (FP) decisions.

- In Test B, 60 people test result was positive and 540 people test result was negative. The diagnostic test to be evaluated yielded 40 TP decisions and 20 FP decisions.

- In Test C, 70 people test result was positive and 530 people test result was negative. The diagnostic test to be evaluated yielded 45 TP decisions and 25 FP decisions.

Questions

1) Please first summarize these data in three separated contingency tables and calculate "Accuracy" for Test A, Test B and Test C.

```
##        observed
## test_A pos neg
##    pos  70  50
##    neg  30 450


##        observed
## test_B pos neg
##    pos  40  20
##    neg  60 480


##        observed
## test_C pos neg
##    pos  45  25
##    neg  55 475
```

```
# accuracy = (TP + TN) / N
n <- 600 # total number of cases
accuracy_A <- sum(diag(table_A)) / n
accuracy_B <- sum(diag(table_B)) / n
accuracy_C <- sum(diag(table_C)) / n
```

The accuracies for each test are summarized in the table below:

```
##   test  accuracy
## 1    A 0.8666667
## 2    B 0.8666667
## 3    C 0.8666667
```

2) Please compare the accuracy of Test A, B and C and tell which test is best and which test is worst?

$$0.87 = Accuracy_A = Accuracy_B = Accuracy_C$$

The accuracies of each diagnostic test are equal. This single metric loses some information, so an ROC curve will better differentiate the three tests.

3) Please draw simple ROC curves for Tests A, B, C. Can you tell which test is best and which test is worst based on the ROC curves?
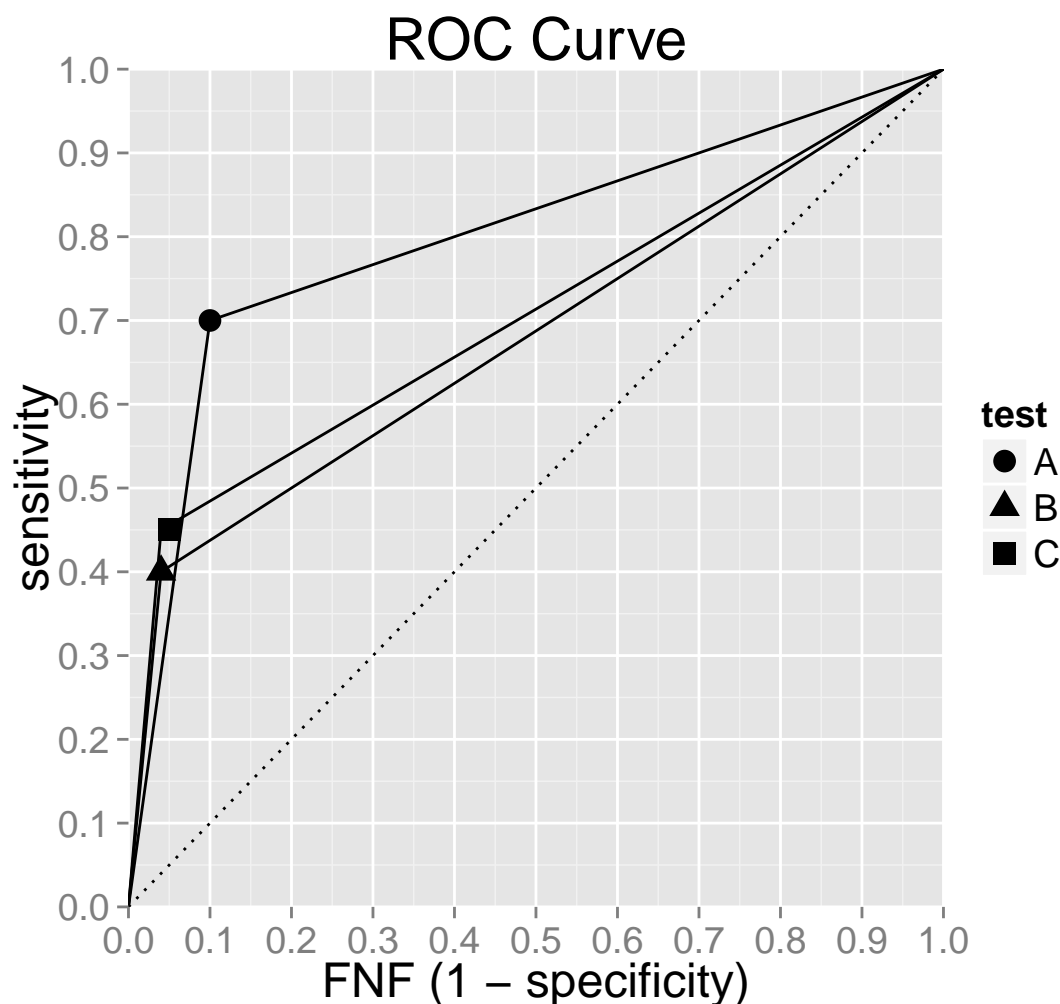
ROC Curve

The ROC Curves give more detailed information for evaluating the three tests. The metrics for each test, including area under the ROC curve are summarized below:

```
##   test sensitivity specificity  FNF  AUC
## 1    A        0.70        0.90 0.10 0.80
## 2    B        0.40        0.96 0.04 0.68
## 3    C        0.45        0.95 0.05 0.70
```

The ROC curves show that **Test A** is the best of three tests, since its area under the curve is largest at $AUC = 0.80$. This means that in a pair of randomly selected individuals, one from the actual positive group and one from the actual negative group, the subject with the more abnormal Test A results will actually be from the positive group with probability 0.80. By similar logic, **Test B** is the least discrimating test, since its area under the curve is the lowest at $AUC = 0.68$.