Assignment 3

5.a) The logistic regression model was created on the entire data set and the misclassification rate is 0.0263.

5.b) The data was split was 70:30 and the misclassification rate calculated was 0.0263.

5.c) with split 80:20 the misclassification rate calculated was 0.0255

with split 60:40 the misclassification rate calculated was 0.028

with split 90:10 the misclassification rate calculated was 0.023

Thus, it can be observed that the error rate is different for different splits.

5.d) When student is added to the model the error rate comes out to be 0.0275. Thus, the error rate does not change much when adding student as a dummy variable to the model.