BS 806

Homework 8

Irene Hsueh

**Question 1**

Part 1

Branch 1: The average FVC of male patients younger than 48.5 years is 574.8

Branch 2: The average FVC of male patients older than 48.5 years is 512.3

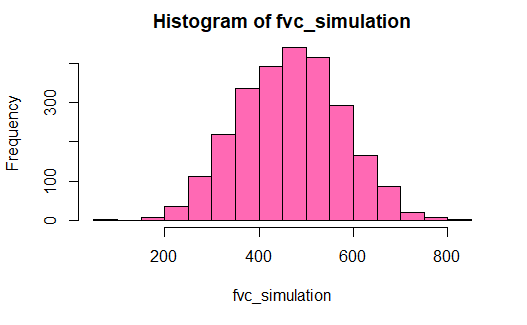
Branch 3: The average FVC of female patients younger than 51.5 years is 444.9

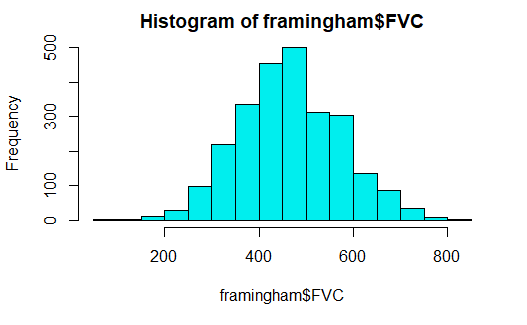
Branch 4: The average FVC of female patients older than 51.5 366.6

Part 2

estimated variance of FVC = 6194

Part 4

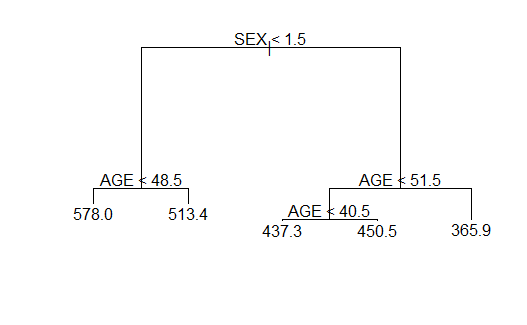




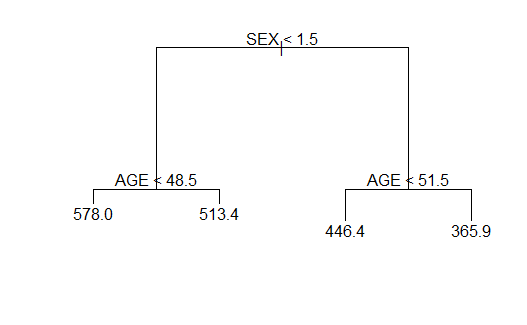
The distributions of the FVC scores of the simulated and real data are both approximately normally distributed.

Part 5

Using the simulated FVC scores and the real data for all other variables, only the variables sex and age were used to create 5 terminal nodes. The residual mean deviance was 5999.



The tree was pruned by cross-validation using k=5. The lowest deviance belonged to trees with 4 and 5 terminal nodes, so the final pruned tree contains 4 terminal nodes.



Branch 1: The average FVC of male patients younger than 48.5 years is 578.8

Branch 2: The average FVC of male patients older than 48.5 years is 513.4

Branch 3: The average FVC of female patients younger than 51.5 years is 446.4

Branch 4: The average FVC of female patients older than 51.5 365.9

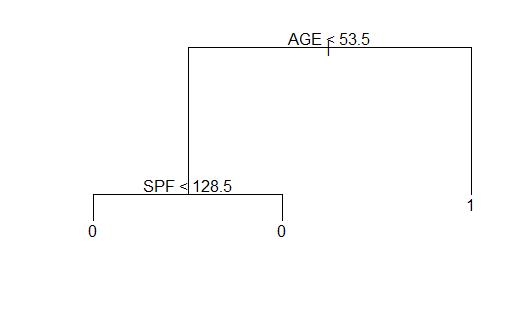
Part 6

The pruned tree from the simulated tree was very similar to the tree given in the problem set. Both trees used the same exact rules to create 4 leaves, and the prediction for FVC scores at each node nearly identical.

**Question 2**

Part A

The variables age and systolic blood pressure were used to create 3 branches. The residual mean deviance was 1.113.



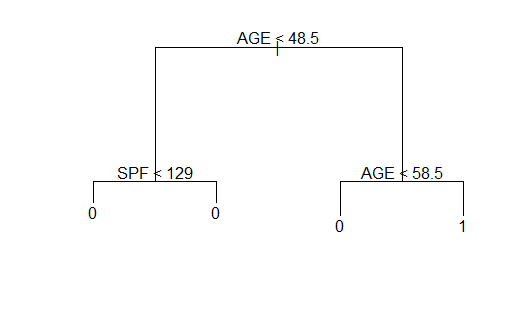
Part B

The probably of death within 20 years for someone over the age of 60 is 0.5567.

Part C

It’s predicted that someone over the age of 60 will be dead within 20 years of measurements.

Part D



The misclassification error of the prediction is 25.53%.