

# STA363Lab6SecA

Jean Ji, Phoebe Yan & Rickey Huang

4/6/2021

## Question 1

*Trees begin with a split on a single feature. Suppose we decided to consider splitting on whether or not the school is a private school. Explain in 1-2 sentences how you would use this feature to create one split, and how you would use the splitting rule to move rows into leaves.*

We would use the splitting rule of “Private == Yes” to create one split which splits the rows into two leaves. The leaf 1 would store the rows whose variable “Private” is “Yes”, and the leaf 2 would store the rows whose variable “Private” is “No”.

## Question 2

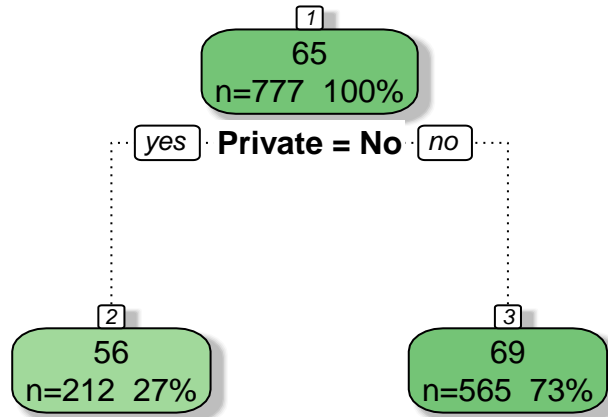
*Without using rpart to build the tree, find the training RSS we would get if we split on whether or not a school is a private school. Show your code.*

```
# Store the number of rows in to n
n <- nrow(college)
# Store the exploratory and response variables
X <- college$Private
Y <- college$Grad.Rate
# Assign rows to leaves
leaf1 <- which(X=="Yes")
leaf2 <- c(1:n)[-leaf1]
# Compute the means and use them for prediction
mean1 <- mean(Y[leaf1])
mean2 <- mean(Y[leaf2])
preds <- rep(0, n)
preds[leaf1] <- mean1
preds[leaf2] <- mean2
# Calculate the training RSS
RSS <- sum((Y - preds)^2)
#Return the training RSS
RSS
```

```
## [1] 203101.6
```

## Question 3

*Now, using the rpart code, create a tree using only Private as a feature. Call tree Tree1. Show a visualization of your tree as your answer.*



Tree1: Regression Tree Using only the Private feature

#### Question 4

Based on your tree, what percent of your training data comes from public schools?

Based on Tree1, 27% of the training data comes from public schools.

#### Question 5

Based on your tree, what graduation rate would you predict for a public school?

The predicting graduation rate for a public school is 56.

#### Question 6

Fit a least squares linear regression model for graduation rate, using whether or not a school is a private school as a feature. Call this model LSLR1. Write out the fitted regression line.

Only using the feature “Private”, the fitted LSLR model has the regression line of  $\widehat{Grad.Rate} = 56.042 + 12.956PrivateYes$

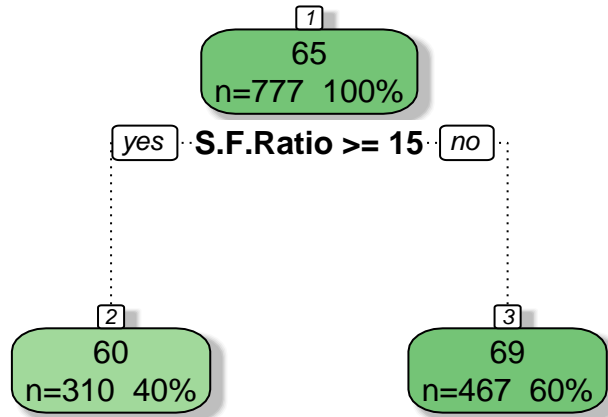
#### Question 7

Based on the LSLR model, what graduation rate would you predict for a public school? Keeping in mind that in the visualization our trees round to the nearest whole number, how do these predictions compare to those you made from the tree?

Based on the regression line of the LSLR model, the predicting graduation rate for a public school would be 56.042. The predictions from both the tree model and the LSLR model is almost the same, since the prediction from the tree model is also 56.042.

#### Question 8

Create a tree using only student faculty ratio as a feature. Use the  $maxdepth = 1$  stopping criterion to make sure that for the moment, the tree only has one split. If you don't do this, the tree will keep growing, and for now, we only want one split. Call tree Tree2, and show a visualization of your tree as your answer.



Tree2: Regression tree using only the S.F.Ratio feature

### Question 9

Based on your tree, what graduation rate would you predict for a school with a student faculty ratio of 10 (1 student to 10 faculty)?

Based on Tree2, the predicting graduation rate for a school with a student faculty ratio of 10 would be 69.

### Question 10

Fit a least squares linear regression model for graduation rate, using student faculty ratio as the only feature. Call this model LSLR2. Write out the fitted regression line.

Only using the “S.F.Ratio”, the LSLR model fitted has the regression line of  $\widehat{Grad.Rate} = 84.2168 - 1.3310S.F.Ratio$

### Question 11

Based on your LSLR model, what graduation rate would you predict for a school with a student faculty ratio of 10 (1 student to 10 faculty)? How does this compare to what you get from a tree?