

# Quiz-01-statistical learning\_v2

April 8, 2023

## 1 Quiz 01 - Statisitcal Learning

In this assessment we would using the **College** data set which can be found in the file **College.csv** on the book website. It contains a number of variables for 777 different universities and colleges in the US.

The variables are \* **Private** : Public/Private indicator \* **Apps** : Number of applications received \* **Accept** : Number of applicants accepted \* **Enroll** : Number of new students enrolled \* **Top10perc** : New students from top 10 % of high school class \* **Top25perc** : New students from top 25 % of high school class \* **F.Undergrad** : Number of full-time undergraduates \* **P.Undergrad** : Number of part-time undergraduates \* **Outstate** : Out-of-state tuition \* **Room.Board** : Room and board costs \* **Books** : Estimated book costs \* **Personal** : Estimated personal spending \* **PhD** : Percent of faculty with Ph.D.'s \* **Terminal** : Percent of faculty with terminal degree \* **S.F.Ratio** : Student/faculty ratio \* **perc.alumni** : Percent of alumni who donate \* **Expend** : Instructional expenditure per student \* **Grad.Rate** : Graduation rate

- (a) Use the `read.csv()` function to read the data into R. Call the loaded data `college`. Make sure that you have the directory set to the correct location for the data

```
[40]: college <- read.csv("College.csv")
```

```
[41]: # hidden test case
```

- (b) Use the `dim()` function to produce the dimensions of the data set.

```
[64]: #dims = ?  
  
dims <- dim(college)  
dims  
  
# your code here
```

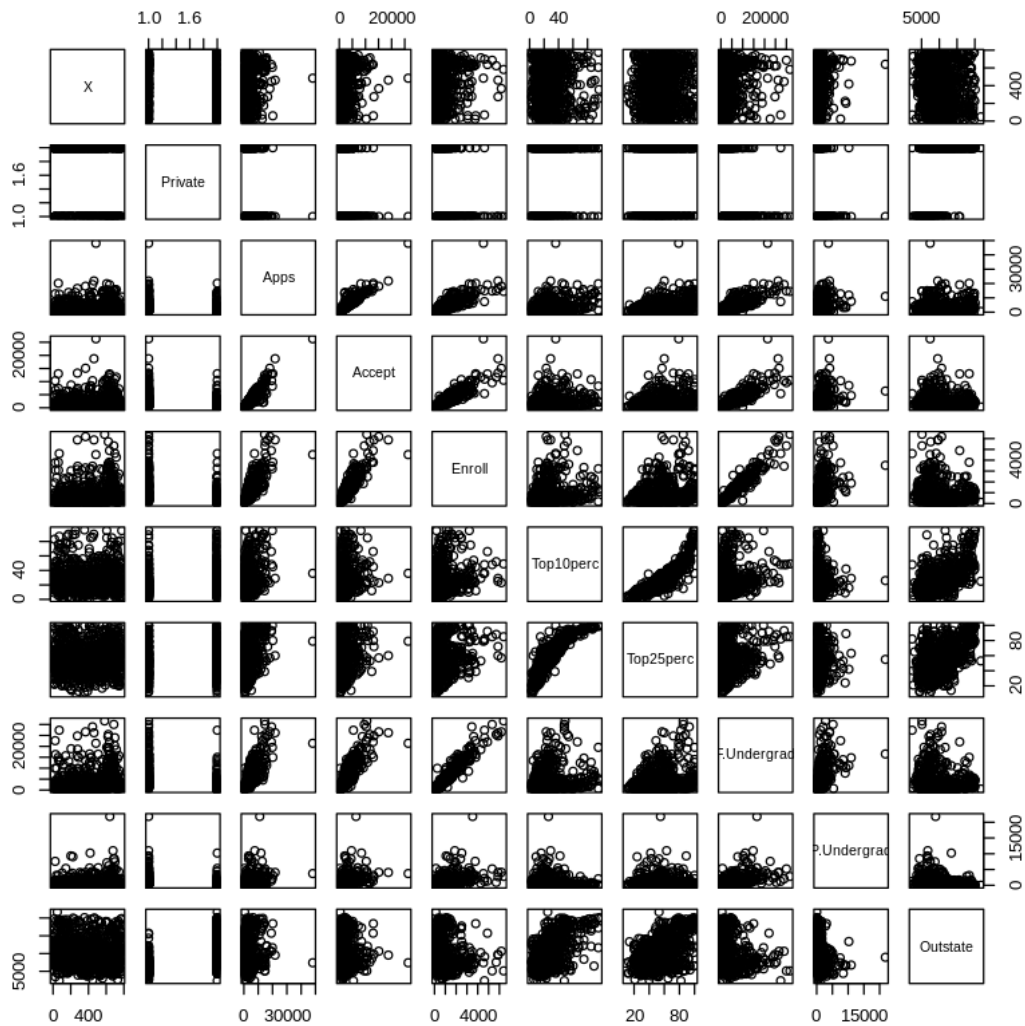
1. 777 2. 22

```
[62]: #hidden tests
```

- (c) Use the `pairs()` function to produce a scatterplot matrix of the first ten columns or variables of the data. Recall that you can reference the first ten columns of a matrix `A` using `A[,1:10]`

```
[69]: #pairs = ?
pairs <- pairs(college[, 1:10])
pairs
# your code here
```

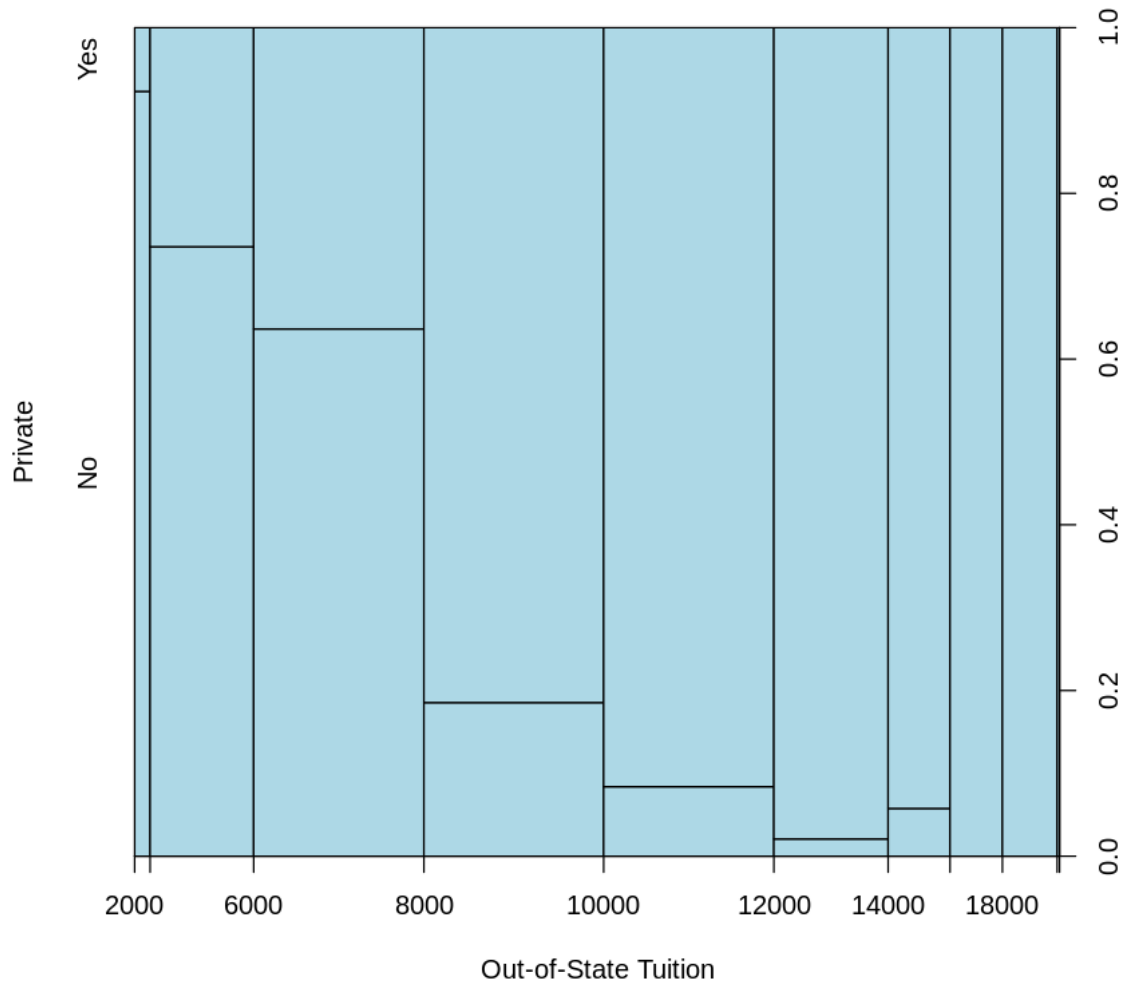
NULL



```
[53]: #hidden test cases
```

- (d) Use the `plot()` function to produce side-by-side boxplots of Outstate versus Private

```
[54]: plot(Private ~ Outstate, data = college, col = "lightblue", xlab = "Out-of-State Tuition", ylab = "Private")
```



- (e) Create a new qualitative variable, called `Elite`, by binning the `Top10perc` variable. Divide universities into two groups based on whether or not the proportion of students coming from the top 10% of their high school classes exceeds 50 %.

`Elite` should contain `Yes` or `No` based on the above condition

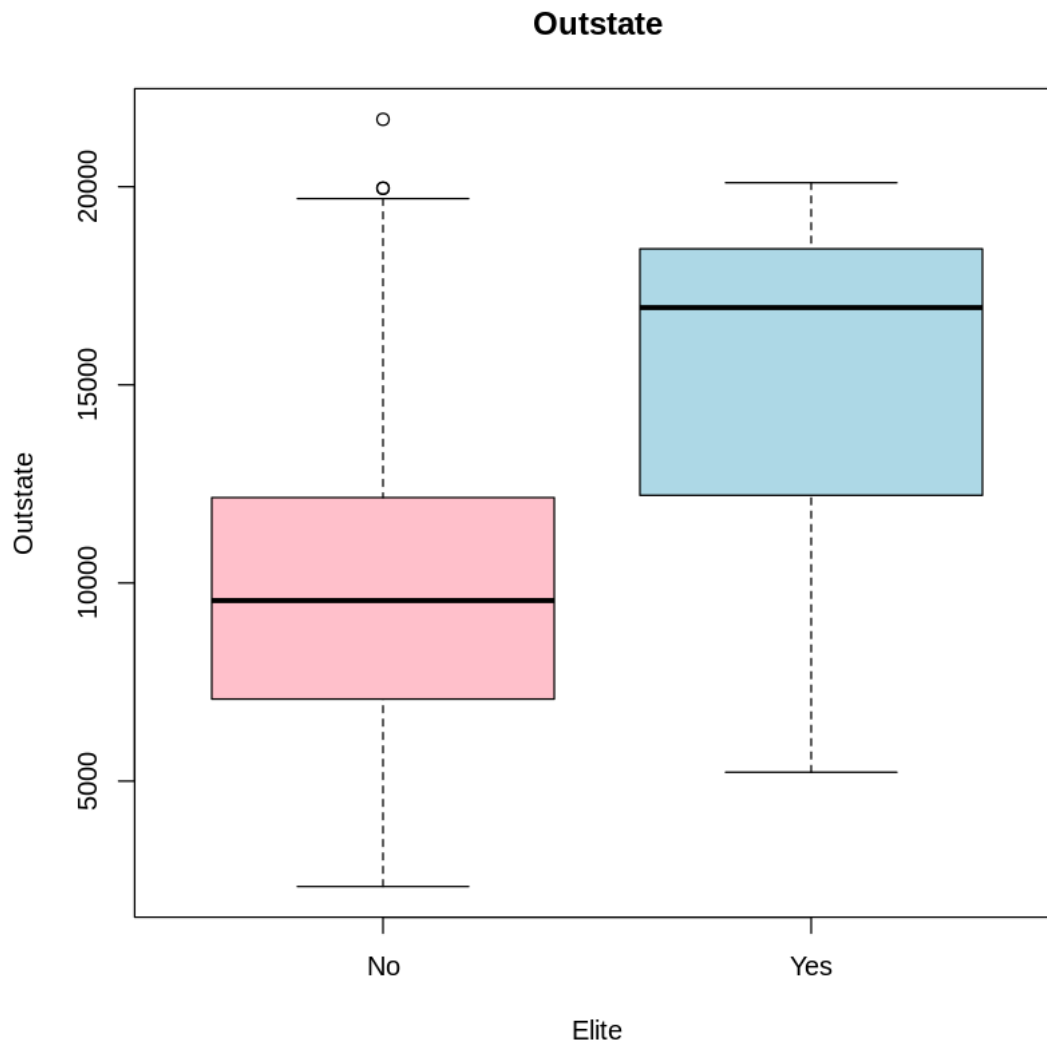
```
[73]: Elite <- ifelse(college$Top10perc > 50, "Yes", "No")
Elite <- factor(Elite)
```

```
[74]: #hidden test cases
```

```
[75]: Elite <- as.factor(Elite)
college <- data.frame(college, Elite)
```

(d) use the `plot()` function to produce side-by-side boxplots of Outstate versus Elite.

```
[68]: boxplot(Outstate ~ Elite, data = college, xlab = "Elite", ylab = "Outstate",
  ↪col = c("pink", "lightblue"), main = "Outstate")
```



```
[ ]:
```