Quiz-01-statistical learning_v2

March 19, 2022

1 Quiz 01 - Statisitical 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

[24]:
```
# your code here
college = read.csv('College.csv')
head(college)
attach(college)
```

A data.frame: 6 × 19

	X <fct>	Private <fct>	Apps <int>	Accept <int>	Enroll <int>	Top10perc <int>	Top25 <int>
1	Abilene Christian University	Yes	1660	1232	721	23	52
2	Adelphi University	Yes	2186	1924	512	16	29
3	Adrian College	Yes	1428	1097	336	22	50
4	Agnes Scott College	Yes	417	349	137	60	89
5	Alaska Pacific University	Yes	193	146	55	16	44
6	Albertson College	Yes	587	479	158	38	62

```
The following objects are masked from college (pos = 3):

    Accept, Apps, Books, Enroll, Expend, F.Undergrad, Grad.Rate,
    Outstate, P.Undergrad, perc.alumni, Personal, PhD, Private,
    Room.Board, S.F.Ratio, Terminal, Top10perc, Top25perc, X
```

[25]:
```
# hidden test case
```

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

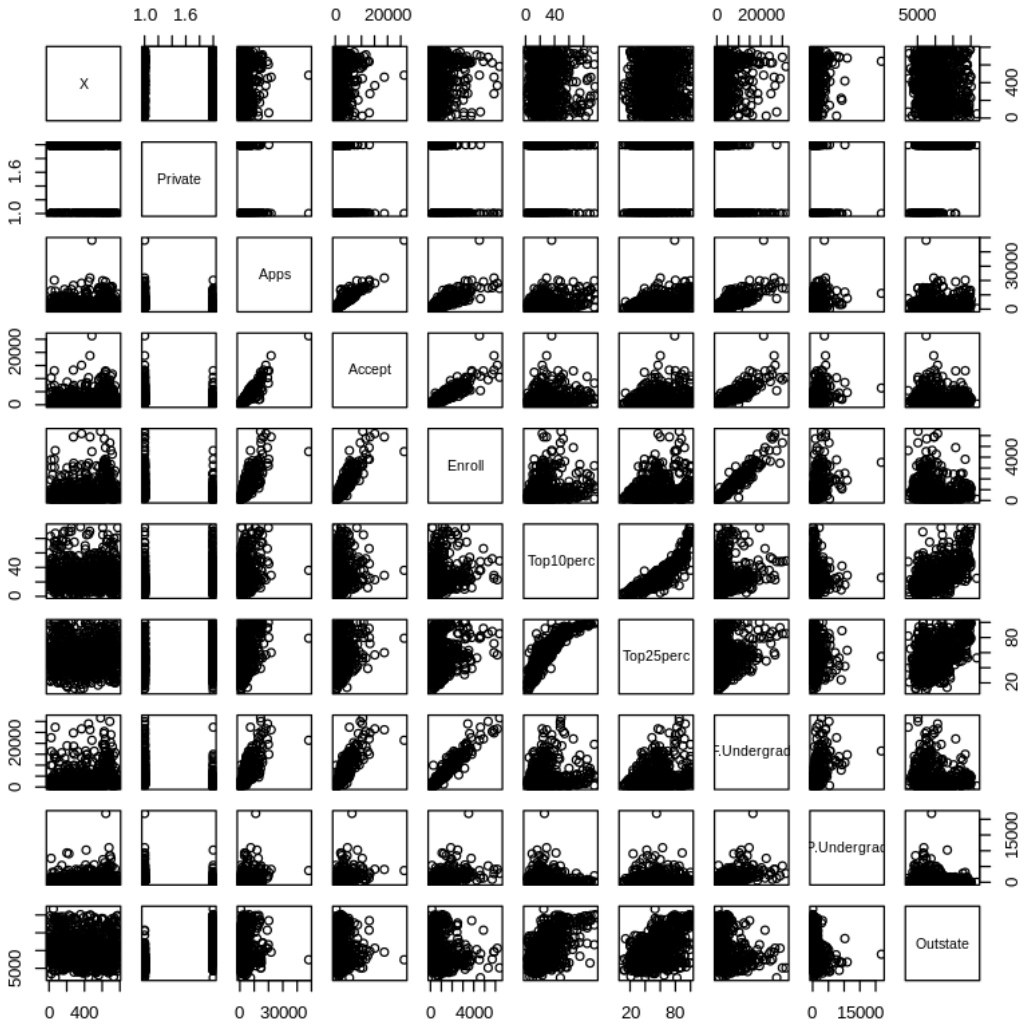
[26]:
```
#dims = ?
# your code here
dims = dim(college)
dims
```

1. 777 2. 19

[27]:
```
#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]`

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



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

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

[ ]:
```
# your code here
```

(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

[ ]:
```
#Elite = ?
# your code here
```

```r
Elite = ifelse(college$Top10perc> 50, 'Yes', 'No')
Elite
```

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

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

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

[ ]:
```r
# your code here
boxplot(Outstate, Elite, names = c("Outstate", "Elite"))
```

[ ]:
```r

```