

exam-2-12

Pradeep Paladugula

7/17/2020

```
library('tidyverse')

## — Attaching packages —

———— tidyverse 1.3.0 —

## ✓ ggplot2 3.3.0      ✓ purrr 0.3.3
## ✓ tibble 2.1.3      ✓ dplyr 0.8.5
## ✓ tidyr 1.1.0       ✓ stringr 1.4.0
## ✓ readr 1.3.1      ✓ forcats 0.4.0

## — Conflicts —

———— tidyverse_conflicts() —
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library("readxl")
appData <- read_excel('Application1.xlsx')
str(appData)

## Classes 'tbl_df', 'tbl' and 'data.frame': 10320 obs. of 4 variables:
## $ budget : num 3 1 8 8 2 5 8 7 7 7 ...
## $ schoolcode : num 6225 8293 8633 8656 789 ...
## $ education score: num 0.09 0.09 0.09 0.09 0.1 0.1 0.1 0.1 0.1 0.1 ...
## $ classsize : num 4 3 3 3 3 3 3 3 3 3 ...

library(car)

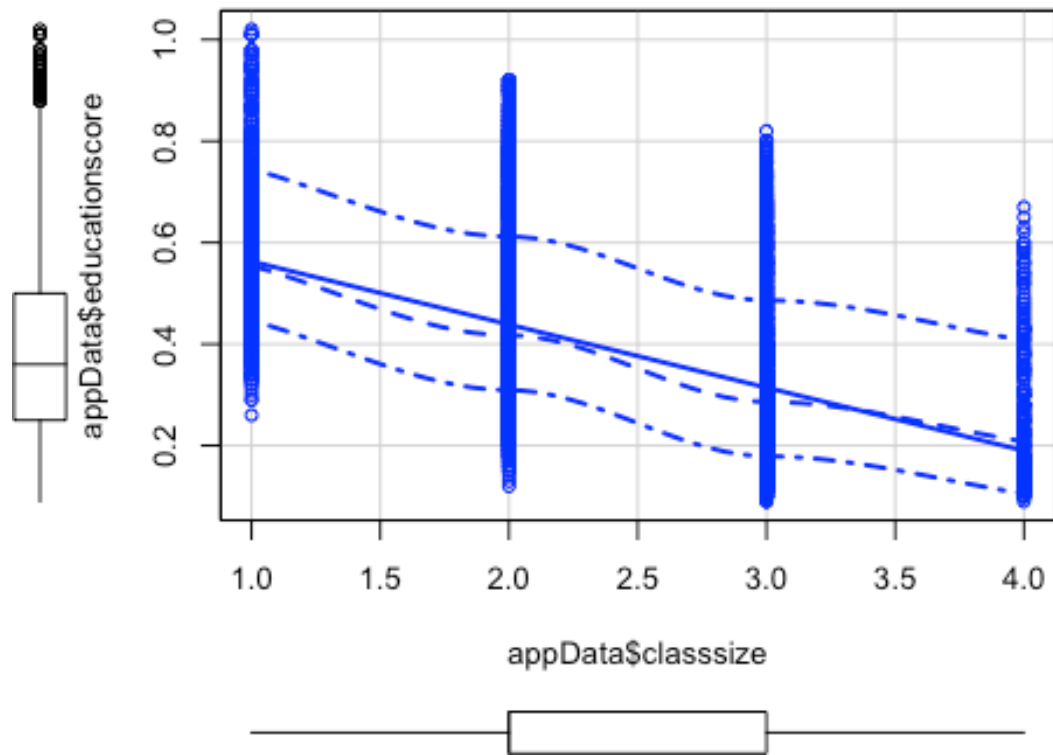
## Loading required package: carData

##
## Attaching package: 'car'

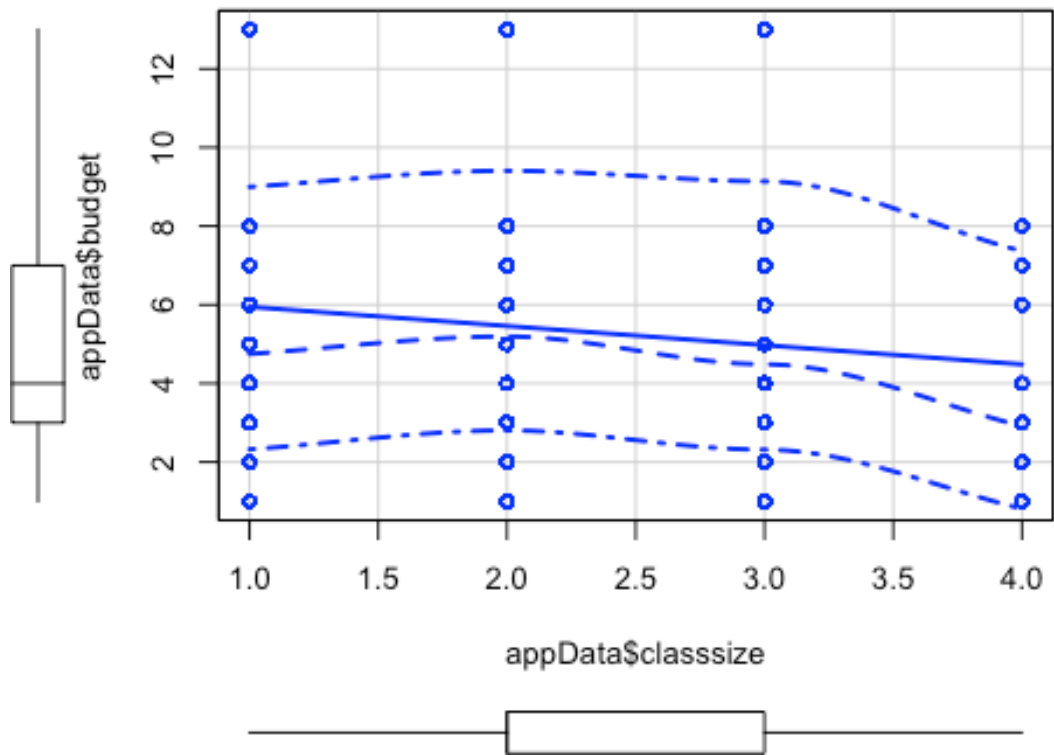
## The following object is masked from 'package:dplyr':
##
## recode

## The following object is masked from 'package:purrr':
##
## some

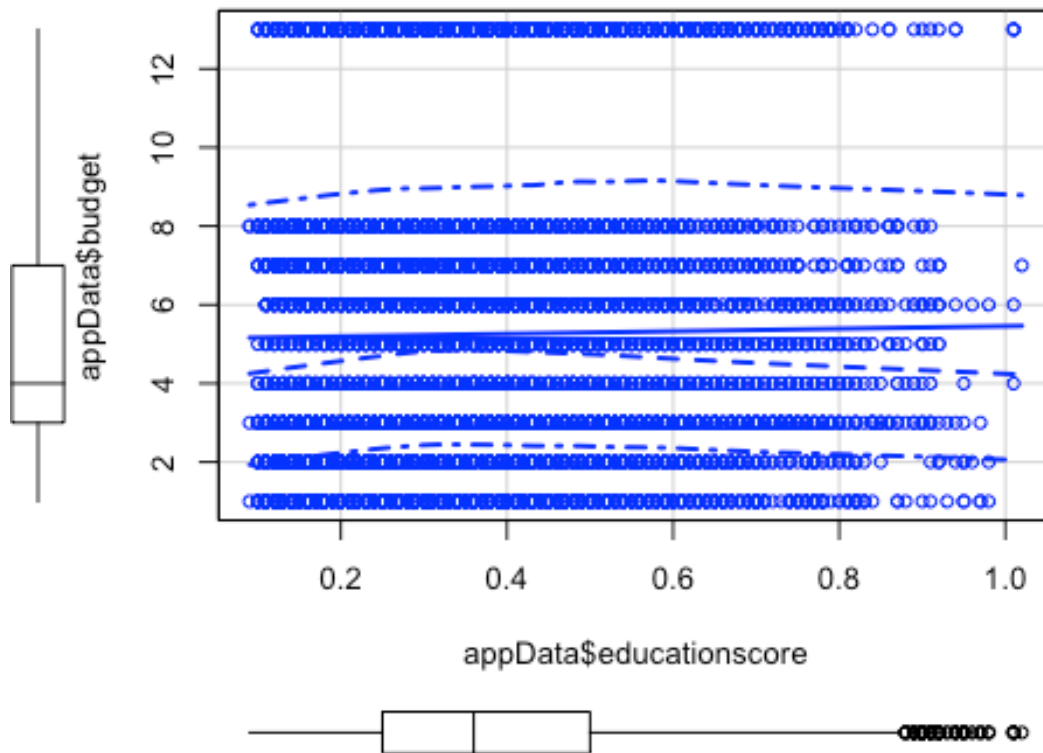
scatterplot(appData$classsize, appData$education score)
```



```
scatterplot(appData$classsize, appData$educationscore)
```



```
scatterplot(appData$educationscore, appData$budget)
```



```
cor(appData$classsize, appData$educationscore, method = "pearson")
## [1] -0.5067745

cor.test(appData$classsize, appData$educationscore, method = "pearson",
conf.level = .95)

##
## Pearson's product-moment correlation
##
## data: appData$classsize and appData$educationscore
## t = -59.713, df = 10318, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.5209745 -0.4922942
## sample estimates:
## cor
## -0.5067745

cor(appData$classsize, appData$budget, method = "pearson")
## [1] -0.1033454
```

```

cor.test(appData$classsize, appData$budget, method = "pearson", conf.level =
.95)

##
## Pearson's product-moment correlation
##
## data: appData$classsize and appData$budget
## t = -10.554, df = 10318, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.12239513 -0.08421953
## sample estimates:
## cor
## -0.1033454

data <- appData[, c("budget", "educationscore", "classsize")]
cor(data)

##                budget educationscore classsize
## budget           1.0000000      0.01698704 -0.1033454
## educationscore  0.01698704      1.00000000 -0.5067745
## classsize       -0.10334539     -0.50677455  1.0000000

library(ppcor)

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
## select

pc <- ppcor(data, method = c("pearson"))
pc

## $estimate
##                budget educationscore classsize
## budget           1.0000000      -0.041268 -0.1099093
## educationscore -0.0412680      1.0000000 -0.5078110
## classsize       -0.1099093     -0.5078111  1.0000000
##
## $p.value
##                budget educationscore classsize
## budget           0.000000e+00  2.748233e-05 4.196688e-29
## educationscore  2.748233e-05  0.000000e+00 0.000000e+00
## classsize       4.196688e-29  0.000000e+00 0.000000e+00
##
## $statistic
##                budget educationscore classsize
## budget           0.000000      -4.195273 -11.23182

```

```
## educationscore -4.195273      0.000000 -59.87415
## classsize      -11.231824     -59.874150   0.00000
##
## $n
## [1] 10320
##
## $gp
## [1] 1
##
## $method
## [1] "pearson"
```

Summary: The correlation test was conducted to test to predeict whether the class size and school budget have siginificant correlations with the students performance. the Pearson's correlation test shows the correlation between the class size and eductionscore ($r = -0.507$, $t(10318) = -59.713$, $p\text{-value} = <0,01$) are statistically significant, whic denotes lesser the classsize higher the education score. the correlation between the class size and budget ($r = -0.103$, $t(10318) = -10.554$, $p\text{-value} = <0,01$) are statistically significant, whic denotes lesser the classsize higher the budget. The correaltion between the classsize and education score while controlling tehe budget ($r = -0.5.8$, $p < 0.01$, $t(10320) = -59.87$) which is also a statistically significant.