hw2_stat123

Koki Itagaki

2023-02-13

#1. The built-in UCBAdmissions data set is a 3-dimensional array that contains #the following information: #• Dimension 1: Admit Admitted, Rejected #• Dimension 2: Gender Male, Female #• Dimension 3: Dept A, B, C, D, E, F #Note: You can find the UCBAdmissions data set using data(). #If you wanted to access the data, you could type in UCBAdmissions[1,1,2] to get #this value. If you wanted to create a table with the number of rejected #students in department A, you could type UCBAdmissions [2, ,1].

```
#(a) Create (and print out) a table that contains all students in department
D
UCB<-data(UCBAdmissions)
#Department D is a 4th of the dimention 3. So I just use hard brancket to
#get all data which students are in the department D
d_data<-UCBAdmissions[, ,4]
d_data
```

```
##           Gender
## Admit      Male Female
##   Admitted  138    131
##   Rejected  279    244
```

```
rowSums(UCBAdmissions)
```

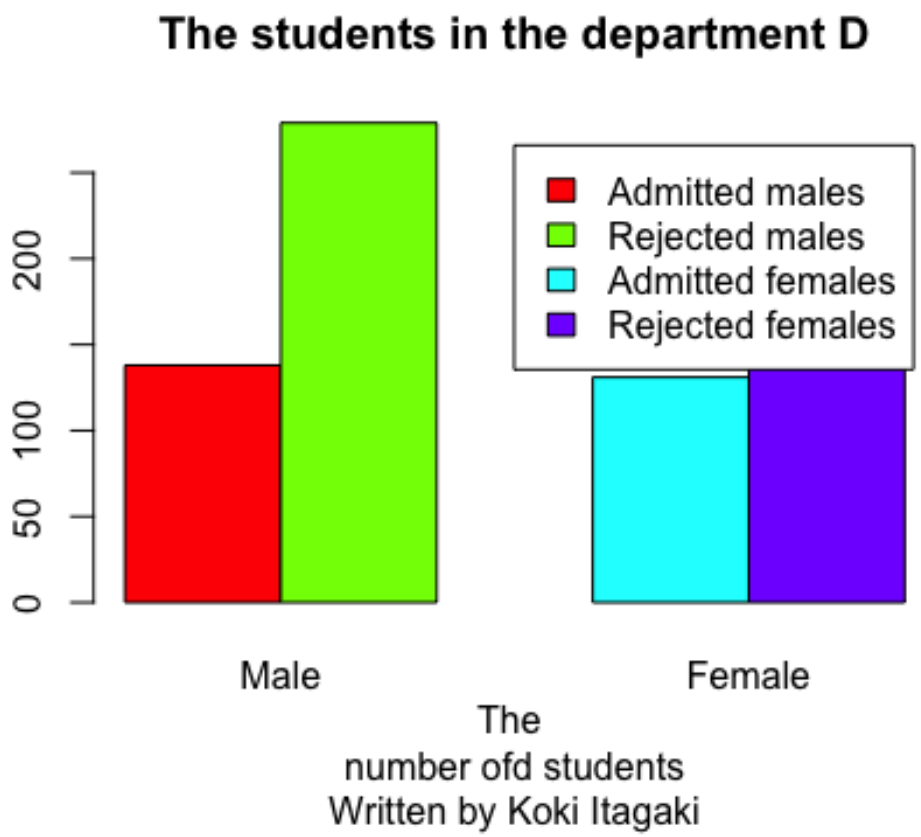
```
## Admitted Rejected
##     1755     2771
```

```
#(b) Create (and print out) a vector called department.D that contains the
#admitted and rejected students.
department.D<-UCBAdmissions [, ,4]
department.D
```

```
##           Gender
## Admit      Male Female
##   Admitted  138    131
##   Rejected  279    244
```

```
#Hint: You may need to use rowSums() on your answer from part (a).

#(c) Create a bar plot displaying the admitted and rejected students in
#Department D. Make sure to include a main title and label your x-axis. Also,
#make sure that each bar is a different color.

barplot(department.D,main = "The students in the department D",xlab = "The
```

```
number ofd students",legend = c("Admitted males", "Rejected males",
"Admitted females","Rejected females"), col = rainbow(4),
sub = "Written by Koki Itagaki",beside = TRUE)
```

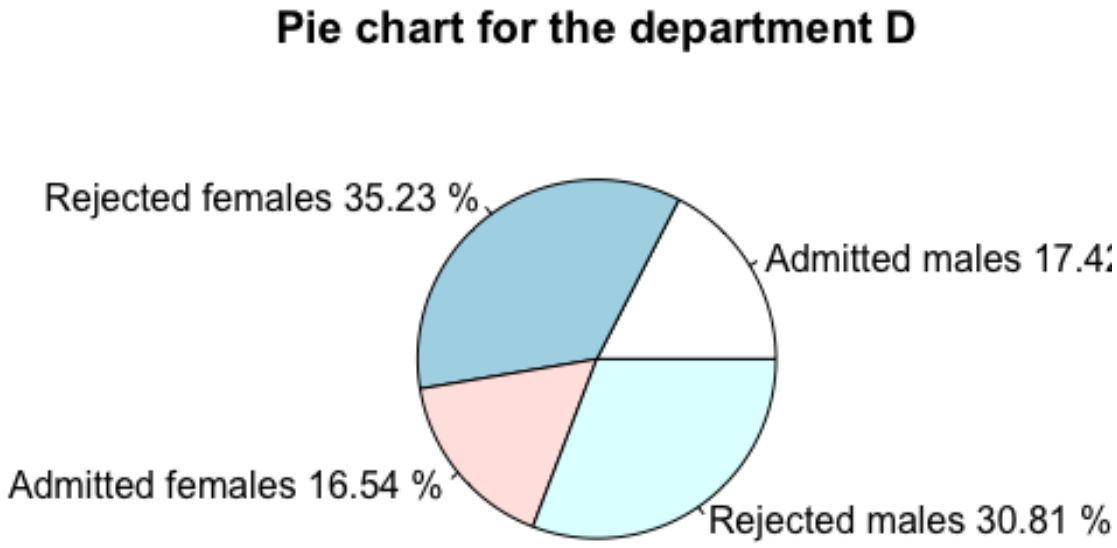



```
#(d) Create (and print out) a vector called admitted.females which contains
#the admitted females in Department D.
#The number of females is the first row and the second column.
admitted.females<-UCBAdmissions [1,2 ,4]

admitted.females

## [1] 131

#(e) Create (and print out) a vector called pct.admitted.females which
contains
#the percentage of admitted females in department D.


pct.admitted.females = round((admitted.females/sum(department.D))*100,
                             digits = 2)
print(paste(pct.admitted.females,"%"))

## [1] "16.54 %"
```

```r
#(f) Create a pie chart that displays the pct.admitted.females data.
#Be sure to include a main title for your pie chart.

pct.departmentD = round((department.D/sum(department.D))*100,
                            digits = 2)
pct.departmentD
```

```
##           Gender
## Admit       Male Female
##   Admitted 17.42  16.54
##   Rejected 35.23  30.81
```

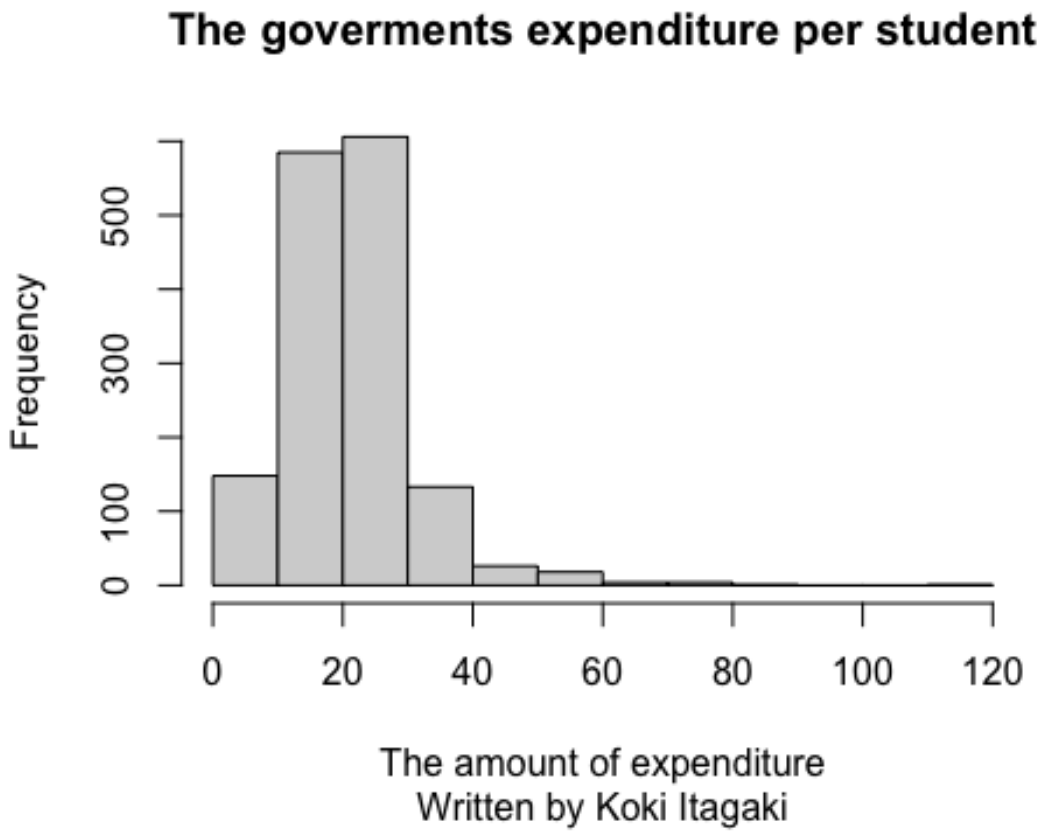
```r
element_d<- paste(c("Admitted males", "Rejected females", "Admitted females",
                    "Rejected males"),pct.departmentD, "%")
pie(pct.departmentD,labels = element_d ,main = "Pie chart for the department
D")
```



#(g) What does the pie chart imply about the number of admitted females #in department D? The pie chart implies that the number of admitted females are the least number compared to the other categories. Also,we could say there is a trend of that more males get admitted compared to females by approximately 1%. Moreover, most of the people who applied got rejected. In total, only about 30% of people get admmited.

#2. The following question deals with the data set
#Government_expenditure_per_student.csv, which you #will need to download from the assignment page.

```
#(a) Create (but do not print) a vector called expenditure containing the
#last column of the dataset named Value.
Value<-
read.csv("/Users/itagakikouki/stat123/Government_expenditure_per_student.csv"
)
expenditure<-Value[,ncol(Value)]


#(b) Create a histogram displaying the distribution of this variable.
#Be sure to have both a main title and a title on your x-axis.
hist(expenditure, main ="The goverments expenditure per student",
     xlab = "The amount of expenditure",sub = "Written by Koki Itagaki")
```




```
#(c) Describe the shape of the distribution (symmetric, left-skewed,
#                                            right-skewed).
print("From the histgram above, we can see that the data loooks symmentric.
However, there are a few extra data on the left side of the peak point.
It is possibly right-skewed, but I would say this is symmentric.")
```

```
## [1] "From the histgram above, we can see that the data loooks
symmentric.\nHowever, there are a few extra data on the left side of the peak
point.\nIt is possibly right-skewed, but I would say this is symmentric."

#(d) Compute the appropriate center value and the corresponding measures
#of variability.

#For the center value, I will find the median.
#I just use the function called median
#To find the mesures of variability, I also get standard deviation by using
#sd function
median(expenditure)

## [1] 20.30278

sd(expenditure)

## [1] 9.98177

#(e) Remove decimals from the vector named expenditure by using round() and
#create a stem plot
re_expenditure<-round(expenditure,0)
stem(re_expenditure)

##
##   The decimal point is 1 digit(s) to the right of the |
##
##    0 |
000000000000013334444444445555555555555555666666666666666677777777+50
##    1 |
00000000000000000000000000000000000000000000011111111111111111+501
##    2 |
000000000000000000000000000000000000000000000000000000000111+537
##    3 |
00000000000000000000000111111111111111111111122222222222222222+62
##    4 | 000111112222333444445667778
##    5 | 011122233344445688
##    6 | 1558
##    7 | 0123
##    8 | 9
##    9 |
##   10 |
##   11 | 6
```

#(c) Describe the shape of the distribution (symmetric, left-skewed, # right-skewed). From the histgram above, we can see that the data loooks symmentric. However, there are a few extra data on the left side of the peak point. It is possibly right-skewed, but I would say this is symmentric.