stat123\_lab5

Koki Itagaki

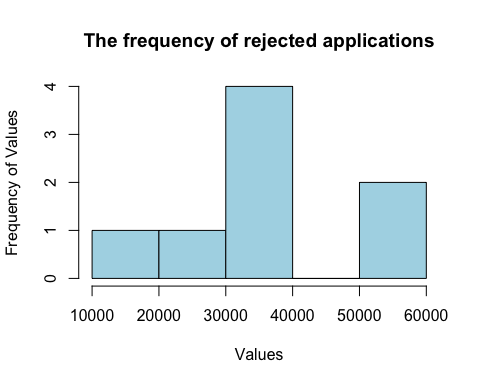
2023-02-14

#The following worksheet is due by 8pm one day after this lab. You can find #the submission dropbox in Brightspace by clicking on Content − > Lab Content. #0.0 Open a new R Markdown file. #1.0 Create the following data frame.

#a) Write the data frame in a CSV file and then read your file.  
dataFrame <- data.frame(Faculty = c("Arts","Science","Education", "Medicine",  
"Nursing", "Business", "Law", "Dentistry"), Application = c(46537,54647,85759,144547,64573,56748,106372,65748), Admitted = c(18615,16394,  
34304,132274,25829,17024,48186,32874), Enrolled = c(18465,16314,34284,122174,  
25629,16894,48116,32734))  
dataFrame

## Faculty Application Admitted Enrolled  
## 1 Arts 46537 18615 18465  
## 2 Science 54647 16394 16314  
## 3 Education 85759 34304 34284  
## 4 Medicine 144547 132274 122174  
## 5 Nursing 64573 25829 25629  
## 6 Business 56748 17024 16894  
## 7 Law 106372 48186 48116  
## 8 Dentistry 65748 32874 32734

#Make dataframe csv file  
#This csv goes to the same file as the r markdown file  
write.csv(dataFrame, "df.csv")  
  
#b) Create a vector called rejected containing the average admitted  
#for each faculty.  
rejected <-dataFrame$Application - dataFrame$Admitted  
  
#c) Create a histogram of the rejected vector. Give the histogram a colour of yo  
#ur liking. Name the x-axis “Values,” and make the title “Frequency of Values.”  
hist(rejected, col = "lightblue", xlab = "Values",  
 main = "The frequency of rejected applications",ylab = "Frequency of Values")



#d) What is an appropriate measure of the center of the distribution   
#(mean or median), and why?  
#Median is the appropriate mesurement bacause the graph is not symmentric and   
#there is a outlier around 0.8 to 1.0.  
#If there are 3 elements: 1,2,100 and get mean and median, these are  
#2 and 33.   
  
#e) Calculate an appropriate measure of the center of the distribution.  
median(rejected)

## [1] 38498.5

#f) Find the first and third quartiles of enrolled applications and save them  
#as q1 and q3,respectively.  
q1<- quantile(dataFrame$Enrolled, 0.25)  
q1

## 25%   
## 18072.25

q3<-quantile(dataFrame$Enrolled,0.75)  
q3

## 75%   
## 37742

#2.0 Download the data set data banking.csv and save it to whatever directory #you are using for this course.

banking<-read.csv("banking.df.csv")  
dim(banking)

## [1] 96 3

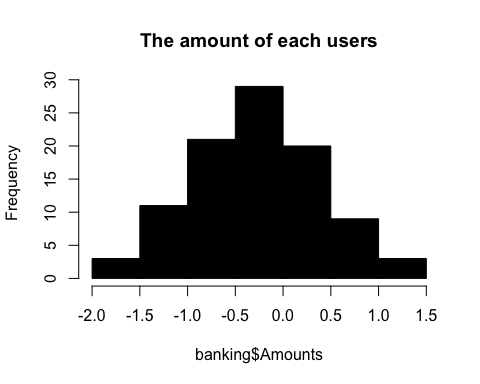
head(banking)

## Users X Amounts  
## 1 management NA 0.71  
## 2 technician NA 0.29  
## 3 entrepreneur NA -0.99  
## 4 blue-collar NA -0.02  
## 5 unknown NA -1.35  
## 6 management NA 0.48

#a) Create a stem plot of the amount vector and use the Amounts vector to   
#set the breaks.  
stem(banking$Amounts)

##   
## The decimal point is at the |  
##   
## -1 | 776  
## -1 | 43322211000  
## -0 | 9999998888877766555555  
## -0 | 4444444443332222222111110000  
## 0 | 11122233333444444  
## 0 | 555667777779  
## 1 | 124

#keep 1 decimal insted of 2 decimals  
banking$Amounts<-round(banking$Amounts, 1)  
#b) Create a histogram of the amount vector with a title. Give the histogram a   
#colour of your liking.  
hist(banking$Amounts, main = "The amount of each users",col = "black")



#c) Does the histogram seem normally distributed?  
#Yes. It looks like the graph is exactly normally distributed.  
#Because it is a belll shape and the distribution on the right side and   
#left side is totally same.  
  
#d) Create a sample from amount vector with sample size seventy and then calcula  
#late the median,mean, variance, and standard deviation of the created sample.  
sample<-sample(banking$Amounts,70)  
median(sample)

## [1] -0.2

mean(sample)

## [1] -0.2028571

var(sample)

## [1] 0.4597019

sd(sample)

## [1] 0.6780132

#e) Create a boxplot for the amount vector.  
boxplot(banking$Amounts, col = "lightgreen")

