ST.310, Spring 2024

Homework #7

Due on Wednesday, 4/3/24

Genevieve Torkornoo

***Directions:***

* Follow the ***homework format guidelines*** shown on the syllabus.
* Do ***not*** fill in your source code and answers on this problem set. Use the homework template.
* You may work with one partner if you wish.
* Upload a single Word file (saved as ***doc*** or ***docx***) on Moodle by ***11:59 PM*** on the due date.
* Late homework will ***not*** be accepted without any legitimate excuses.

Load the UsingR package. The package includes all textbook datasets. Show your R command lines and outputs for each question.

**Problem#2.49.** [Page 83]

The npdb data set contains malpractice-award information. The variable amount is the size of malpractice awards in dollars. Find the mean and median award amount. What percentile is the mean? Can you explain why this might be the case? [10 pts]

data(npdb)

meanAmount<- mean(npdb$amount)

medianAmount <- median(npdb$amount)

percentile\_mean <- ecdf(npdb$amount)(meanAmount) \* 100

cat("Mean amount:", meanAmount, "\n")

cat("Median amount:", medianAmount, "\n")

cat("Percentile of the mean:", percentile\_mean, "%\n")

Mean amount: 166257.2

Median amount: 37500

Percentile of the mean: 74.90069 %

The percentile of the mean is 74% which means that the percentage of the awards are in the top percentage of the population

**Problem#2.56.** [Page 84]

The Michelson data set records 100 measurements by Michelson of the speed of light (in the variable velocity). Make a quantile-normal graph and discuss if the graphic shows the points falling on a straight line. [15 pts]

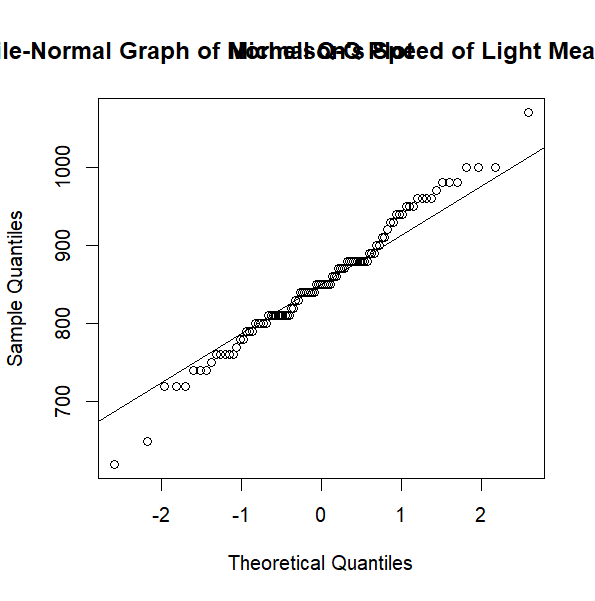
data(Michelson)

qqnorm(Michelson$velocity)

qqline(Michelson$velocity)

title(main = "Quantile-Normal Graph of Michelson's Speed of Light Measurements")

Yes the graph does the point falling on a fairly straight line. Towards the top of the graph, the plots begin to separate from the line.



**Problem#2.58.** [Page 84]

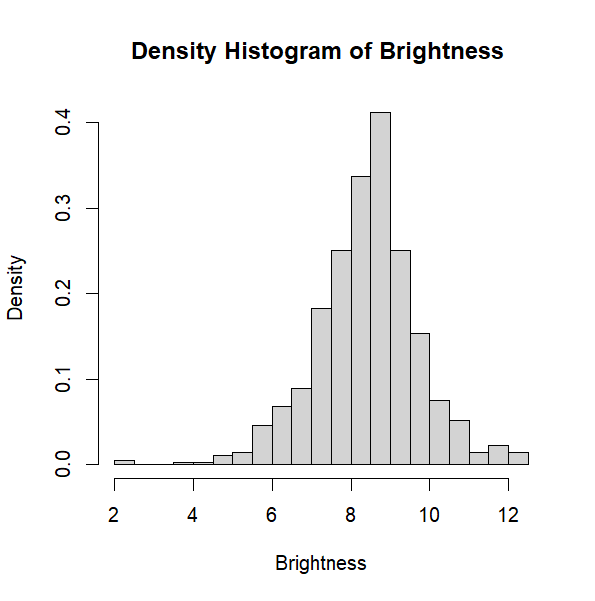
The brightness data set contains the brightness for 966 stars in a sector of the sky. It comes from the Hipparcos Catalogue. Make a density histogram of the data. Describe the shape of the distribution. [15 pts]

data(brightness)

histogram <- hist(brightness, breaks = 30, freq = FALSE, main = "Density Histogram of Brightness ", xlab = "Brightness")

summary(brightness)

Min. 1st Qu. Median   
 2.070 7.702 8.500   
 Mean 3rd Qu. Max.   
 8.418 9.130 12.430



The graph is skewed to the left of the graph which means that majority of the stars are mildly bright (8-10)

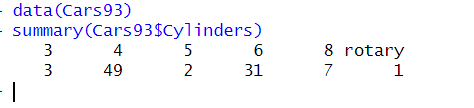
There is one outlier which means compared to the other stars this star shines the least brightest compared to other stars

**Problem#2.62.** [Page 87]

The Cylinders variable in the Cars93 data set records the number of cylinders in a factor. What kind of summary does R compute for factors? Look at summary(Cars93$Cylinders) to see. [10 pts]

data(Cars93)

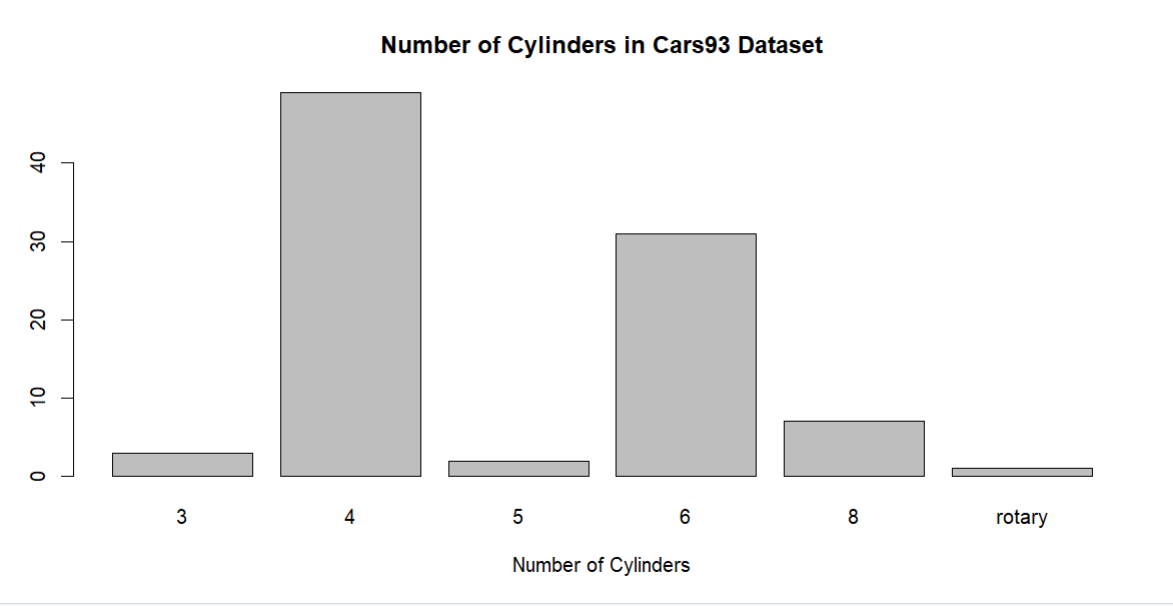
summary(Cars93$Cylinders)



**Problem#2.64.** [Page 87]

Make a vertical barplot (NOT a dotchart) of the Cylinders variable in the Cars93 data set. [10 pts]

barplot(table(Cars93$Cylinders), main = "Number of Cylinders in Cars93 Dataset", xlab = "Number of Cylinders")



**Problem#3.9. [Page 102]**

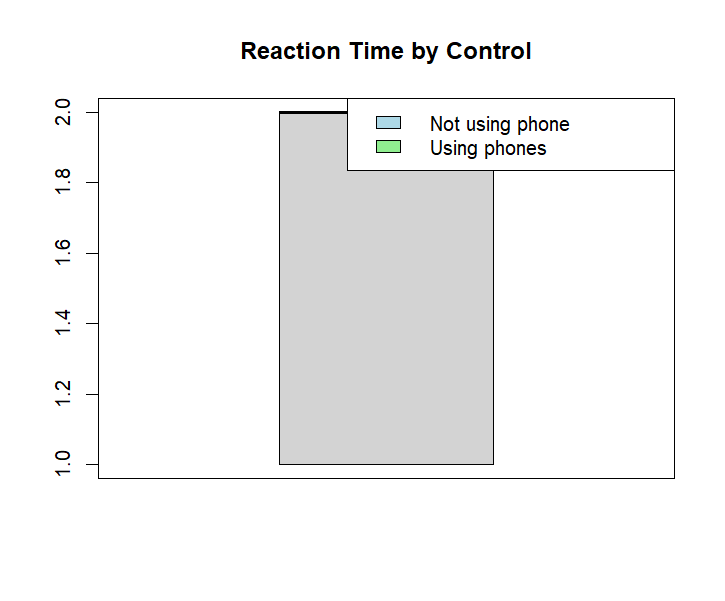
The use of a cell phone while driving is often thought to increase the chance of an accident. The data set reaction.time is simulated data on the time it takes to react to an external event while driving. S**ubjects with control == "C" are not using a cell phone, and those with control == "T" are.** Their time to respond to some external event is recorded in seconds.

Create side-by-side boxplots of the variable reaction.time for the two values of control. Compare the centers and spreads. [20 pts]

data(reaction.time)

boxplot(reaction.time $ control data= reaction.time, main="Reaction Time by Control")

legend("topright", legend = c("Not using phone", "Using phones"), fill =c("lightblue","lightgreen"))



summary(reaction.time[reaction.time$control == "C", ])

age gender control  
 16-24: 4 F: 8 C:20   
 25+ :16 M:12 T: 0   
   
   
   
   
 time   
 Min. :1.245   
 1st Qu.:1.314   
 Median :1.371   
 Mean :1.390   
 3rd Qu.:1.448   
 Max. :1.595

summary(reaction.time[reaction.time$control == "T", ])

age gender control  
 16-24:16 F:22 C: 0   
 25+ :24 M:18 T:40   
   
   
   
   
 time   
 Min. :1.293   
 1st Qu.:1.380   
 Median :1.454   
 Mean :1.446   
 3rd Qu.:1.505   
 Max. :1.615

**Problem#3.11.** [Page 102]

The data set stud.recs contains 160 SAT scores for incoming college students stored in the variables sat.v and sat.m. Produce layered density plots of the data. Do the two data sets appear to have the same center? Then make a quantile-quantile plot. Do the data sets appear to have the same shape? [20 pts]

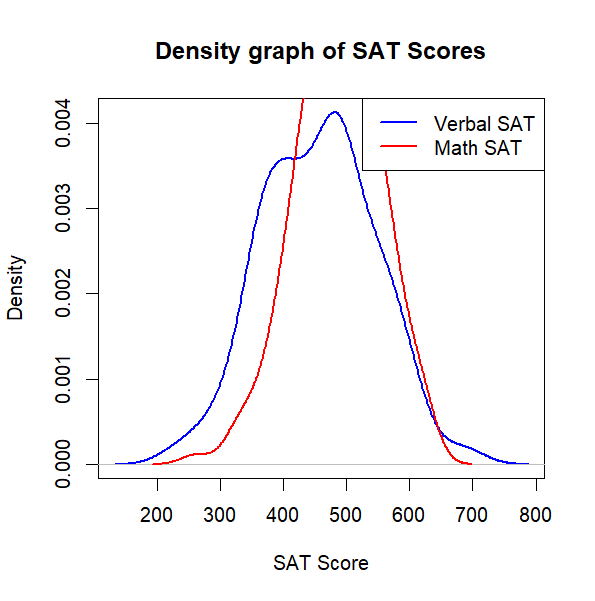
data(stud.recs)

plot(density(stud.recs$sat.v), main = "Density graph of SAT Scores", xlab = "SAT Score", col = "blue", lwd = 2)

lines(density(stud.recs$sat.m), col = "red", lwd = 2)

legend("topright", legend = c("Verbal SAT", "Math SAT"), col = c("blue", "red"), lty = 1, lwd = 2)

The centers of the graphs are also very similar to each other.



meanSat\_v <- mean(stud.recs$sat.v)

meanSat\_m <- mean(stud.recs$sat.m)

cat("Mean of Verbal SAT scores:", meanSat\_v, "\n")

cat("Mean of Math SAT scores:", meanSat\_m, "\n")

Mean of Verbal SAT scores: 455.8438

Mean of Math SAT scores: 485.9375

qqPlot(stud.recs$sat.v, stud.recs$sat.m, main = "Quantile-Quantile Plot of SAT Scores", xlab = "Verbal SAT", ylab = "Math SAT", col = "blue", pch = 20)

The graph has a form of a straight line with a few outliers

