R Assignment 1

Alireza Sheikh-Zadeh, PhD

Nicholas Small

Document format: Follow the instructions given on the web page. Always review your solution word document before submission.

Plagiarism: You are not allowed to share your write-up with your peers. It's okay to advise your peers about how to solve problem, but you never share your own write-up.

Problem 1: 34 points

Problem 2: 46 points

Format: 20 points

Problem 1 (34 points)

The production of beer is a multibillion-dollar worldwide industry. The dataset in the following link include the alcohol per volume and calories of famous beer brands.

```
mydata <- read.csv("http://tiny.cc/isqs5347-beer")
```

a. Compute the following statistics for variables "Alcohol" and "Calories": mean, median, variance, standard deviation, Q1 & Q3, and interquartile range. (14 points)

```
# Listed are the Mean, Median, Variance, Standard Deviation, and Summary Data

colMeans(mydata[sapply(mydata, is.numeric)])

##   Alcohol  Calories
##  4.955827 42.220472

apply(mydata, 2, median)

##                                              Alcohol
##                                               "4.93"
##                                             Calories
##                                                 "43"
##                                                brand
## "McEwans Scotch Ale Scottish & Newcastle (Scotland)"

var(mydata$Alcohol)

## [1] 0.8045737

var(mydata$Calories)
```

```
## [1] 63.07799
```

```
sd(mydata$Alcohol)
```

```
## [1] 0.8969803
```

```
sd(mydata$Calories)
```

```
## [1] 7.942165
```

```
summary(mydata)
```

```
##     Alcohol         Calories
##  Min.   :2.290   Min.   :19.00
##  1st Qu.:4.510   1st Qu.:39.50
##  Median :4.930   Median :43.00
##  Mean   :4.956   Mean   :42.22
##  3rd Qu.:5.265   3rd Qu.:45.00
##  Max.   :9.500   Max.   :83.00
##
##                                                    brand
##  Michelob Classic Dark Beer Anheuser Busch (USA)       :  2
##  Amstel Light Bier Amstel Brouwerij B.V. (Holland)     :  1
##  Anchor Porter Anchor (USA)                            :  1
##  Anchor Steam Beer Anchor (USA)                        :  1
##  Anheuser Busch Natural Light Beer Anheuser Busch (USA):  1
##  Asahi Draft Beer Asahi (Japan)                        :  1
##  (Other)                                               :120
```

```
IQR(mydata$Alcohol)
```

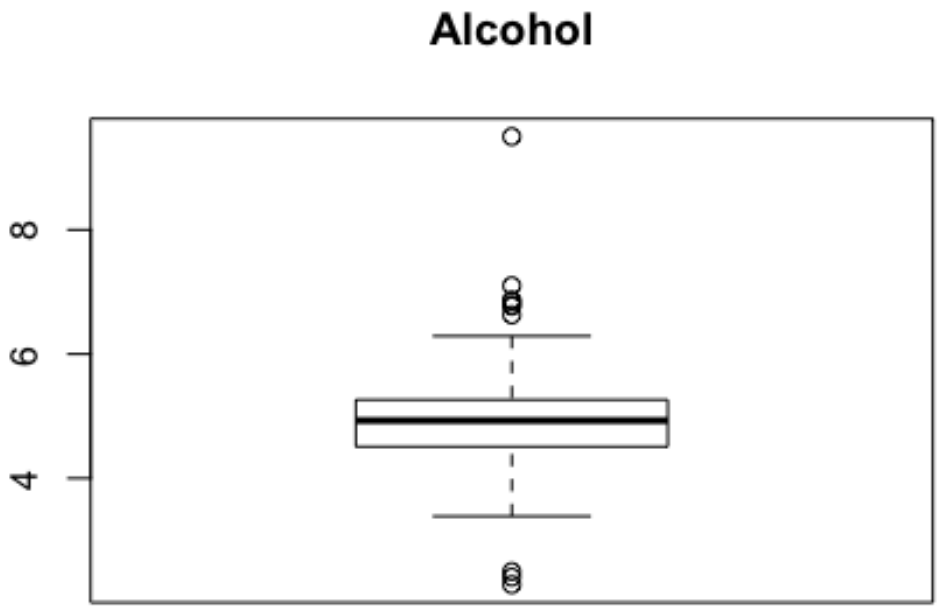
```
## [1] 0.755
```

```
IQR(mydata$Calories)
```

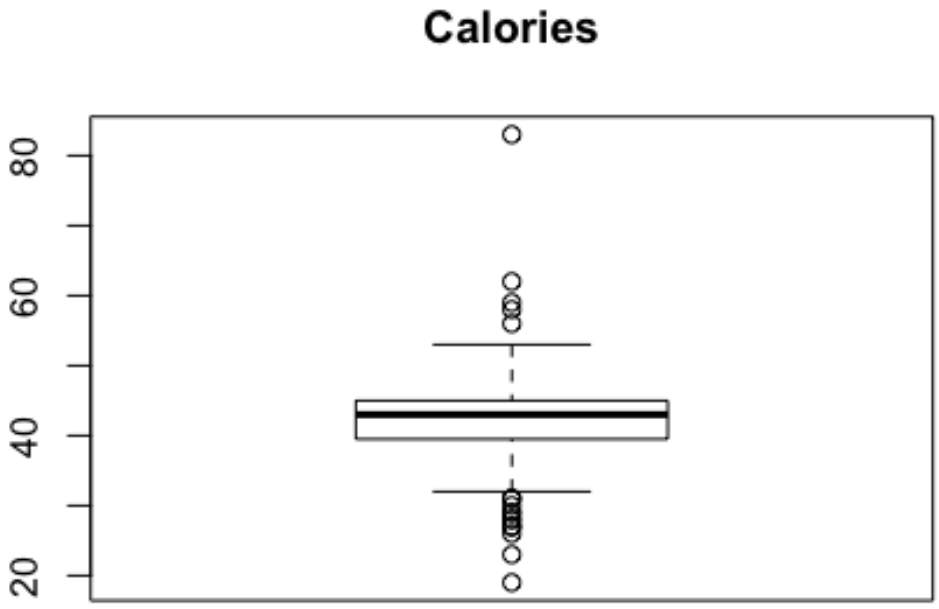
```
## [1] 5.5
```

b. Create separate boxplots for variables “Alcohol” and “Calories”. Are there any outliers for each variable (8 points).

```
# Listed are the BoxPlot for Alcohol and Calories.
# There are several outliers on each box plot
boxplot(mydata$Alcohol, main = "Alcohol")
```



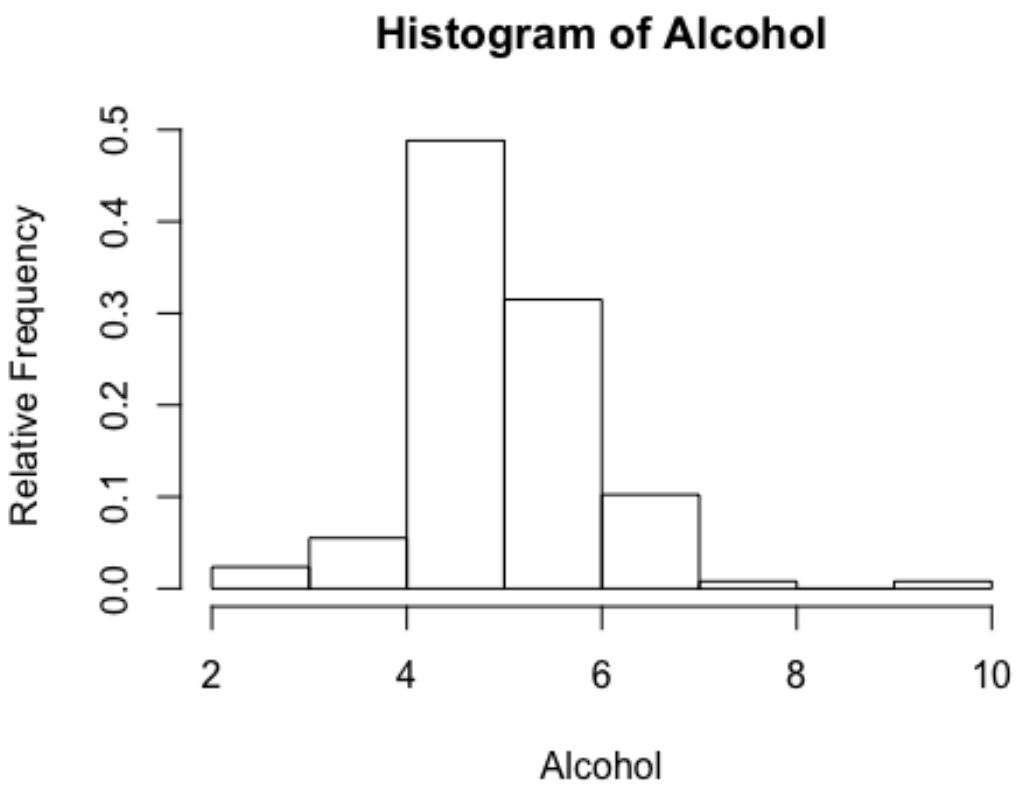

```r
boxplot(mydata$Calories, main = "Calories")
```



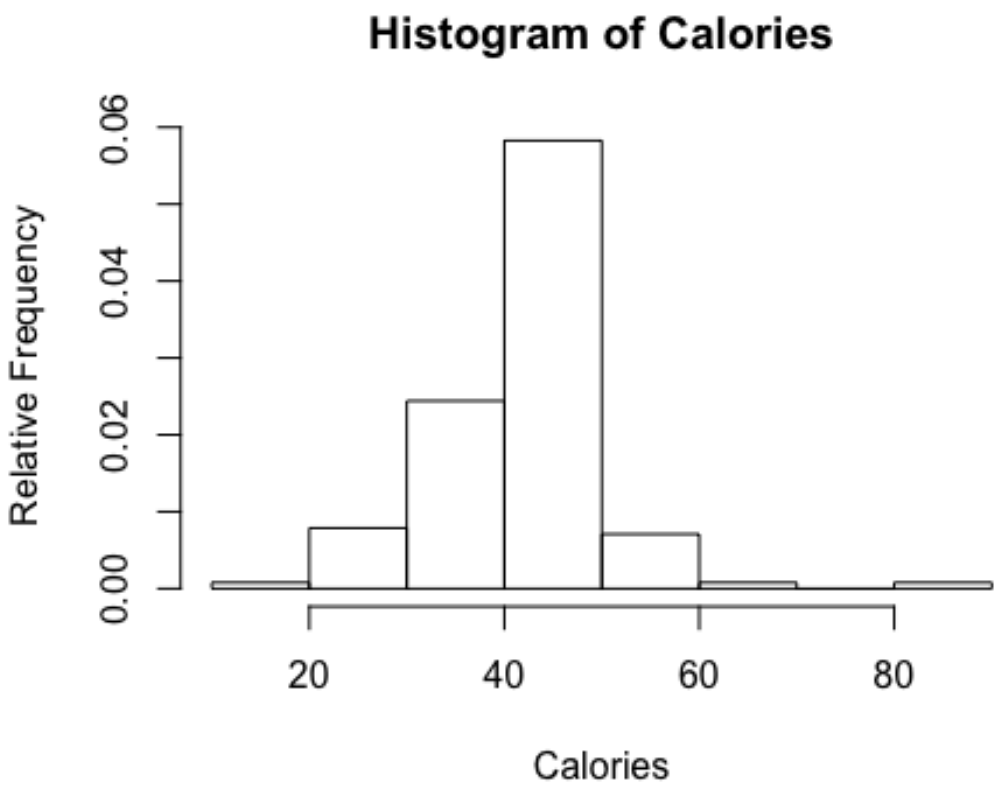
c. Create separate histograms for variables “Alcohol” and “Calories”. Make sure the y-axes presents the relative frequency (a value between 0 and 1). (6 points)

```
# Histograms for Alcohol and Calories.

hist(mydata$Alcohol, freq = FALSE, ylab = "Relative Frequency", main =
"Histogram of Alcohol", xlab = "Alcohol" )
```




```r
hist(mydata$Calories, freq = FALSE, ylab = "Relative Frequency", main =
"Histogram of Calories", xlab = "Calories" )
```



d. Are variables "Alcohol" and "Calories" skewed or symmetrical? If skewed, in which direction? What that means, explain. (6)

```
# Alcohol is slightly skewed to the right with a high peak and Calories is
slightly skewed to the left with a high peak.

library(e1071)
skewness(mydata$Alcohol, na.rm = FALSE, type = 3)

## [1] 0.8563446

skewness(mydata$Calories)

## [1] 0.662515

kurtosis(mydata$Alcohol)

## [1] 5.20623

kurtosis(mydata$Calories)

## [1] 5.108726
```

Problem 2 (46 points)

Use the TTU graduate student exit survey data.

```
grad <- read.csv("http://westfall.ba.ttu.edu/isqs6348/Rdata/pgs.csv", header
= T)
attach(grad)
```

Two variables of interest are "FacTeaching", a 1,2,3,4,5 rating of teaching at TTU by the student, and "COL", the college from which the student graduated.

a. Construct the contingency table showing counts of students in all combinations of these two variables. (10 points)

```
# This is Contingency table
a <- table(COL, FacTeaching)
a
```

```
##       FacTeaching
## COL      1   2   3   4   5
##   AG     4  15  26  78  56
##   AR     3   4   6  16   4
##   AS    12  24 124 290 171
##   BA     9  28  44 116  66
##   DUAL   0   0   2   0   0
##   ED     3   6  26 113  93
##   EN     5  36  65 168  86
##   GR     0   3   8  27  15
##   HS     1   5  17  41  33
##   MC     0   0   3  25   6
##   VPA    4   7  10  37  44
```

b. Construct a contingency table showing the proportion (probability) of students in all combinations of these two variables (5 points). Round results by three decimals (1 points).

```
# This is a Continguency Table for Teaching Rating and Colleges
a <- a/nrow(grad)
a
```

```
##       FacTeaching
## COL              1            2            3            4            5
##   AG   0.0019980020 0.0074925075 0.0129870130 0.0389610390 0.0279720280
##   AR   0.0014985015 0.0019980020 0.0029970030 0.0079920080 0.0019980020
##   AS   0.0059940060 0.0119880120 0.0619380619 0.1448551449 0.0854145854
##   BA   0.0044955045 0.0139860140 0.0219780220 0.0579420579 0.0329670330
##   DUAL 0.0000000000 0.0000000000 0.0009990010 0.0000000000 0.0000000000
##   ED   0.0014985015 0.0029970030 0.0129870130 0.0564435564 0.0464535465
##   EN   0.0024975025 0.0179820180 0.0324675325 0.0839160839 0.0429570430
##   GR   0.0000000000 0.0014985015 0.0039960040 0.0134865135 0.0074925075
##   HS   0.0004995005 0.0024975025 0.0084915085 0.0204795205 0.0164835165
```

```
##   MC   0.0000000000 0.0000000000 0.0014985015 0.0124875125 0.0029970030
##   VPA  0.0019980020 0.0034965035 0.0049950050 0.0184815185 0.0219780220
```

c. What is the probability that a randomly selected student is from college of business administration (BA)? We call this the marginal probability, P(COL=BA). (5 points)

```
# This is a Marginal Probability of a random student is selected from BA
rowSums(a)

##          AG          AR          AS          BA        DUAL          ED
## 0.089410589 0.016483516 0.310189810 0.131368631 0.000999001 0.120379620
##          EN          GR          HS          MC         VPA
## 0.179820180 0.026473526 0.048451548 0.016983017 0.050949051

# P(COL=BA) = 0.131
```

d. What is the probability that a randomly selected student is from BA and rates the teaching quality by 5? We call this the joint probability, P(COL = BA and FacTeaching =5). (5 points)

```
# This is a Joint Probability of BA and FacTeaching 5
# P(COL=BA, = 5) = 0.0329670330
```

e. Given that a randomly selected student is from BA, what is the probability that he/she rates the teaching quality by 5? We call this the conditional probability, P(FacTeaching = 5 | COL=BA). (5 points)

```
# This is a Conditional Probability of a randoomly selected BA student and
FacTaching 5
# Conditional probability:  0.25160305
```

f. Given that a randomly selected student is from college of education (ED), what is the probability that he/she rates the teaching quality by 5? In other words P(FacTeaching = 5 | COL=ED)? What is your conclusion about the difference between the quality of teaching in BA and ED. (5 points)

```
# This is a Conditional Probability of a random student from ED and
FacTeaching 5
# Conditional probability:  0.38589350
# Students rate ED Teaching higher
```

g. What is the probability that a randomly selected student is fully happy about the teaching quality at TTU, hence rates FacTeaching = 5? We call this the marginal probability, P(FacTeaching=5). (5 points)

```
# This is Marginal Probability of a selected student and Facteaching 5
w <- colSums(a)
w

##          1          2          3          4          5
## 0.02047952 0.06393606 0.16533467 0.45504496 0.28671329

# Marginal Probability = 0.28671329
```

h. Given that a randomly selected student rates the teaching quality by 5, what is the probability that he/she is graduated from BA? The P(COL=BA | FacTeaching = 5). (5 points)

```
# This is Conditional Probability of a randomly selected student graduated
from BA | Facteaching 5
# Conditional probability:  0.11495937
```