IST772 Problem Set 4

Michael A. d'Amore

The homework for week 4 is based is based on exercises 7-10 on page 66, but with changes as noted in the text in this notebook (i.e., follow the problems as given in this document and not the textbook).

Attribution statement: (choose only one) 1. I did this homework by myself, with help from the book and the professor

```
wine <- read.csv(file.choose())
```

Chapter 4, Exercise 7

The wine data set available on Blackboard contains data for properties of different wines (see https://archive.ics.uci.edu/dataset/109/wine for documentation). The wines are labelled in the class variable as 1, 2 or 3. Run the summary() command on the dataset and explain the output. Create a histogram of the hue content for wine 1 (1 pt). As a reminder about R syntax, here is one way that you can access the class 1 hue data:

```
library(ggplot2)

wine$hue[wine$class==1]
summary(wine)
class_1_wines <- wine[wine$class == 1, ]

ggplot(class_1_wines, aes(x = hue)) +
  geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black") +
  labs(title = "Histogram of Hue for Class 1 Wines", x = "Hue", y = "Frequency") +
  theme_minimal()
```

Using the dplyr package, you can instead write:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
wine %>% filter(class == 1) %>% select(hue)
```

```
##     hue
## 1  1.04
## 2  1.05
## 3  1.03
## 4  0.86
## 5  1.04
## 6  1.05
## 7  1.02
## 8  1.06
## 9  1.08
## 10 1.01
## 11 1.25
## 12 1.17
## 13 1.15
## 14 1.25
## 15 1.20
## 16 1.28
## 17 1.07
## 18 1.13
## 19 1.23
## 20 0.96
## 21 1.09
## 22 1.03
## 23 1.11
## 24 1.09
## 25 1.12
## 26 1.13
## 27 0.92
## 28 1.02
## 29 1.25
## 30 1.04
## 31 1.19
## 32 1.09
## 33 1.23
## 34 1.25
## 35 1.10
## 36 1.04
## 37 1.09
## 38 1.12
## 39 1.18
## 40 0.89
## 41 0.95
## 42 0.91
## 43 0.88
## 44 0.82
## 45 0.88
## 46 0.87
## 47 1.04
## 48 0.91
## 49 1.07
## 50 1.12
```

```
## 51 1.12
## 52 1.24
## 53 1.01
## 54 1.13
## 55 0.92
## 56 0.98
## 57 0.94
## 58 1.07
## 59 0.89
```

(Note that a select function is defined in multiple packages, so if you want to be sure you're using the one from the dplyr library, call dplyr::select.)

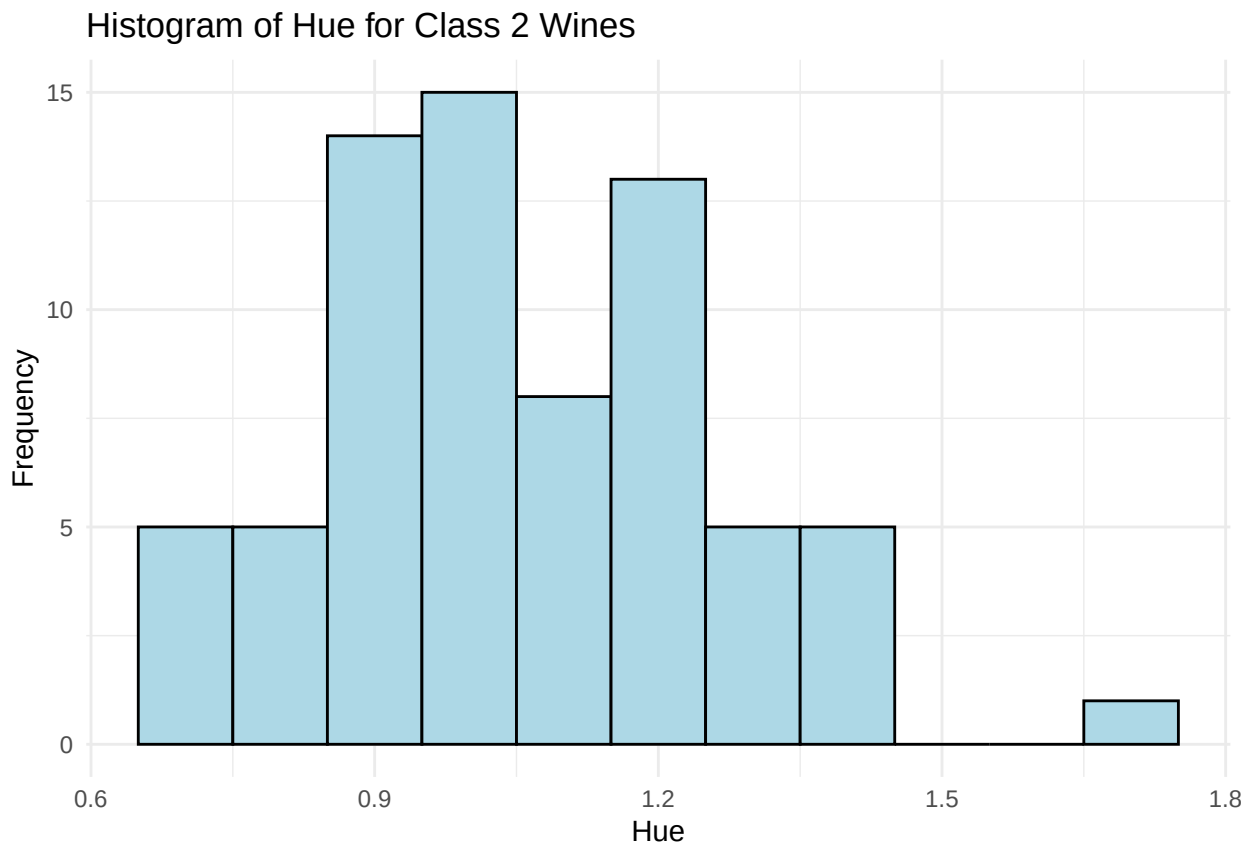
Also create histograms of the hue level for wines 2 and 3. What can you say about the differences in the hue level by looking at the histograms? (1 pt)

```
library(ggplot2)
```

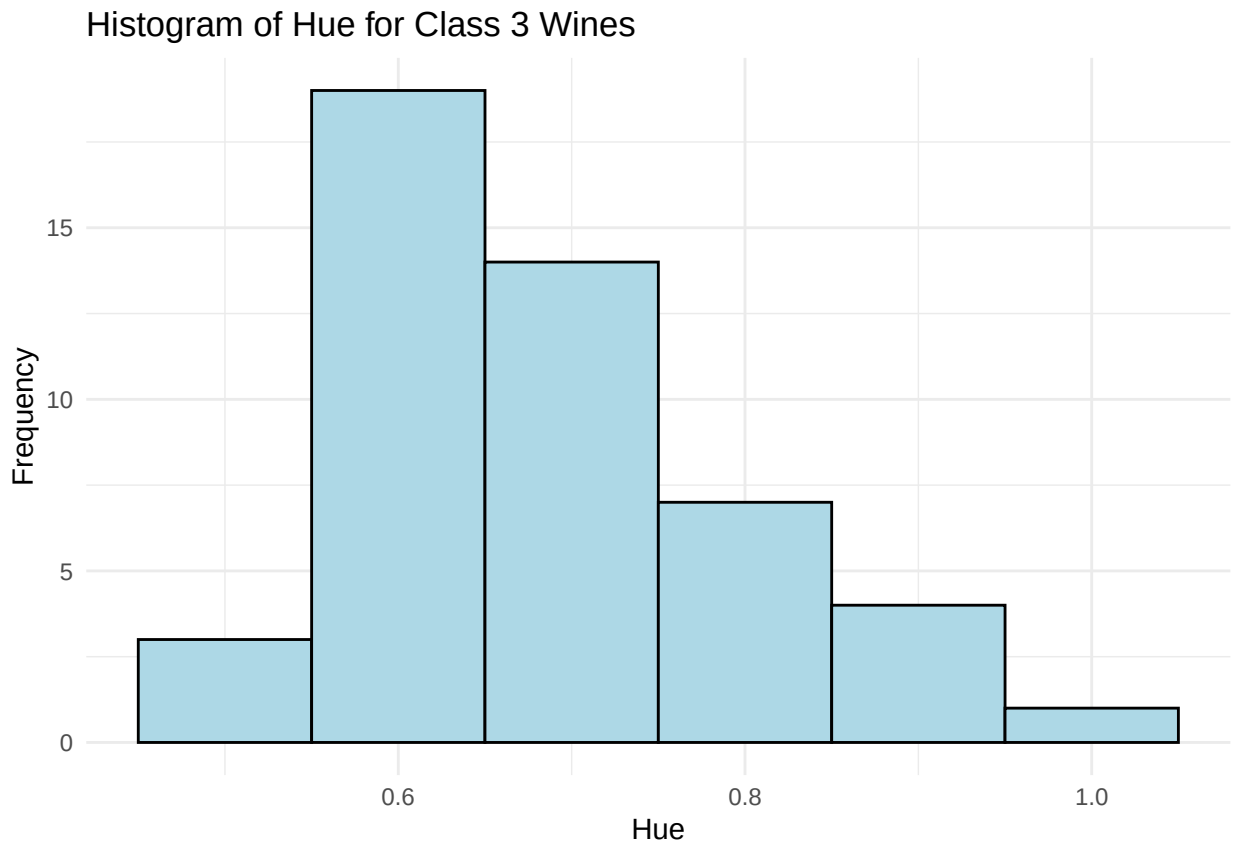
```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
class_2_wines <- wine[wine$class == 2, ]
class_3_wines <- wine[wine$class == 3, ]

ggplot(class_2_wines, aes(x = hue)) +
  geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black") +
  labs(title = "Histogram of Hue for Class 2 Wines", x = "Hue", y = "Frequency") +
  theme_minimal()
```



```r
ggplot(class_3_wines, aes(x = hue)) +
  geom_histogram(binwidth = 0.1, fill = "lightblue", color = "black") +
  labs(title = "Histogram of Hue for Class 3 Wines", x = "Hue", y = "Frequency") +
  theme_minimal()
```




```r
#Class 1 and 2 wines have hue values that are more spread out and range into higher hue values,
#while Class 3 wines are more concentrated around lower hue values.

#Class 3 wines have a more skewed distribution, concentrated in the lower end of the hue spectrum, whic
#Class 2 wines have a wider range of hue values and appear more variable compared to class 1 and class

#My interpretation is that class 1 corresponds to red wine lower hue color, darker colors absorb more l
#The broader hue distribution class 2 (peaking around 0.9 but extending to 1.8) suggests these wines mi
#Class 3 is by elimination white but also they have a higher hue color corresponding to the lighte colo
```

Chapter 4, Exercise 8

Create a boxplot (or violin plot) of the hue data, using the model "hue ~ class" so they are plotted side-by-side. (1 pt) What can you say about the differences in the hue by looking at the boxplots for the different wines? (1 pt)

```r
library(ggplot2)
ggplot(wine, aes(x = factor(class), y = hue)) +
```

```
  geom_violin(trim = FALSE, fill = "lightblue") +
  labs(title = "Violin Plot of Hue by Wine Class", x = "Wine Class", y = "Hue") +
  theme_minimal()
```


```
#As I stated in my previous answer my hypothesis is further proved how 2 classes are simmilar and their

#Class 1 wines have a more consistent hue, representing white wines with light and consistent coloring.
#Class 2 wines have a broader hue range, which indicates rosé wines or other wines with intermediate hu
#Class 3 wines have lower hue values, these are red wines, known for having deeper and darker hues.
```

Chapter 4, Exercise 9

Run a t-test to compare the means of classes 1 and 2 in the wine data and report the confidence interval. (1 pt) Give an interpretation of the confidence interval. (1 pt) Make sure to include a carefully worded statement about what the confidence interval implies with respect to the population mean difference between the wines (specifically, if the groups are different or not). (1 pt)

```
t_test_result <- t.test(hue ~ class, data = wine, subset = class %in% c(1, 2))

# Display the results of the t-test
print(t_test_result)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  hue by class
## t = 0.20211, df = 114.74, p-value = 0.8402
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -0.05062450  0.06212891
## sample estimates:
## mean in group 1 mean in group 2
##        1.062034        1.056282
```

```
#The t-test results show no statistically significant difference in the mean hue between class 1 and cl

#Although my theory still stands and my new hypothesis is the wines used in the class 2 are really ligh
```

Chapter 4, Exercise 10

Run a t-test to compare the means of wines 1 and 3 in the wine data. (1 pt) Report and interpret the confidence interval. (1 pt + 1 pt for statement about means)

```
# Perform a t-test between class 1 and class 3 wines for hue
t_test_1_3 <- t.test(hue ~ class, data = wine, subset = class %in% c(1, 3))

# Print the t-test results
print(t_test_1_3)
```

```
##
##  Welch Two Sample t-test
##
## data:  hue by class
## t = 16.916, df = 101.3, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 1 and group 3 is not equal to 0
## 95 percent confidence interval:
##  0.3348447 0.4238064
## sample estimates:
## mean in group 1 mean in group 3
##       1.0620339       0.6827083
```

```
#The 95% confidence interval for the difference in means is [0.3348, 0.4238], indicating that the mean
#Here theres a big difference between hues 1 and 3 my theory stands about the type of hue colors of the
```