

Homework 4

Kwasi Mensah

2022-11-03

Author: Kwasi Mensah Homework_Number: 4 Due Date: "2022-05-11" Output:
pdf_document Attribution statement:

I did the homework by myself, with help from the book and the professor #R Markdown
#Run these three functions to get a clean test of homework code

```
dev.off() #Clear the graph window

## null device
##          1

cat('\014') #Clear the console

rm(list = ls()) #Clear user objects from the environment
```

#R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#####
#####
```

7. The built-in PlantGrowth data set contains three different groups, each representing a different plant food diet (you may need to type `data(PlantGrowth)` to activate it). The group labeled "ctrl" is the control group, while the other two groups are each a different type of experimental treatment. Run the `summary()` command on PlantGrowth and explain the output.

Create a histogram of the ctrl group. As a hint about R syntax, here is one way that you can access the ctrl group data:

```
PlantGrowthweight[PlantGrowthgroup=='ctrl']
```

Also create histograms of the trtl and trt2 groups. What can you say about the differences in the groups by looking at the histograms?

```
# Load the Plant Growth data set
```

```
data(PlantGrowth)
```

```
?PlantGrowth
```

```
## starting httpd help server ... done
```

```
# Print a summary of the Plant Growth data
```

```
summary(PlantGrowth)
```

```
##      weight      group  
## Min.   :3.590   ctrl:10  
## 1st Qu.:4.550   trt1:10  
## Median :5.155   trt2:10  
## Mean   :5.073  
## 3rd Qu.:5.530  
## Max.   :6.310
```

#The summary function summarize the values in a a data frame. It shows there are 3 groups with 10 values each, (ctrl,trt1,& trt2). In addition it shows the mean,5.073, median, 5.155, the minimum and maximum value(3.590,6.310) and the 1st and 3rd quartiles, 4.550 & 5.530.

```
# Break the weights down by group
```

```
ctrl.weights = PlantGrowth$weight[PlantGrowth$group == 'ctrl']
```

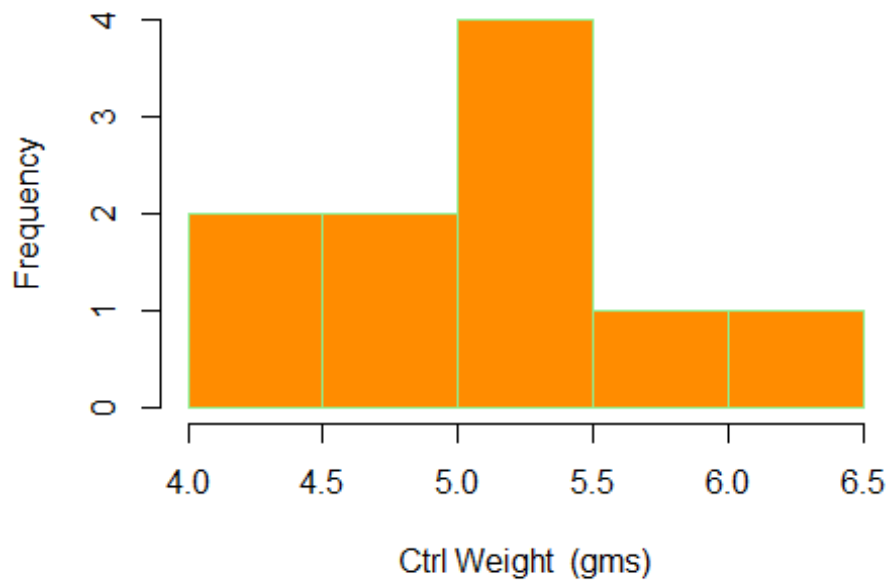
```
trt1.weights = PlantGrowth$weight[PlantGrowth$group == 'trt1']
```

```
trt2.weights = PlantGrowth$weight[PlantGrowth$group == 'trt2']
```

```
# Histogram of Control group
```

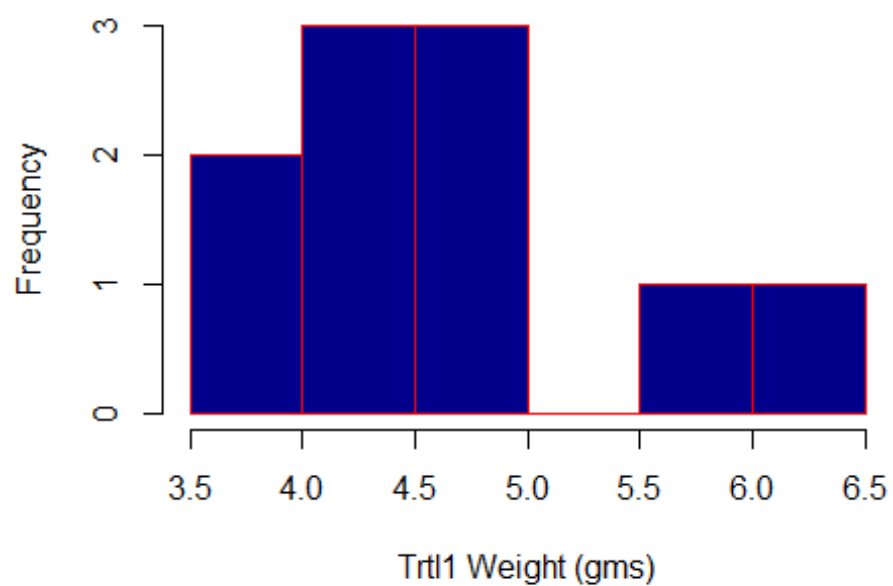
```
hist(ctrl.weights, main="Histogram of Control Group Plant Weights"  
  , ylab = "Frequency"  
  , xlab="Ctrl Weight (gms)"  
  , col="darkorange"  
  , border = "lightgreen"  
  , breaks=5)
```

Histogram of Control Group Plant Weights



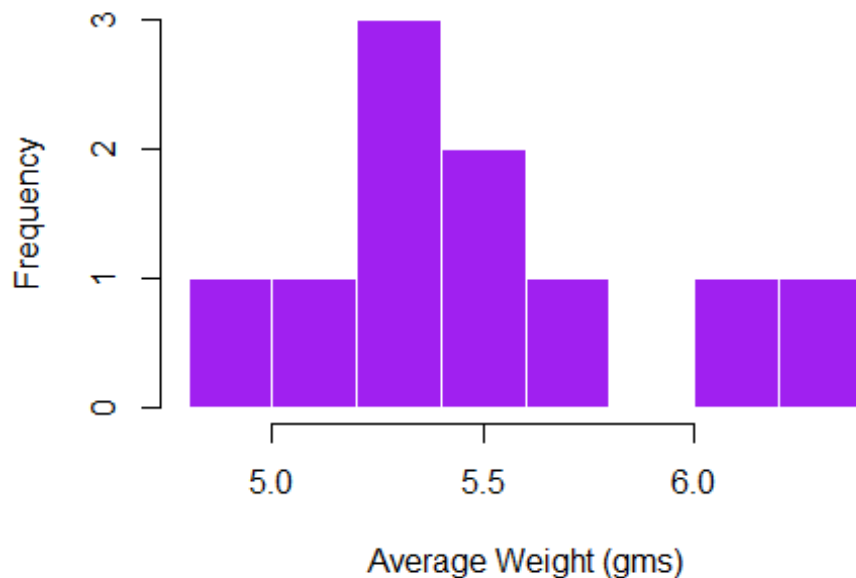
```
# Histogram of Treatment1 & Treatment2  
hist(trt1.weights, main="Histogram of TRT1 Group Plant Weights"  
      , ylab = "Frequency"  
      , xlab="Trt1 Weight (gms)"  
      , col="darkblue"  
      , border = "red"  
      , breaks=5)
```

Histogram of TRT1 Group Plant Weights



```
hist(trt2.weights,main="Histogram of TRT2 Group Plant Weights"  
  , ylab = "Frequency"  
  , xlab="Average Weight (gms)"  
  , col="purple"  
  , border = "white"  
  , breaks=5)
```

Histogram of TRT2 Group Plant Weights

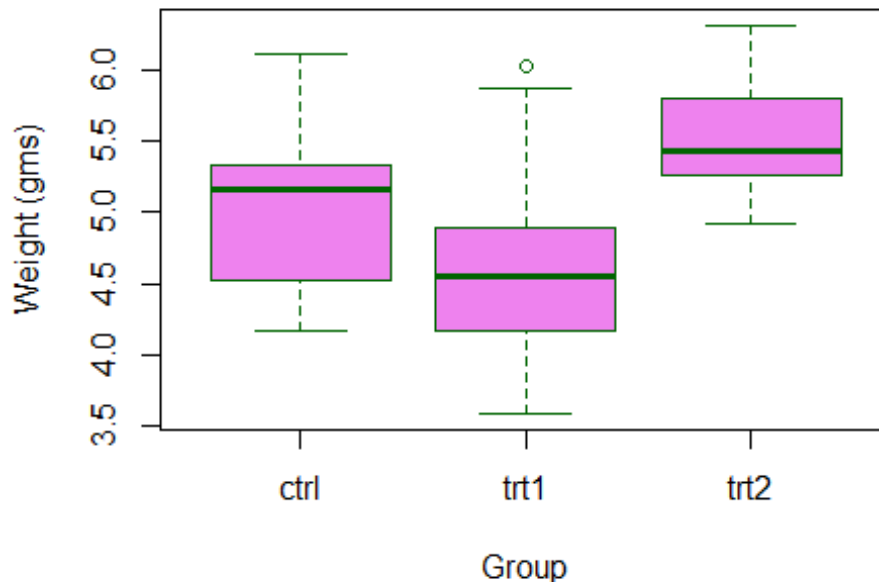


#The Graphs show that the Treatment1 group frequently produces plants that weight less and Treatment2 group frequently produce plants with more weight compare to the other 2 groups. Treatment2 data is more widely dispersed at the upper and lower levels compared to the other two histograms. Both the Control group and the Treatment1 histograms are more skewed to the Left.

8. Create a boxplot of the plant growth data, using the model "weight - group." What can you say about the differences in the groups by looking at the boxplots for the different groups?

```
# Create a boxplot of weights by group
boxplot(weight ~ group, PlantGrowth
, main="Boxplot Chart of Plant Weights by Group"
, xlab="Group"
, ylab="Weight (gms)"
, col = "violet"
, border = "darkgreen")
```

Boxplot Chart of Plant Weights by Group



The Graph shows that the median weight for the Treatment2 group is higher than both the Treatment1 & Control groups. It may be possible the the Treatment2 results in the dried weight of plants being greater than the other treatment vs the control group. The graph also seems to show that Treatment1 group results in plants with less weight as it has the lowest median as well as 1st quartile and minimum value. In addition there is an outlier among Treatment1. The ranges in Treatment1 overlap some of the ranges in the control group.

9. Run a t-test to compare the means of ctrl and trt1 in the PlantGrowth data. Report and interpret the confidence interval. Make sure to include a carefully worded statement about what the confidence interval implies with respect to the population mean difference between the ctrl and trt1 groups.

Generate a t test on the means between the control group and diet trt1
t.test(ctrl.weights, trt1.weights)

```
##
## Welch Two Sample t-test
##
## data: ctrl.weights and trt1.weights
## t = 1.1913, df = 16.524, p-value = 0.2504
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2875162 1.0295162
## sample estimates:
```

```
## mean of x mean of y
##      5.032      4.661
```

If we do many replications we are 95% confident that the population mean maybe between, -0.2875162, and 1.0295162.

10. Run a t-test to compare the means Of ctrl and trt2 in the PlantGrowth data. Report and interpret the confidence interval. Generate a t test on the means between the control group and diet trt2.

```
t.test(ctrl.weights, trt2.weights)
```

```
##
## Welch Two Sample t-test
##
## data: ctrl.weights and trt2.weights
## t = -2.134, df = 16.786, p-value = 0.0479
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.98287213 -0.00512787
## sample estimates:
## mean of x mean of y
##      5.032      5.526
```

The whole confidence interval is on negative side so it does not overlap with 0. If we do many replications we are 95% confident that the population mean maybe between, -0.98287213 and -0.00512787.