Matthew L. Pergolski
IST 772
Dr. Block
10/24/2021

Homework 3

Beginning Statement

"I produced the material below with no assistance [direct quote from IST 772 class syllabus]." Note: Homework questions from the book may have been copied/pasted into the document for both the student and viewer's convenience.

The homework for Week 4 is exercises 7-10 on page 66.

Homework Question 7

Question:
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:

PlantGrowth$weight[PlantGrowth$group=="ctrl"]

Also create histograms of the trt1 and trt2 groups. What can you say about the differences in the groups by looking at the histograms?

Answer/Student Response:
The summary command was run and the following was observed:

```
>           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
```

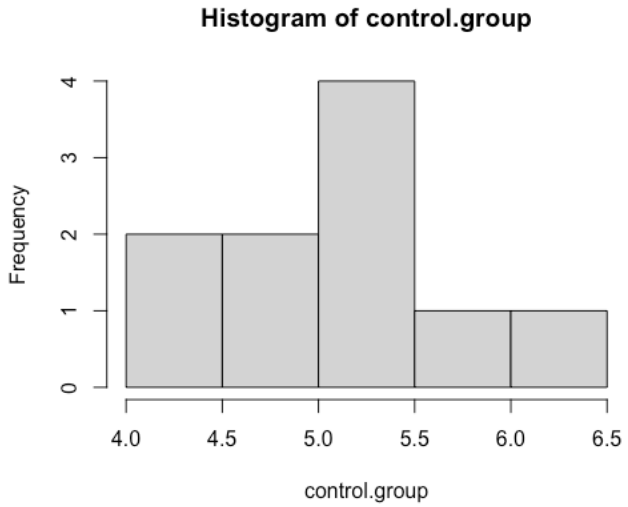
The summary command shows that the minimum weight value is 3.590 units while the max is 6.310; the mean is 5.073 and median is 5.155.

This conveys that the distribution of weight is slightly positively skewed between all control groups (since the means are not the same -- the mean is more sensitive to outliers).

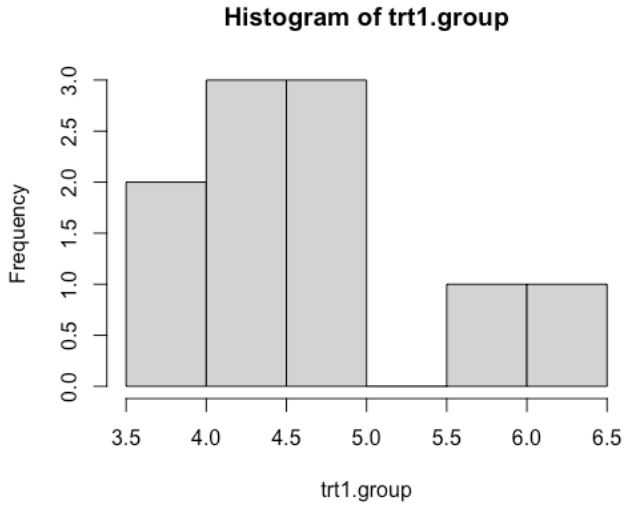
Histograms were taken of the different control groups:

Control group

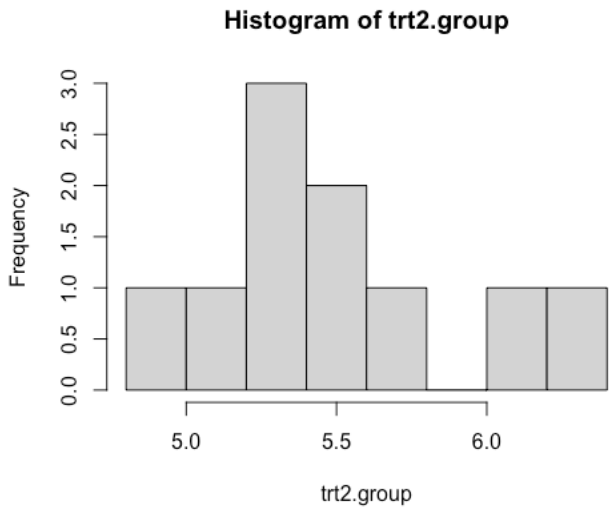
control.group <- PlantGrowth$weight[PlantGrowth$group=="ctrl"], hist(control.group) – see R-Code file for more information



Trt1
PlantGrowth$weight[PlantGrowth$group=="trt1"], hist(trt1.group) – see R-Code file for more information



Trt2
trt2.group <- PlantGrowth$weight[PlantGrowth$group=="trt2"], hist(trt2.group) – see R-Code file for more information



In looking at all of the histograms for the different groups, I'd say that they all have 'tails' extending towards the right-side of the diagrams, which suggests a positive skew; none of these would be considered a uniform distribution, as there is a bit of a central tendency in each of the control groups.

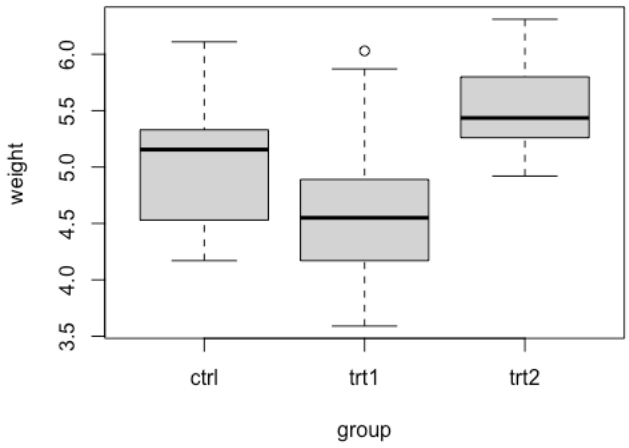
Homework Question 8

Question:
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?

Answer/Student Response:
The following code was ran:

```
boxplot(weight ~ group, data = PlantGrowth)
```

The following was displayed as output:



None of the boxplots are uniform (i.e., none of them look the same and are not positioned similarly in the diagram). Because of this, we can see that none of the median values are the same, and the lower bounds are considerably off as well. The upper bounds are closer, but still visually not equal. All in all, the central tendency and dispersion of these independent samples are quite different from each other. Trt2 distribution is higher in every dimension, while the trt1 distribution has the lowest dimensions out of the three; lastly, ctrl is about in the middle of the other two, in terms of distribution metrics, comparatively.

Homework Question 9

Question:
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.

Answer/Student Response:
The following output was observed from the t.test code:

```
>           t.test(control.group, trt1.group)

        Welch Two Sample t-test

data:  control.group and trt1.group
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
```

A 95% confidence interval of -0.2875162 and 1.0295162 was observed; we can conclude that, if this test was replicated 100 times, 95 of these replications would contain the true population mean and be represented within this interval of -0.2875162 and 1.0295162 -- or, at least, that's what we can estimate. The point estimate, which is the difference of the two means listed in the t.test (i.e., 5.032 - 4.661 = 0.371), would be the metric of choice to compare to this interval range.

Our range of uncertainty lies between the upper limit and the lower band. The wider the interval, the more uncertain we are in this specific test. We can not say for certain the true population value would be within this range (in the long run), since inferential statistics can not necessarily 'prove' anything for sure -- but we can use this information to be as confident as possible about determining the true population mean (or getting close to it, at least – which may or may not be the point-estimate) from the data we have.

Homework Question 10

Question:
10. Run a t-test to compare the means of ctrl and trt2 in the PlantGrowth data. Report and interpret the confidence interval.

Answer/Student Response:
The following code was ran:

```
>           t.test(control.group, trt2.group)

        Welch Two Sample t-test

data:  control.group and trt2.group
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
```

A 95% confidence interval of -0.98287213 and -0.00512787 was observed; we can conclude that, if this test was replicated 100 times, 95 of these replications would contain the true population mean and be represented within this interval of -0.98287213 and -0.00512787 -- or, at least, that's what we can estimate. The point estimate, which is the center point of the two means listed in the t.test (i.e., 5.032 - 5.526 = -0.4939), would be the metric of choice to compare to this interval range.

Our range of uncertainty lies between the upper limit and the lower band. The wider the interval, the more uncertain we are in this specific test. We can not say for certain the true population value would be within this range (in the long run), since inferential statistics can not necessarily 'prove' anything for sure, but we can use this information to be as confident as possible about determining the true population mean (or getting close to it, at least – which may or may not be the point-estimate) from the data we have.