

Attribution Statement:

1. Homework 4 by Parin Patel: I did this homework by myself, with help from the book and the professor. In addition I used the following websites to help with the r code for the boxplot help:
 - a. <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>

Exercises:

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.

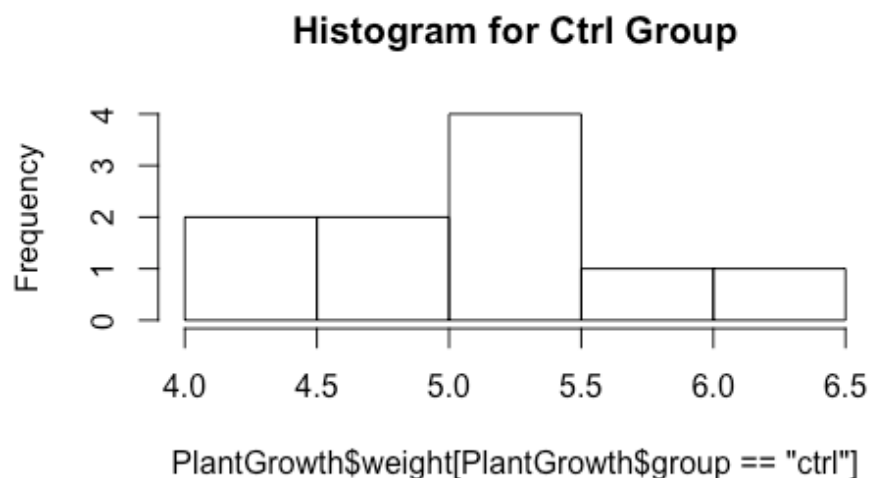
`PlantGrowth$weight[PlantGrowth$group=="ctrl"]`

Also create a histogram of the tr1 and tr2 groups. What can you see about the differences in the groups by looking at the histograms?

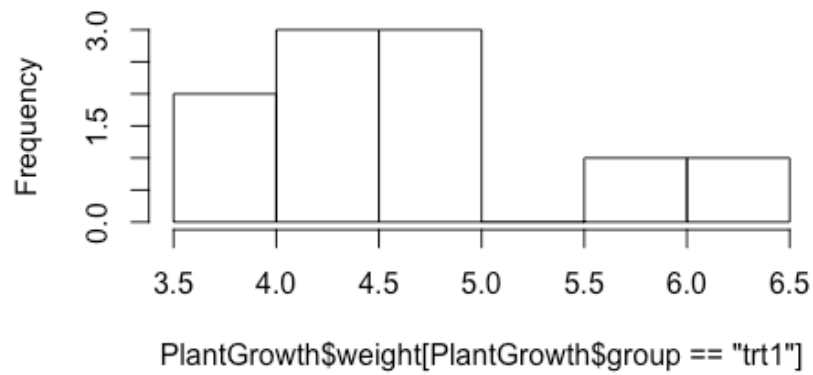
```
> ##7.
> summary(PlantGrowth)
      weight      group
Min.   :3.590   ctrl:10
1st Qu.:4.550   trt1:10
Median :5.155   trt2:10
Mean    :5.073
3rd Qu.:5.530
Max.    :6.310
>
```

The numeric variables in the dataset are weight and group. Within the group variable, there are three categories. One control group and two treatment groups. Between all three groups, there are 30 datapoints, or cases. Additionally, the variable weights is recorded for each group. The mean is the average number in the set of data. The median is the middle number in the dataset. The mean weight of all the groups is 5.073, while the min is 3.590 and the max is 6.310. The interquartile range of the weight variable ranges from 4.550 for Q1 (25th percentile) and 5.530 for Q3 (75th percentile).

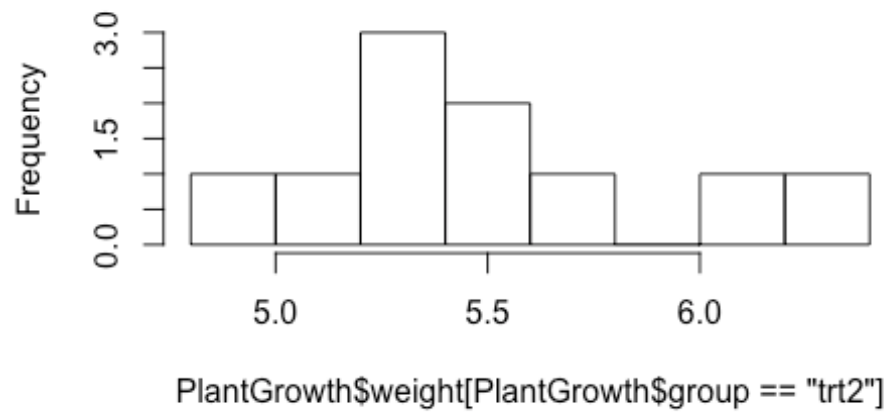
After looking at all three histograms below, you can say that the weights of the treatment groups vary quite differently from that of the control group. The control group is nearly normally distributed. And has one central peak around the mean and median weights. Looking at its histogram, it can be expected that the mean and median of the ctrl group is similar to that of the overall PlantGrowths weights data. This is unlike the first treatment group, which has a more positive distribution. Therefore, the weights of the treatment group are clustered to the left, lower bins. Based on this, it can be expected that the mean of the first treatment group is lower than that of the control and overall PlantGrowth weights mean. In treatment group 1, a significant number of datapoints are below the mean. However, treatment group 2 seems to have a overall more negatively skewed distribution. It's important to note that while the histogram below, at first, does seem to have a positive, left-leaning distribution, when comparing the x-axis bins to the other two groups, it is clear that the bins for the treatment 2 histogram has higher values than those shown for the control and treatment 1 groups. Therefore, it can be expected that the mean of the treatment 2 group will be higher than that for the treatment 1 group, control group, and the overall PlantGrowth weights variable.



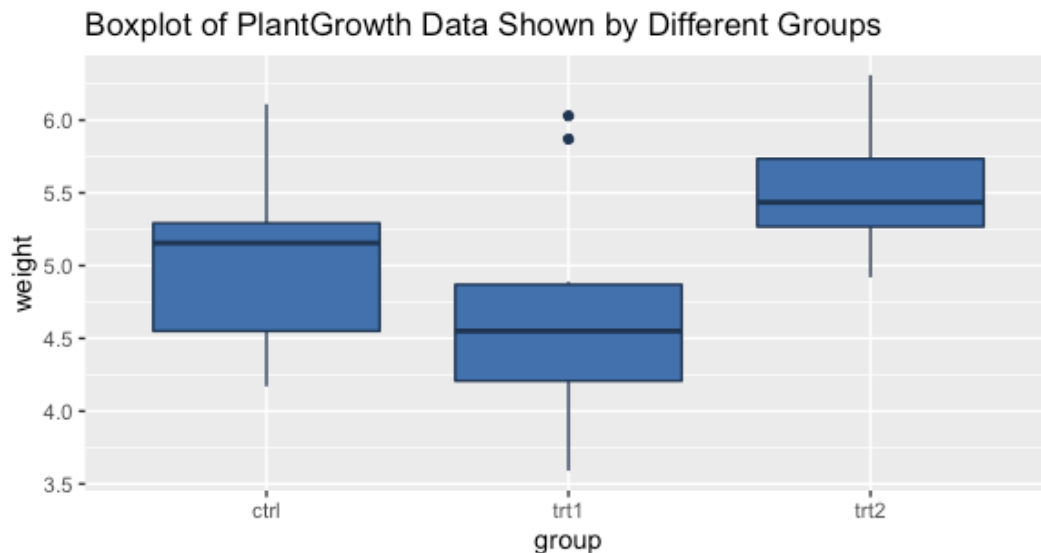
Histogram for Trt1 Group



Histogram for Trt2 Group



8. Create a boxplot of the PlantGrowth data, using the model “weight~growth.”. What can you say about the differences in the groups by looking at the boxplot for the different groups?



Based on the boxplot, we can say confirm some of the statements made in question 7 above. First, the means for trt1 is lower than the mean for the control group. Also, the mean for the trt2 group is higher than the control and trt1. In addition, the range for trt1 contains much lower weight values than trt2 and the control, overall. However, the boxplot shows that trt2 contains two outliers that are likely resulting in greater variability for that group. It is interesting to note that the median for the trt1 group is the only one out of the three groups to seem more centered within the range.

It can also be observed that the range between Q1 (25th percentile) and Q2 (75th percentile) of trt2 is much smaller than the control and trt1 groups, therefore, we would expect the data to appear more centralized, and consolidated around the mean.

9. Run a t-test to compare the means of ctrl1 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.

Parin Patel
Homework 4
10/27/2019

```
> #t-test
> t.test(PlantGrowth$weight[PlantGrowth$group=='ctrl'], PlantGrowth$weight[PlantGrowth$group=='trt1'])

Welch Two Sample t-test

data: PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[PlantGrowth$group == "trt1"]
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
```

The p-value is 0.25. Since the $p\text{-value} > 0.05$, and the null hypothesis is likely to be true, and we cannot reject it. We are 95% confident that there is no significant difference between the population means of the ctrl and trt1 groups. The maximum difference between the two means can be as low as -0.288 and as high as 1.030.

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

```
> CtrlMean - Trt2Mean
[1] -0.455
> #t-test
> t.test(PlantGrowth$weight[PlantGrowth$group=='ctrl'], PlantGrowth$weight[PlantGrowth$group=='trt2'])

Welch Two Sample t-test

data: PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[PlantGrowth$group == "trt2"]
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 p-value is 0.0479. Since the $p\text{-value} < 0.05$, and the null hypothesis is rejected, and is unlikely to be true. We are 95% confident that there is a difference between the population means of the ctrl and trt2 groups. The maximum difference between the two means can be as low as -0.983 and as high as -0.005. The plants seem to have a larger weight under treatment 2 than under the control.

Parin Patel
Homework 4
10/27/2019

Appendix A: Final Script:

```
install.packages("animation")  
install.packages("ggplot2")  
library("animation")  
library("ggplot2")
```

```
data("PlantGrowth")
```

```
##7.
```

```
summary(PlantGrowth) #summary of PlantGrowth  
#Histogram for Control Group  
histctrl<-hist(PlantGrowth$weight[PlantGrowth$group=='ctrl'], main = "Histogram for Ctrl  
Group")  
histctrl  
#Histogram for TR1  
histTRT1<-hist(PlantGrowth$weight[PlantGrowth$group=='trt1'], main = "Histogram for Trt1  
Group")  
histTRT1  
#Histogram for TR2  
histTRT2<-hist(PlantGrowth$weight[PlantGrowth$group=='trt2'], main = "Histogram for Trt2  
Group")  
histTRT2
```

```
##8 Boxplot of groups, weights
```

```
fill <- "#4271AE"  
line <- "#1F3552"  
boxplot1 <- ggplot(PlantGrowth, aes(x=group, y=weight)) +  
  geom_boxplot(fill=fill, colour=line) +  
  labs(title="Boxplot of PlantGrowth Data Shown by Different Groups")  
boxplot1
```

```
##9. t-test to compare means of ctrl1 and trt1
```

```
#sample mean calc
```

```
CtrlMean<-mean(  
sample(PlantGrowth$weight[PlantGrowth$group=='ctrl'],size=10,replace=TRUE) ) #ctrl
```

IST772

Parin Patel

Homework 4

10/27/2019

```
Trt1Mean<-mean(
sample(PlantGrowth$weight[PlantGrowth$group=='trt1'],size=10,replace=TRUE) ) #trt1
#mean difference
CtrlMean - Trt1Mean
```

```
#t-test
t.test(PlantGrowth$weight[PlantGrowth$group=='ctrl']
,PlantGrowth$weight[PlantGrowth$group=='trt1'])
```

##10. t-test to compare means of ctrl1 and trt2

```
#sample mean calc
CtrlMean<-mean(
sample(PlantGrowth$weight[PlantGrowth$group=='ctrl'],size=10,replace=TRUE) ) #ctrl
Trt2Mean<-mean(
sample(PlantGrowth$weight[PlantGrowth$group=='trt2'],size=10,replace=TRUE) ) #trt2
#mean difference
CtrlMean - Trt2Mean
```

```
#t-test
t.test(PlantGrowth$weight[PlantGrowth$group=='ctrl']
,PlantGrowth$weight[PlantGrowth$group=='trt2'])
```

Appendix B: Final Output;

```
> ##7.
> summary(PlantGrowth) #summary of 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
> #Histogram for Control Group
> histctrl<-hist(PlantGrowth$weight[PlantGrowth$group=='ctrl'], main = "Histogram for Ctrl
Group")
> histctrl
$breaks
[1] 4.0 4.5 5.0 5.5 6.0 6.5
```

IST772

Parin Patel
Homework 4
10/27/2019

\$counts

[1] 2 2 4 1 1

\$density

[1] 0.4 0.4 0.8 0.2 0.2

\$mids

[1] 4.25 4.75 5.25 5.75 6.25

\$xname

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

\$equidist

[1] TRUE

attr("class")

[1] "histogram"

> #Histogram for TR1

> histTRT1<-hist(PlantGrowth\$weight[PlantGrowth\$group=='trt1'], main = "Histogram for Trt1 Group")

> histTRT1

\$breaks

[1] 3.5 4.0 4.5 5.0 5.5 6.0 6.5

\$counts

[1] 2 3 3 0 1 1

\$density

[1] 0.4 0.6 0.6 0.0 0.2 0.2

\$mids

[1] 3.75 4.25 4.75 5.25 5.75 6.25

\$xname

[1] "PlantGrowth\$weight[PlantGrowth\$group == \"trt1\"]"

\$equidist

[1] TRUE

attr("class")

[1] "histogram"

> #Histogram for TR2

IST772

Parin Patel

Homework 4

10/27/2019

```
> histTRT2<-hist(PlantGrowth$weight[PlantGrowth$group=='trt2'], main = "Histogram for Trt2 Group")
```

```
> histTRT2
```

```
$breaks
```

```
[1] 4.8 5.0 5.2 5.4 5.6 5.8 6.0 6.2 6.4
```

```
$counts
```

```
[1] 1 1 3 2 1 0 1 1
```

```
$density
```

```
[1] 0.5 0.5 1.5 1.0 0.5 0.0 0.5 0.5
```

```
$mids
```

```
[1] 4.9 5.1 5.3 5.5 5.7 5.9 6.1 6.3
```

```
$xname
```

```
[1] "PlantGrowth$weight[PlantGrowth$group == \"trt2\"]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
>
```

```
>
```

```
>
```

```
> ##8 Boxplot of groups, weights
```

```
> fill <- "#4271AE"
```

```
> line <- "#1F3552"
```

```
> boxplot1 <- ggplot(PlantGrowth, aes(x=group, y=weight)) +
```

```
+ geom_boxplot(fill=fill, colour=line) +
```

```
+ labs(title="Boxplot of PlantGrowth Data Shown by Different Groups")
```

```
> boxplot1
```

```
>
```

```
>
```