Homework Chapter 4 - Mark Cappiello

1. 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:

PlantGrowthgroup==”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?

data("PlantGrowth")  
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 provides a statistical summary of the variables in the PlantGrowth dataset. For the weight varaible it shows the minimum, 1st quartile, median, mean, 3rd quartile, and maximum values. For the group variable, the frequency of each group (ctrl, trt1, trt2) is returned

Explanation of the summary() Output

Weight: ( I looked but could not find the unit of measure )

Min.: The minimum weight of the plants = 3.590

1st Qu.: The first quartile = 4.550 – 25% of the data falls below this value

Median: The median weight = 5.155 – the middle value of the dataset. 50% falls below this

Mean: The average weight of the plants. 5.073 The sum of the values divided by the count.

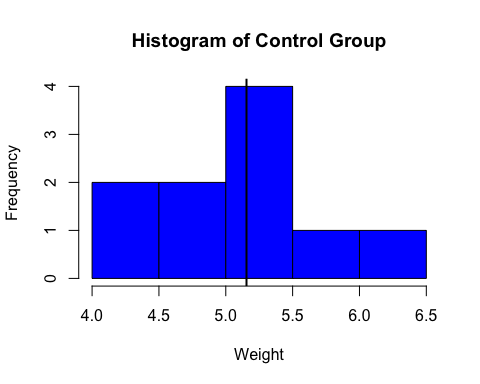
3rd Qu.: The third quartile = 5.530 – 75% of the data is below this value.

Max.: The maximum weight of the plants = 6.310

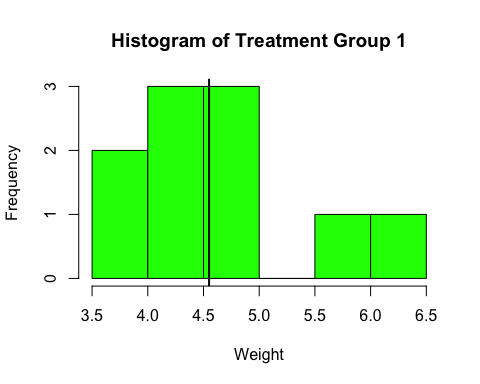
Group:

The frequency count of each group (ctrl, trt1, trt2), each group has 10 observations.

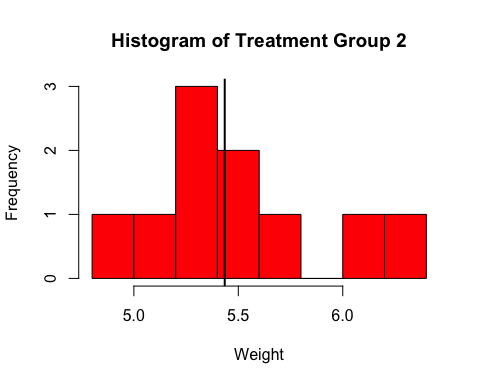
# Histogram of the control group  
ctrl\_median <- median(PlantGrowth$weight[PlantGrowth$group == "ctrl"])  
hist(PlantGrowth$weight[PlantGrowth$group == "ctrl"],   
 main = "Histogram of Control Group",   
 xlab = "Weight",   
 col = "blue",   
 border = "black")  
abline(v = ctrl\_median, col = "black", lwd = 2)



# Histogram of the trt1 group  
trt1\_median <- median(PlantGrowth$weight[PlantGrowth$group == "trt1"])  
hist(PlantGrowth$weight[PlantGrowth$group == "trt1"],   
 main = "Histogram of Treatment Group 1",   
 xlab = "Weight",   
 col = "green",   
 border = "black")  
abline(v = trt1\_median, col = "black", lwd = 2)



# Histogram for the trt2 group  
trt2\_median <- median(PlantGrowth$weight[PlantGrowth$group == "trt2"])  
hist(PlantGrowth$weight[PlantGrowth$group == "trt2"],   
 main = "Histogram of Treatment Group 2",   
 xlab = "Weight",   
 col = "red",   
 border = "black")  
abline(v = trt2\_median, col = "black", lwd = 2)

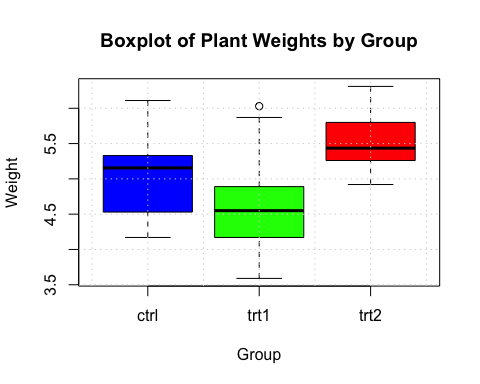


Observations

The ctrl group is close to being a normal distribution but having only a 10 point sample makes that difficult. Ctrl’s spread is in between the other two spreads. Trt1 has a larger spread and trt2 has a smaller spread. Both trt1 and trt2 to be right-skewed.

1. 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?

# Load the dataset  
data("PlantGrowth")  
  
# Create the boxplot  
boxplot(weight ~ group, data = PlantGrowth,  
 main = "Boxplot of Plant Weights by Group",  
 xlab = "Group",  
 ylab = "Weight",  
 col = c("blue", "green", "red"))  
grid()



Interpretation of the Boxplot

The boxplot shows the following for groups ctrl, trt1 and trt2:

Median: The line inside the box represents the median weight of the plants in each group.

Interquartile Range (IQR): The box represents the range between the first quartile (25th percentile) and the third quartile (75th percentile).

Whiskers are the lines extending from the box represent the range of the data but not including outliers.

Outliers are data points outside the whiskers and are plotted as individual points, like the small circle above trt1.

Output and Interpretation

The median weight for trt1 is lower than the ctrl group, a possible a negative effect of the trt1.

The median weight for trt2 is higher than the ctrl group, a possible a positive effect of the trt2.

Trt2 has a larger IQR this indicates more variability in the treatment effect.

The control group has a narrower IQR, indicating more consistent weights compared to the treatment groups.

1. 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.

# Extract the weights for ctrl and trt1 groups and run t-test  
ctrl\_weights <- PlantGrowth$weight[PlantGrowth$group == "ctrl"]  
trt1\_weights <- PlantGrowth$weight[PlantGrowth$group == "trt1"]  
t\_test\_result <- t.test(ctrl\_weights, trt1\_weights)  
  
print(t\_test\_result)

##   
## 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

Summary of the t-test Results

t-value: 1.1913

degrees of freedom (df): 16.524

p-value: 0.2504

alternative hypothesis: true difference in means is not equal to 0

95% confidence interval for the difference in means: -0.2875162 to 1.0295162

mean of ctrl\_weights (mean of x): 5.032

mean of trt1\_weights (mean of y): 4.661

Interpretation

t-value (1.913): The t-value measures the size of the difference relative to the variation in the sample data. A higher absolute t-value indicates a more significant difference between the groups. A t-value of 1.1913 says that there is a moderate difference between the means of the ctrl and trt1 groups.

degrees of freedom (16.524): Degrees of freedom reflect the number of independent values or quantities which can be assigned to a statistical distribution. This value indicates a moderate sample size,

p-value (0.2504): The p-value is the probability of observing the data if the null hypothesis is true, meaning that there is no difference in the means between the two groups observed. A p-value of 0.2504 is greater than 0.05 also known as the common significance level. This says that the result is not statistically significant. And it means that there is not enough evidence to reject the null hypothesis.

Alternative hypothesis: The alternative hypothesis says that the true difference in means is not equal to 0.

The 95% confidence interval is between -0.2875162 and 1.0295162 The confidence interval gives a range of values where we can be 95% confident that the difference in means lies between. Since this interval includes 0, it means that ther is no significant difference.

Sample estimates are mean of ctrl, x = 5.032 and mean of trt1, y = 4.661

Conclusion

The t-test results suggest that there is no significant difference between the mean weights of the ctrl group (mean = 5.032) and the trt2 group (mean = 4.661).

The p-value (0.2504) is greater than 0.05 indicates that the observed difference is not statistically significant.

The 95% confidence interval for the difference in means (-0.2875162 to 1.0295162) includes 0, supporting the conclusion that the true mean difference could be 0.

Based on these t-test results, there isn’t sufficient evidence to conclude that trt2 had significant effect on plant weights compared to the ctrl group. So any observed difference in the sample means could be due to random chance instead of the effect of trt2.

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

# Load the dataset  
data("PlantGrowth")  
  
# Extract the weights for ctrl and trt2 groups  
ctrl\_weights <- PlantGrowth$weight[PlantGrowth$group == "ctrl"]  
trt2\_weights <- PlantGrowth$weight[PlantGrowth$group == "trt2"]  
  
# Run the t-test  
t\_test\_result <- t.test(ctrl\_weights, trt2\_weights, var.equal = TRUE)  
  
# Print the t-test result  
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: ctrl\_weights and trt2\_weights  
## t = -2.134, df = 18, p-value = 0.04685  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.980338117 -0.007661883  
## sample estimates:  
## mean of x mean of y   
## 5.032 5.526

Summary of the t-test Results

t-value: -2.134

degrees of freedom (df): 18

p-value: 0.04685

alternative hypothesis: true difference in means is not equal to 0

95% confidence interval for the difference in means: -0.980338117 to -0.007661883

mean of ctrl\_weights (mean of x): 5.032

mean of trt2\_weights (mean of y): 5.526

Interpretation

t-value (-2.134): A t-value of -2.134 suggests a moderate difference between the means of the ctrl and trt2 groups. The negative sign shows that the mean of the ctrl group is less than the mean of the trt2 group.

Degrees of freedom (df = 18): Says there were 20 observations minus 2 for the two group means

p-value (0.04685): A p-value of 0.04685 is just below the common significance level of 0.05, suggesting that the result is statistically significant. This means there is enough evidence to reject the null hypothesis and conclude that there is a significant difference in means between the ctrl and trt2 groups.

Alternative hypothesis: The alternative hypothesis says that the true difference in means is not equal to 0.

95% confidence interval: -0.980338117 to -0.007661883

The confidence interval does not include 0 and both limits are negative, it indicates that the true mean of the ctrl group is likely less than the true mean of the trt2 group.

Sample estimates are mean of ctrl, x = 5.032 and mean of trt2, y = 5.526

Conclusion

The t-test results indicate that there is a significant difference between the mean weights of the ctrl group (mean = 5.032) and the trt2 group (mean = 5.526).

The p-value (0.04685) is less than 0.05, suggesting that the observed difference is statistically significant.

The 95% confidence interval (-0.980338117 to -0.007661883) does not include 0, supporting the conclusion that the true mean difference is likely not zero and that trt2 group’s mean is higher.

Based on this t-test, there is sufficient evidence to conclude that trt2 has a significant effect on weights compared to the ctrl group. The mean weight of the trt2 is significantly higher than that of the control group, indicating a positive effect of the treatment on weights.