

Mini-Project 3

Q1: Fuji Apples

1. Descriptive Summary

This dataset represents a sample of 24 apple trees planted in an experimental forest in Idaho in 2012. The trees were divided into three equal rows of eight. Each row was assigned one of the following treatments: “overlapping arm,” “tipped arm,” and “tall spindle.” The main question to consider from the experiment was “Do the data suggest that the branch configuration affects the fruit weights?” This dataset contains a mere three columns: *the index*, *weight.per.fruit*, and *treatment*. The weights for *overlapped_arm* had a minimum value of 141.7, a first quartile of 167.6, a median of 184.2, a mean of 181.8, a third quartile of 199.7, and a maximum of 218.3. The weights for *tipped_arm* had a minimum value of 175.6, a first quartile of 207.5, a median of 248.0, a mean of 240.0, a third quartile of 265.3, and a maximum of 291.8. The weights for *tall_spindle* had a minimum value of 184.0, a first quartile of 223.4, a median of 294.8, a mean of 276.7, a third quartile of 322.5, and a maximum of 360.5. For comparative visuals of the univariate summaries, please refer to Figures 1 and 2. Looking at the boxplots in Figure 1, there appears to be much variation between the three treatment groups. *tall_spindle*, on average, looks to have the highest mean and median weights, followed by *tipped_arm* and then *overlapped_arm*.

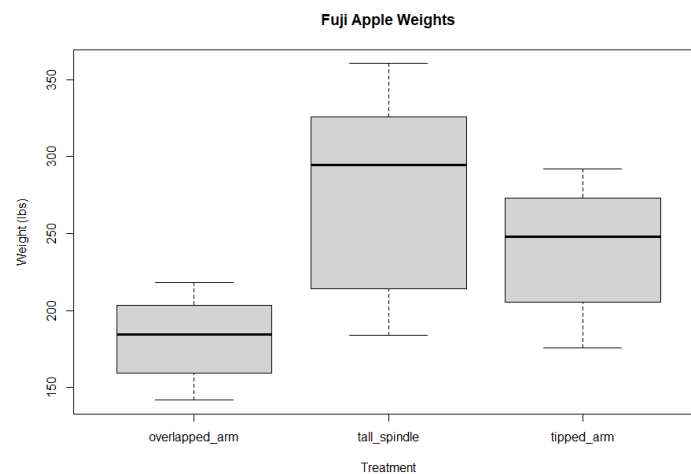


Figure 1: Box Plot of Apple Weights by Treatment Group

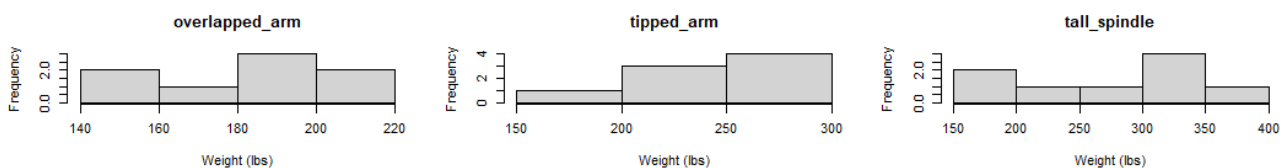


Figure 2: Histogram of Apple Weights by Treatment Group

2. Inferential Analysis

For the inferential analysis component, I decided to conduct both an analysis of variance (ANOVA) test and Tukey's honest significant difference (HSD). When creating the confidence interval for both tests, I decided to use a confidence level of 95%. For the ANOVA test in Figure 3, The null hypothesis was $H_0 : \mu_1 = \mu_2 = \mu_3$, the alternative hypothesis was that not all sample means are equal ($H_A : \mu_1 \neq \mu_2 \neq \mu_3$), the test statistic was 8.0712, and the p-value was 0.00251. Based on the results of the test, we reject the null hypothesis because the p-value of 0.00251 is less than the alpha of 0.05. There is sufficient evidence to conclude that the treatment means differ.

Now that we know the sample means between the treatment groups differ, the next question to consider is "Which treatment groups have statistically significant differences?" To answer this question, I created a Tukey's HSD to compare treatment groups to one another. Based on the results in Figure 4, the only groups which have statistically significant differences is the *tall_spindle-overlapped_arm* treatment groups. This is because the adjusted p-value of 0.0018721 is less than the alpha of 0.05.

Analysis of Variance Table

```
Response: weight.per.fruit
      Df Sum Sq Mean Sq F value Pr(>F)
treatment  2  36668  18333.9   8.0712 0.00251 **
Residuals 21   47702   2271.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: ANOVA Test

```
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = lm_apples)

$treatment
              diff            lwr            upr            p adj
tall_spindle-overlapped_arm  94.93935  34.873543 155.00515 0.0018721
tipped_arm-overlapped_arm    58.19761  -1.868195 118.26341 0.0586586
tipped_arm-tall_spindle     -36.74174 -96.807543  23.32407 0.2924339
```

Figure 4: Tukey HSD

3. Conclusion

The data does indeed suggest that branch configuration affects the fruit weights. This is based on the weight variation between the three treatment groups, the ANOVA test, and Tukey's HSD. The boxplots and histograms in Figures 5 and 6 demonstrated this great variation that exists between three treatment groups, the key takeaway being the material difference in mean and median between the groups. The ANOVA test's ruling to reject the null hypothesis also supports this conclusion, as does the statistically significant difference between the *tall_spindle* and *overlapped_arm* treatment groups. Looking at the normal quantile plot in Figure 5, the pattern of the plot is very close to the line, so it is reasonable to assume normality in the data. However, the widespread variation in the plot of residuals by treatment group in Figure 6 indicates that the assumption of equal variances does not seem to be reasonable.

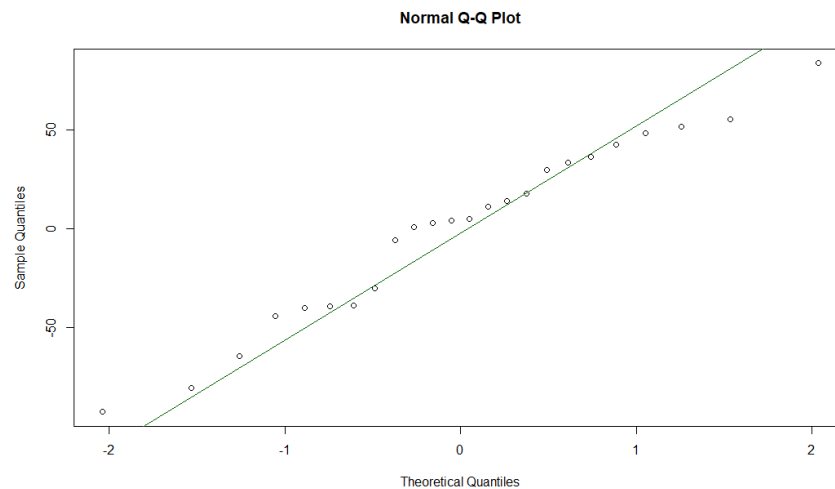


Figure 5: Normal Quantile Plot of the Data

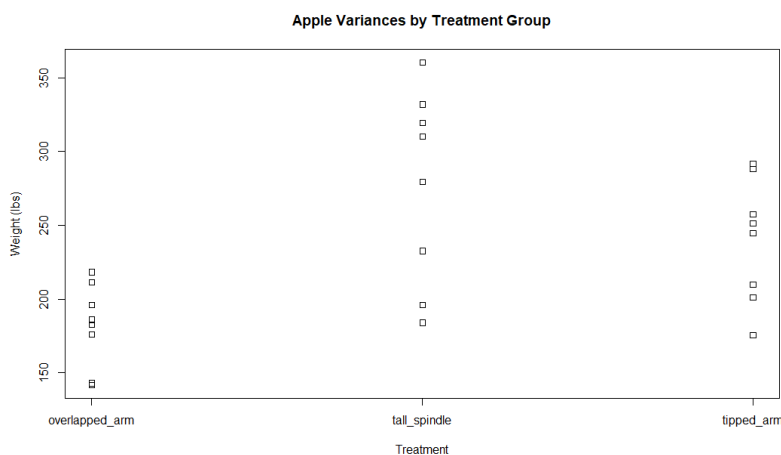


Figure 6: Residual Analysis of the Data

Q2: Instructor Evaluations

1. Descriptive Summary

This dataset contains 463 end of semester professor evaluations from the University of Texas at Austin. From this data, researchers found evidence that physical appearance is associated with higher evaluation scores. Some further questions to consider based on this data include “What factors besides the appearance-related variables are associated with score?” and “Is there evidence that the evaluation scores measures might reflect the influence of non-teaching related characteristics?” For this test of association, I decided to analyze four non-appearance-related variables. They include *cls_did_eval*, *age*, *ethnicity*, and *cls_level*. *cls_did_eval* is the number of students in that class who completed the evaluation, *age* is the age of the instructor, *ethnicity* is the ethnicity of the professor (minority or not minority), and *cls_level* is the level of the class (lower or upper). *Cls_did_eval* had a minimum value of 5, a first quartile of 15, a median of 23, a mean of 36.62, a third quartile of 40, and a maximum of 380. *age* had a minimum value of 29, a first quartile of 42, a median of 48, a mean of 48.37, a third quartile of 57, and a maximum of 73. For visuals of these summary statistics, please refer to the boxplots in Figure 9. For *ethnicity*, there were 64 minority and 399 non-minority, or 13.82% and 86.18% respectively. For *cls_level*, there were 157 lower and 306 upper, or 33.91% and 66.09% respectively. The scatterplots in Figures 10 and 11 show the *score* variable compared *cls_did_eval* and *score* and *age*. For *score* vs. *cls_did_eval*, there is a weak, positive association, with the two variables having a Pearson correlation of 0.0628. For *score* vs. *age*, there is a weak, negative association, with the two variables having a Pearson correlation of -0.1070.

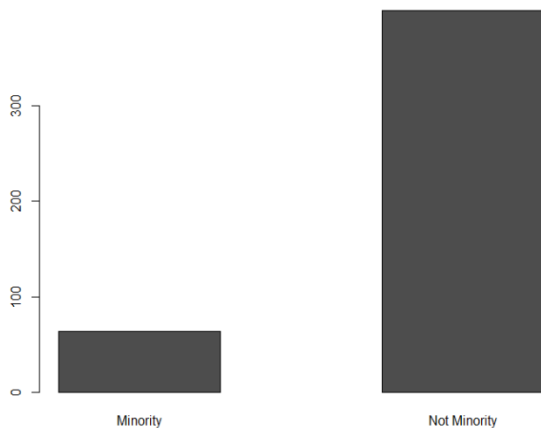


Figure 7: Distribution of Minority to Not Minority Instructors

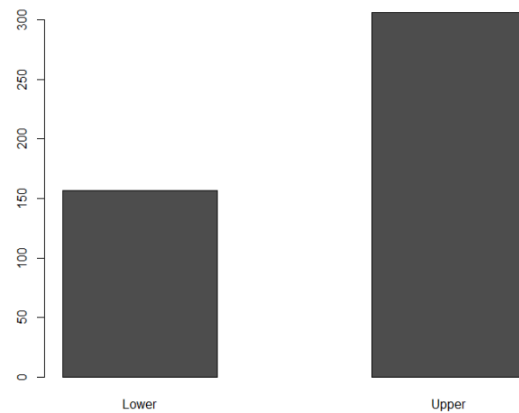


Figure 8: Distribution of Lower to Upper Class Evaluations

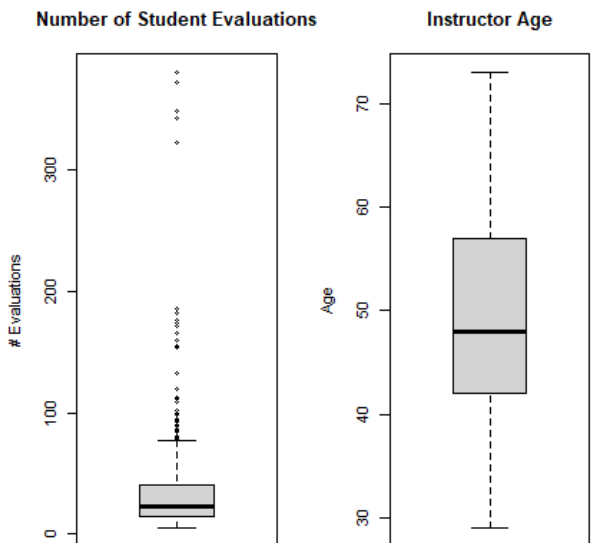


Figure 9: Boxplots of *cls_did_eval* and *age*

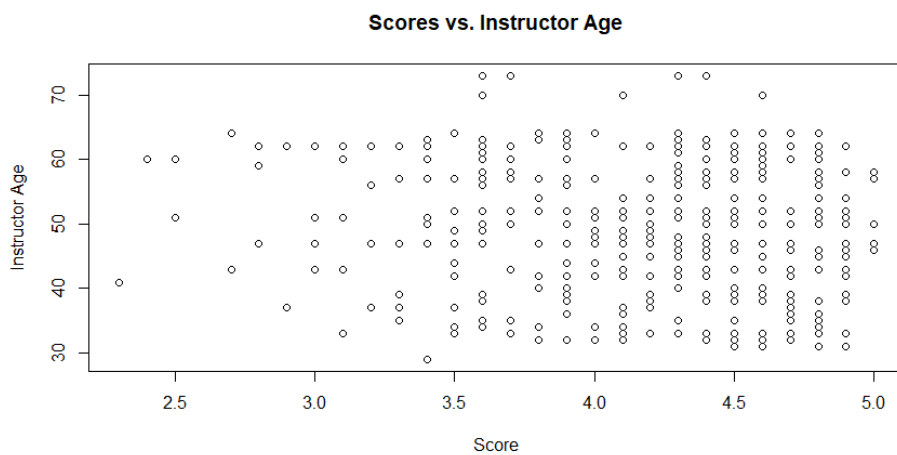


Figure 10: Scatterplot of scores and *age*

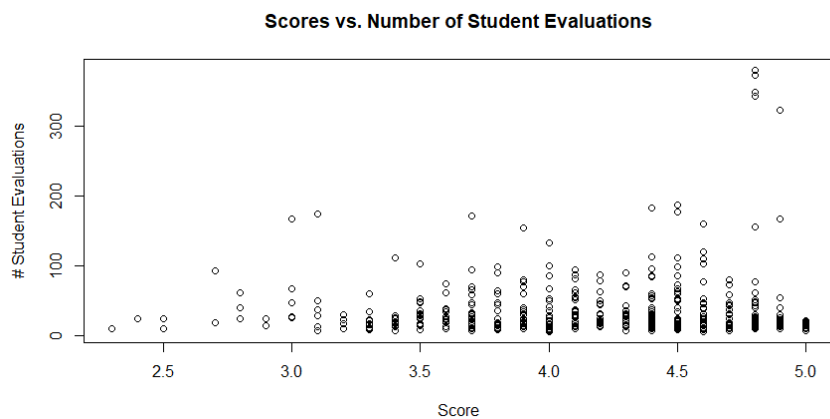


Figure 11: Scatterplot of score and *cls_did_eval*

2. Inferential Analysis

I decided to conduct a two sample t-test with the *ethnicity* and *cls_level* variables. For *ethnicity*, I divided the data into two separate samples: one for minority and one for not minority. Similarly, for *cls_level*, I also divided the data into two separate samples, this time for lower and upper. From there, I calculated the summary statistics seen in Figures 12 and 13. When conducting the hypothesis test for *ethnicity*, I decided to use a confidence level of 95%. The null hypothesis was $H_0 : \mu_1 - \mu_2 = D_0$, the alternative hypothesis was $H_A : \mu_1 - \mu_2 \neq D_0$, the test statistic was -1.54 (for unequal variances), and the p-value was 0.0637. Based on the results of the t-test, we fail to reject the null hypothesis because the p-value of 0.0637 is greater than the alpha of 0.05. There is not sufficient evidence to suggest that the mean scores between minority scores and non-minority scores differ.

When conducting the hypothesis test for *cls_level*, I decided to use a confidence level of 95%. The null hypothesis was $H_0 : \mu_1 - \mu_2 = D_0$, the alternative hypothesis was $H_A : \mu_1 - \mu_2 \neq D_0$, the test statistic was 1.7243 (for unequal variances), and the p-value was apparently 0.9571. Based on the results of the t-test, we fail to reject the null hypothesis because the p-value of 0.9571 is greater than the alpha of 0.05. There is not sufficient evidence to suggest that the mean scores between lower and upper level classes differ.

Ethnicity	mean	standard deviation	n
minority	4.0719	0.5816	64
not minority	4.1912	0.5365	399

Figure 12: Summary Statistics of ethnicity Samples

Class Level	mean	standard deviation	n
lower	4.2382	0.5925	157
upper	4.1422	0.5151	306

Figure 13: Summary Statistics of cls_level Samples

I decided to create two simple linear regressions (SLRs) with the final two variables (*cls_did_eval* and *age*). For *cls_did_eval*, I got a regression of $\text{SCORE} = 0.0007589 \times \text{CLS_DID_EVAL} + 4.1469347$ (Figure 14). The null hypothesis for this SLR was $H_0 : \beta_1 = 0$, the alternative hypothesis was $H_A : \beta_1 > 0$, the alpha was 0.01, the test statistic was 2.586536, and the p-value was 0.995. Based on these results, we fail to reject the null hypothesis. Sufficient evidence does not exist to indicate that the slope of the regression line is greater than 0.

For *age*, I got a regression of $\text{SCORE} = -0.005938 \times \text{AGE} + 4.461932$ (Figure 15). The null hypothesis for this SLR was $H_0 : \beta_1 = 0$, the alternative hypothesis was $H_A : \beta_1 < 0$, the alpha was 0.01, the test statistic was 2.586536, and the p-value was 0.005. Based on these results, we reject the null hypothesis. Sufficient evidence exists to indicate that the slope of the regression line is less than 0.

```
call:
lm(formula = score ~ cls_did_eval, data = evals)

Residuals:
    Min       1Q   Median       3Q      Max
-1.8545 -0.3595  0.1303  0.4269  0.8485

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.1469347  0.0325682  127.331  <2e-16 ***
cls_did_eval  0.0007589  0.0005616   1.351   0.177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5434 on 461 degrees of freedom
Multiple R-squared:  0.003946, Adjusted R-squared:  0.001786
F-statistic: 1.827 on 1 and 461 DF, p-value: 0.1772
```

Figure 14: SLR Using cls_did_eval

```
call:
lm(formula = score ~ age, data = evals)

Residuals:
    Min       1Q   Median       3Q      Max
-1.9185 -0.3531  0.1172  0.4172  0.8825

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.461932  0.126778  35.195  <2e-16 ***
age          -0.005938  0.002569  -2.311  0.0213 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5413 on 461 degrees of freedom
Multiple R-squared:  0.01146, Adjusted R-squared:  0.009311
F-statistic: 5.342 on 1 and 461 DF, p-value: 0.02125
```

Figure 15: SLR Using age

3. Conclusion

Based on my analysis of the four non-appearance-related variables, there was evidence to conclude that only *age* was associated with *score*. The scatterplot in Figure 10, gives an initial assessment of a weak, negative association between *score* and *age*, having a Pearson correlation of -0.1070. While this was true, it was really the SLR which provided the strongest evidence for this association. The decision to reject the null hypothesis was a result of statistical significance in the slope of the regression line being negative.

As for the other three variables, there was no evidence of association with *score*. According to the scatterplot for *cls_did_eval*, the variable had a weak, positive association with *score* and a Pearson correlation of 0.0628. The hypothesis test of the SLR using *cls_did_eval* did not support this idea of a mild association between the two variables, as there was no statistically significant evidence that the slope of the regression line was greater than zero. The two sample t-test for *cls_level* did not have any evidence to suggest that the mean scores differed between lower and upper level classes. The decision to fail to reject the null hypothesis meant that the level of the class did not have a material impact on the score that the instructor received. The two sample t-test for *ethnicity* also did not provide any evidence that the mean scores differed between instructors of minority and non-minority. Similar to *cls_level*, the decision to fail to reject the null hypothesis meant that the ethnicity of the instructor did not have a material impact on the score they received.

Looking at the normal quantile plots in Figures 16 and 17, the patterns of the plots run very close to the lines, so it is reasonable to assume normality in the data for *cls_did_eval* and *age*. However, this sentiment was not the same for the residual analysis. Looking at Figures 18 and 19, the distributions of the residuals is not random, which leads us to reject the assumption of equal variances.

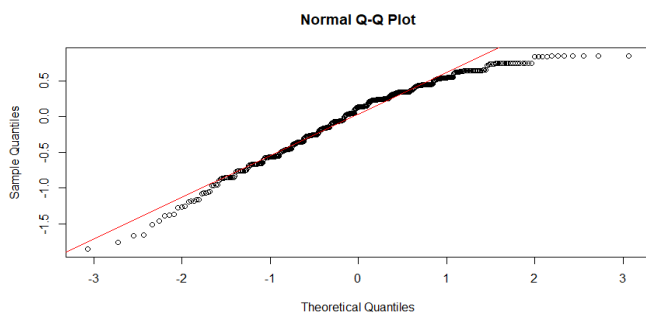


Figure 16: Normal Q-Q Plot of *cls_did_eval*

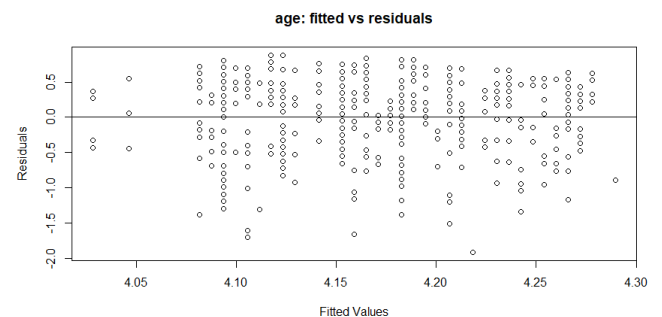


Figure 18: Residual Analysis of *cls_did_eval*

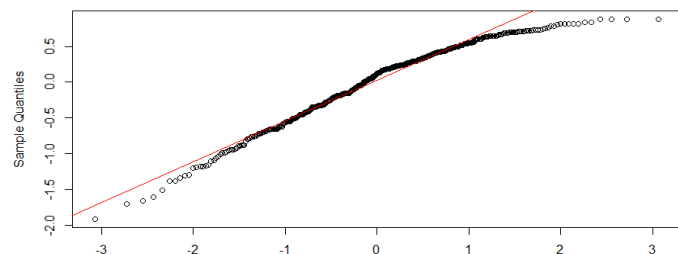


Figure 17: Normal Q-Q Plot of *age*

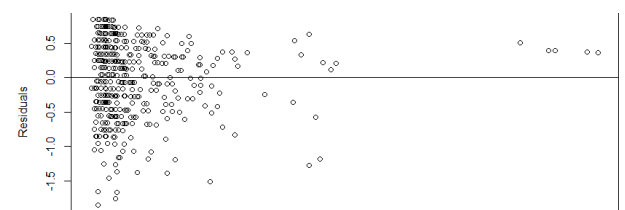


Figure 19: Residual Analysis of *age*