**2021025081 DESQUITADO, Yuichi**

**2021011911 GACASA, Ymanuel Josh**

**APM1111 Statistical Theory - Formative Assessment 8**

**Dataset and the Problem**

The dataset *PlantGrowth* from R [*library(datasets)*] talks about a study conducted about plants that were given two different treatment conditions and their weights after being dried out.

This dataset is composed of three columns, such as their V1 (indexing), independent variable (group: Control, Treatment 1, Treatment 2), and the dependent variable (weight).

From this dataset, we may investigate to see if such treatments will be significant to incur differences among the tested specimens.

As for this study, the formulated hypotheses (null and alternate), which are the following:

: All group population means are equal (i.e., )

: At least one group population mean is different (i.e., they are not all equal)

**Checking Assumptions**

**Assumption #1:** You have one dependent variable that is measured at the continuous level.

**Remark** - This dependent variable was called “*weight*” in the dataset. This dependent variable is continuous.

**Assumption #2:** You have one independent variable that consists of three categorical, independent groups.

**Remark** - The independent variable in the dataset is called “group”, which is composed of Control, Treatment 1, and Treatment 2, which refers to the variation of treatment done by the researchers.

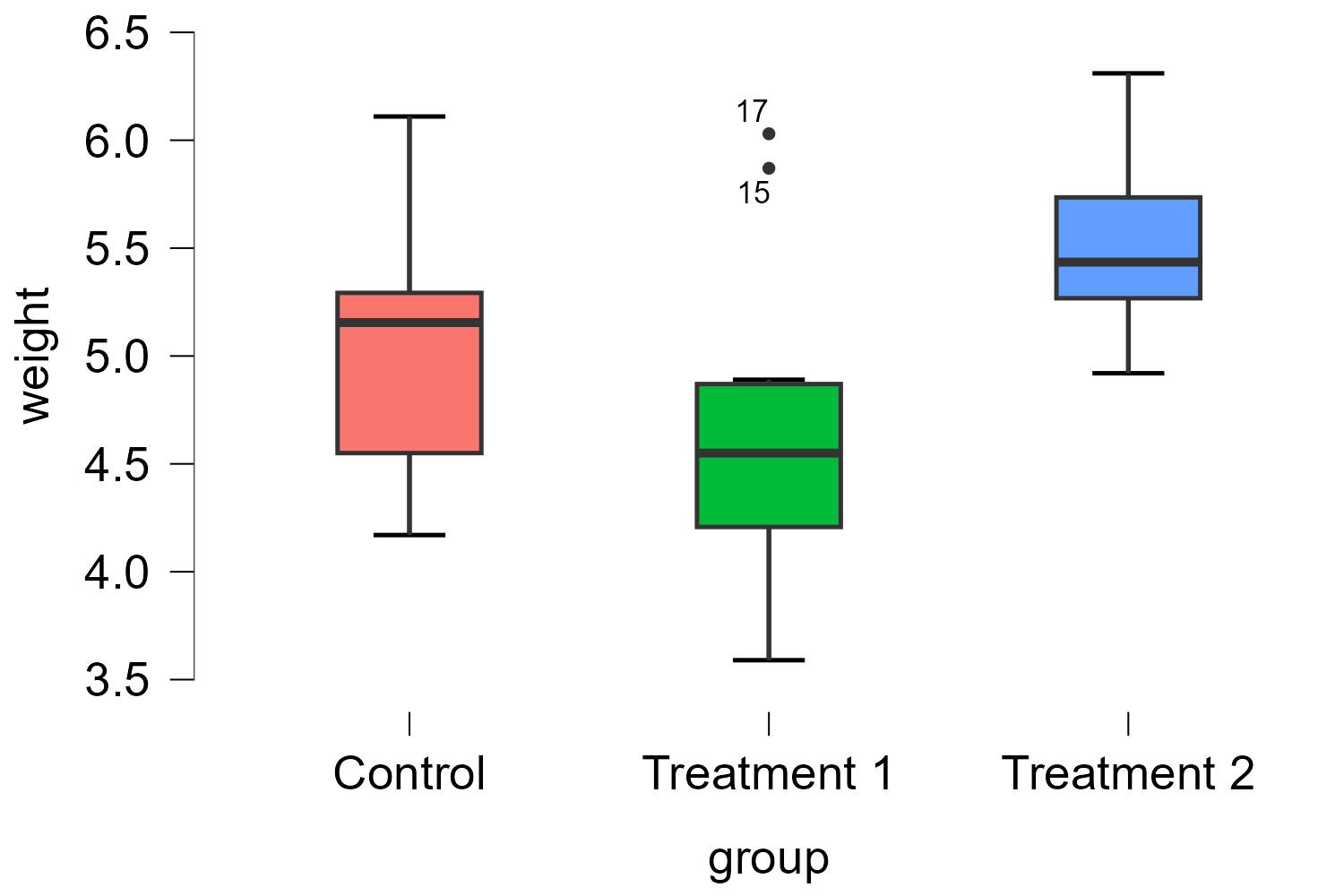
**Assumption #3:** You should have independence of observations.

**Remark** - Each observation is independent of each other as there is no relationship between the observations in each group of the observation or between the groups themselves.

**Assumption #4:** There should be no significant outliers in the three or more groups of your independent variable in terms of the dependent variable.

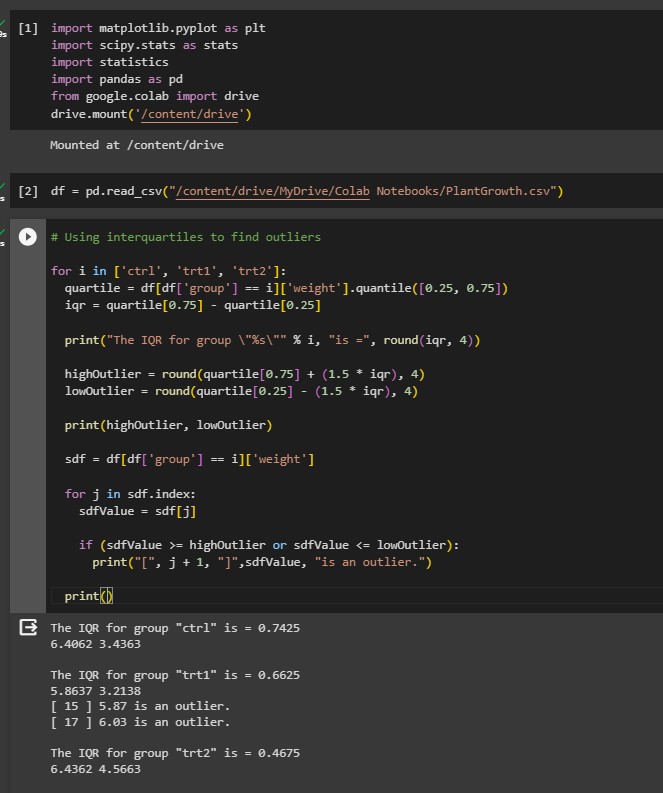
There are many ways to check for outliers. For this paper, we would be employing two ways, visual using boxplot and using interquartile.

**Boxplot**

****

*Figure 1. Boxplot of the dataset PlantGrowth*

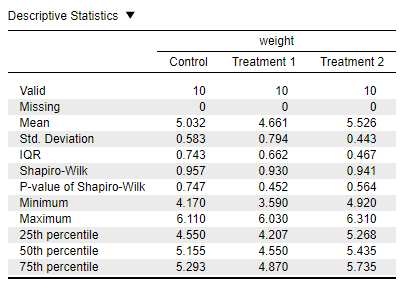
**Interquartile**

****

*Image 1. Program for identifying outliers using IQR*

**Remark** - There were significant outliers in the dataset, namely data entry 15 (*Treatment 1, 5.87*) and 17 (*Treatment 1, 6.03*) as assessed by visual inspection of boxplot and usage of interquartile. Such outliers will not be removed from the dataset in the continuation of the computation for ANOVA as they are not extreme outliers.

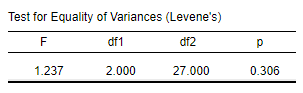
**Assumption #5:** Your dependent variable should be approximately normally distributed for each group of the independent variable.



*Table 1. Descriptive statistics of PlantGrowth dataset with Shapiro-Wilk’s Test P-Value*

**Remark** - The weights are approximately normally distributed for each of the treatment conditions, as assessed by Shapiro-Wilk’s Test, *p > 0.05*.

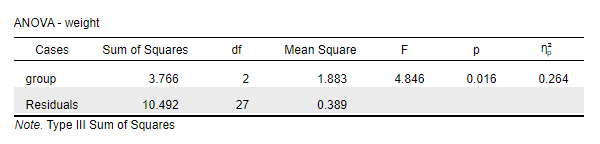
**Assumption #6:** You have homogeneity of variances (i.e., the variance of the dependent variable is equal in each group of your independent variable).



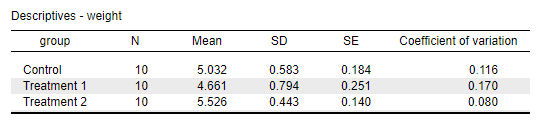
*Table 2. Levene’s Test for Equality of Variance for PlantGrowth dataset*

**Remark** - There was homogeneity of variances of the weights for all treatment conditions, as assessed by Levene’s Test for Equality of Variance, *p > 0.05*.

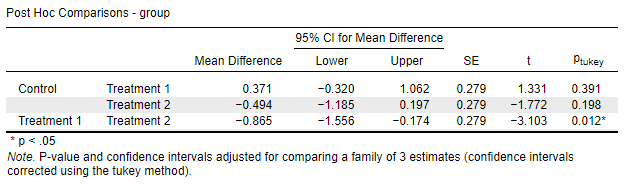
**Computation**



*Table 3. Result of ANOVA for PlantGrowth dataset*

**

*Table 4. Descriptives of the PlantGrowth dataset*

**

*Table 5. Post Hoc Comparison for PlantGrowth dataset*

**Reporting**

A one-way ANOVA was conducted to determine if there is a significant difference between treatment conditions to the weight of the plants. Specimens were divided into three groups: Control (*n = 10*), Treatment 1 (*n = 10*), and Treatment 2 (*n = 10*).

There were outliers under Treatment 1, namely (*Row 15 = 5.87*) and (*Row 17 = 6.03*). The said outliers were identified using visual inspection via boxplot and interquartile computation (*Control IQR = 0.7425*, *Treatment 1 IQR = 0.6625*, *Treatment 2 IQR = 0.4675*).

The dataset is normally distributed for each group, as assessed by Shapiro-Wilk’s Test (*p > 0.05*).

The dataset is homogeneous via Levene’s Test (*p = 0.306 > 0.05*).

Data is presented as mean standard deviation. The weights were statistically significantly different between different treatment conditions, *F(2, 27) = 4.846, p = 0.016 < 0.05, = 0.264*.

Weights decreased from Control ( *= 5.032, = 0.583*) to Treatment 1 ( *= 4.661, = 0.794*), then increased to Treatment 2 ( *= 5.526, = 0.443*), in that order.

Turkey post hoc analysis showed that the mean increased from Treatment 1 to Treatment 2 (*0.865, 95% CI [0.174, 1.556]*), meaning a statistically significant (*p = 0.012 < 0.05*).

**In totality, having be the *p-value = 0.016*, it makes the dataset statistically significantly different, which gives us a full motivation to reject the null hypothesis in favor or alternative hypothesis.**