Plant Growth dataset using R: One Way ANOVA

Assumptions

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

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

Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group of the independent variable or among 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.

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

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

Null and Alternative Hypotheses

Null hypothesis: There is no significant difference on weight between treatment groups.

Introduction

Alternative hypothesis: There is a significant difference on weight between treatment groups.

Dataset and Problem

This analysis using R aims to explore and investigate the different plant weights across the three treatment groups: crt, trt1, and trt2, in the built-in *PlantGrowth* data set. The objective is to assess whether there are any statistically significant differences in the mean plant weights among the three treatment groups. Each group would represent a type of treatment given to the plants and how it affected its weight based on the weight given after a period of time.

Assumptions:

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

Remark. In the analysis, the dependent variable is present, which is the "weight of the plants". Since this dependent variable is at continuous level, assumption 1 has been fulfilled.

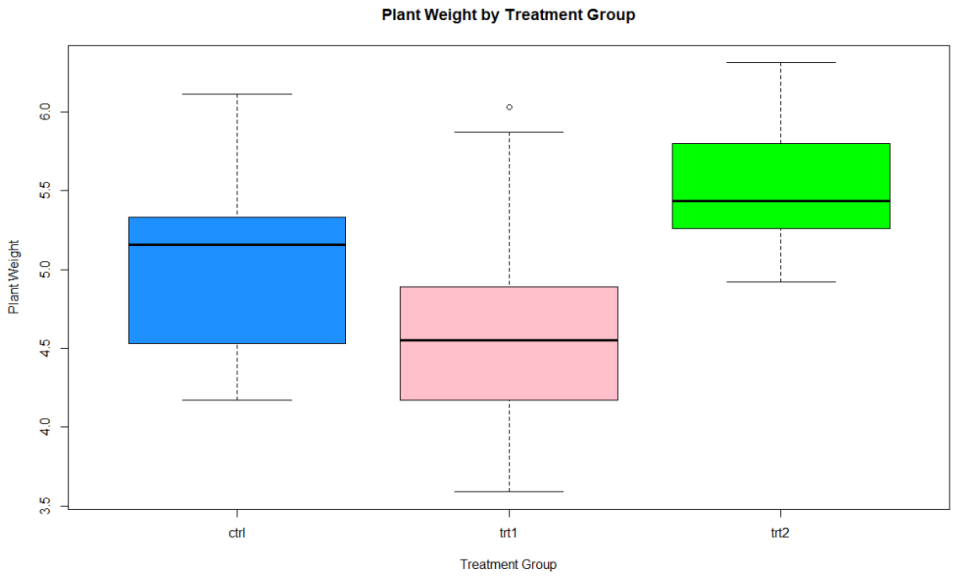
Assumption #2: You have one independent variable that consists of two identical independent groups.

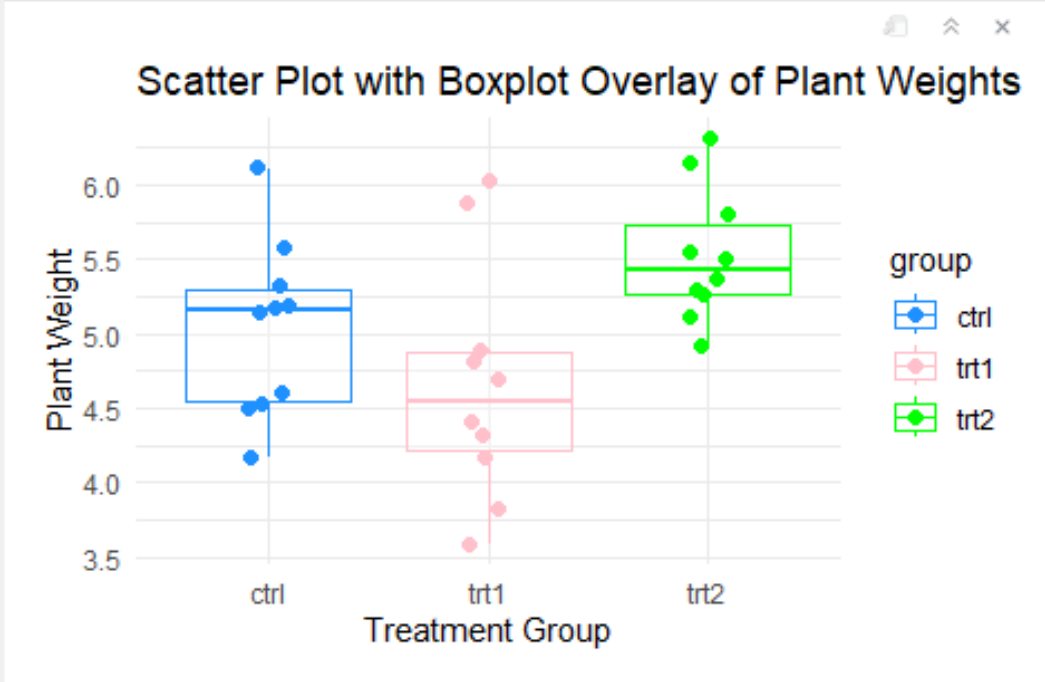
Remark. The independent variable in the analysis is the control group which consists of three independent groups: ctrl, trt1, trt2, which is the basis of the yield of the weight of plants.

Assumption #3: You should have independence of observations.

Remark. Each independent group does not correlate to each other, and there is no significant relationship between the observation of each group.

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





Remark. In the independent group: trt1, there is an outlier where two dependent variables are at relatively high weight at around 6.0, where the range of weight in group trt1 ranges at around 3.5 to 5.0. But in accordance, since this is not an extreme outlier as it still ranges near the range, we accept the assumption.

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

group <fctr>	Valid <int>	Mode <chr>	Median <dbl>	Mean <dbl>	Std. Deviation <dbl>	Variance <dbl>	Skewness <dbl>	Std. Error of Skewness <dbl>
ctrl	10	4.17	5.155	5.032	0.5830914	0.3399956	0.2311020	0.6870429
trt1	10	3.59	4.550	4.661	0.7936757	0.6299211	0.4744580	0.6870429
trt2	10	4.92	5.435	5.526	0.4425733	0.1958711	0.4847153	0.6870429

Kurtosis <dbl>	Std. Error of Kurtosis <dbl>	Minimum <dbl>	Maximum <dbl>	25th Percentile <dbl>	50th Percentile <dbl>	90th Percentile <dbl>
-1.116799	1.549193	4.17	6.11	4.5500	5.155	5.633
-1.104735	1.549193	3.59	6.03	4.2075	4.550	5.886
-1.160361	1.549193	4.92	6.31	5.2675	5.435	6.166

group <fctr>	Shapiro_Wilk_p <dbl>
ctrl	0.7474734
trt1	0.4519440
trt2	0.5642519

Remark. Since all groups have p > 0.05, ss assessed by the Shapiro-Wilk's normality test, the weight of the plants is is approximately normally distributed.

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

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  1.1192 0.3412
      27
```

Remark. The p-value at 0.3412 shows that there is a homogeneity of variance for all independent groups, as assessed by Levene's test of homogeneity of variances.

Computations:

```
            Df Sum Sq Mean Sq F value Pr(>F)
group        2  3.766  1.8832   4.846 0.0159 *
Residuals   27 10.492  0.3886
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Reporting.

A one-way ANOVA was conducted to determine if the weight of the plants' changes based on the respective independent groups: ctrl, trt1, trt2. As assessed by both the box plot and scatterplot, there were no extreme outliers in the three independent groups. In addition, data was also found out to be normally distributed through the Shapiro-Wilk normality (p > .05); and there exists a homogeneity of variance through Levene's test of homogeneity. Since the p-value (0.0159) is less than the standard significance level of 0.05, it indicates that there is a **statistically significant difference** in plant weights among the treatment groups.