Plant Growth

Applied Biostatistics – First Assignment

Juraj Korcek Lucia Montero Sanchis

1 Introduction

The "PlantGrowth" dataset consists of results collected in an experiment to compare crop yields obtained under a control and two different treatment conditions. The yields were measured in the dried weight of plants.

We carry out an exploratory analysis and apply Analysis of Variance (ANOVA) on this data to determine whether the treatments had a significant effect, by verifying if any of the group weights has a significantly different mean.

2 Exploratory Data Analysis

This dataset contains 30 observations for the two variables:

- **Weight**: represents the dried weight of plants. It is a numerical variable, continuous and positive.
- **Group**: categorical, nominal variable. There are three categories: Control (ctrl), treatment-1 (trt1) and treatment-2 (trt2). There are 10 observations per category.

Figure 1 shows the combined box plot and dot plot for each Group variable category. We can observe an outlier in the treatment-1 group.

3 Comparison of means

As we want to compare means of three groups of one factor, the one-way ANOVA is a reasonable approach, because it is capable of comparing several means at once.

3.1 Analysis of Variance (ANOVA)

The null hypothesis (H) claims that all the means that are being compared are equal. If H is rejected, the alternative hypothesis (A) means that at least one mean is significantly different:

$$H : \mu_{\text{ctrl}} = \mu_{\text{trt1}} = \mu_{\text{trt2}} \qquad A : \exists \mu_{\text{i}} \neq \mu_{\text{j}}$$

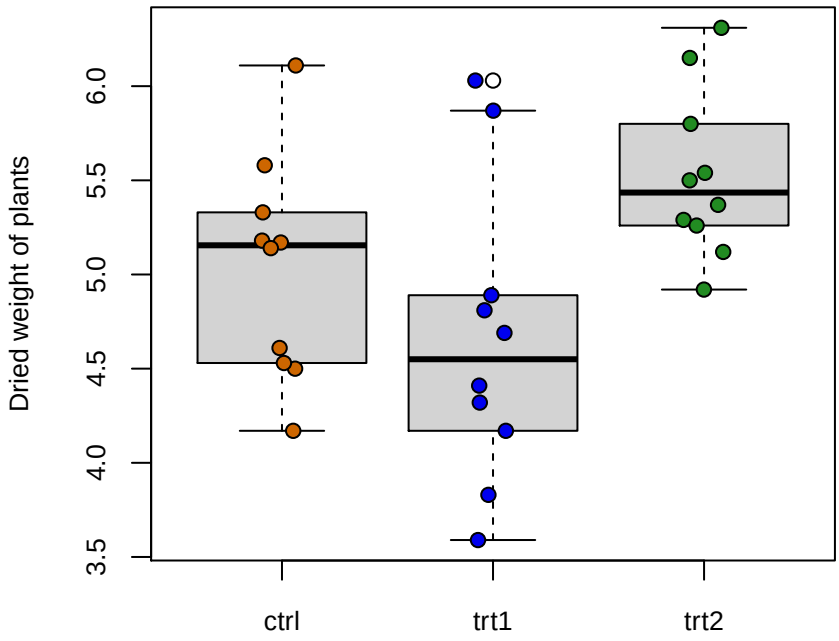


Figure 1: Box plot combined with dot plot for each Group category

If H is rejected, ANOVA does not specify which means are different. Therefore, an additional analysis would be necessary in such a case.

3.1.1 Assumptions

ANOVA makes the following assumptions about the probability distribution of the observations:

- **Normality** – The first assumption is that the data is normally distributed within each group or category. In our case we do not have a sufficiently large number of observations and therefore we need to verify this for the quantitative variable Weight.

  We can do this by looking at the Normal Q-Q plot for the residuals (Figure 2) and the boxplots (Figure 1). In the Q-Q plot there is a curvature at one of the ends which corresponds to a long tail. In spite of this, we can still observe that the residuals are mostly normally distributed.

  In the boxplots graph we can see that the boxplot for the control group is not exactly symmetric, suggesting that the data in this group is not completely normally distributed, whereas for the two treatment groups they seem mostly symmetric and therefore normally distributed.

- **Homoscedasticity** – We then need to verify that the variances in each population are equal (*homogeneity of variance*). In Figure 2a, we see that the variability of the residuals is similar if we do not consider the outliers. We can also see in the boxplots in Figure 1 that the variability is similar for the control and the treatment-1 groups, while it is slightly smaller for the treatment-2 group. For heteroscedastic data there are alternative approaches available such as ANOVA with Welch correction;

however, we use basic ANOVA because if it was not for the outliers the data would be homoscedastic.

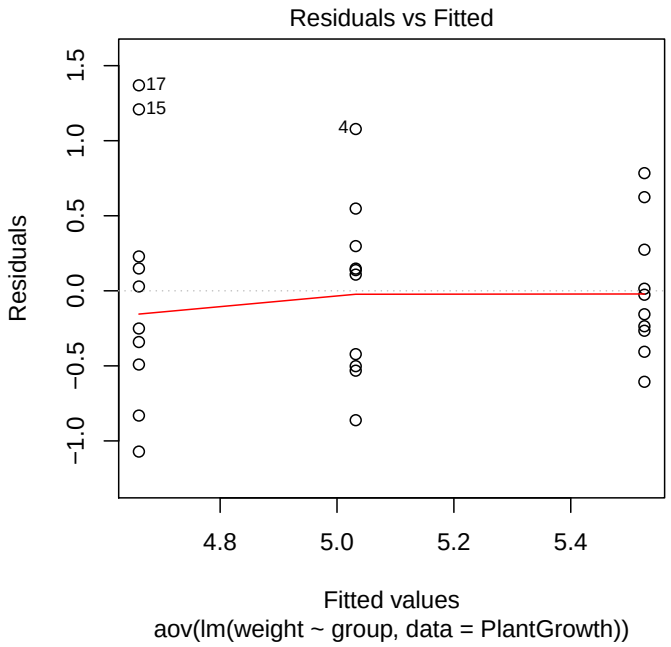
- **Independence** – We do not observe anything that suggests lack of independence in the observations. We assume randomization of plants to treatments, meaning that the plants were randomly assigned to one of the three groups.

3.1.2 Result

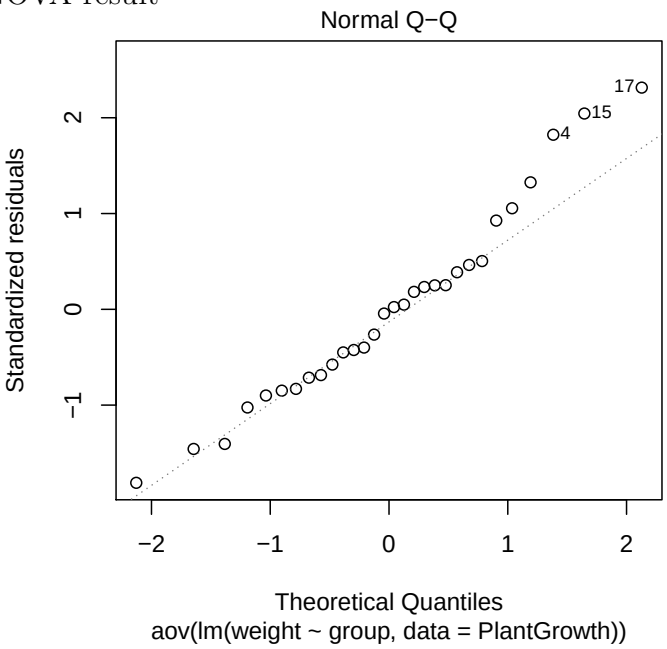
In Table 1 one can see that the p-value obtained is 0.0159. Since this p-value is smaller than 0.05, for significance level $\alpha = 0.05$ we reject the null hypothesis H that all the means are equal. According to the alternative hypothesis A, there is at least one significantly differing group mean. The next section describes how we find out which one (or which ones).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	3.77	1.88	4.85	0.0159
Residuals	27	10.49	0.39		

Table 1: ANOVA result



(a) Per-group variance of residuals



(b) Residuals distribution compared to normal distribution

Figure 2: Statistics of ANOVA residuals

3.2 Tukey's Honest Significant Difference test

To determine which group or groups have a significantly different mean, we need to carry out pairwise comparisons. For this purpose, we use Tukey's Honest Significant Difference test. It computes, for every possible pair of treatment groups, the 95% confidence interval for the difference between each pair of means ($\mu_i - \mu_j$) and the adjusted p-values $p\,adj$. The output is shown in Table 2. We can see that the only significant difference between groups

is for treatment 1 and treatment 2, since this is the only adjusted p-value smaller than 0.05. These results are plotted in Figure 3. If the confidence interval for a difference of means does not include the 0, it means that the two means being compared are significantly different at given confidence level. According to the plot the only statistically significant difference at 95% confidence level is between the treatment-1 and treatment-2 groups, since this is the only interval not including the 0.

	diff	lwr	upr	p adj
trt1-ctrl	-0.37	-1.06	0.32	0.39
trt2-ctrl	0.49	-0.20	1.19	0.20
trt2-trt1	0.86	0.17	1.56	0.01

Table 2: Tukey's Honest Significant Difference for the comparison of Group categories

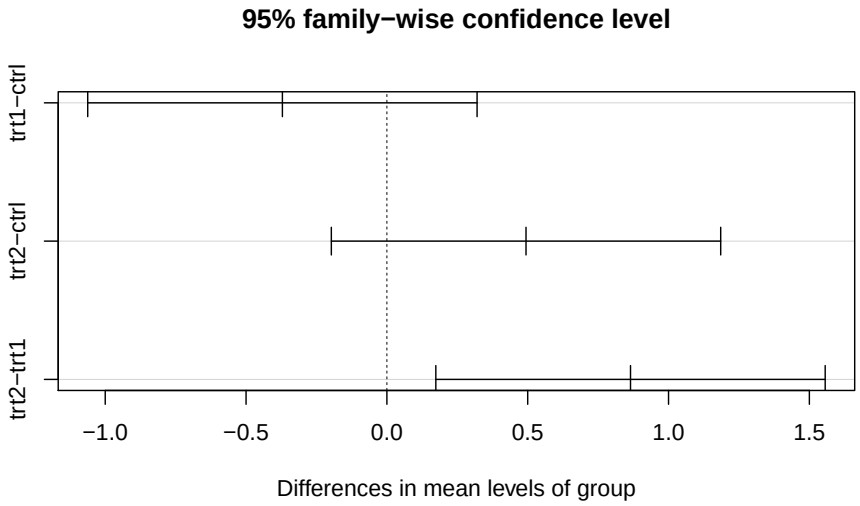


Figure 3: Tukey's Honest Significant Difference

4 Conclusions

Based on the analysis of the data, we conclude that treatments 1 and 2 are significantly different at 5% level and treatment 2 seems to be better than treatment 1. We come to this conclusion based on the result obtained from Tukey's Honest Significant Difference test, where we see that the mean of the treatment-2 group is significantly larger than the mean of the treatment-1 group. However, neither of the treatments is significantly different from the control group. Therefore, we conclude that neither of the treatments had a significant effect on the weight of plants.