Lima, Paul Carlos I T.

Rascano, Jan Celine B.

Formative Assessment #8

APM1111 - Statistical Theory

INTRODUCTION

Using the PlantGrowth dataset, the aim of this study is to determine if there is a significant difference in the weight of plants among different treatment groups using the PlantGrowth dataset.

METHODS

For this investigation, the PlantGrowth dataset—which can be found in the R datasets package—was used. The variable of interest in the dataset is the weight of the plants, and there are three treatment groups: "ctrl," "trt1," and "trt2."

To find out if there are any statistically significant variations in plant weights between the treatment groups, a one-way analysis of variance (ANOVA) was performed. The idea that there is no discernible weight difference between treatment groups is the null hypothesis that is being investigated.

ASSUMPTIONS

Assumption #1

The one dependent variable is measured at the continuous level.

The dependent variable is called “weight”. This dependent variable is continuous level.

Assumption #2

The one independent variable consists of three categorical, independent groups.

The independent variable is a group which is categorized into independent groups: ‘ctrl’, ‘trt1’, and ‘trt2’.

Assumption #3

Normality: There should be an approximate normal distribution for the residuals, which are the discrepancies between the observed and predicted values. In the case of smaller sample numbers, this assumption is particularly crucial. The dependent variable (weight) is normally distributed on all independent variables (groups).

Descriptive Statistics

	weight		
	ctrl	trt1	trt2
Valid	10	10	10
Missing	0	0	0
Mean	5.03	4.66	5.53
Std. Deviation	0.58	0.79	0.44
Skewness	0.32	0.66	0.67
Std. Error of Skewness	0.69	0.69	0.69
Kurtosis	−0.23	−0.20	−0.32
Std. Error of Kurtosis	1.33	1.33	1.33
Shapiro-Wilk	0.96	0.93	0.94
P-value of Shapiro-Wilk	0.75	0.45	0.56
Minimum	4.17	3.59	4.92
Maximum	6.11	6.03	6.31

Assumption #4

Homogeneity of Variances: At every level of the independent variable, the residuals' variances should be about identical. This implies that for any group, the distribution of data points should be about constant. The variance of the dependent variables (weight) are equal in each group of the independent variables (group). Thus we have homogeneity of variances.

Test for Equality of Variances (Levene's)

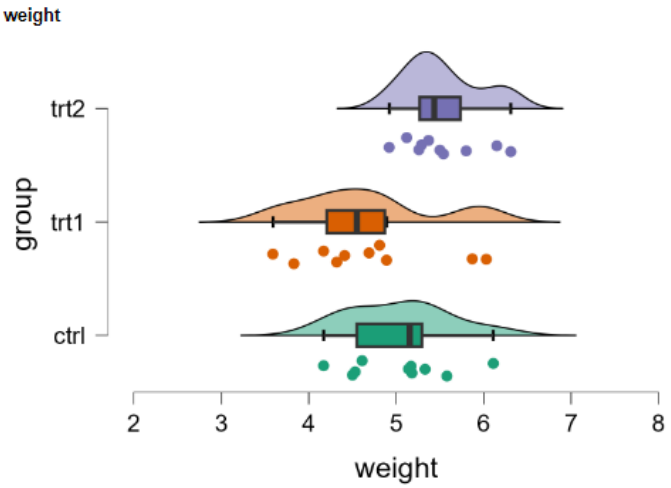
F	df1	df2	p
1.24	2.00	27.00	0.31

Assumption #5

Independence of Observations: Every group's observations ought to stand alone from one another. This implies that the dependent variable values for one group shouldn't be affected by the dependent variable values for other groups. Thus based on the results there is independence of observations.

Assumption #6

Absence of Significant Outliers: The outcomes shouldn't be significantly impacted by any notable outliers. Outliers have the potential to change the mean and have an impact on the variance homogeneity and normality assumptions. Thus, there are no significant outliers in the three groups of the independent variable (group) in terms of the dependent variable (weight).



NULL AND ALTERNATIVE HYPOTHESIS

For Null Hypothesis:

The mean plant weights in each treatment group do not differ significantly. This can be stated mathematically as:

$H_0: \mu_{ctrl} = \mu_{trt1} = \mu_{trt2}$

where μ_{ctrl}, μ_{trt1}, μ_{trt2} represent the population means of plant weights for the "ctrl," "trt1," and "trt2" treatment groups, respectively.

For Alternative Hypothesis:

Between two or more of the treatment groups, there is a notable variation in the average plant weights. This can be stated mathematically as:

H_1: At least one pair of means is different

This alternate hypothesis encapsulates the notion that there exists diversity in plant weights among the treatment groups within the framework of the PlantGrowth dataset.

In conclusion, the alternative hypothesis proposes that there is a difference in at least one pair of treatment groups, whereas the null hypothesis asserts that there is no difference in mean plant weights. We can ascertain whether there is sufficient evidence to reject the null hypothesis in favor of the alternative by using the one-way ANOVA test.

COMPUTATION

ANOVA - weight

Cases	Sum of Squares	df	Mean Square	F	p	η^2	η_p^2
group	3.77	2	1.88	4.85	0.02	0.26	0.26
Residuals	10.49	27	0.39				

Note. Type III Sum of Squares

Descriptives

Descriptives - weight

group	N	Mean	SD	SE	Coefficient of variation
ctrl	10	5.03	0.58	0.18	0.12
trt1	10	4.66	0.79	0.25	0.17
trt2	10	5.53	0.44	0.14	0.08

Post Hoc Comparisons - group

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
ctrl	trt1	0.37	−0.32	1.06	0.28	1.33	0.39
	trt2	−0.49	−1.19	0.20	0.28	−1.77	0.20
trt1	trt2	−0.86	−1.56	−0.17	0.28	−3.10	0.01

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

REPORTING/CONCLUSION

Treatment group main effects on plant weights ($F(2,27)=4.846$, $p=0.0159$) were found to be statistically significant by using a one-way ANOVA on the PlantGrowth dataset. This suggests that the mean plant weights of the treatment groups varied from one another. The mean weights for the "ctrl," "trt1," and "trt2" groups were 5.073, 4.661, and 5.527, respectively, according to descriptive statistics. The average plant weights for each treatment are shown by these values.

Comprehending these distinctions holds pragmatic consequences for uses like farming or test design. It implies that the treatment selected may have an effect on plant growth, offering important information for making decisions. Although the research yields valuable insights, it is imperative to recognize its limits, namely the very limited sample size. The results may only be applicable to the particular circumstances in which the study was carried out. To sum up, the analysis advances our knowledge of how various treatments affect plant growth. Larger sample sizes or alternative experimental configurations could be taken into account in future studies to strengthen the validity and relevance of these results.