1B ANOVA

Afundar, Audrie Lex L.

2022018211

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

The null hypothesis for a one-way analysis of variance is:

H_0: There is no significant difference on weight between treatment groups

H_A: There is a significant difference in weight between at least one pair of treatment groups

Dataset and problem

The PlantGrowth dataset from R is a dataset that calculates the weight of the plants to 3 different treatment groups: control, treatment 1 and treatment 2. The problem is would there be a correlation between the treatment on the plants in terms of weight or would there be no significant difference between them.

Checking of Assumptions:

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

Remark. The PlantGrowth dataset has a dependent variable called weight. This represents the weight of the plants and is also at the continuous level.

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

Remark. The PlantGrowth dataset has 3 independent variables: ctrl, trt1, trt2. These represents the 3 types of treatment done in the plants

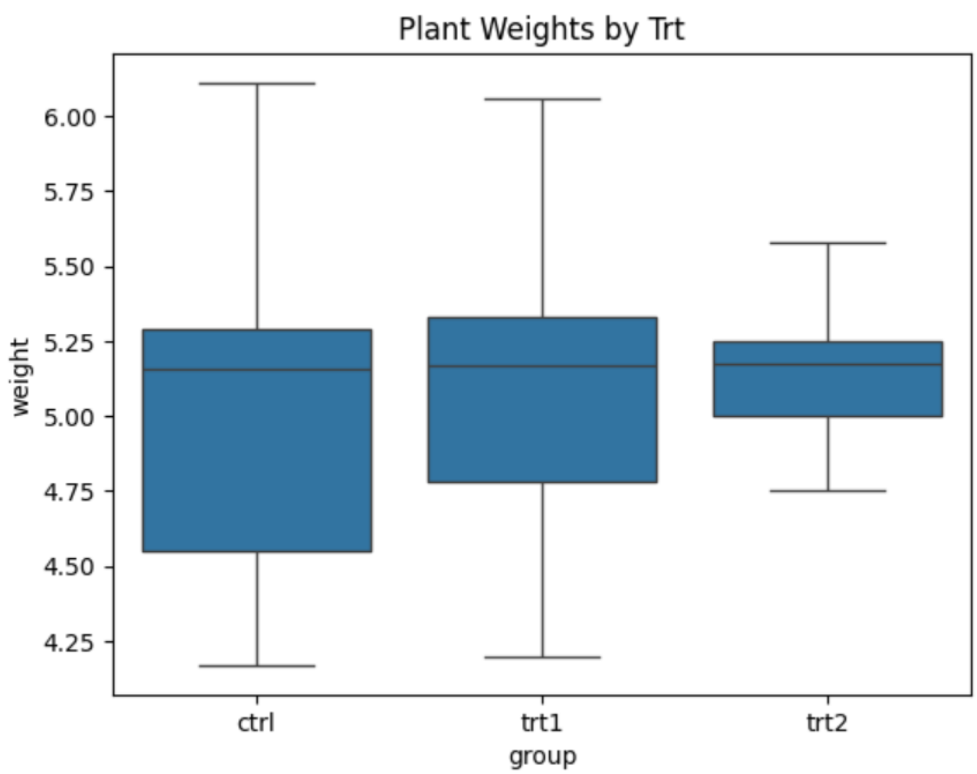
Assumption #3: You should have independence of observations, which means that there is no relationship between the observations in each group of the independent variable or among the groups themselves.

Remark. Each observation is independent to each other as there are no relationships between the 3 independent variables.

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 4

```
[8]: sns.boxplot(x='group', y='weight', data=plant_growth_df)
plt.title('Plant Weights by Trt')
plt.show()
```



Remark. There are no significant outliers between the 3 types of treatment.

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

```
Descriptive Statistics:
```

	ctrl	trt1	trt2
Valid	10	10	10
Mode	[4.17, 4.5, 4.53, 4.61, 5.14, 5.17, 5.18, 5.33...	[4.2, 4.32, 4.75, 4.88, 5.16, 5.18, 5.27, 5.35...	[5.0]
Median	5.155	5.17	5.175
Mean	5.032	5.065	5.144
Std. Deviation	0.583091	0.552213	0.231862
Variance	0.339996	0.304939	0.05376
Skewness	0.320975	-0.006273	0.184911
Std. Error of Skewness	0.687043	0.687043	0.687043
Kurtosis	-0.229125	0.114014	0.473919
Std. Error of Kurtosis	1.549193	1.549193	1.549193
Minimum	4.17	4.2	4.75
Maximum	6.11	6.06	5.58
25th Percentile	4.55	4.7825	5.0
50th Percentile	5.155	5.17	5.175
90th Percentile	5.633	5.538	5.346

```
Normality Test Results (Shapiro-Wilk p-values):
ctrl: p-value = 0.7475
trt1: p-value = 0.8235
trt2: p-value = 0.8985
```

Remark. The dataset is normally distributed based on the Shapiro-Wilk test on the 3 treatment groups, $p > 0.05$.

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

```
stat, p_value = stats.levene(
    plant_growth_df[plant_growth_df['group'] == 'ctrl']['weight'],
    plant_growth_df[plant_growth_df['group'] == 'trt1']['weight'],
    plant_growth_df[plant_growth_df['group'] == 'trt2']['weight']
)

print("\nHomogeneity of Variances (Levene's test):")
print(f"Levene Statistic: {stat:.4f}, p-value: {p_value:.4f}")
```

```
Homogeneity of Variances (Levene's test):
Levene Statistic: 2.0301, p-value: 0.1509
```

Remark. There was homogeneity of variances of the dependent variable for all treatment type, as assessed by Levene's test of homogeneity of variances, p = 0.1509.

Computations:

Computations

```
anova_results = stats.f_oneway(
    plant_growth_df[plant_growth_df['group'] == 'ctrl']['weight'],
    plant_growth_df[plant_growth_df['group'] == 'trt1']['weight'],
    plant_growth_df[plant_growth_df['group'] == 'trt2']['weight']
)

print("\nOne-Way ANOVA Results:")
print(f"F-statistic: {anova_results.statistic:.4f}, p-value: {anova_results.pvalue:.4f}")
```

```
One-Way ANOVA Results:
F-statistic: 0.1422, p-value: 0.8681
```

Reporting

A one-way ANOVA was conducted to determine if there is significant difference between the weights in 3 different treatments. Participants were classified into three groups: control, treatment1 and treatment 2. There were no outliers, as assessed by visual inspection of boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .120$). Since One-Way ANOVA resulted into $p>.05$, post hoc is no longer necessary. Therefore, there is no significant difference between the weights in 3 different treatments as proven by the One-Way ANOVA ($p >.05$) [Failure to reject the null hypothesis].