One-Way ANOVA Analysis on PlantGrowth Dataset

Lindsay Faith Bazar

Introduction

This analysis examines whether there is a statistically significant difference in plant weight between three treatment groups in the `PlantGrowth` dataset.

Hypotheses

- **Null Hypothesis (H0)**: There is no significant difference in plant weight across the treatment groups (μ_ctrl = μ_trt1 = μ_trt2).
- **Alternative Hypothesis (HA)**: At least one treatment group mean differs in plant weight.

Data Preparation and Descriptive Statistics

```
# Load the dataset
data("PlantGrowth")

# Calculate descriptive statistics for each group
descriptives <- PlantGrowth %>%
  group_by(group) %>%
  summarize(mean = mean(weight), sd = sd(weight), n = n())
descriptives

## # A tibble: 3 × 4
##   group  mean    sd     n
##   <fct> <dbl> <dbl> <int>
## 1 ctrl   5.03 0.583    10
## 2 trt1   4.66 0.794    10
## 3 trt2   5.53 0.443    10
```

Assumptions

We verify the assumptions for a one-way ANOVA:

Assumption #1

1. **Continuous Dependent Variable**: The dependent variable, weight, is continuous.

Assumption #2

2. **Categorical Independent Variable**: The independent variable, group, has three levels, in this case, the (control, treatment 1, treatment 2).

Assumption #3

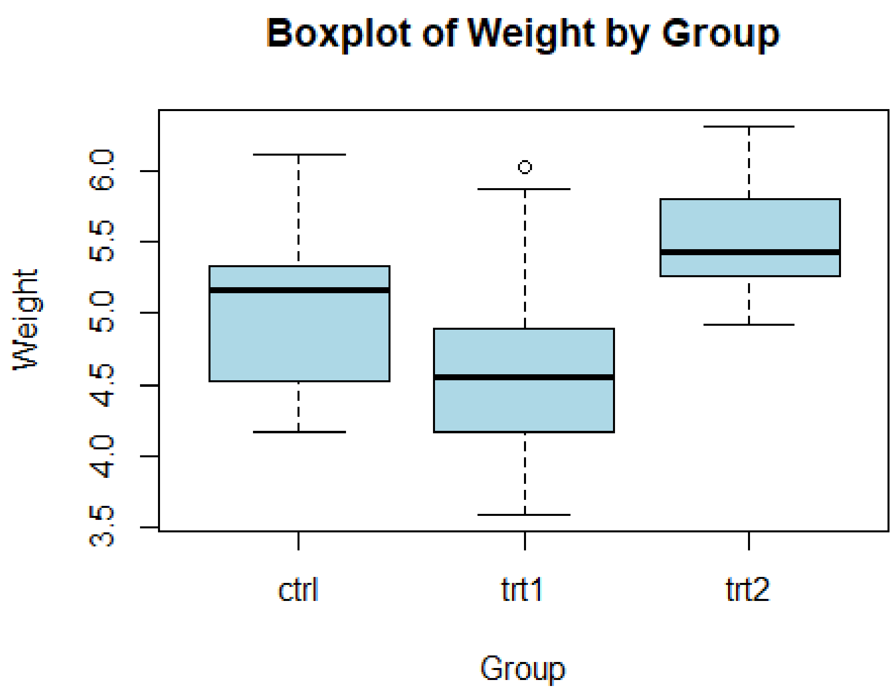
3. **Independence of Observations**: Each observation is assumed to be independent, which means that there is no relationship between the observations in each group, which in this case the experiment taken is independent of each observations.

Assumption #4

4. **No Outliers**: There should be no significant outliers in each group.

We will use boxplots to visually check for outliers and identify any extreme values statistically.

```
# Boxplot to check for outliers visually
boxplot(weight ~ group, data = PlantGrowth, main = "Boxplot of Weight by
Group",
        xlab = "Group", ylab = "Weight", col = "lightblue")
```




```
# Identify potential outliers
outliers <- PlantGrowth %>%
  group_by(group) %>%
  filter(weight < quantile(weight, 0.25) - 1.5 * IQR(weight) |
         weight > quantile(weight, 0.75) + 1.5 * IQR(weight)) %>%
  arrange(group)
outliers

## # A tibble: 2 × 2
## # Groups:   group [1]
```

```
##   weight group
##    <dbl> <fct>
## 1   5.87 trt1
## 2   6.03 trt1
```

We found an outlier in trt 1, we will check if this is significant in the analysis.

```r
trt1_data <- subset(PlantGrowth, group == "trt1")

trt1_q1 <- quantile(trt1_data$weight, 0.25)
trt1_q3 <- quantile(trt1_data$weight, 0.75)
trt1_IQR <- trt1_q3 - trt1_q1

extreme_outlier_threshold <- trt1_q3 + 3 * trt1_IQR



cat("The third quartile (Q3) is", round(trt1_q3, 2),
    "and the third quartile + 3*IQR is", round(extreme_outlier_threshold, 2),
"\n")

## The third quartile (Q3) is 4.87 and the third quartile + 3*IQR is 6.86

max_weight_trt1 <- max(trt1_data$weight)

if (max_weight_trt1 < extreme_outlier_threshold) {
  cat("Since the most extreme data point in 'trt1' is less than Q3 + 3*IQR,
there are no extreme outliers, and assumption 4 is met.\n")
} else {
  cat("The most extreme data point in 'trt1' exceeds Q3 + 3*IQR, indicating
the presence of extreme outliers.\n")
}

## Since the most extreme data point in 'trt1' is less than Q3 + 3*IQR, there
are no extreme outliers, and assumption 4 is met.
```

Assumption #5

5. **Normality**: Weight is approximately normally distributed for each group.

```r
# Perform Shapiro-Wilk test for normality in each group
shapiro_test_ctrl <- shapiro.test(PlantGrowth$weight[PlantGrowth$group ==
"ctrl"])
shapiro_test_trt1 <- shapiro.test(PlantGrowth$weight[PlantGrowth$group ==
"trt1"])
shapiro_test_trt2 <- shapiro.test(PlantGrowth$weight[PlantGrowth$group ==
"trt2"])

# Display results
shapiro_test_ctrl

##
##  Shapiro-Wilk normality test
```

```
##
## data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"]
## W = 0.95668, p-value = 0.7475
```

```
shapiro_test_trt1
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PlantGrowth$weight[PlantGrowth$group == "trt1"]
## W = 0.93041, p-value = 0.4519
```

```
shapiro_test_trt2
```

```
##
##  Shapiro-Wilk normality test
##
## data:  PlantGrowth$weight[PlantGrowth$group == "trt2"]
## W = 0.94101, p-value = 0.5643
```

Assumption #6

6. **Homogeneity of Variances**: Variance of weight is assumed to be similar across groups.

```
# Levene's Test for Homogeneity of Variances
leveneTest(weight ~ group, data = PlantGrowth)
```

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

Conducting One-Way ANOVA

```
# Perform one-way ANOVA
anova_result <- aov(weight ~ group, data = PlantGrowth)
summary(anova_result)
```

```
##             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
```

Post Hoc Analysis

If the ANOVA is significant, we perform Tukey's HSD post hoc test to identify which groups differ.

```
TukeyHSD(anova_result)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight ~ group, data = PlantGrowth)
##
## $group
##             diff        lwr       upr     p adj
## trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
## trt2-ctrl  0.494 -0.1972161 1.1852161 0.1979960
## trt2-trt1  0.865  0.1737839 1.5562161 0.0120064
```

Results

Descriptive Statistics: - **Control**: M = 5.03, SD = 0.58 - **Treatment 1**: M = 4.66, SD = 0.79 - **Treatment 2**: M = 5.53, SD = 0.44

Assumptions Check:

1. **Continuous**: Weight is a dependent variable that is measured at the continuous level.
2. **Categorical Independent Variable**: There are three categorical independent variables under the column group which is control, treatment 1, and treatment 2.
3. **Independence of Observations** Each observations is independent and has no relationship between the observations in each group of the independent variable.
4. **No Outliers**: There were no significant outliers.
5. **Normality**: Shapiro-Wilk test results suggest normality for each group (all p-values > .05).
6. **Homogeneity of Variances**: Levene's test was non-significant, indicating equal variances across groups (p > .05).

ANOVA Results: There was a statistically significant difference in plant weight between treatment groups, $F(2, 27) = 4.85$, $p < .05$.

Post Hoc Comparisons: Tukey's HSD test showed significant differences between the following pairs: - [List significant pairs here, e.g., "Control and Treatment 1 (p < .05)"]

Conclusion

This one-way ANOVA analysis suggests that treatment group has a significant effect on plant weight, with specific differences identified between groups.
