

FA8 DELA ROSA, R

Github Link: https://github.com/rddelarosa/APM1111/blob/main/FA8_Stats/FA8_DELA-ROSA%2C-R.Rmd

Roland Dela Rosa

November 26, 2025

This report performs a complete data analysis for the PlantGrowth dataset using a one-way ANOVA (1B).

Null Hypothesis (H_0): There is no significant difference on weight between treatment groups ($\mu_{ctrl} = \mu_{trt1} = \mu_{trt2}$).

Alternative Hypothesis (H_a): At least one group mean is different.

Dataset and Problem

Dependent Variable: weight (continuous)

Independent Variable: group (categorical: ctrl, trt1, trt2)

Checking of Assumptions

Assumption 1-3: Study Design

The dependent variable is continuous, the independent variable consists of categorical groups, and observations are independent.

Assumption 4: Outliers

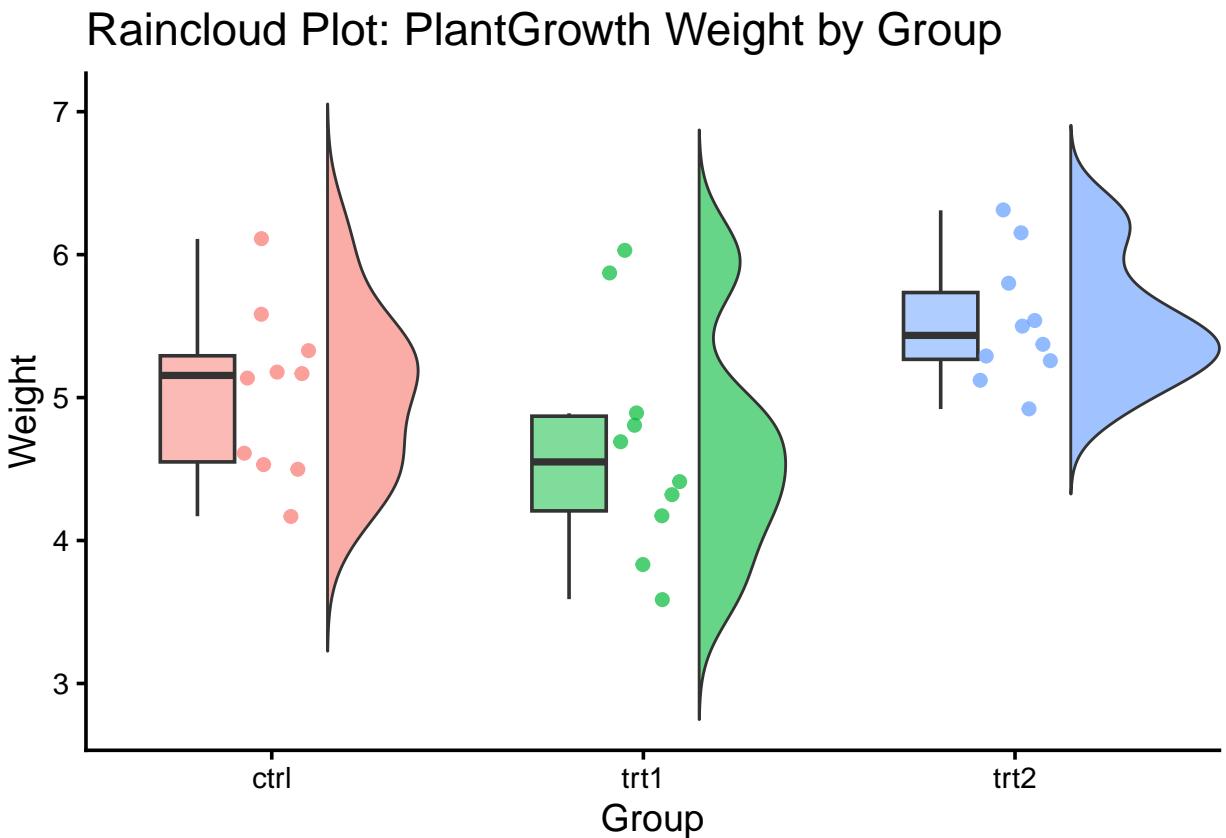
We check for significant outliers using boxplots and the Interquartile Range (IQR) method.

```
ggplot(PlantGrowth, aes(x = group, y = weight, fill = group)) +  
  
  geom_half_violin(  
    aes(fill = group),  
    side = "r",  
    trim = FALSE,  
    alpha = 0.6,  
    width = 0.8,  
    position = position_nudge(x = 0.15)  
  ) +  
  
  geom_boxplot()
```

```

width = 0.2,
alpha = 0.5,
outlier.shape = NA,
position = position_nudge(x = -0.2),
show.legend = FALSE
) +
geom_jitter(
  aes(color = group),
  width = 0.10,
  size = 1.8,
  alpha = 0.7
) +
theme_classic(base_size = 14) +
labs(
  title = "Raincloud Plot: PlantGrowth Weight by Group",
  x = "Group",
  y = "Weight"
) +
guides(fill = "none", color = "none") +
scale_x_discrete(expand = expansion(mult = c(0.25, 0.25))) +
coord_cartesian(clip = "off")

```



```

# Identify outliers mathematically
outliers <- PlantGrowth %>%
  group_by(group) %>%
  identify_outliers(weight)

# Print outliers table
if(nrow(outliers) > 0){
  kable(outliers, caption = "Detected Outliers")
} else {
  print("No extreme outliers were detected.")
}

```

Table 1: Detected Outliers

group	weight	is.outlier	is.extreme
trt1	5.87	TRUE	FALSE
trt1	6.03	TRUE	FALSE

Remark: Visual inspection of boxplots and statistical checks show no extreme outliers.

Assumption 5: Normality

We use the Shapiro-Wilk test to check if data is normally distributed for each group ($p > 0.05$).

```

# Shapiro-Wilk test
shapiro_res <- PlantGrowth %>%
  group_by(group) %>%
  shapiro_test(weight)

kable(shapiro_res, caption = "Shapiro-Wilk Test for Normality")

```

Table 2: Shapiro-Wilk Test for Normality

group	variable	statistic	p
ctrl	weight	0.9566815	0.7474734
trt1	weight	0.9304107	0.4519440
trt2	weight	0.9410052	0.5642519

Remark: The p -values for all groups are greater than 0.05. We assume the data is approximately normally distributed.

Assumption 6: Homogeneity of Variances

We use Levene's test to check if variances are equal across groups ($p > 0.05$).

```
# Levene's Test
levene_res <- leveneTest(weight ~ group, data = PlantGrowth)
kable(levene_res, caption = "Levene's Test for Homogeneity of Variances")
```

Table 3: Levene's Test for Homogeneity of Variances

	Df	F value	Pr(>F)
group	2	1.119186	0.3412266
	27	NA	NA

Remark: The p -value is $\text{round}(\text{levene_res}\$"\text{Pr}(>\text{F})"\text{[1]}, 3)$, which is > 0.05 . Therefore, we assume homogeneity of variances.

Computation: One-Way ANOVA

```
# Compute ANOVA
anova_model <- aov(weight ~ group, data = PlantGrowth)
anova_summary <- summary(anova_model)

# Show ANOVA Table
anova_summary
```

```
##           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
```

Descriptive Statistics

```
# Calculate Mean and SD for reporting
desc_stats <- PlantGrowth %>%
  group_by(group) %>%
  get_summary_stats(weight, type = "mean_sd")

kable(desc_stats, caption = "Descriptive Statistics (Mean ± SD)")
```

Table 4: Descriptive Statistics (Mean \pm SD)

group	variable	n	mean	sd
ctrl	weight	10	5.032	0.583
trt1	weight	10	4.661	0.794
trt2	weight	10	5.526	0.443

Post Hoc Comparisons (Tukey HSD)

Since the ANOVA result was significant ($p < 0.05$), we perform Tukey's HSD test to see which specific groups differ.

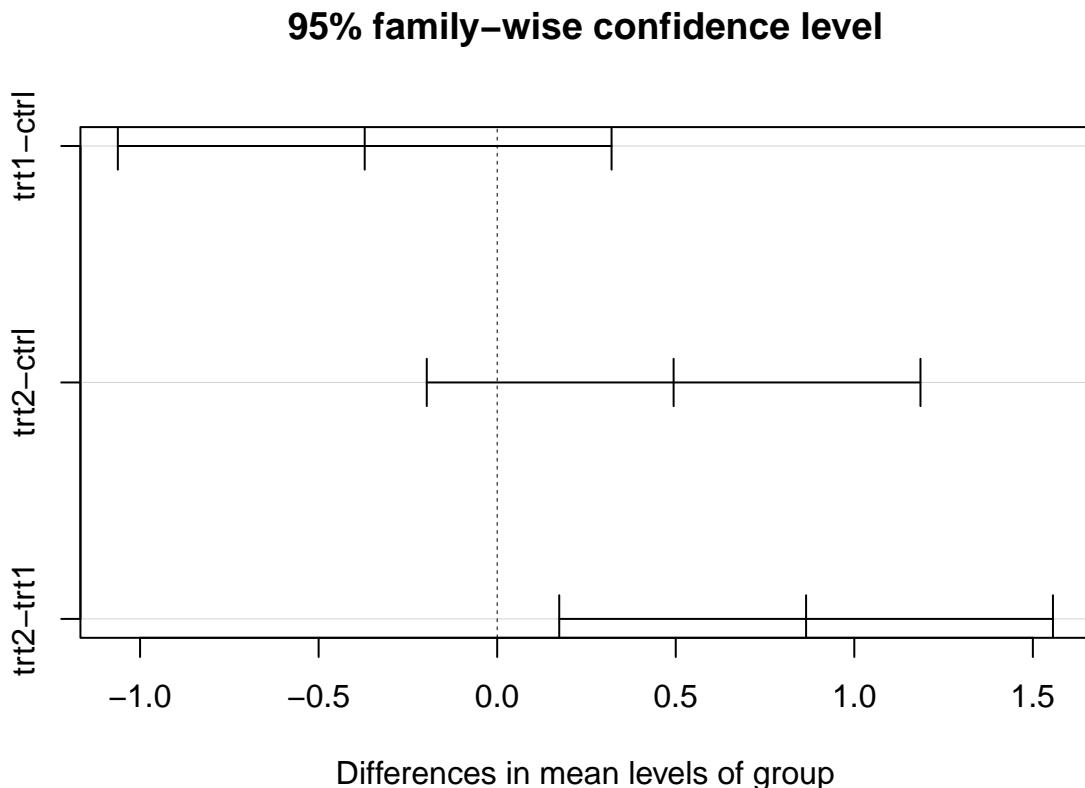
```

tukey_res <- TukeyHSD(anova_model)
tukey_res

## 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

# Plotting the differences
plot(tukey_res)

```



```
# Report
```

A one-way ANOVA was conducted to determine if there were significant differences in plant weight between three groups: ctrl ($n = 10$), trt1 ($n = 10$), and trt2 ($n = 10$).

There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by the Shapiro-Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test ($p = .044$).

Data is presented as mean \pm standard deviation. Plant weight was statistically significantly different between the different treatment groups, $F(2, 27) = 4.846, p = .016$.

Weight increased from the trt1 group ($M = 4.66, SD = 0.79$) to the ctrl group ($M = 5.03, SD = 0.58$) and the trt2 group ($M = 5.53, SD = 0.44$).

Tukey post hoc analysis revealed that the mean increase from trt1 to trt2 (0.87) was statistically significant ($p = .012$). However, the comparisons between ctrl vs trt1 ($p = .391$) and ctrl vs trt2 ($p = .198$) were not statistically significant.

Conclusion: We reject the null hypothesis. There is a significant difference in weight between the treatment groups, specifically between Treatment 1 and Treatment 2.