Two-way ANOVA Analyzing Randomized Block Design

Cheng Peng

STA200 Statistics II

Contents

1 Introduction

The data from the CRD (Completely Randomized Design) were analyzed using a one-way ANOVA procedure. Since the treatment is a factor with different groups (representing different doses), it influences the response variable. One-way ANOVA, also called one-factor ANOVA, focuses on testing whether there is a treatment effect. In statistical terms, this is equivalent to testing the null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_t.$$

Under this null hypothesis, the variance between treatments and the variance within treatments are approximately the same. Thus, we assess the equality of means across treatments by analyzing these variances, as summarized in the classical ANOVA table.

The key information MS_B and MS_W used to constructed the test statistic is based on the SS_B and SS_W extracted from the total sum of squares.

$$SS_T = SS_B + SS_W.$$

Two fundamental assumptions for the one-way ANOVA analysis are

- Populations is normally distributed
- Variances of treatment groups are equal

Under the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_t$

$$F = \frac{SS_B/t-1}{SS_W/n-t} \to F_{t-1,\ n-t}$$

where t is the number of treatment groups and n is the sample size.

The RBD designs add one additional factor variable. We are interested in assessing the effects in both the additional factor and the treatment effect. If the RBD has replicates, we also assess the potential interaction effect. Because of the additional factor, the decomposition of the total sum of square becomes more complex although the logic and mathematics derivation are the same as the one-way ANNOVA.

Before discussing the construction of ANOVA for replicated RBD data, we first look at an example to see what are practical questions we can answer based on the data collected from an replicated RBD.

Plant Grouth: a botanist wants to know whether or not **plant growth** is influenced by **sunlight exposure** and **watering frequency**. She plants 40 seeds and lets them grow for two months under different conditions for sunlight exposure and watering frequency. After two months, she records the height of each plant. The results are shown below:

	Sunlight Exposure			
Watering Frequency	None	Low	Medium	High
Daily	4.8	5	6.4	6.3
	4.4	5.2	6.2	6.4
	3.2	5.6	4.7	5.6
	3.9	4.3	5.5	4.8
	4.4	4.8	5.8	5.8
Weekly	4.4	4.9	5.8	6
	4.2	5.3	6.2	4.9
	3.8	5.7	6.3	4.6
	3.7	5.4	6.5	5.6
	3.9	4.8	5.5	5.5

Clearly, this is a balanced RBD with replications. The data structure required for analysis with software programs will be discussed later.

2 Hypotheses and Testing Methods

We will formulate the hypotheses based on the practical questions to be addressed based on replicated RBD data and set up the two-way ANOVA table for testing the hypothesis along with fundamental assuumptions required for the two-way ANOVA.

2.1 Setting-Up Hypotheses

As we see in the **Plant Growth** example, the practical questions is whether **sunlight exposure** and **watering frequency** affect **plant grouwth**. The is practical question has several sub-questions from analytic point of view. For convenience, let call **sunlight exposure** Factor A and **watering frequency** Factor B. The related hypotheses associated with the original practical question are

- **Main Effect of Factor A**:

H_0 : All levels of Factor A have the same mean effect. vs H_a : At least one level of Factor A has a different mean effect.

- **Main Effect of Factor B**:

H_0 : All levels of Factor B have the same mean effect. vs H_a : At least one level of Factor B has a different mean effect.

- **Interaction Effect (A × B)**:

H_0 : There is no interaction between Factor A and Factor B. vs H_a : There is a significant interaction effect.

3 Two-way ANOVA

Two-way ANOVA is used to analyze the effects of two independent categorical variables (factors) on a continuous dependent variable while also assessing the interaction between these factors. Unlike one-way ANOVA, which examines only one factor, two-way ANOVA allows researchers to determine whether the two factors independently influence the outcome and whether their combined effect (interaction) is significant.

3.1 Two-Way ANOVA Table Structure

Because two factors involved in the analysis, the mathematical decomposition of the total square into more components for testing the above three hypotheses is much more complex than the one-way ANOVA algebra is not **difficult**. We will not derive the components in the ANOVA table based on the raw data values. Instead, we focus on the structure of the ANOVA table and how to use the key statistics in the table to test the above hypotheses. We use software programs to generate the ANOVA table. We first introduce some notations to help understand the ANOVA table.

- Number of categories in **factor A**: a
- Number of categories in **factor B**: b
- Total sample size: N
- **Total Sum of Squares**: SST
- **Sum of Squares for Factor A**: SSA (Variation due to **factor A**).
- **Sum of Squares for Factor B**: SSB (Variation due to **factor B**).
- **Sum of Squares for Interaction between Factor A and Factor B**: $SSAB$ (Variation due to the interaction between **Factor A** and **Factor B**).
- **Residual Sum of Squares**: SSE (Unexplained variation - error).

Mathematically

$$SST = SSA + SSB + SSAB + SSE$$

Each **Sum of Squares** has its own **degrees of freedom**, next, we list the degrees of freedom os each the **sum of squares**:

- DF.A = $a - 1$
- DF.B = $b - 1$
- DF.AB = $(a - b) \times (b - 1)$ (interaction term)
- DF.E = $N - ab$
- DF.T = $N - 1$

Note that $N - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1) + N - ab$.

With the above notations, we introduce the two-way ANOVA table

Source	df	SS	MS (SS/df)	F-Statistic (MS Effect / MS Error)	p-value
Factor A	$a-1$	SSA	MSA	$F_A = \frac{MSA}{MSE}$	p_A
Factor B	$b-1$	SSB	MSB	$F_B = \frac{MSB}{MSE}$	p_B
Interaction	$(a-1)(b-1)$	SSAB	MSAB	$F_{AB} = \frac{MSAB}{MSE}$	p_AB
Residual (Error)	$N-ab$	SSE	MSE	-	-
Total	$N-1$	SST	-	-	-

The first three rows of the table contains information related to testing the above three hypotheses. In one-way ANOVA, we used F-distribution to find p-value for statistical decision under some assumptions as mentioned in the first section. We also need to make some assumptions in order to specify the distributions of F statistics. Next, we discuss the assumptions of two-way ANOVA and testing the hypotheses.

3.2 Assumptions of Two-Way ANOVA

As we know in the one-way ANOVA table, the F value follows an F distribution under some assumptions. In the two-way ANOVA, we also need the following assumptions to make inference of the three F statistics:

- **Normality**: The dependent variable should be normally distributed within **each combination of factor levels**.
- **Homogeneity of Variance (Homoscedasticity)**: The variances across groups defined by **each combination of factor levels** should be equal. In practice, we usally perform a formal test to check this assumption.
- **Independence of Observations**: Data points should be independent, meaning the measurement of one subject does not influence another. **This assumtion is hard to perform a test for independence**. We usually justify the data collection process.
- **Random Sampling**: Data should be collected using a random sampling method. This is justified from the data collection process.

Violations of these assumptions may require transformations or non-parametric alternatives like the Kruskal-Wallis test.

Assume that all above assumptions are met. Next we discuss the distribution of the test statistics for testing the three hypotheses.

- Effect of **Factor A**: under the following null hypothesis,

$$H_0 : \text{All levels of Factor A have the same mean effect.}$$

The test statistic

$$F_A = \frac{MS_A}{SS_E} \to F_{a-1,\ n-ab}.$$

The p-value can be found from $F_{a-1,\ n-ab}$.

- Effect of **Factor B**: under the following null hypothesis,

$$H_0 : \text{All levels of Factor B have the same mean effect.}$$

The test statistic

$$F_B = \frac{MS_B}{MS_E} \to F_{b-1,\ n-ab}$$

The p-value can be found from $F_{b-1,\ n-ab}$.

- Interactive Effect between **Factor A** and **Factor B**: under the following null hypothesis,

$$H_0 : \text{There is no interaction between Factor A and Factor B.}$$

The test statistic

$$F_{AB} = \frac{MS_{AB}}{MS_E} \to F_{(a-b)(b-1),n-ab}$$

The p-value can be found from $F_{(a-1)(b-1),\ n-ab}$.

Remarks: understanding the structure of the ANOVA table is crucial.

- If the first two columns are given, you are expected to derive the rest of the columns including the p-values in the last column using R command `pf(F, df1, df2, lower.tail = FLASE)` (F table is not recommended).
- The interaction effect is not available if working with RBD data with no replicates.

3.3 Performing ANOVA Using R

The R function `aov()` will be used to perform ANOVA test. If the given data table is in wide format, we need to reformat it into a long table. This involves placing all response values in a single column and using two separate columns to label each response value with its associated factors.

The R function `rep()` is useful for generating patterned data. Its syntax is `rep(data.value, times = n)`, which means it **replicates** `data.value` n times. The `data.value` argument can be either a **single value** or a **vector of values**. Below are some examples demonstrating how to use `rep()` to create patterned data.

```
rep(5, times = 3)    # replicate 5 three times
```

```
| [1] 5 5 5
```

```
rep("a", 9)          # replicate lower case letter "a" 9 times
```

```
| [1] "a" "a" "a" "a" "a" "a" "a" "a" "a"
```

```r
rep(c(1,3,5), 3)    # replicate vector (1,3,5) three times
```

```
| [1] 1 3 5 1 3 5 1 3 5
```

```r
rep(rep(c("A", "B"), 2), 3) # nested replicate rep(c("A", "B")) three times
```

```
|  [1] "A" "B" "A" "B" "A" "B" "A" "B" "A" "B" "A" "B"
```

Next, we work on the wide table in the example given in Section 1.

```r
## first define 4 column vectors
None = c(4.8, 4.4, 3.2, 3.9, 4.4, 4.4, 4.2, 3.8, 3.7, 3.9)
Low = c(5.0, 5.2, 5.6, 4.3, 4.8, 4.9, 5.3, 5.7, 5.4, 4.8)
Medium = c(6.4, 6.2, 4.7, 5.5, 5.8, 5.8, 6.2, 6.3, 6.5, 5.5)
High = c(6.3, 6.4, 5.6, 4.8, 5.8, 6.0, 4.9, 4.6, 5.6, 5.5)
## place the above values in a single column
growth <- c(None, Low, Medium, High)
## define sunlight labels
sunlight <- c(rep("None", length(None)), rep("Low", length(Low)),
              rep("Medium", length(Medium)), rep("High", length(High)))
## Watering patterns
## The inner rep() returns the pattern of first column
## The outer rep() replicate the rest of the columns
watering <-rep(c(rep("Daily", 5), rep("Weekly", 5)),4)
## Store variables in a data frame
growthData <- data.frame(growth = growth, sunlight = sunlight, watering = watering)
## check the data frame
#growthData
```

We now perform the ANOVA using the data frame.

```r
# we fit an ANOVA with interactive effect
growth.aov <- aov(growth ~ sunlight*watering, data = growthData)
summary(growth.aov)
```

```
|                   Df Sum Sq Mean Sq F value  Pr(>F)    
| sunlight           3 18.765   6.255  23.049 3.9e-08 ***
| watering           1  0.000   0.000   0.001   0.976    
| sunlight:watering  3  1.011   0.337   1.242   0.311    
| Residuals         32  8.684   0.271                    
| ---
| Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Formal Summary of ANOVA: Two formats of write-up are recommended.

Format 1

The above two-way ANOVA revealed a significant main effect of **sunlight** ($F(3, 32) = 23.049$), $p \approx 0$), indicating that the plant heights (growth) were significantly higher when being exposed to moresunlight. Neither the main effect of **watering frequency** ($F(1, 32) = 0.01, p = 0.976$) nor the **sunlight** and **watering frequency** interaction ($F(1, 32) = 1.242, p = 0.311$) was statistically significant.

Format 2

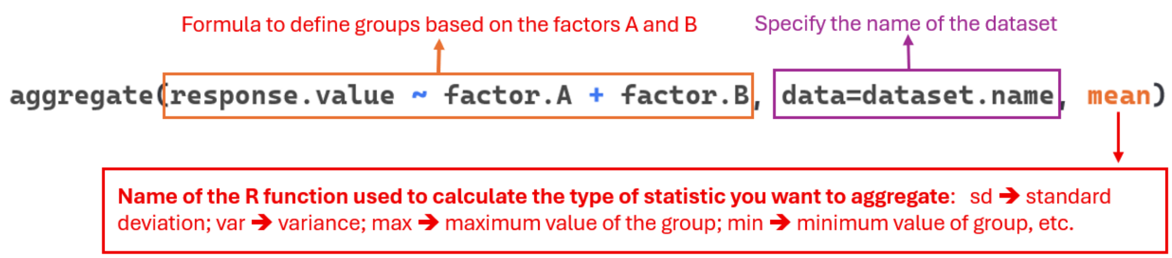
A two-way ANOVA was conducted, yielding the following findings:

- **Watering Frequency**: No significant effect on plant growth (F(1, 32) = 0.01, $p = 0.976$).
- **Sunlight Exposure**: Significant effect on plant growth (F(3, 32) = 23.049, $p < 0.001$). We will conduct a post-hoc comparisons (e.g., Tukey's HSD) would determine which sunlight levels differ.

- **Interaction (Watering × Sunlight)**: No significant interaction (F(3, 32) = 1.242, p = 0.311), indicating that the effect of sunlight does not depend on watering frequency.

4 Post-hoc Test

The interpretation of post hoc tests becomes more complex when two or more factors are included in an ANOVA, especially when interaction effects are present. In such cases, it is often useful to create line plots showing the means of the primary factor at each level of the secondary factor. As an illustrative example, we will use the data frame created in the previous section to calculate the means of groups defined by the combination of `watering` and `sunlight`. The R function `aggregate()` can be used to compute these group means. In general, `aggregate()` is used to summarize the values within groups using descriptive statistics such as the mean, variance, maximum, minimum, etc. The syntax of aggregate() is shown below.




```
## We use the data frame "growthData" created in previous section
grp.means <- aggregate(growth ~ watering + sunlight, data = growthData, mean)
## checking the resulting table
#grp.means
## split into daily means and weekly means
## using which() to find the rows in which watering is "Daily" and "Weekly" respectively
daily.id <- which(grp.means$watering == "Daily")
weekly.id <- which(grp.means$watering == "Weekly")
## subset of means of sunlight at its individual levels
daily.avg <- grp.means[daily.id, 3]
weekly.avg <- grp.means[weekly.id, 3]
### draw line plots of
```

We have not formally introduce R base plot function with details. Next, we briefly introduce base R function `plot()`:

```
plot(x, y, type = "b", xaxt = "n", xlab = "Category", ylab = "Value", main = "Custom X-Axis Labels")
```

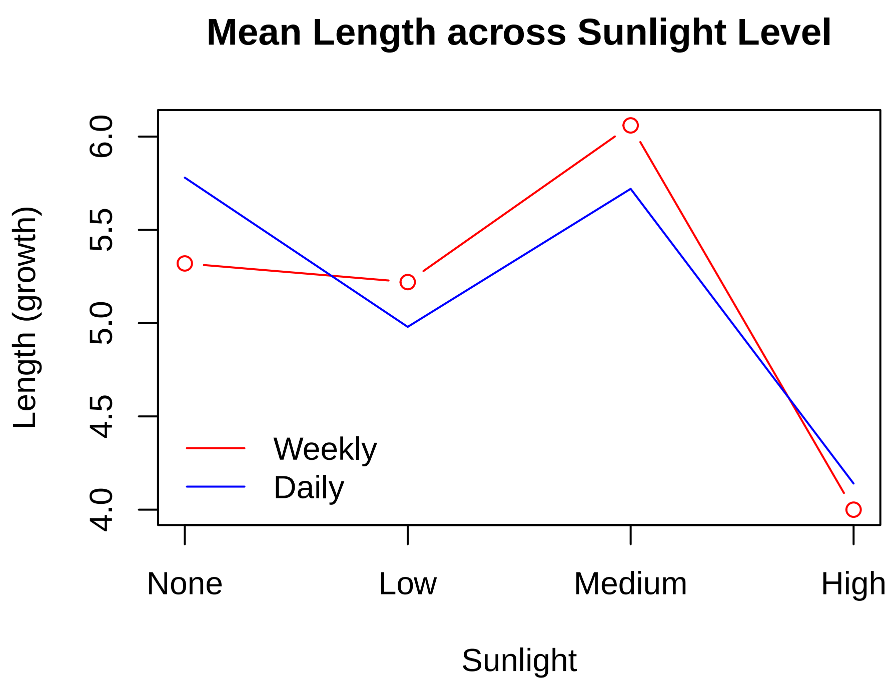
where * `x` horizontal x-axis * `y` vertical y-axis * `type = "b"` draws both points and lines. * `xlab=` the label of x-axis * `ylab` = the label of y-axis * `main=` the title of the plot * `xaxt = "n"` suppresses the default x-axis. * `axis(side = 1, ...)` adds a customized horizontal axis.

The code below draws a line plot of the mean sunlight levels for each watering level.

```
# Custom tick mark positions and labels
x.ticks <- c(1,2,3,4)                              # tick marks / location of tick labels
x.labels <- c("None", "Low", "Medium", "High")

# Create the plot without x-axis (xaxt = "n")
plot(c(1,2,3,4), weekly.avg,
     type = "b",
     xaxt = "n",
     xlab = "Sunlight",
     ylab = "Length (growth)",
```

```r
      col = "red",           # color of the line graph
      lty = 1,               # line type. 1 = solid line
      main = "Mean Length across Sunlight Level")
## add a line plot of weekly watering
lines(c(1,2,3,4), daily.avg,
       col = "blue",
       lty = 1)
# Add custom x-axis
axis(side = 1, at = x.ticks, labels = x.labels)
## add a legend to specify watering level
legend("bottomleft",                # location for the legend
        c("Weekly", "Daily"),       # names of line plot, Caution: keep the order of the names!
        lty = c(1,1),
        col = c("red", "blue"),
        bty = "n"                   # don't include box of the legend
        )
```



If the two line graphs are parallel (i.e., there is NO intersection), the factors `watering` and `sunlight` do not have interactive effect!. Otherwise, the two factors do have an interactive effect!

Graphical Interpretations:

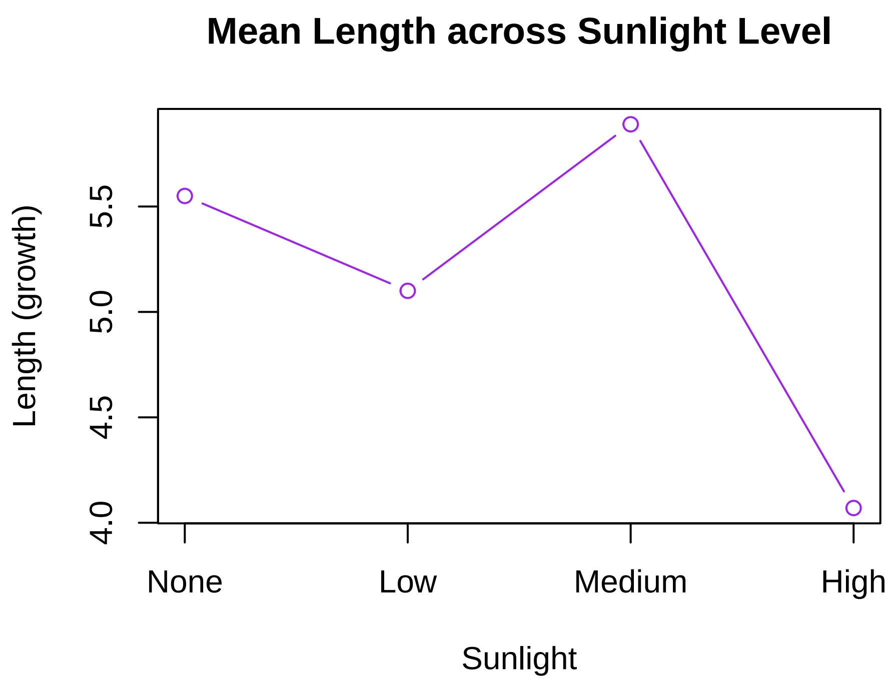
- *Main Effect* interpret `sunlight` patterns (trends) at each `watering` level:
  - When watered weekly, the plant grew best with **medium sunlight** and poorest with the **high sunlight**.
  - The same growth patterns were observed when atering daily.
- **Interactive effect** cross comparison between **watering frequencies**
  - The plants grew better with `Low` and `Medium` sunlight when **watered weekly**
  - However, the plant grew better with with `None` and `High` sunlight when **watered daily**

The following subsections provide detailed numerical comparisons with difference and associated confidence intervals.

4.1 ANOVA with No Interactive Effect

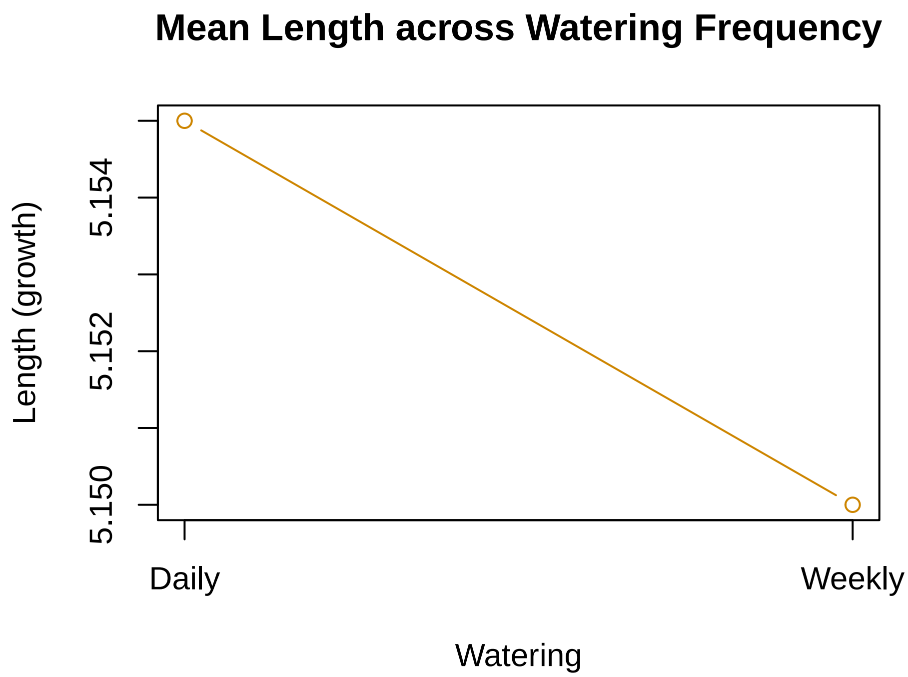
If there is no interaction effect, we interpret the main effects of the factor variables independently. In the **plant growth** analysis, the key factor is **sunlight**, and the secondary factor is **watering**. We compare the main effects of the key and secondary factors separately. For example:

- **Sunlight (Key Factor)**: Compare mean outcomes across sunlight levels, ignoring watering. "*High sunlight yields higher growth than low sunlight, averaged over all watering levels.*



The above line plot shows that **the growth rate generally decreases as sunlight increases, with the exception of the medium sunlight level, at which the rate peaks**.

- **Watering (Secondary Factor)**: Compare means across watering levels, ignoring sunlight. *Frequent watering increases growth, but this effect is smaller than sunlight's effect.*



The plot above shows that **the plant grew better when watered daily**.

4.2 ANOVA with Interactive Effect

After finishing ANOVA test, we report multiple comparions between groups associated with the **significant factors**. If a two-way ANOVA reveals a significant interaction effect, it means the impact of one factor (e.g., Sunlight Exposure) depends on the level of the other factor (e.g., Watering Frequency). To interpret this interaction properly, we must:

- Compare groups **within combinations** (e.g., "`Daily + Low Sunlight`" vs. "`Weekly + Medium Sunlight`").
- Control for inflated Type I error rates (false positives) due to **multiple comparisons** by using **simultaneous methods** such Tukey's HSD to adjust p-values to maintain the family-wise error rate (FWER) at a set level (e.g., 5%).

In the above **plant growtth** dataset, the ANOVA test suggests insignificance of the interactive effect between **sunlight** and **watering**. The only significant factor is **sunlight**. For illustration, we perform **simultaneous comparisons** for individual factors and their interactions using Tukey's HSD.

The general idea in interpreting multiple comparisons is to focus on the key factor (such as sunlight) and breakdown the secondary factor (watering)

```
anova.interact <- aov(growth ~ sunlight*watering, data = growthData)
TukeyHSD(anova.interact)  # Compare sunlight levels within Daily watering
```

```
|   Tukey multiple comparisons of means
|     95% family-wise confidence level
|
| Fit: aov(formula = growth ~ sunlight * watering, data = growthData)
```

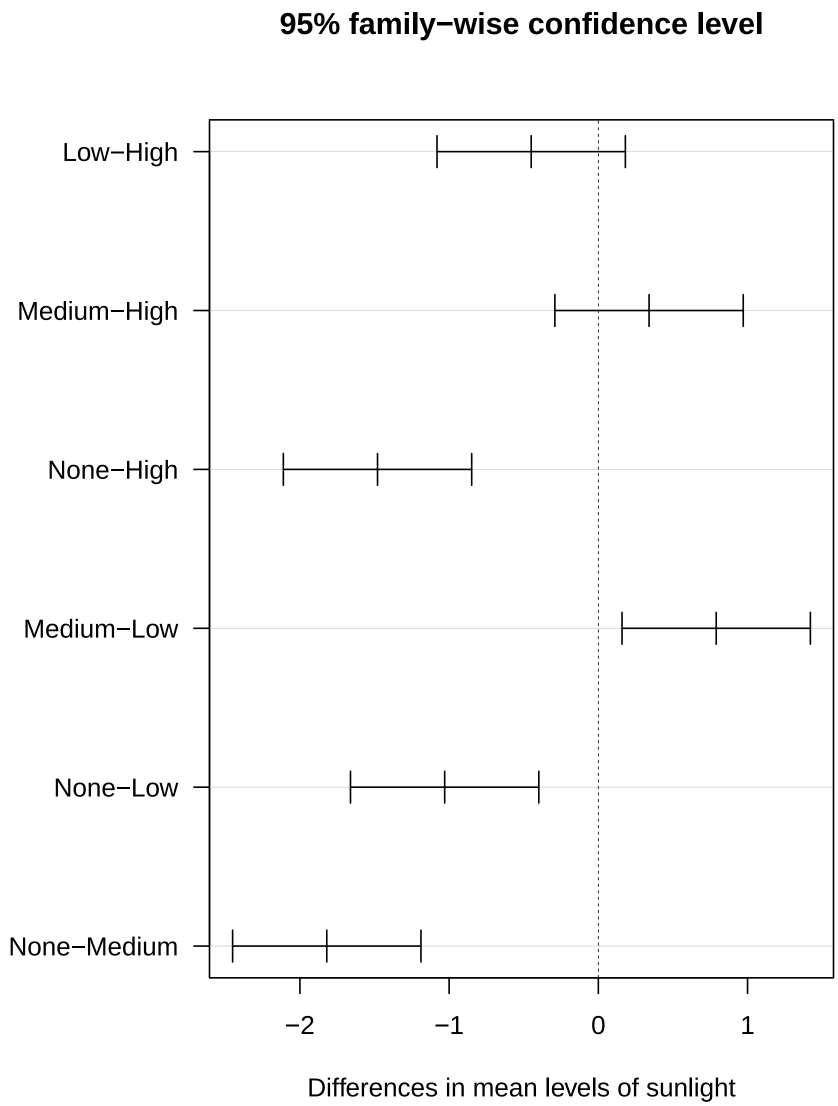
```
|
| $sunlight
|              diff        lwr        upr     p adj
| Low-High    -0.45 -1.0811999  0.1811999 0.2353814
| Medium-High  0.34 -0.2911999  0.9711999 0.4730539
| None-High   -1.48 -2.1111999 -0.8488001 0.0000023
| Medium-Low   0.79  0.1588001  1.4211999 0.0095962
| None-Low    -1.03 -1.6611999 -0.3988001 0.0005862
| None-Medium -1.82 -2.4511999 -1.1888001 0.0000000
|
| $watering
|                diff        lwr       upr    p adj
| Weekly-Daily -0.005 -0.3405535 0.3305535 0.975975
|
| $`sunlight:watering`
|                             diff         lwr         upr     p adj
| Low:Daily-High:Daily       -0.80 -1.86724928  0.26724928 0.2626201
| Medium:Daily-High:Daily    -0.06 -1.12724928  1.00724928 0.9999996
| None:Daily-High:Daily      -1.64 -2.70724928 -0.57275072 0.0005069
| High:Weekly-High:Daily     -0.46 -1.52724928  0.60724928 0.8523023
| Low:Weekly-High:Daily      -0.56 -1.62724928  0.50724928 0.6873529
| Medium:Weekly-High:Daily    0.28 -0.78724928  1.34724928 0.9884511
| None:Weekly-High:Daily     -1.78 -2.84724928 -0.71275072 0.0001516
| Medium:Daily-Low:Daily      0.74 -0.32724928  1.80724928 0.3529276
| None:Daily-Low:Daily       -0.84 -1.90724928  0.22724928 0.2118374
| High:Weekly-Low:Daily       0.34 -0.72724928  1.40724928 0.9657630
| Low:Weekly-Low:Daily        0.24 -0.82724928  1.30724928 0.9954055
| Medium:Weekly-Low:Daily     1.08  0.01275072  2.14724928 0.0456827
| None:Weekly-Low:Daily      -0.98 -2.04724928  0.08724928 0.0905656
| None:Daily-Medium:Daily    -1.58 -2.64724928 -0.51275072 0.0008466
| High:Weekly-Medium:Daily   -0.40 -1.46724928  0.66724928 0.9217373
| Low:Weekly-Medium:Daily    -0.50 -1.56724928  0.56724928 0.7925032
| Medium:Weekly-Medium:Daily  0.34 -0.72724928  1.40724928 0.9657630
| None:Weekly-Medium:Daily   -1.72 -2.78724928 -0.65275072 0.0002546
| High:Weekly-None:Daily      1.18  0.11275072  2.24724928 0.0219109
| Low:Weekly-None:Daily       1.08  0.01275072  2.14724928 0.0456827
| Medium:Weekly-None:Daily    1.92  0.85275072  2.98724928 0.0000450
| None:Weekly-None:Daily     -0.14 -1.20724928  0.92724928 0.9998578
| Low:Weekly-High:Weekly     -0.10 -1.16724928  0.96724928 0.9999854
| Medium:Weekly-High:Weekly   0.74 -0.32724928  1.80724928 0.3529276
| None:Weekly-High:Weekly    -1.32 -2.38724928 -0.25275072 0.0073419
| Medium:Weekly-Low:Weekly    0.84 -0.22724928  1.90724928 0.2118374
| None:Weekly-Low:Weekly     -1.22 -2.28724928 -0.15275072 0.0161418
| None:Weekly-Medium:Weekly  -2.06 -3.12724928 -0.99275072 0.0000134
```

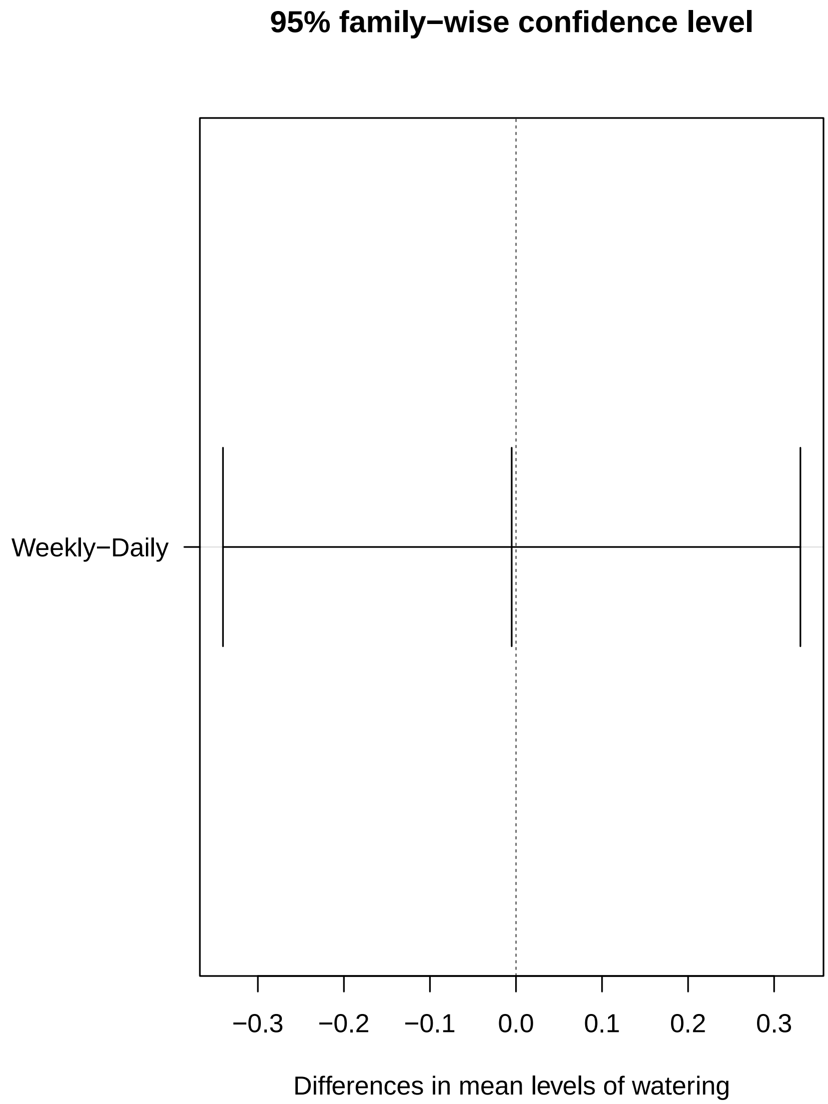
The above simultaneous multiple comparisons are presented in the form of confidence intervals for the differences, as well as adjusted p-values to ensure that the overall p-value is not inflated. As the number of combinations of factor levels increases, interpreting the confidence intervals becomes more difficult. To aid interpretation, we provide a visual representation of the outputted confidence intervals.

The first two plots correspond to the first two sets of confidence intervals. The interpretation of these confidence tables and figures is straightforward. However, the number of intervals associated with interaction effects increases as the number of factor levels grows, making interpretation more complex. In the plant growth example, the two factors result in 28 different group comparisons! Practically meaningful comparisons focus on differences in growth rates under the same sunlight conditions but with different watering frequencies.

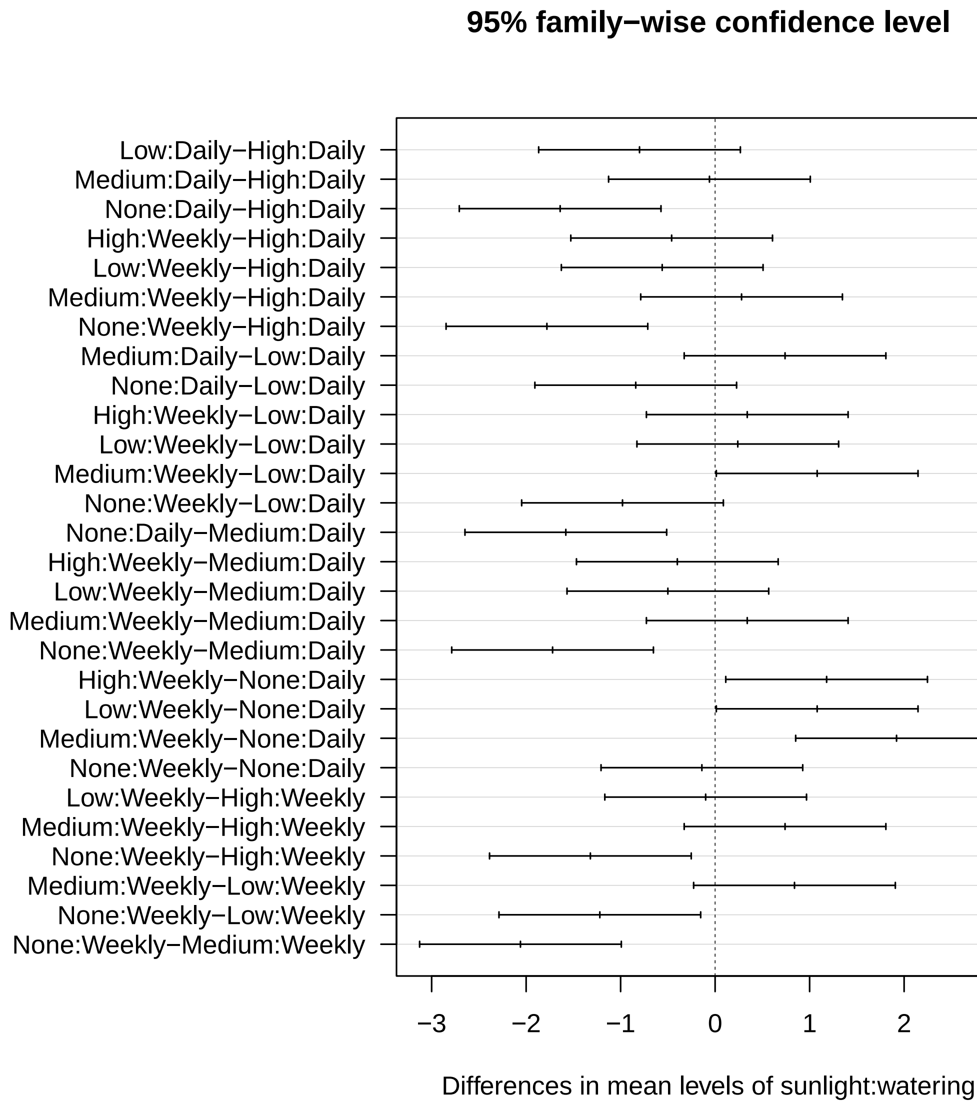
```r
par(mai=c(1.5,3,1,1))           # Makes room on the plot for the group names
plot(TukeyHSD(anova.interact),  # name of the TukeyHSD() object
     cex.lab = 0.6,             # adjust the font size of the labels of the
                                # vertical axis

     las = 1)
```

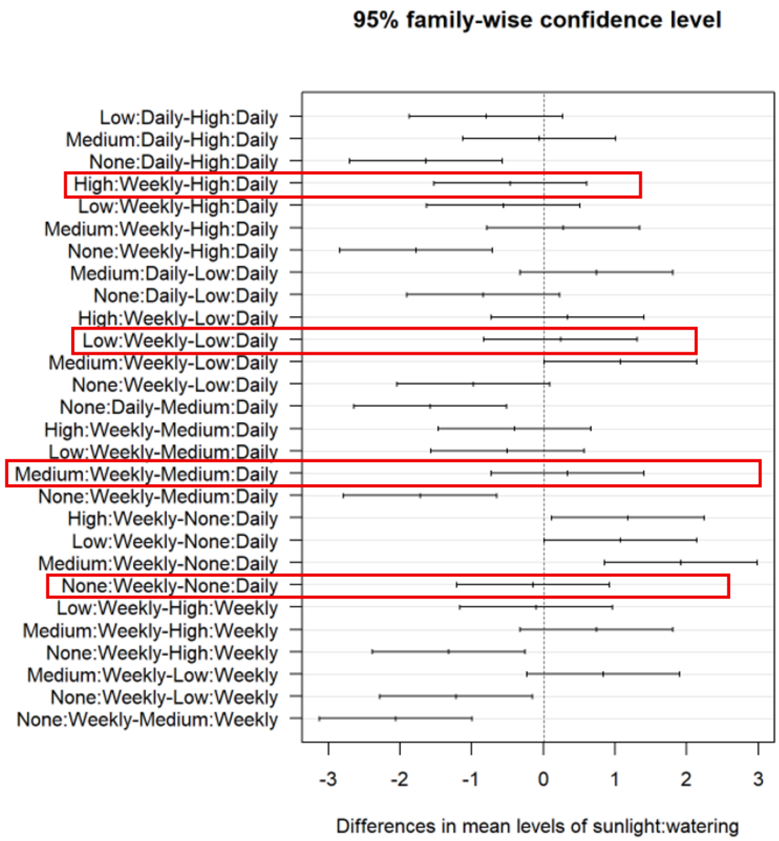




95% family−wise confidence level
Weekly−Daily
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
Differences in mean levels of watering



The across comparion we should focus on is hightlighted in the following figure.



5 Practice Exercise

Using the following **reaction time** data set collected from a replicate RBD experiment to perform a comprehensive ANOVA analysis and post-hoc tests. The objective of the replicated RBD is to assess the `beer` and `caffeine` effects of `responsive time`.

	no caffeine	caffeine
no beer	2.24, 1.62, 1.48, 1.70, 1.06, 1.39, 2.69, 0.28, 2.24, 1.15, 1.53, 2.43	0.62, 1.72, 1.75, 1.84, 1.30, 1.52, 1.31, 1.63, 1.91, 1.33, 0.84, 0.45
beer	1.71, 2.19, 2.27, 2.35, 2.47, 2.07, 2.56, 2.35, 1.50, 2.63, 2.48, 1.94	2.05, 1.51, 1.65, 2.68, 2.06, 1.80, 2.68, 1.93, 1.29, 1.93, 1.35, 1.37