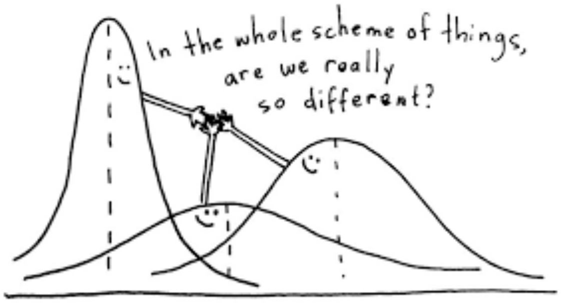
# R ANOVA Methods



## 1. Introduction

Creating an Analysis of Variance (ANOVA) model is a technique to test whether the mean value of different groups of data are statistically significantly different. Another way of saying this is that ANOVA tests whether the variation among groups is larger than the variation within groups; if it is larger, then the groups are significantly different. The assumptions for an ANOVA are:

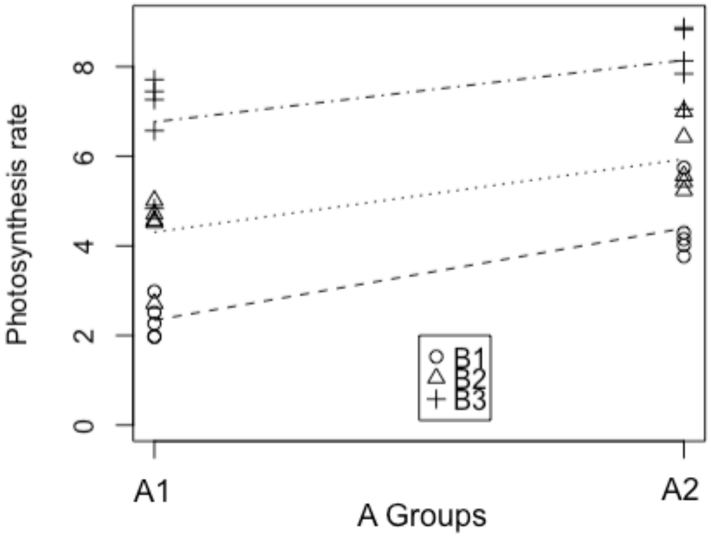
1) Independence: the data collected are independent observations
2) Normality: the residuals (error) are normally distributed (i.e., bell curve)
3) Equality of variances: the variance within groups should be roughly equal

ANOVA compares the variance attributed by different factors (or groups) by using the F-test statistic, which is the ratio of the variance between treatments to the variance within treatments. R then calculates the probability (p-value) of a value of the F-test statistic that is greater than or equal to the observed value of the F-test statistic. The null hypothesis (that the mean of the groups are not different) can be rejected if this probability is less than or equal to the significance level (typically $p<0.05$).

## 2. Conceptual Example

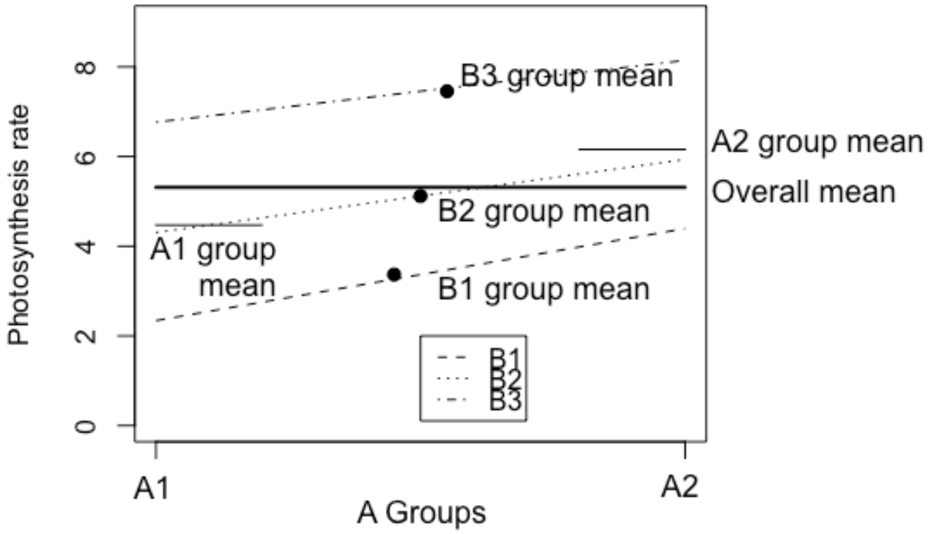
Let's look at how ANOVA works with a conceptual example with two experimental factors. In this example, let's say we do an experiment where we measure the rates of photosynthesis under two fertilizer treatments (A1 and A2) for three species of plants (B1, B2, and B3). We replicated our experiment by measuring five plants within each species within each fertilizer treatment (i.e., there are 5 B1 plants that received A1 treatment, 5 B2 plants that received A1 treatment, etc.). In this experiment for our ANOVA, we would say that there are two "groups": the A group, which represents fertilizer treatment, and the B group, which represents the three different species. For our ANOVA analysis, we want to test whether fertilizer treatment created an increase in photosynthesis, and we also want to determine whether this response was different for the three species.

We should first look at our data:

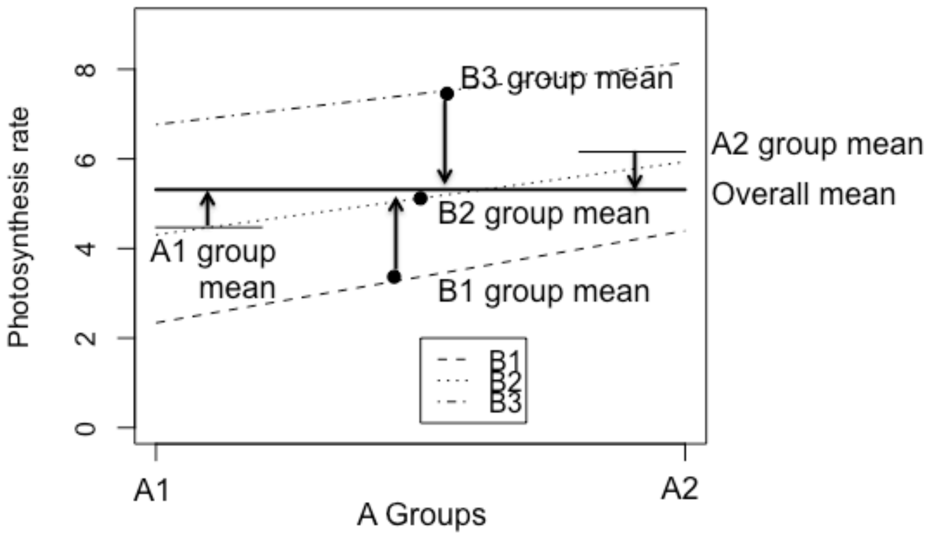


In this example, we see our five replicate measurements for each group, and the dotted lines connect the means for the A1 and A2 treatments by species.

Now let's have a closer look at patterns within groups:

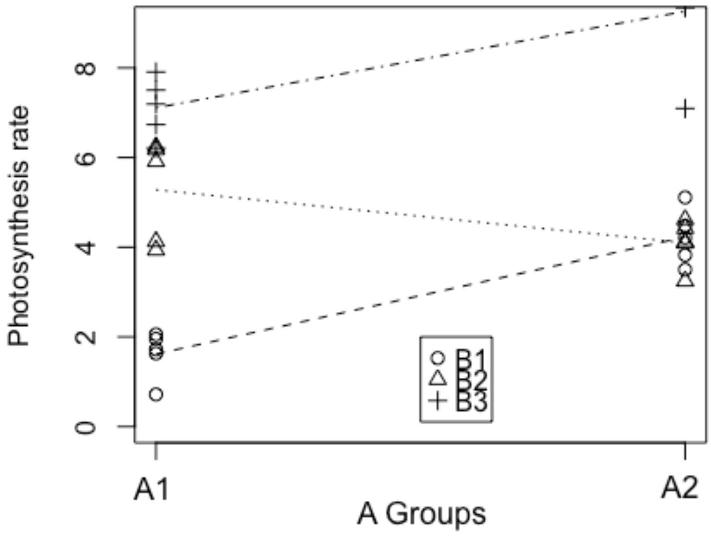


This slightly modified plot no longer shows the raw data, but highlights the overall mean, the A1 and A2 group means, and the B1, B2, and B3 group means. In ANOVA, we are analyzing how the "effect" of being in one particular group is different from the overall mean. We measure the effect size of a particular group as the distance from the overall mean:



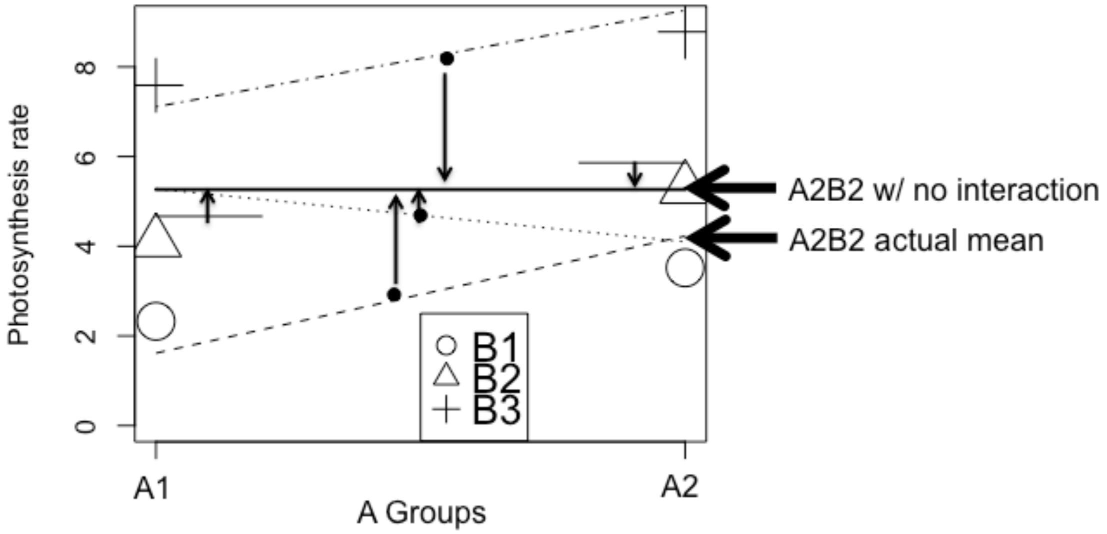
In this graph, each of the arrows represents the effect of being in a particular group, where the effect size is equal to the overall mean subtracted from the group mean. So for example, the B1 effect is negative because it is lower than the overall mean and the B3 effect is positive because it is higher than the overall mean. We can reconstruct the total effect of being in a particular experimental group (i.e., A1B1) by adding the overall mean, the A1 effect, and the B1 effect.

In this example so far, it doesn't look like there are interacting effects between the A and B groups. We can tell this because the slopes of the dotted lines in the previous graphs were all relatively similar. However, let's say that our data from the same experimental design looked like this:



Now we can see that the B2 species behaved very differently from B1 and B3 in the A2 treatment. In this case, we would say that there is an interacting effect between the A and B groups. Another way of saying this is that in order to predict the response to the A treatments, we need to know the B species group.

We can further corroborate the presence of an interaction effect between A and B groups if we reconstruct the total effect of each experimental group (A1B1, A1B2, etc.) as the sum of the overall mean, the A effects, and the B effects. These are plotted as the large symbols on the following graph:



Large symbols = effect with no interaction
Ex.) A2B2 triangle = overall mean + A2 effect + B2 effect

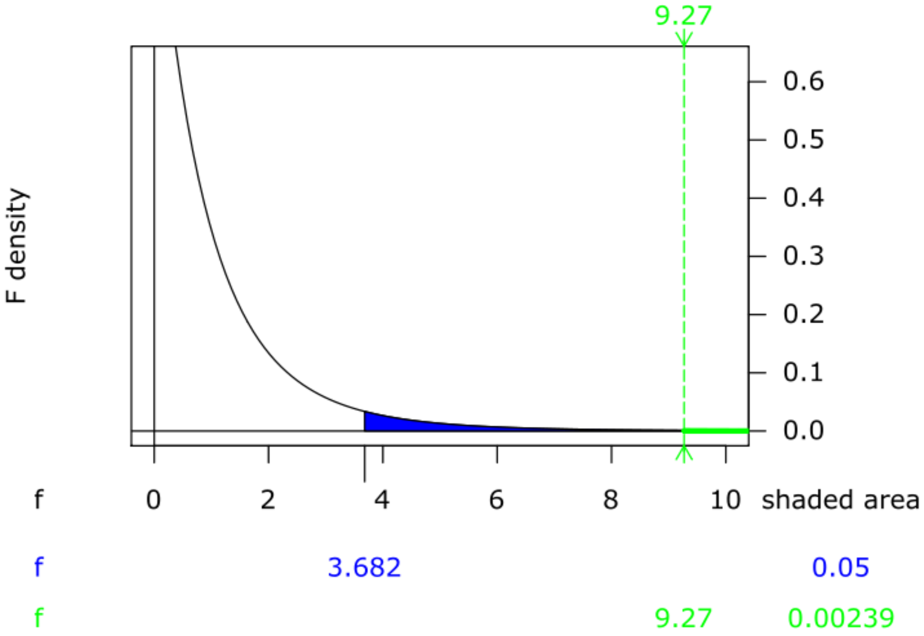
Particularly for the A2B2 group, we see that the reconstructed experimental group effect is fairly far from the A2B2 actual group mean (dotted line). This tells us that our ANOVA model without interaction isn't fully capturing the patterns in our data, and that we should think about including an interaction between A and B for more explanatory power within our model.

### 3. Testing for Statistical Significance

So we now know conceptually how to think about these different group means, but how can we tell whether groups and/or interactions in our ANOVA models are significant? We use the F-test statistic to determine whether the ratio of between-group variability to within-group variability exceeds a critical threshold (the p-value) of the F distribution, beyond which we can be relatively confident that our results were not random. (see https://en.wikipedia.org/wiki/F-test for details)

The shape of the F distribution depends on the degrees of freedom (df) between groups and the df within groups. The between-groups df is equal to one less than the number of groups: df_b = g-1. The within-groups df is equal to the number of groups times one less than the number of observations: df_w = g*(n-1). The shape of the F distribution is determined by these two parameters, where we can know what F-test statistic value we need in order to be confident that the differences are not random.

The following graph highlights the significance level (0.05) in blue on an F-distribution where df_b = 2 and df_w = 15. The significance level is equal to the probability density colored in blue on the graph. The critical F-statistic value (3.682) where the probability density equals 0.05 is also written in blue. In green are the results from an actual dataset, where the F-test statistic is shown as the green dotted line (9.27). Since 9.27 is greater than the critical value, 3.682, we can reject the null hypothesis that the means were not significantly different. The p-value for the differences between groups is written in green, and is equal to the green probability density shaded in the graph.



Example from: https://en.wikipedia.org/wiki/F-test

The experimental F-test statistic (9.27 in green in the above graph) is calculated as the ratio of the mean-squared value between groups to the ratio of the mean-squared value within groups. I won't go into more details on that, but know that these two values in the ratio are calculated from a ratio of the difference between the group mean from the overall mean (the things highlighted on the previous series of graphs with theoretical data) divided by the degrees of freedom.

## 4. How to do this in R

R automates many of the steps involved with running an ANOVA model. The structure of an ANOVA model is the same as a linear regression model in R, but in an ANOVA the predictor variables (x variables) are categories, rather than continuous variables. To run through an example in R, let's use the data from the previous Conceptual Example that I used to make the graphs highlighting the group differences. You should be able to copy and paste the following code to set up a data frame with theoretical data with five replicates in each of the A and B groups[1]:

```
# Example 1
# Randomly generate 5 replicate data points for the six groups
A1B1 = rnorm(5, mean=2, sd=1)
A1B2 = rnorm(5, mean=5, sd=1)
A1B3 = rnorm(5, mean=7, sd=1)
A2B1 = rnorm(5, mean=4, sd=1)
A2B2 = rnorm(5,mean=6, sd=1)
A2B3 = rnorm(5,mean=9, sd=1)

A.group = c(rep("A1",15),rep("A2",15)) # A Group names
B.group = c(rep(c(rep("B1",5),rep("B2",5),rep("B3",5)),2)) #B Group names

# Paste together data frame
exp.dat = data.frame(ps = c(A1B1,A1B2,A1B3,A2B1,A2B2,A2B3), A.group, B.group)
exp.dat[1:10,] # Look at first 10 rows
```

### *4.1 One-factor ANOVA Model*

The columns within this data frame indicate photosynthesis measurement, A treatment group (moisture levels), and B treatment group (species). Let's set up a one-factor ANOVA to get us started, by first examining the A group effects:

```
aov.1 = lm(ps ~ A.group, data=exp.dat) #run ANOVA model and save it to aov.1

summary(aov.1) #look at summary output of ANOVA model
```

The summary() function generates a table with statistics that we can use to evaluate the ANOVA model effects:

```
Call:
lm(formula = ps ~ A.group, data = exp.dat)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9521 -1.9688 -0.7419  1.9422  5.0487

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6689     0.6604   7.069 1.09e-07 ***
A.groupA2     1.1929     0.9340   1.277    0.212 (p-value for A effect)
```

[1] **WARNING**: Because this code generates pseudo-data, your results will look a little different from mine, and will look different each time you re-run the code to randomly generate data points.

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.558 on 28 degrees of freedom
Multiple R-squared:  0.05505, Adjusted R-squared: 0.0213 (variance explained)
F-statistic: 1.631 on 1 and 28 DF,  p-value: 0.212 (p-value for whole fit)
```

The main things to focus on in the summary output are the p-values for the factors (in bold blue; p<0.05 is significant), the $R^2$ for the overall fit to the data (bold green; $R^2 = 1$ is perfect fit), and the p-value for the overall fit (bold red; p<0.05 is significant, this will be different from the factor p-values for ANOVAs with >1 factor). From this summary, we can see that the means between groups A1 and A2 were not significantly different (i.e., p>0.05).

### *4.2 Two-factor ANOVA Model*

Let's now take a look at the ANOVA model with both A and B factors included:

```
aov.2 = lm(ps ~ A.group + B.group, data=exp.dat)
summary(aov.2) #look at summary output of ANOVA model
```

```
Call:
lm(formula = ps ~ A.group + B.group, data = exp.dat)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.0407 -0.8140 -0.1208  0.6788  2.1481 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   2.3251     0.4512   5.153 2.25e-05 ***
A.groupA2     1.1929     0.4512   2.644  0.01371 *  
B.groupB2     1.7679     0.5526   3.199  0.00361 ** 
B.groupB3     5.2635     0.5526   9.524 5.79e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.236 on 26 degrees of freedom
Multiple R-squared:  0.7952,  Adjusted R-squared:  0.7716 
F-statistic: 33.65 on 3 and 26 DF,  p-value: 4.195e-09
```

From this summary, we see that both the A group and B group means were significantly different and explained a substantial amount of variability within the data ($R^2 = 0.77$). Remember that the A groups were not previously statistically significant. This highlights one potential downside to using ANOVA models: with increasing numbers of variables, marginally significant variables can become significant in more complex models as the residuals (error) within the models becomes smaller. For this reason, it is useful to also compare the relative effect sizes of the various factors. As you remember from our previous graphical Conceptual Example, we can do this by taking the difference between the Overall Mean and the Group Mean:

```
#A1.effect = mean(A1 photosynthesis) – mean(overall photosynthesis)
A1.effect = mean(exp.dat$ps[exp.dat$A.group=="A1"]) - mean(exp.dat$ps)
A1.effect
[1] -0.596444

A2.effect = mean(exp.dat$ps[exp.dat$A.group=="A2"]) - mean(exp.dat$ps)
A2.effect
[1] 0.596444


#B1.effect = mean(B1 photosynthesis) – mean(overall photosynthesis)
B1.effect = mean(exp.dat$ps[exp.dat$B.group=="B1"]) - mean(exp.dat$ps)
B1.effect
[1] -2.34378

B2.effect = mean(exp.dat$ps[exp.dat$B.group=="B2"]) - mean(exp.dat$ps)
B2.effect
[1] -0.5759122

B3.effect = mean(exp.dat$ps[exp.dat$B.group=="B3"]) - mean(exp.dat$ps)
B3.effect
[1] 2.919692
```

In this example, we see that some of the B effects are much larger than the A effects. For example the B3 effect is about four times higher than the A2 effect. It's useful to report both the p-values and whether they are significant in addition to the effect size (as long as there aren't too many groups) to understand both whether the effect was statistically significance, and if so, what it's size of impact was on our y variable (photosynthesis).

After you determine that there are significant p-values for effects in an ANOVA model, you need to conduct a post-hoc test. We need to do this because the ANOVA test tells whether there is a significant difference among the groups, but it doesn't indicate which specific groups differed. One good option for a post-hoc test is the Tukey Honestly Significant Difference test (commonly called the Tukey HSD test). In R you can use the function TukeyHSD() to conduct this test on an ANOVA R object. This test does pairwise comparisons between all the groups to determine which are signficiantly different from each other.

```
#Run Tukey HSD test on output of ANOVA 2 model (aov.2)
#Note: if we had used aov() rather than lm() to create the aov.2 model, we
#could have left out the aov() part of this function
TukeyHSD(aov(aov.2))

Tukey multiple comparisons of means
```

```
    95% family-wise confidence level

Fit: aov(formula = aov.2)

$A.group
          diff       lwr     upr     p adj
A2-A1 1.192888 0.2653963 2.12038 0.0137136

$B.group
          diff       lwr      upr     p adj
B2-B1 1.767867 0.3946486 3.141086 0.0097683
B3-B1 5.263471 3.8902526 6.636690 0.0000000
B3-B2 3.495604 2.1223852 4.868823 0.0000031
```

In this case, we see that all the groups are significantly different in our post-hoc comparison (p<0.05 for all pairwise comparisons). Thus, we can conclude that all the A and B effects are indeed significantly different.

### *4.3 Two-factor ANOVA model with interactions*

We also might be interested in testing whether there are interacting effects among the two groups within our ANOVA model. We can test this by:

```
#asterisk indicates interactions
aov.3 = lm(ps ~ A.group*B.group, data=exp.dat)
summary(aov.3) #look at summary output of ANOVA model

Call:
lm(formula = ps ~ A.group * B.group, data = exp.dat)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.16604 -0.39364  0.09426  0.42789  1.65038 

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)           1.6168     0.3905   4.140  0.00037 ***
A.groupA2             2.6096     0.5523   4.725 8.36e-05 ***
B.groupB2             3.6633     0.5523   6.633 7.35e-07 ***
B.groupB3             5.4932     0.5523   9.946 5.46e-10 ***
A.groupA2:B.groupB2  -3.7908     0.7811  -4.853 6.03e-05 ***
A.groupA2:B.groupB3  -0.4594     0.7811  -0.588  0.56192    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8733 on 24 degrees of freedom
Multiple R-squared:  0.9056,  Adjusted R-squared:  0.8859 
F-statistic: 46.05 on 5 and 24 DF,  p-value: 1.585e-11
```

In looking at our summary table, we can determine whether the interaction effects were significant (for the interaction among group A1 and B2). If there are no interaction effects where p<0.05, we can conclude our analysis. However, in this example some of the interactions were significant. This is a case that warrants further investigation to understand the interaction effects.

```
TukeyHSD(aov(aov.3))
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = aov.3)

$A.group
          diff       lwr   upr     p adj
A2-A1 1.192888 0.5347761 1.851 0.0010111

$B.group
          diff      lwr      upr     p adj
B2-B1 1.767867 0.792596 2.743139 0.0003933
B3-B1 5.263471 4.288200 6.238743 0.0000000
B3-B2 3.495604 2.520333 4.470875 0.0000000

$`A.group:B.group`
                  diff        lwr       upr     p adj
A2:B1-A1:B1  2.6096177  0.9019560 4.3172793 0.0010519
A1:B2-A1:B1  3.6632697  1.9556081 5.3709314 0.0000100
A2:B2-A1:B1  2.4820828  0.7744211 4.1897445 0.0018588
A1:B3-A1:B1  5.4931637  3.7855021 7.2008254 0.0000000
A2:B3-A1:B1  7.6433967  5.9357351 9.3510584 0.0000000
A1:B2-A2:B1  1.0536520 -0.6540096 2.7613137 0.4217357
A2:B2-A2:B1 -0.1275349 -1.8351965 1.5801268 0.9998971
A1:B3-A2:B1  2.8835461  1.1758844 4.5912077 0.0003085
A2:B3-A2:B1  5.0337790  3.3261174 6.7414407 0.0000000
A2:B2-A1:B2 -1.1811869 -2.8888486 0.5264747 0.3020168
A1:B3-A1:B2  1.8298940  0.1222324 3.5375557 0.0307489
A2:B3-A1:B2  3.9801270  2.2724653 5.6877887 0.0000026
A1:B3-A2:B2  3.0110809  1.3034193 4.7187426 0.0001745
A2:B3-A2:B2  5.1613139  3.4536522 6.8689756 0.0000000
A2:B3-A1:B3  2.1502330  0.4425713 3.8578946 0.0080028
```

In this case, we see that there are some interaction effects among groups that are not significant in the post-hoc test. We should be careful when reporting our results to include these results from the post-hoc analysis flag interaction effects where $p<0.05$ as significant, and report those that weren’t significant as well.