

One-Way and Two-Way Analysis of Variance (ANOVA)

Lecture #11

Adi Sarid

Tel-Aviv University

updated: 2020-01-11

Reminder from Previous Lecture (One-Way ANOVA)

- We discussed experiment design and the motivation to analysis of variance
- A method to compare means in different groups (treatments)
- Leveraging the variance of the model vs. the variance of the errors (hence Analysis of Variance, ANOVA)

$$\circ Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n$$

- Fixed factors: when the factor by which we are comparing the populations, τ_i , has fixed levels. The a treatments are chosen
- Random factors: when the factor τ_i is a random variable, the a treatments are a random sample, conclusions are to be extended

Reminder from Previous Lecture (cont.)

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad i = 1, \dots, j = 1, \dots, n$$

The value $\mu_i = \mu + \tau_i$ is the mean value of the i th treatment.

We assume that the errors $\epsilon_{i,j}$ are normally and independently distributed $N(0, \sigma^2)$.

For **fixed factors** we are interested in the following test:

- $H_0: \tau_1 = \dots = \tau_a = 0$
- $H_1: \exists i \mid \tau_i \neq 0$

For **random factors** we are interested in the following test

- $H_0: \sigma_\tau^2 = 0$
- $H_1: \sigma_\tau^2 > 0$

The Sum of Squares

We use the sum of squares equation: $SS_T = SS_{\text{Treatments}} + SS_E$:

$$\sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 = n \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^a \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

In both fixed and random factor cases, we ended up with the same F-test statistic:

$$F_0 = \frac{SS_{\text{Treatments}} / (a - 1)}{SS_E / [a(n - 1)]} = \frac{MS_{\text{Treatments}}}{MS_E}$$

With $a - 1$ and $a(n - 1)$ degrees of freedom.

We use an upper-tail, one-sided critical region and reject H_0 if $f_0 > f_{\alpha, a-1, a(n-1)}$.

The ANOVA Table

The corresponding ANOVA table, for a Single-Factor Experiment, Fixed or Random-Effects Model:

Source of Variation	Sum of Squares	df	Mean Squares	F_0
Treatments	$SS_{\text{Treatments}}$	$a - 1$	$MS_{\text{Treatments}}$	$\frac{MS_{\text{Treatments}}}{MS_E}$
Error	SS_E	$a(n - 1)$	MS_E	
Total	SS_T	$an - 1$		

For an unbalanced experiment (each treatment has varying group size, n_i) we would have:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{N}$$

$$SS_{\text{Treatments}} = \sum_{i=1}^a \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{N}$$

And $SS_E = SS_T - SS_{\text{Treatment}}$

ANOVA Example

```
mtcars2 <- mtcars %>%
  mutate(cyl_fct = factor(cyl))
mtcars_anova <- aov(formula = mpg ~ cyl_fct, data = mtcars2)
summary(mtcars_anova)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## cyl_fct    2  824.8   412.4    39.7 4.98e-09 ***
## Residuals  29   301.3    10.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(mtcars_anova)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = mpg ~ cyl_fct, data = mtcars2)
##
## $cyl_fct
##           diff          lwr          upr      p adj
## 6-4  -6.920779 -10.769350 -3.072086  0.0003424
## 8-4 -11.563636 -14.770779 -8.3564942  0.0000000
## 8-6  -4.642857  -8.327583 -0.9581313  0.0112287
```

Usually, ANOVA will be followed by a multiple comparisons procedures, that will help identify which factors contribute to the variation.

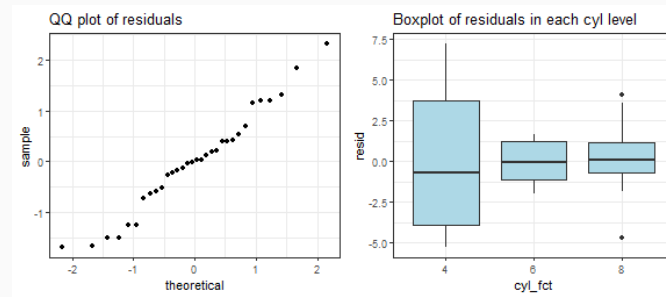
E.g., Tukey's HSD (`TukeyHSD`), and Dunnett's test (package `multcomp:glht`).

Verifying ANOVA assumptions

The ANOVA model assumes that observations are normally and independently distributed, with the same variance for each treatment.

This can be verified using hypothesis tests on the residuals, or viewing a proper plot (e.g., qqplot). The following illustrates that some assumptions are invalid in the previous example (which?)

```
mtcars2_resid <- mtcars2 %>%
  mutate(resid = mtcars_anova$residuals)
boxplot_resid <- ggplot(mtcars2_resid, aes(x = cyl_fct, y = resid)) +
  geom_boxplot(fill = "lightblue") +
  theme_bw() + ggtitle("Boxplot of residuals in each cyl level")
qqplot_resid <- ggplot(mtcars2_resid, aes(sample = (resid - mean(resid))/sd(resid))) +
  geom_qq() +
  theme_bw() + ggtitle("QQ plot of residuals")
cowplot::plot_grid(qqplot_resid, boxplot_resid)
```



Determining the Sample Size

The selection of the sample size is based on the difference we want to detect, and at what test power:

$$1 - \beta = P(\text{Reject } H_0 | H_1) = P(F_0 > f_{1-\alpha, a-1, a(n-1)} | H_1)$$

The effect size represents the differences in means between groups:

$$ES = \frac{\mu_{\text{experiment}} - \mu_{\text{control}}}{s}$$

```
pwr::pwr.anova.test(k = 3, n = NULL, f = 0.25, sig.level = 0.05, power = 0.8)
```

```
##
##   Balanced one-way analysis of variance power calculation
##
##           k = 3
##           n = 52.3966
##           f = 0.25
##   sig.level = 0.05
##   power = 0.8
##
## NOTE: n is number in each group
```


Random-Effects Model - Example

When manufacturing food and drugs (medication) we usually aim for having a high degree of consistency, so that different batches produce the same product (quality, concentration, etc.).

Consistency is not trivial to produce: production lines malfunction, shifts change, and machinery is sometimes replaced.

The random effects model can be used to determine if consistency of the production lines should be rejected.

The model - Olive Oil Acidity

- We are manufacturing olive oil, and each batch should have an acidity level of 0.5%
- The factory wants to design an experiment which will test if the acidity is consistent.
- Questions: What is our factor? What are the factor levels?

Example - Olive Oil Acidity

The batch is our factor. It has many levels and therefore we treat it as a random effect.

$$\text{acidity} = \mu + \tau_{\text{batch}} + \epsilon_{\text{batch},j}$$

```
oil_manufacturing <- tibble(  
  batch = factor(c("A", "A", "A", "A", "B", "B", "B", "B",  
                  "C", "C", "C", "C", "D", "D", "E", "E")),  
  acidity =  
    c(0.45, 0.49, 0.51, 0.50, 0.49, 0.60, 0.55, 0.41, 0.36,  
      0.50, 0.51, 0.49, 0.48, 0.50, 0.51, 0.08, 0.20))  
  
oil_aov <- aov(formula = acidity ~ batch, data = oil_manufacturing)  
summary(oil_aov)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)      
## batch      4 0.21707 0.05427   13.43 0.000219 ***  
## Residuals 12 0.04851 0.00404                  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(oil_aov)
```

```
## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##  
## Fit: aov(formula = acidity ~ batch, data = oil_manufacturing)  
##  
## $batch  
##      diff      lwr      upr    p adj  
## B-A -0.0055 -0.1414408 0.1304408 0.9999276  
## C-A 0.0075 -0.1357942 0.1507942 0.9997988  
## D-A 0.0175 -0.1579988 0.1929988 0.9974654  
## E-A -0.3475 -0.5229988 -0.1720012 0.0003058  
## C-B 0.0130 -0.1229408 0.1489408 0.9978453
```

Example - Olive Oil Acidity (cont.)

We can use Dunnett's test to specify the exact contrasts we would like to examine:

```
suppressWarnings(suppressMessages(library(multcomp)))
glht(oil_aov,
  linfct = mcp(batch = c("E-A=0", "E-B=0", "E-C=0", "E-D=0"))) %>%
  summary()

##
##      Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
##
## Fit: aov(formula = acidity ~ batch, data = oil_manufacturing)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## E - A == 0 -0.34750    0.05506  -6.311  <0.001 ***
## E - B == 0 -0.34200    0.05319  -6.429  <0.001 ***
## E - C == 0 -0.35500    0.05506  -6.448  <0.001 ***
## E - D == 0 -0.36500    0.06358  -5.741  <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Randomized Complete Block Design

Sometimes we have an additional *nuisance* factor which is not the aim of the study. For example, imagine a clinical research where a treatments are used in b medical centers.

It might be that the fact that there are varying medical centers in itself influences the results.

If the selection of treatments is randomized across medical centers, the experiment is called a randomized complete block design.

	Blocks					
Treatments	1	2	...	b	Totals	Averages
1	y_{11}	y_{12}	...	y_{1b}	$y_{1\cdot}$	$\bar{y}_{1\cdot}$
\vdots			\vdots			\vdots
a	y_{a1}	y_{a2}	...	y_{ab}	$y_{a\cdot}$	$\bar{y}_{a\cdot}$
Totals	$y_{\cdot 1}$	$y_{\cdot 2}$...	$y_{\cdot b}$	$y_{\cdot \cdot}$	
Averages	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$...	$\bar{y}_{\cdot b}$		

Randomized Complete Block Design (cont.)

The formula for this design is given by:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad i \in \{1, \dots, a\}, j \in \{1, \dots, b\}$$

Where

- μ is the overall mean.
- τ_i is the effect of the i th treatment.
- β_j is the effect of the j th block.
- ϵ_{ij} is the error term $\epsilon_{ij} \sim N(0, \sigma)$

We assume that the treatments and blocks are fixed factors, and that they are deviations from the overall mean, so $\sum_{i=1}^a \tau_i = 0$ and $\sum_{j=1}^b \beta_j = 0$. We test the hypothesis:

- $H_0: \tau_1 = \dots = \tau_a = 0$
- $H_1: \exists i: \tau_i \neq 0$

Sum of Squares Identity in the Randomized Complete Block Design

The sum of squares identity can be broken into

$$\sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 = b \sum_{i=1}^a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^b (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i,j} (y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$$

Symbolically stated as

$$SS_T = SS_{\text{Treatments}} + SS_{\text{Blocks}} + SS_E$$

With degrees of freedom:

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

Randomized Complete Block Design Hypothesis Test

Set:

$$MS_{\text{Treatments}} = \frac{SS_{\text{Treatments}}}{a - 1}, \quad MS_{\text{Blocks}} = \frac{SS_{\text{Blocks}}}{b - 1}, \quad MS_E = \frac{SS_E}{(a - 1)(b - 1)}$$

The mean sum of squares is given by:

- $E(MS_{\text{Treatments}}) = \sigma^2 + \frac{b \sum_{i=1}^a \tau_i^2}{a - 1}$
- $E(MS_{\text{Blocks}}) = \sigma^2 + \frac{a \sum_{j=1}^b \beta_j^2}{b - 1}$
- $E(MS_E) = \sigma^2$

If H_0 is true and all $\tau_i = 0$ then:

$$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$$

Is F -distributed with $a - 1$ and $(a - 1)(b - 1)$ degrees of freedom.

Example for Randomized Complete Block Design - ANOVA

An experiment was performed to determine the effect of four chemicals on the fabric strength (Example 13.5 from Montgomery).

```
fabric_strength <- read_csv("https://raw.githubusercontent.com/adisarid/intro_statistics_R/master/lectures/data/montgomery_13.5_fabric_strength.csv", col_types = cols()) %>%
  pivot_longer(cols = -chemical, names_to = "fabric_sample", values_to = "strength") %>% mutate(chemical = factor(chemical))

fabric_aov <- aov(formula = strength ~ chemical + fabric_sample, data = fabric_strength)
summary(fabric_aov)

##           Df Sum Sq Mean Sq F value    Pr(>F)
## chemical    3 18.044    6.015    75.89 4.52e-08 ***
## fabric_sample 4  6.693    1.673    21.11 2.32e-05 ***
## Residuals   12  0.951    0.079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

glt(fabric_aov, linfct = mcp(chemical = "Tukey")) %>%
  summary()

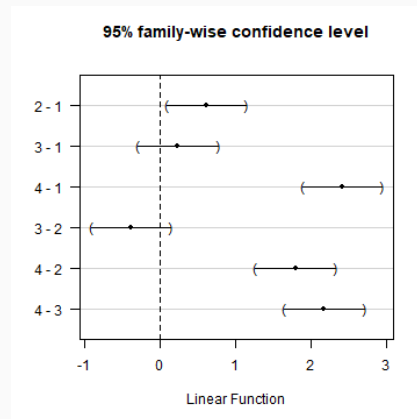
##
##      Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: aov(formula = strength ~ chemical + fabric_sample, data = fabric_strength)
##
## Linear Hypotheses:
##           Estimate Std. Error t value Pr(>|t|)
## 2 - 1 == 0    0.620     0.178   3.482  0.0203 *
## 3 - 1 == 0    0.240     0.178   1.348  0.5523
## 4 - 1 == 0    2.420     0.178  13.592 <0.001 ***
## 3 - 2 == 0   -0.380     0.178  -2.134  0.1973
## 4 - 2 == 0    1.800     0.178  10.110 <0.001 ***
```


Visualizing Multiple Comparisons

The `multcomp` package also includes a basic visualization to illustrate the differences in means.

The chart can be invoked by using `plot` on a `glht` object.

```
plot(glht(fabric_aov, linfct = mcp(chemical = "Tukey")))
```



Two-Way ANOVA (Design of Experiments with Several Factors)

So far we treated a single factor experiment using one-way ANOVA, with a minor generalization towards a block design with a nuisance factor.

We are going to generalize this to experiments with several factors, using two-way ANOVA (Montgomery Ch.14).

For example, the population are students with:

- Faculty - a factor describing the faculty the student came from
- Semester - a factor describing which semester the student is currently at
- Y - the dependent variable is leisure time (free time)

We would like to explore the leisure time as a function of faculty, semester, and the interaction between them.

Factorial Experiments

A factorial experiment is an experiment with all possible combinations of factor levels (all combinations of faculties and semesters). For example:

- Include Exact Sciences, Social Sciences,
- Students from semester 1 and semester 6

We will investigate all possible combinations of faculties and semesters in our analysis.

Question: What is the difference between the two examples?

Example 1	S.1	S.6	Example 2	S.1	S.6
------------------	------------	------------	------------------	------------	------------

Exact	5	2.5	Exact	5	2.5
-------	---	-----	-------	---	-----

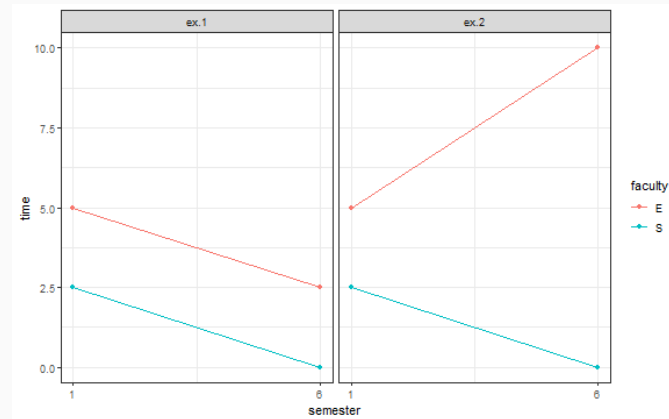
Social	2.5	0	Social	2.5	10
--------	-----	---	--------	-----	----

Which example illustrates:

- Main effects of Faculty/Semester?
- Interaction effects between Faculty and Semester?

Illustration of Interaction

```
leisure_time_ex1 <- tibble(faculty = c("E", "E", "S", "S"),
                           semester = c(1, 6, 1, 6),
                           time = c(5, 2.5, 2.5, 0),
                           ex = "ex.1")
leisure_time_ex2 <- mutate(leisure_time_ex1, ex = "ex.2")
leisure_time_ex2$time[2] <- 10
leisure <- bind_rows(leisure_time_ex1, leisure_time_ex2)
ggplot(leisure, aes(x = semester, color = faculty, y = time)) +
  geom_point() + geom_line() +
  facet_wrap(~ ex) +
  scale_x_continuous(breaks = c(1,6)) +
  theme_bw()
```



Warning: Non-Factorial Experiments

Changing the factors one at a time might miss interaction effects.

Example 2 S.1 S.6

Exact	5	2.5
-------	---	-----

Social	2.5	10
--------	-----	----

For example if we are conducting an experiment to see which students has the most leisure time, in a non-factorial experiment, checking factors one at a time instead we would:

- Compare semester 1 and semester 6 for exact sciences, and detect that semester 1 has more leisure time
- Then, compare Exact versus social sciences in semester 1 only
- Deduce that Exact sciences students in semester 1 has the most leisure time
- Miss the interaction of Faculty and Semester (Social science at semester 6 has the most leisure time).

Two-Factor Factorial Experiments

Fixed factors: in each group of factor A and factor B we randomly sample n observations. In total, we end up with abn . The sampling order is completely random, hence the experiment is of *completely randomized design*.

We describe the relationship by the following linear statistical model:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

For $i = 1, \dots, a$ $j = 1, \dots, b$ $k = 1, \dots, n$, where

- τ_i is the effect of the i th level of factor A
- β_j is the effect of the j th level of factor B
- $(\tau\beta)_{ij}$ is the effect of the interaction between A and B
- ϵ_{ijk} is the random error component, $N(0, \sigma^2)$

Question: what is the difference from the single factor block design?

Statistical Hypothesis in the Fixed Two Factor Design

We will examine three distinct hypothesis:

Factor A

- $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$
- $H_1: \exists i: \tau_i \neq 0$

Factor B

- $H_0: \beta_1 = \beta_2 = \dots = \beta_a = 0$
- $H_1: \exists i: \beta_i \neq 0$

Interaction of A and B

- $H_0: (\tau\beta)_{11} = (\tau\beta)_{12} = \dots = (\tau\beta)_{ab} = 0$
- $H_1: \exists i, j: (\tau\beta)_{ij} \neq 0$

Sum of Squares Breakdown in the Fixed Two Factor Design

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

Where

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y} \dots)^2$$

$$SS_A + SS_B = bn \sum_{i=1}^a (\bar{y}_{i..} - \bar{y} \dots)^2 + an \sum_{j=1}^b (\bar{y}_{.j.} - \bar{y} \dots)^2$$

$$SS_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y} \dots)^2$$

$$SS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2$$

The Test Statistics for a Fixed Two Factor Design

The mean square errors are given by:

$$MS_A = \frac{SS_A}{a-1}, \quad MS_B = \frac{SS_B}{b-1}, \quad MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}, \quad MS_E = \frac{SS_E}{ab(n-1)}$$

The sizes $MS_A, MS_B, MS_{AB}, MS_E$ are all unbiased estimators for σ^2 given the null hypothesis.

Similar to what we've seen in the single factor experiment design (in the previous lecture), it can be shown that:

$$E(MS_A) = E\left[\frac{SS_A}{a-1}\right] = \sigma^2 + \frac{bn \sum_{i=1}^a \tau_i^2}{a-1}, \quad E[MS_E] = \sigma^2$$

Hence we can use the following test statistics:

$$F_0 = \frac{MS_A}{MS_E}, \quad F_0 = \frac{MS_B}{MS_E}, \quad F_0 = \frac{MS_{AB}}{MS_E}$$

Two-Way ANVOA Table Example

Example 14.5 from Montgomery: Adhesion Force by Primer Type (1-4) and Application Method (Dipping and Spraying)

```
montgomery14.5 <- read_csv("https://raw.githubusercontent.com/adisarid/intro_statistics_R/a6b19adadb793317e53f8529252bd4bc0df557a7/lectures/data/montgomery_14.5_adhesion_force.csv", col_types = "ffn")
glimpse(montgomery14.5)

## Observations: 18
## Variables: 3
## $ primer_type    <fct> 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3...
## $ application_method <fct> Dipping, Dipping, Dipping, Spraying, Sprayi...
## $ adhesion_force  <dbl> 4.0, 4.5, 4.3, 5.4, 4.9, 5.6, 5.6, 4.9, 5.4...

adhesion_aov <- aov(formula = adhesion_force ~ primer_type + application_method + primer_type*application_method,
  data = montgomery14.5)

summary(adhesion_aov)

##              Df Sum Sq Mean Sq F value    Pr(>F)
## primer_type      2  4.581    2.291   27.858 3.10e-05 ***
## application_method 1  4.909    4.909   59.783 5.36e-06 ***
## primer_type:application_method 2  0.241    0.121    1.466    0.269
## Residuals       12  0.987    0.082
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two-Way ANOVA: Multiple Comparisons

```
TukeyHSD(adhesion_aov)

## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = adhesion_force ~ primer_type + application_method + primer_type * application_method, data = montgomery14.5)
##
## $primer_type
##      diff      lwr      upr      p adj
## 2-1  0.9000000  0.4583303  1.3416697  0.0004100
## 3-1 -0.2833333 -0.7250030  0.1583364  0.2409687
## 3-2 -1.1833333 -1.6250030 -0.7416636  0.0000323
##
## $application_method
##      diff      lwr      upr      p adj
## Spraying-Dipping 1.044444  0.7499289  1.33896  5.4e-06
##
## $'primer_type:application_method'
##      diff      lwr      upr      p adj
## 2:Dipping-1:Dipping  1.033333e+00  0.24692384  1.8197428  0.0084706
## 3:Dipping-1:Dipping -4.333333e-01 -1.21974283  0.3530762  0.4724075
## 1:Spraying-1:Dipping  1.033333e+00  0.24692384  1.8197428  0.0084706
## 2:Spraying-1:Dipping  1.800000e+00  1.01359051  2.5864095  0.0000645
## 3:Spraying-1:Dipping  9.000000e-01  0.11359051  1.6864095  0.0220643
## 3:Dipping-2:Dipping -1.466667e+00 -2.25307616 -0.6802572  0.0004605
## 1:Spraying-2:Dipping -8.881784e-16 -0.78640949  0.7864095  1.0000000
## 2:Spraying-2:Dipping  7.666667e-01 -0.01974283  1.5530762  0.0575657
## 3:Spraying-2:Dipping -1.333333e-01 -0.91974283  0.6530762  0.9913119
## 1:Spraying-3:Dipping  1.466667e+00  0.68025717  2.2530762  0.0004605
## 2:Spraying-3:Dipping  2.233333e+00  1.44692384  3.0197428  0.0000070
## 3:Spraying-3:Dipping  1.333333e+00  0.54692384  2.1197428  0.0010825
## 2:Spraying-1:Spraying  7.666667e-01 -0.01974283  1.5530762  0.0575657
## 3:Spraying-1:Spraying -1.333333e-01 -0.91974283  0.6530762  0.9913119
## 3:Spraying-2:Spraying -9.000000e-01 -1.68640949 -0.1135905  0.0220643
```