# STAT 509: Statistics for Engineers

## Chapters 13-14: Single-/Two-Factor Factorial Design

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

# Chapters 13-14: Single-/Two-Factor Factorial Design

Learning Objectives:

1. Design and conduct engineering experiments involving one factor or two factors with an arbitrary number of levels
2. Understand how the analysis of variance is used to analyze the data from these experiments
3. Use multiple comparison procedures to identify specific differences between means

# Introduction

Experiments are a natural part of the engineering and scientific decision-making process. Suppose, for example, that a civil engineer is investigating the effects of different curing methods on the mean compressive strength of concrete. The experiment would consist of making up several test specimens of concrete using each of the proposed curing methods and then testing the compressive strength of each specimen. The data from this experiment could be used to determine which curing method should be used to provide maximum mean compressive strength.

If there are only two curing methods of interest, this experiment could be designed and analyzed using the statistical hypothesis methods for two samples introduced in Chapter 10. That is, the experimenter has a single factor of interest—curing methods—and there are only two levels of the factor. The t-test can be used to decide if the two means differ.

# Introduction

Many single-factor experiments require that more than two levels of the factor be considered. For example, the civil engineer may want to investigate five different curing methods. In this chapter, we show how the **analysis of variance** (frequently abbreviated **ANOVA**) can be used for comparing means when there are more than two levels of a single factor. We also discuss randomization of the experimental runs and the important role this concept plays in the overall experimentation strategy.

# Example: Tensile Strength

A manufacturer of paper is interested in improving the product's tensile strength. Product engineering believes that tensile strength is a function of the hardwood concentration in the pulp and that the range of hardwood concentrations of practical interest is between 5 and 20%. A team of engineers responsible for the study decides to investigate four levels of hardwood concentration: 5%, 10%, 15%, and 20%. They decide to make up six test specimens at each concentration level by using a pilot plant. All 24 specimens are tested on a laboratory tensile tester in random order. The data from this experiment are

| Hardwood Concentration (%) | Observations | | | | | | Totals | Averages |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | | |
| 5 | 7 | 8 | 15 | 11 | 9 | 10 | 60 | 10.00 |
| 10 | 12 | 17 | 13 | 18 | 19 | 15 | 94 | 15.67 |
| 15 | 14 | 18 | 19 | 17 | 16 | 18 | 102 | 17.00 |
| 20 | 19 | 25 | 22 | 23 | 18 | 20 | 127 | 21.17 |
| | | | | | | | 383 | 15.96 |

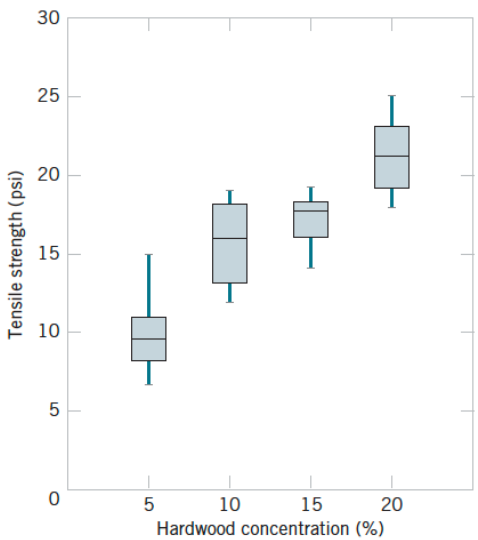
# Example: Tensile Strength

This is an example of a completely randomized **single-factor** experiment with **four** levels of the factor. The levels of the factor are sometimes called **treatments**, and each treatment has six observations or replicates. **The role of randomization in this experiment is extremely important**. By randomizing the order of the 24 runs, the effect of any nuisance variable that may influence the observed tensile strength is approximately balanced out.

For example, suppose that there is a warm-up effect on the tensile testing machine; that is, the longer the machine is on, the greater the observed tensile strength. If all 24 runs are made in order of increasing hardwood concentration (that is, all six 5% concentration specimens are tested first, followed by all six 10% concentration specimens, etc.), any observed differences in tensile strength could also be due to the warm-up effect.

# Example: Tensile Strength

If possible, the first step is always to graphically analyze the data from a designed experiment. The figure below presents box plots of tensile strength at the four hardwood concentration levels. This figure indicates that changing the hardwood concentration has an effect on tensile strength; specifically, higher hardwood concentrations produce higher observed tensile strength. Furthermore, the distribution of tensile strength at a particular hardwood level is reasonably symmetric, and the variability in tensile strength does not change dramatically as the hardwood concentration changes.




Graphical interpretation of the data is always useful. Box plots show the variability of the observations **within** a treatment (factor level) and the variability **between** treatments. We now discuss how the data from a single-factor randomized experiment can be analyzed statistically.

# One-Factor Factorial Experiments

Suppose that we have $a$ different levels (or treatments) of a single factor that we wish to compare. Say $Y_{ij}$ represents the $j$th observation taken under treatment $i$ for $j = 1, \ldots, n$ and $i = 1, \ldots, a$. We describe the observations by the linear statistical model

$$Y_{ij} = \mu_i + \epsilon_{ij} = \mu + \tau_i + \epsilon_{ij},$$

$j = 1, \ldots, n$ and $i = 1, \ldots, a$, where

- $\mu$ is a parameter common to all treatments called the **overall mean**
- $\tau_i$ is a parameter associated with the $i$th treatment called the $i$th **treatment effect** and satisfies a constraint that

$$\sum_{i=1}^{a} \tau_i = 0,$$

- $\epsilon_{ij} \sim N(0, \sigma^2)$ is a random error component.

, The unknown parameters are $\mu_i = \mu + \tau_i$ for $i = 1, \ldots, a$ and $\sigma^2$.

# Analysis of Variance

We are interested in testing the equality of the $a$ treatment means $\mu_1, \mu_2, \ldots, \mu_a$; i.e., $H_0 : \mu_1 = \mu_2 = \cdots = \mu_a$ versus $H_1$: at least two of $\mu_i$'s are unequal. This is equivalent to testing the hypotheses

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \text{ versus } H_1 : \tau_i \neq 0 \text{ for at least one } i.$$

if the null hypothesis is true, each observation consists of the overall mean $\mu$ plus a realization of the random error component $\epsilon_{ij}$. This is equivalent to saying that all $N$ observations are taken from a normal distribution with mean $\mu$ and variance $\sigma^2$. Therefore, if the null hypothesis is true, changing the levels of the factor has no effect on the mean response.

We use an analysis of variance to conduct the test.

# Analysis of Variance

Let

$$y_{i\cdot} = \sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{i\cdot} = y_{i\cdot}/n, \quad i = 1, 2, \ldots, a$$

$$y_{\cdot\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}, \quad \bar{y}_{\cdot\cdot} = y_{\cdot\cdot}/N$$

where $N = an$ is the total number of observations and the "dot" subscript notation implies summation over the subscript that it replaces.

The total variability in the entire data is described by the **total sum of squares**

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

Then we partition the total variability into two component parts.

## ANOVA Sum of Squares Identity

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

or symbolically

$$SS_T = SS_{treatments} + SS_E$$

and degrees of freedom can be partitioned as

$$an - 1 = a - 1 + a(n-1)$$

or

$$df_{Total} = df_{Treatments} + df_{Error}.$$

The $SS_{treatments}$ is known as the **treatment sum of squares**. The $SS_E$ is the **error sum of squares**.

# Analysis of Variance

The ratio

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}$$

is called the **mean square for treatments** and the **mean square for error** is

$$MS_{Error} = \frac{SS_{Error}}{a(n-1)}.$$

We have

$$E[MS_{Treatments}] = \sigma^2 + \underbrace{\frac{n\sum_{i=1}^{a}\tau_i^2}{a-1}}_{\geq 0},$$

$$E[MS_{Error}] = \sigma^2.$$

Only if $H_0 : \tau_1 = \cdots = \tau_a = 0$ is true, $E[MS_{Treatments}] = E[MS_{Error}]$; otherwise, $E[MS_{Treatments}] > E[MS_{Error}]$.

# Analysis of Variance

To test

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \text{ versus } H_1 : \tau_i \neq 0 \text{ for at least one } i,$$

our test statistic is

$$F_0 = \frac{MS_{Treatments}}{MS_{Error}}.$$

Reject $H_0$ at significance level $\alpha$ if

$$F_0 > f_{a-1,a(n-1),alpha},$$

or if $P$-value is less than $\alpha$.

**TABLE 13.3** **Analysis of Variance for a Single-Factor Experiment, Fixed-Effects Model**

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

# A Multiple Comparison Following the ANOVA

When the null hypothesis $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ is rejected in the ANOVA, we know that some of the treatment or factor-level means are different. However, the ANOVA does not identify which means are different. Methods for investigating this issue are called **multiple comparisons methods**. Many of these procedures are available. R package provide a simple function `TukeyHSD` to identify which pairs of treatments have different means.

# Example: Tensile Strength

Consider the paper tensile strength experiment described previously. This experiment is a completely randomized design.

(a) Test the hypothesis that different hardwood concentrations do not affect the mean tensile strength of the paper.

(b) If different hardwood concentrations affects the mean tensile strength of the paper, identify which pairs of treatments have different means.

**Solution:** The four treatments are labeled by the concentration level. Let $\tau_1, \ldots, \tau_4$ denote the treatment effects. We wish to test

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_4 = 0 \text{ versus } H_1 : \tau_i \neq 0 \text{ for at least one } i,$$

# Example: Tensile Strength

**Solution (continued):** We read the data

```
> my_data=read.csv("https://raw.githubusercontent.com/Harrindy/StatEngine/
                   master/Data/TensileStrength.csv")
> head(my_data,2)
  Concentration Strength
1             5        7
2             5        8
# We need to tell R Concentration levels are the factors
> my_data$Concentration=as.factor(my_data$Concentration)

# Run ANOVA test
> res.aov=aov(Strength~Concentration,data=my_data)
> summary(res.aov)
             Df Sum Sq Mean Sq F value   Pr(>F)
Concentration  3  382.8  127.60   19.61 3.59e-06 ***
Residuals     20  130.2    6.51
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We see the $P$-value is significantly smaller than any commonly used $\alpha$. Thus, we reject $H_0$.

# Example: Tensile Strength

**Solution (continued):** Now Part (b):

```
> TukeyHSD(res.aov,conf.level=0.95)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Strength ~ Concentration, data = my_data)

$Concentration
            diff          lwr       upr     p adj
10-5   5.666667  1.54410408  9.789229 0.0051108
15-5   7.000000  2.87743741 11.122563 0.0006501
20-5  11.166667  7.04410408 15.289229 0.0000015
15-10  1.333333 -2.78922925  5.455896 0.8022275
20-10  5.500000  1.37743741  9.622563 0.0065966
20-15  4.166667  0.04410408  8.289229 0.0470251
```

For each pair, it gives a 95% confidence interval of $\mu_i - \mu_j$ by considering mutiple comparison. All we need to focus on is the adjusted $P$-value. Only the pair between the 10% and 15% concentration treatments does not have a significance mean difference.

# Two-Factor Factorial Experiments

In practice, the experiment may invovle two factors, say $A$ and $B$. There are $a$ levels of factor $A$ and $b$ levels of factor $B$ (shown as)

| | | Factor $B$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | $\cdots$ | $b$ | Totals | Averages |
| | 1 | $y_{111}, y_{112}, \cdots, y_{11n}$ | $y_{121}, y_{122}, \cdots, y_{12n}$ | | $y_{1b1}, y_{1b2}, \cdots, y_{1bn}$ | $y_{1\cdot\cdot}$ | $\bar{y}_{1\cdot\cdot}$ |
| **Factor $A$** | 2 | $y_{211}, y_{212}, \cdots, y_{21n}$ | $y_{221}, y_{222}, \cdots, y_{22n}$ | | $y_{2b1}, y_{2b2}, \cdots, y_{2bn}$ | $y_{2\cdot\cdot}$ | $\bar{y}_{2\cdot\cdot}$ |
| | $\vdots$ | | | | | | |
| | $a$ | $y_{a11}, y_{a12}, \cdots, y_{a1n}$ | $y_{a21}, y_{a22}, \cdots, y_{a2n}$ | | $y_{ab1}, y_{ab2}, \cdots, y_{abn}$ | $y_{a\cdot\cdot}$ | $\bar{y}_{a\cdot\cdot}$ |
| Totals | | $y_{\cdot 1\cdot}$ | $y_{\cdot 2\cdot}$ | | $y_{\cdot b\cdot}$ | $y_{\cdots}$ | |
| Averages | | $\bar{y}_{\cdot 1\cdot}$ | $\bar{y}_{\cdot 2\cdot}$ | | $\bar{y}_{\cdot b\cdot}$ | | $\bar{y}_{\cdots}$ |

The experiment has n replicates, and each replicate contains all $ab$ treatment combinations. The observation in the $ij$-th cell for the $k$th replicate is denoted by $y_{ijk}$. In performing the experiment, the $abn$ observations would be run in random order. Thus, like the single-factor experiment, the two-factor factorial is a *completely randomized design*.

# Two-Factor Factorial Experiments

The observations may be described by the linear statistical model

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

for $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$, and $k = 1, 2, \ldots, n$, where

- $\mu$ is the overall mean effect
- $\tau_i$ is the effect of the $i$th level of factor $A$
- $\beta_j$ is the effect of the $j$th level of factor $B$
- $(\tau\beta)_{ij}$ is the **effect of the interaction** between $A$ and $B$
- $\epsilon_{ijk} \sim N(0, \sigma^2)$ is a random error component.

We have a constraint:

$$\sum_{i=1}^{a} \tau_i = 0, \sum_{j=1}^{b} \beta_j = 0, \sum_{i=1}^{a} (\tau\beta)_{ij} = 0, \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

We are interested in testing the hypotheses of no main effect for factor $A$, no main effect for $B$, and no $AB$ interaction effect.

# Two-Factor Factorial Experiments

The hypotheses that we test are as follows:

1. No main effect of factor $A$:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0 \text{ versus } H_1 : \text{ at least one } \tau_i \neq 0.$$

2. No main effect of factor $B$:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0 \text{ versus } H_1 : \text{ at least one } \beta_j \neq 0.$$

3. No interaction:

$$H_0 : (\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab} = 0 \text{ versus}$$
$$H_1 : \text{ at least one } (\tau\beta)_{ij} \neq 0.$$

# Analysis of Variance

The tests can be done via analysis of variance.

$$
\begin{aligned}
& SS_T(\text{total sum of squares}) \\
= \; & SS_A(\text{sum of squares for factor } A) \\
& + SS_B(\text{sum of squares for factor } B) \\
& + SS_{AB}(\text{sum of squares for the interaction between } A \text{ and } B) \\
& + SS_E(\text{error sum of squares}).
\end{aligned}
$$

The degrees of freedom are

$$
\underbrace{abn-1}_{df_{Total}} = \underbrace{(a-1)}_{df_A} + \underbrace{(b-1)}_{df_B} + \underbrace{(a-1)(b-1)}_{df_{AB}} + \underbrace{ab(n-1)}_{df_{Error}}.
$$

# Analysis of Variance

The mean squares are

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

And their expectations are

$$E[MS_A] = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}, \quad E[MS_B] = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$
$$E[MS_{AB}] = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}, \quad E[MS_E] = \sigma^2.$$

# Test statistics

1. $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ versus $H_1$ : at least one $\tau_i \neq 0$, we use
$$F_0 = \frac{MS_A}{MS_E}, \text{ reject } H_0 \text{ if } F_0 > f_{a-1,ab(n-1),\alpha}.$$

2. $H_0 : \beta_1 = \beta_2 = \cdots = \beta_b = 0$ versus $H_1$ : at least one $\beta_j \neq 0$, we use
$$F_0 = \frac{MS_B}{MS_E}, \text{ reject } H_0 \text{ if } F_0 > f_{b-1,ab(n-1),\alpha}.$$

3. $H_0 : (\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab} = 0$ versus $H_1$ : at least one $(\tau\beta)_{ij} \neq 0$, we use
$$F_0 = \frac{MS_{AB}}{MS_E}, \text{ reject } H_0 \text{ if } F_0 > f_{(a-1)(b-1),ab(n-1),\alpha}.$$

Always use the $P$-value approach!

# ANOVA Table

Here is the structure of an ANOVA table most statistical software produces:

**TABLE 14.4 ANOVA Table for a Two-Factor Factorial, Fixed-Effects Model**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ |
|---|---|---|---|---|
| $A$ treatments | $SS_A$ | $a-1$ | $MS_A = \frac{SS_A}{a-1}$ | $\frac{MS_A}{MS_E}$ |
| $B$ treatments | $SS_B$ | $b-1$ | $MS_B = \frac{SS_B}{b-1}$ | $\frac{MS_B}{MS_E}$ |
| Interaction | $SS_{AB}$ | $(a-1)(b-1)$ | $MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$ | $\frac{MS_{AB}}{MS_E}$ |
| Error | $SS_E$ | $ab(n-1)$ | $MS_E = \frac{SS_E}{ab(n-1)}$ | |
| Total | $SS_T$ | $abn-1$ | | |

Note: if $H_0$ is rejected, we still do not know which pairs are significant different in means. Again, we can use `TukeyHSD`.

# Example: Aircraft Primer Paint

Aircraft primer paints are applied to aluminum surfaces by two methods: dipping and spraying. The purpose of using the primer is to improve paint adhesion, and some parts can be primed using either application method. The process engineering group responsible for this operation is interested in learning whether three different primers differ in their adhesion properties. A factorial experiment was performed to investigate the effect of paint primer type and application method on paint adhesion. For each combination of primer type and application method, three specimens were painted, then a finish paint was applied and the adhesion force was measured. The data from the experiment are shown

| Primer Type | Dipping | Spraying | $y_{i\cdot\cdot}$ |
|---|---|---|---|
| 1 | 4.0, 4.5, 4.3 | 5.4, 4.9, 5.6 | 28.7 |
| 2 | 5.6, 4.9, 5.4 | 5.8, 6.1, 6.3 | 34.1 |
| 3 | 3.8, 3.7, 4.0 | 5.5, 5.0, 5.0 | 27.0 |
| $y_{\cdot j\cdot}$ | 40.2 | 49.6 | $89.8 = y_{\cdots}$ |

# Example: Aircraft Primer Paint

**Solution:** Run analysis of variance:

```
> my_data=read.csv("https://raw.githubusercontent.com/Harrindy/StatEngine/
                   master/Data/AdhesionForce.csv")
> head(my_data,2)
  Primer Method Adhesion
1      1    Dip      4.0
2      1    Dip      4.5

# We need to tell R Primer and Method are the two factors
> my_data$Primer=as.factor(my_data$Primer)
> my_data$Method=as.factor(my_data$Method)

# Run ANOVA test
> res.aov=aov(Adhesion~Primer+Method+Primer:Method,data=my_data)
> summary(res.aov)
              Df Sum Sq Mean Sq F value   Pr(>F)
Primer         2  4.581   2.291  27.858 3.10e-05 ***
Method         1  4.909   4.909  59.703 5.36e-06 ***
Primer: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
```

# Example: Aircraft Primer Paint

**Solution:** Based on the three $P$-values, we have

1. for the test on the factor $A$ (Primer Type): $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ versus $H_1$ : at least one $\tau_i \neq 0$, the $P$-value is $3.10e-05$ much smaller than any commonly used $\alpha$, we reject $H_0$ (Primer Type affects the response).
2. for the test on the factor $B$ (Method: Dipping or Spraying): $H_0 : \beta_1 = \beta_2 = 0$ versus $H_1$ : at least one $\beta_j \neq 0$, the $P$-value is $5.36e-06$ much smaller than any commonly used $\alpha$, we reject $H_0$ (Method affects the response).
3. For the test on the interaction, we have $P$-value $0.269$ is quite large, we conclude that the data do not provided sufficient evidence to reject the null hypothesis (the interaction might not affect the response).

# Example: Aircraft Primer Paint

**Solution:** Because the data concluded Primer Type affects the response (at least two Primer Types produce different means), but we do not know which two. To find out, we can apply the `TukeyHSD`

```
> TukeyHSD(res.aov,'Primer')
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Adhesion ~ Primer + Method + Primer:Method, data = my_data)

$Primer
          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
```

We see that only the pair, Primer types 3 and 1, do not provide significantly different means.

# Summary

One can also apply the R program to more than 2-factors factorial designs! The steps are the same.

```
# Example 14.2 in the textbook, where we have three factors.
# Each has two levels.
# Read data
> my_data=read.csv("https://raw.githubusercontent.com/Harrindy/StatEngine/
                    master/Data/CodedSurfaceRoughness.csv")

# We need to tell R all the factors
> head(my_data,2)
> my_data$Feed=as.factor(my_data$Feed)
> my_data$Depth=as.factor(my_data$Depth)
> my_data$Angle=as.factor(my_data$Angle)

# Run ANOVA to find P-value fot the tests
> res.aov=aov(Roughness~Feed+Depth+Angle+
            Feed:Depth+Feed:Angle+Depth:Angle+
            Feed:Depth:Angle,data=my_data)
> summary(res.aov)

# If rejecting H0, TukeyHSD finds which pairs caused the difference.
> TukeyHSD(res.aov,'Feed')
```