[MACHINE LEARNING](https://stepupanalytics.com/category/machine-learning/)[REGRESSION](https://stepupanalytics.com/category/regression/)

Two Way ANOVA Calculation By Hand

<https://stepupanalytics.com/two-way-anova-calculation-by-hand-analysis-of-variance/>

[**MOHAMMAD SAJID**](https://stepupanalytics.com/author/sajid/)Nov 6, 2016 [0](https://stepupanalytics.com/two-way-anova-calculation-by-hand-analysis-of-variance/#respond)

* FACEBOOK
* TWITTER
* GOOGLE+

The two-way ANOVA compares the mean differences between groups that have been split between two independent variables (called factors). The primary purpose of a two-way ANOVA is to understand if there is an interaction between the two independent variables on the dependent variable. For example, you may want to determine whether there is an interaction between physical activity level(IV) and gender(IV) on blood cholesterol concentration(DV) in children.

**The interaction term in a two-way ANOVA informs you whether the effect of one of your independent variables on the dependent variable is the same for all values of your other independent variable (and vice versa).**

There some assumptions to do Two way ANOVA or we can say that these are the conditions for Two way ANOVA

* **Assumption #1:** Your **dependent variable** should be measured at the **continuous** level (i.e., they are **interval** or **ratio** variables).
* **Assumption #2:** Your **two independent variables** should each consist of **two or more categorical**, **independent groups**.
* **Assumption #3:** You should have **the independence of observations**, which means that there is no relationship between the observations in each group or between the groups themselves.
* **Assumption #4:** There should be **no significant outliers**. Outliers are data points within your data that do not follow the usual pattern
* **Assumption #5:** Your **dependent variable** should be **approximately normally distributed for each combination of the groups of the two independent variables**.
* **Assumption #6:** There needs to be **the homogeneity of variances for each combination of the groups of the two independent variables**.

Two way ANOVA calculation by hand:

We will do two way ANOVA for example, let’s start the calculation
Example: Suppose you want to determine whether the brand of laundry detergent used and the temperature affects the amount of dirt removed from your laundry. To this end, you buy two detergents with the different brand (“Super” and “Best”) and choose three different temperature levels (“cold”, “warm” and “hot”).

Then you divide your laundry randomly into “6\*r” pile of equal size and assign each ‘r’ piles into the combination of (“super” and “Best”) and (“cold”, “warm” and “hot”). In this example, we are interested in [testing](http://stepupanalytics.com/hypothesis-testing/%22%20%5Ct%20%22_blank)the Null Hypothesis.

**H(οD) = The amount of dirt removed does not depend on the type of detergent.**
**H(οT) = The amount of dirt removed does not depend on the temperature.**

The example has two factors(factor detergent, factor temperature) at a=2(Super and Best) and b=3(cold, warm and hot) levels. Thus, there are a\*b = 3\*2=6 different combination of detergent and temperature with each combination. There are r=4 loads. (r is called the number of replicates). This sums up to “n=a\*b\*r”=24=2\*3\*4 loads in total.

The amounts of Y(ijk) of dirt removed when washing sub pile k(k=1,2,3,4) with detergent i(i=1,2) at temperaturej(j=1,2,3) are recorded in table below:-

|  |  |  |  |
| --- | --- | --- | --- |
|  | **cold** | **warm** | **hot** |
|  Super | 4 | 7 | 10 |
|  | 5 | 9 | 12 |
|  | 6 | 8 | 11 |
|  | 5 | 12 | 9 |
|  |  |  |  |
|  Best | 6 | 13 | 12 |
|  | 6 | 15 | 13 |
|  | 4 | 12 | 10 |
|  | 4 | 12 | 13 |

Solution:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **cold** | **warm** | **hot** | **M(d) [Y(i)]** |
|  Super | 4 | 7 | 10 |  |
|  | 5 | 9 | 12 |  |
|  | 6 | 8 | 11 |  |
|  | 5 | 12 | 9 |  |
|  | **mean(Yij)=5** | **mean(Yij)=9** | **mean(Yij)=10.5 ~10** | **8** |
|  |  |  |  |  |
|  Best | 6 | 13 | 12 |  |
|  | 6 | 15 | 13 |  |
|  | 4 | 12 | 10 |  |
|  | 4 | 12 | 13 |  |
|  | **mean(Yij)=5** | **mean(Yij)=13** | **mean(Yij)=12** | **10** |
|  |  |  |  |  |
| **M(t)[Y(j)]** | 5 | 11 | 11 | 9 |

We have calculated all the means like detergent mean(Md), temperature means (Mt) and mean of every group combination.

Now what we only have to do is calculate the sum of squares(ss) and degree of freedom(df) for temperature, detergent and interaction between factor and levels.

First calculate the SS(within)/df(within) we have already know how to calculate SS(within)/df(within) in [one way ANOVA](http://stepupanalytics.com/analysis-of-variance-anova/) we calculated this but in two way anova the formula is different

**STEP 1:** **Formula for calculation of SS(within) is:**
**Yijk is the elements in the groups.**
**Y‾(ij) is mean of combinations**
When we put the values and do calculations with this formula we will get SS(within) is

                                      = (4 − 5)² + (5 − 5)² + (6 − 5)² + (5 − 5)²

                                            +(7 − 9)² + (9 − 9)² + (8 − 9)² + (12 − 9)²
· · · · · ·
                                                          +(12 − 12)² + (13 − 12)² + (10 − 12)² + (13 − 12

                                                                       = 38
**Calculate the df(within):**
                                                         df(within) = (r-1)\*a\*b = 3\*2\*3 = 18

**Calculate MS(within):**
MS(within) = SS(within)/df(within) = 38/18 = 2.1111

**STEP 2: Calculate SS(detergent) and df(detergent) and MS(detergent)**
**Y¯(i) is the mean of detergent**
**Y¯ is the total mean detergent and temperature**
 = 4\*3[(8-9)²+(10-9)²]
                                                                          = 24
**Calculate df(detergent):**
df(detergent) = a-1= 2-1 = 1

**Calculate MS(detergent):**
MS(detergent) = SS(detergent)/df(detergent)
                                                                 = 24/1= 24

**STEP 3:** **Calculate the SS(temperature), df(temperature) and MS(temperature)**
**Y¯(i) is the mean of detergent**
**Y¯ is the total mean detergent and temperature**

  = 4\*2\*[(5 − 9)² + (11 − 9)² + (11 − 9)²]
                                                                          = 192
**Calculate df(temperature):**
df(temperature) = b-1 = 3-1 = 2

**Calculate MS(temperature):**
MS(temperature) = SS(temperature)/df(temperature)
                                                                       = 192/2 = 81

**STEP 4:** **Calculate SS(interaction), df(interaction) and MS(interaction)**

**Y‾(ij) is mean of combinations**

**Y¯(i) is the mean of detergent**

**Y¯(j) is the mean of temperature**

**Y¯ is the total mean detergent and temperature**

**Calculate SS(interaction):**

  = 4 ×(5 − 8 − 5 + 9)² + (9 − 8 − 11 + 9)² + (110 − 8 − 11 + 9)² + · · · + (12 − 11 − 10 + 9)²
                                                                       = 12
**Calculate df(interaction):**
df(interaction) = (a-1)\*(b-1) = (2-1)\*(3-1) = 2

**Calculate MS(interaction):**
 MS(interaction) = SS(interaction)/df(interaction)

                                                                            = 12/2

                                                                            = 6

**Its time to calculate the F-test: Calculate critical F-value**

MS(detergent)/MS(within) ~  F(df(detergent), df(within))

MS(temperature)/MS(within) ~ F(df(temperature), df(within))

MS(interaction)/MS(within) ~ F(df(interaction), df(within))

***If you found the F-value less than the critical F-value then you will not be able to reject the null hypothesis I explained and how to and from where to calculate the [critical F-value](http://web.anglia.ac.uk/numbers/biostatistics/one_way_anova/local_folder/critical_values.html).***

If you have doubts and found any errors please ping me on [linkedin.com](https://www.linkedin.com/in/irrfankhan) or shoot a comment.