

# ANOVA: Complete guide to Statistical Analysis & Applications (Updated 2024)

[EXCEL](#)[INTERMEDIATE](#)[STATISTICS](#)[TECHNIQUE](#)

## Introduction

In this tutorial, I'll introduce you to anova, its objectives, statistical tests, test examples, statistical analysis, and the different ANOVA techniques used for making the best decisions. We'll take a few cases and try to understand the techniques for getting the results. We will also be leveraging the use of Excel to understand these concepts.



You must know the basics of anova statistics to understand this topic. Knowledge of t-tests and [Hypothesis testing](#) would be an additional benefit. And we believe the best way to learn statistics is by doing. That's how we follow in the '[Introduction to Data Science](#)' course, where we provide a comprehensive introduction to descriptive and inferential statistics.

## Learning Objectives

- In this tutorial, we will learn about Anova and its different types.
- You will familiarize yourself with the different terminologies associated with Anova.
- You will also learn how to calculate Anova In Microsoft Excel.

## Table of contents

- [Introduction](#)
- [What is Analysis of Variance \(ANOVA\)?](#)
  - [Analysis of Variance \(ANOVA\)](#)
- [Terminologies Related to ANOVA](#)
  - [Calculation B/W Group Variability](#)
- [One Way ANOVA or Single Factor Anova](#)
  - [Figure For three different Classes](#)
  - [Step-by-Step to Perform One-Way ANOVA With Post-hoc Test](#)
- [Two-Way ANOVA](#)

- [Two Factors of ANOVA](#)
- [F-Statistic for Each Hypothesis](#)
- [Multi-Variate ANOVA \(MANOVA\)](#)
- [Conclusion](#)
- [Frequently Asked Questions](#)

## What is Analysis of Variance (ANOVA)?

ANOVA stands for Analysis of Variance. It is a statistical method used to analyze the differences between the means of two or more groups or treatments. It is often used to determine whether there are any statistically significant differences between the means of different groups.

[ANOVA](#) compares the variation between group means to the variation within the groups. If the variation between group means is significantly larger than the variation within groups, it suggests a significant difference between the means of the groups.

ANOVA calculates an F-statistic by comparing between-group variability to within-group variability. If the F-statistic exceeds a critical value, it indicates significant differences between group means.

ANOVA is used to compare treatments, analyze factors impact on a variable, or compare means across multiple groups.

Types of [ANOVA](#) include one-way (for comparing means of groups) and two-way (for examining effects of two independent variables on a dependent variable).

Buying a new product or testing a new technique but not sure how it stacks up against the alternatives? It's an all too familiar situation for most of us. Most options sound similar to each other, so picking the best out of the lot is a challenge.

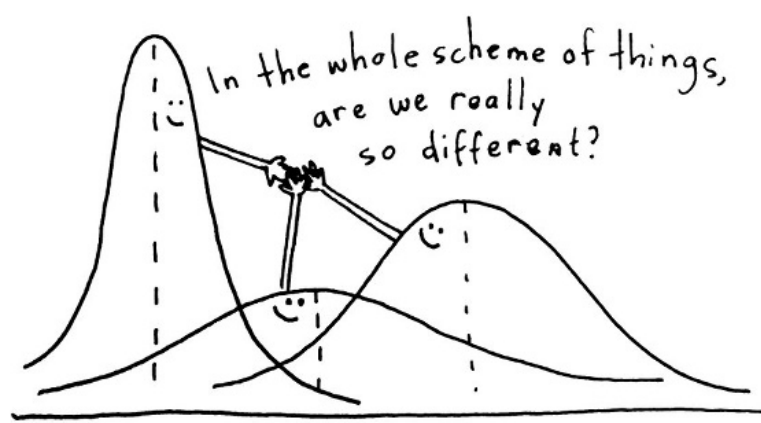
Consider a scenario where we have three medical treatments for patients with similar diseases. Once we have the test results, one approach is to assume that the treatment which took the least time to cure the patients is the best among them. What if some of these patients had already been partially cured, or if any other medication was already working on them?

In order to make a confident and reliable decision, we will need evidence to support our approach. This is where the concept of ANOVA comes into play.

## Analysis of Variance (ANOVA)

A common approach to figuring out a reliable treatment method would be to analyze the days the patients took to be cured. We can use a statistical technique to compare these three treatment samples and depict how different these samples are from one another. Such a technique, which compares the samples based on their means, is called ANOVA.

Analysis of variance (ANOVA) is a statistical technique used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples. We can use [ANOVA](#) to prove/disprove whether all the medication treatments were equally effective.



Source: Questionpro

Another measure to compare the samples is called a t-test. When we have only two samples, t-test, and ANOVA give the same results. However, using a t-test would not be reliable in cases with more than 2 samples. If we conduct multiple t-tests for comparing more than two samples, it will have a compounded effect on the error rate of the result.

## Terminologies Related to ANOVA

Before we start with the ANOVA applications, I would like to introduce some common terminologies used in the technique.

### Grand Mean

Mean is a simple or arithmetic average of a range of values. There are two kinds of means that we use in ANOVA calculations, which are separate sample means

$$(\mu_1, \mu_2 \text{ \& } \mu_3)$$

and the grand mean

$$(\mu).$$

. The grand mean is the mean of sample means or the mean of all observations combined, irrespective of the sample.

### Hypothesis

Considering our above medication example, we can assume that there are 2 possible cases – either the medication will have an effect on the patients or it won't. These statements are called Hypothesis. A hypothesis is an educated guess about something in the world around us. It should be testable either by experiment or observation.

Just like any other kind of hypothesis that you might have studied in statistics, ANOVA also uses a Null hypothesis and an Alternate hypothesis. The Null hypothesis in [ANOVA](#) is valid when all the sample means are equal, or they don't have any significant difference. Thus, they can be considered as a part of a larger set of the population. On the other hand, the alternate hypothesis is valid when at least one of the sample means is different from the rest of the sample means. In mathematical form, they can be represented as:

$$H_o : \mu_1 = \mu_2 = \cdots = \mu_L$$

*Null hypothesis*

$$H_1 : \mu_l \neq \mu_m$$

*Alternate hypothesis*

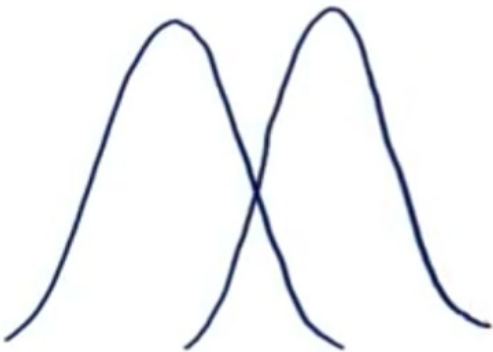
Where

$$\mu_l \text{ and } \mu_m$$

belonging to any two sample means out of all the samples considered for the test. In other words, the null hypothesis states that all the sample means are equal or the factor did not have any significant effect on the results. Whereas, the alternate hypothesis states that at least one of the sample means is different from another. But we still can't tell which one specifically. For that, we will use other methods that we will discuss later in this article.

## Between Group Variability

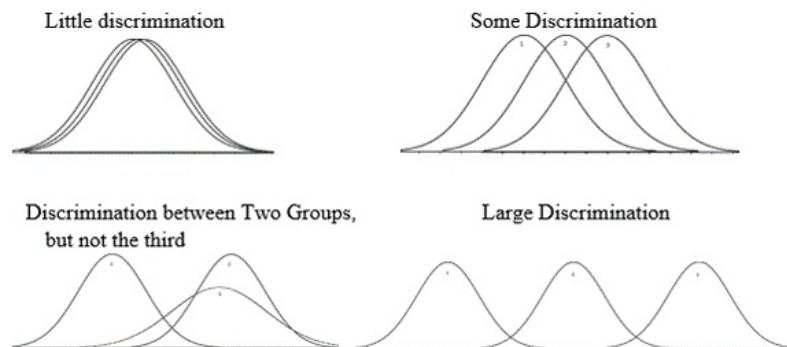
Consider the distributions of the below two samples. As these samples overlap, their individual means won't differ by a great margin. Hence the difference between their individual and grand means won't be significant enough.



Now consider these two sample distributions. As the samples differ from each other by a big margin, their individual means would also differ. The difference between the individual means and grand mean would, therefore, also be significant.



Such variability between the distributions is called **Between-group variability**. It refers to variations between the distributions of individual groups (or levels) as the values within each group differ. Each sample is examined, and the difference between its mean and grand mean is calculated to calculate the variability. If the distributions overlap or are close, the grand mean will be similar to the individual means, whereas if the distributions are far apart, the difference between means and grand mean would be large.

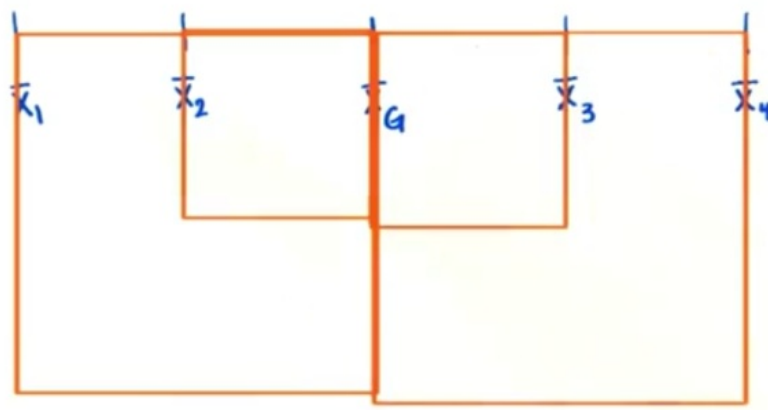


Source: Psychstat – Missouri State

## Calculation B/W Group Variability

We will calculate **Between Group Variability** just as we calculate the standard deviation. Given the sample means and Grand mean, we can calculate it as follows:

$$\frac{1}{N_1} \quad \frac{1}{N_2} \quad \frac{1}{N_G} \quad \frac{1}{N_3} \quad \frac{1}{N_4}$$



Source: Udacity

We also want to weigh each squared deviation by the sample size. In other words, a deviation is given greater weight if it's from a larger sample. Hence, we'll multiply each squared deviation by each sample size and add them. This is called the **sum-of-squares for between-group variability**

( $SS_{\text{between}}$ ).

$$SS_{\text{between}} = n_1(\bar{x}_1 - \bar{x}_G)^2 + n_2(\bar{x}_2 - \bar{x}_G)^2 + n_3(\bar{x}_3 - \bar{x}_G)^2 + \dots + n_k(\bar{x}_k - \bar{x}_G)^2$$

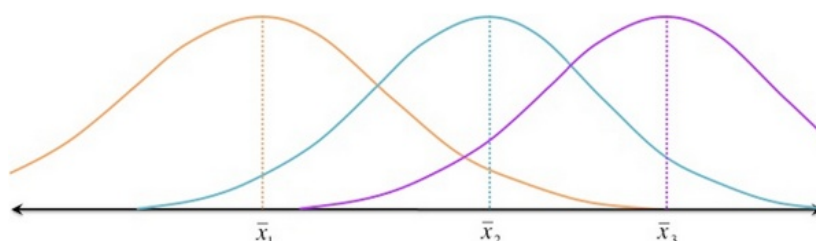
We must do one more thing to derive a good measure of between-group variability. Again, recall how we calculate the sample standard deviation.

We find the sum of each squared deviation and divide it by the degrees of freedom. For our between-group variability, we will find each squared deviation, weigh them by their sample size, sum them up, and divide by the degrees of freedom ( $k - 1$ ), which in the case of between-group variability is the number of sample means ( $k$ ) minus 1.

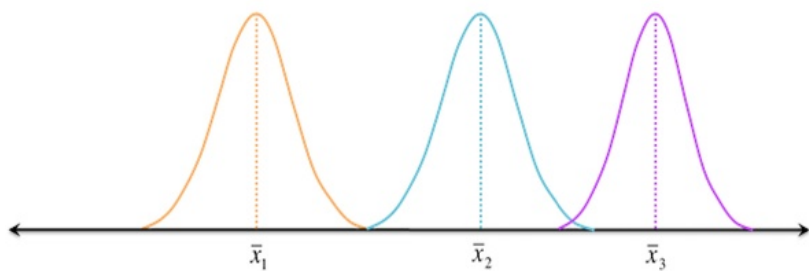
$$MS_{\text{between}} = \frac{n_1(\bar{x}_1 - \bar{x}_G)^2 + n_2(\bar{x}_2 - \bar{x}_G)^2 + n_3(\bar{x}_3 - \bar{x}_G)^2 + \dots + n_k(\bar{x}_k - \bar{x}_G)^2}{k - 1}$$

## Within Group Variability

Consider the given distributions of three samples. As the spread (variability) of each sample increases, their distributions overlap, and they become part of a big population.



Now consider another distribution of the same three samples but with less variability. Although the means of samples are similar to those in the above image, they seem to belong to different populations.



Source: TurntheWheelsandBox

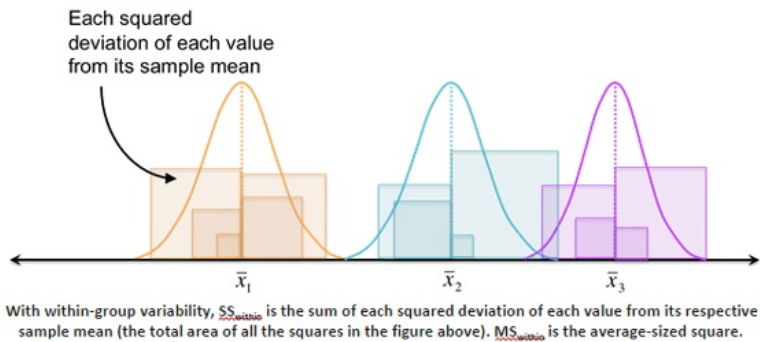
### Variations Within – Group

Such variations within a sample are denoted by **Within-group variation**. It refers to variations caused by differences within individual groups (or levels), as not all the values within each group are the same. Each sample is looked at on its own, and variability between the individual points in the sample is calculated. In other words, no interactions between samples are considered.

We can measure **Within-group variability** by looking at how much each value in each sample differs from its respective sample mean. So first, we'll take the squared deviation of each value from its respective sample mean and add them up. This is the **sum of squares for within-group variability**.

$$\begin{aligned} SS_{\text{within}} &= \sum (x_{i1} - \bar{x}_1)^2 + \sum (x_{i2} - \bar{x}_2)^2 + \dots + \sum (x_{ik} - \bar{x}_k)^2 \\ &= \sum (x_{ij} - \bar{x}_j)^2 \end{aligned}$$

Note:  $x_{i1}$  is the  $i$ th value from the first sample,  $x_{i2}$  is the  $i$ th value from the second sample, and so on all the way to  $x_{ik}$ , the  $i$ th value from the  $k$ th sample.  $x_{ij}$  is therefore the  $i$ th value from the  $j$ th sample.



Source: TurntheWheelsandBox

Like between-group variability, we then divide the sum of squared deviations by the **degrees of freedom**

to find a less-biased estimator for the average squared deviation (essentially, the average-sized square from the figure above). Again, this quotient is the mean square, but for within-group variability:

$$MS_{\text{within}}$$

. This time, the degrees of freedom is the sum of the sample sizes (N) minus the number of samples (k). Another way to look at degrees of freedom is that have the total number of values (N) and subtract 1 for each sample:

$$df_{\text{within}} = (n_1 - 1) + (n_2 - 1) + \dots (n_k - 1) = n_1 + n_2 + n_3 + \dots + n_k - k(1) = N - k$$

$$MS_{\text{within}} = \sum (x_{ij} - \bar{x}_j)^2 / (N - k)$$

## F-Statistic (F-test)

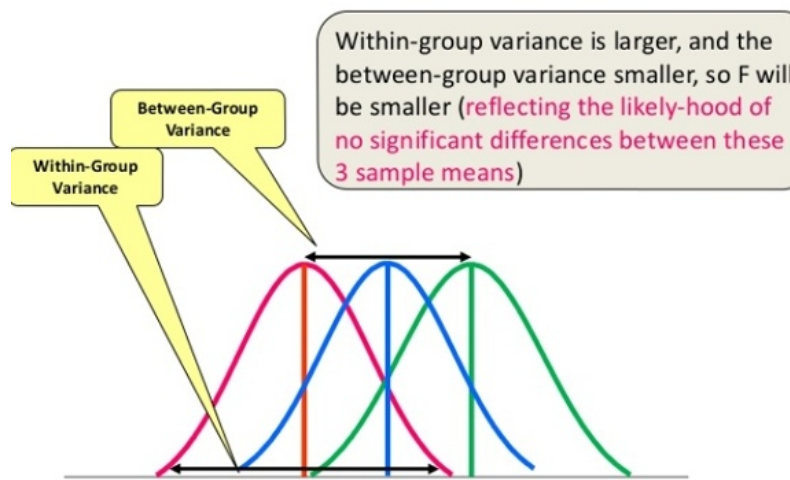
*The statistic that measures whether the means of different samples are significantly different is called the F-Ratio.* The lower the F-Ratio, the more similar will the sample means be. In that case, we cannot reject the null hypothesis.

### **F = Between-group variability / Within-group variability**

This above formula is pretty intuitive. The numerator term in the F-statistic calculation defines the between-group variability. As we read earlier, the sample means to grow further apart as between-group variability increases. In other words, the samples are likelier to belong to different populations.

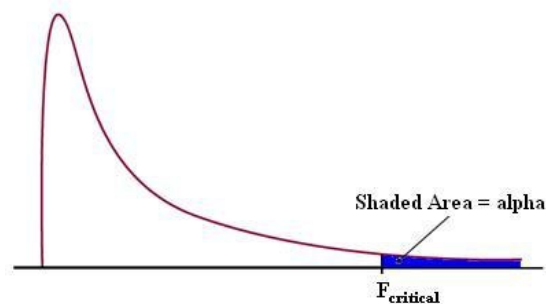
This F-statistic calculated here is compared with the F-critical value for concluding. In terms of our medication example, if the value of the calculated F-statistic is more than the F-critical value (for a specific  $\alpha$ /significance level), then we reject the null hypothesis and can say that the treatment had a significant effect.





Source: Dr. Asim's Anatomy Cafe

Unlike the z and t-distributions, the F-distribution has no negative values because between and within-group variability are always positive due to squaring each deviation.



Source: Statistics How To

Therefore, there is only one critical region in the right tail (shown as the blue-shaded region above). If the F-statistic lands in the critical region, we can conclude that the means are significantly different, and we reject the null hypothesis. Again, we must find the critical value to determine the cut-off for the critical region. We'll use the [F-table](#) for this purpose.

We need to look at different F-values for each alpha/significance level because the F-critical value is a function of two things:

$df_{within}$

and

$df_{between}$

# One Way ANOVA or Single Factor Anova

As we now understand the basic terminologies behind ANOVA, let's dive deep into its implementation using a few examples.

A recent study claims that music in a class enhances the concentration and consequently helps students absorb more information. As a teacher, your first reaction would be skepticism.

What if it affected the results of the students in a negative way? Or what kind of music would be a good choice for this? Considering all this, it would be immensely helpful to have some proof that it actually works.

## Figure For three different Classes

To figure this out, we implemented it on a smaller group of randomly selected students from three different classes. The idea is similar to conducting a survey. We took three different groups of ten randomly selected students (all of the same age) from three different classrooms. Each classroom was provided with a different environment for students to study. Classroom A had constant music being played in the background, classroom B had variable music being played, and classroom C was a regular class with no music playing. After one month, we conducted a test for all the three groups and collected their test scores. The test scores that we obtained were as follows:

|                          | Test scores of students (out of 10) |   |   |   |   |   |   |    |   |   |
|--------------------------|-------------------------------------|---|---|---|---|---|---|----|---|---|
| Class A (constant sound) | 7                                   | 9 | 5 | 8 | 6 | 8 | 6 | 10 | 7 | 4 |
| Class B (variable sound) | 4                                   | 3 | 6 | 2 | 7 | 5 | 5 | 4  | 1 | 3 |
| Class C (no sound)       | 6                                   | 1 | 3 | 5 | 3 | 4 | 6 | 5  | 7 | 3 |

Now, we will calculate the means and the Grand mean.  
So, in our case,

$$\mu_1 = 7, \mu_2 = 4, \mu_3 = 4.3 \text{ \& } \mu = 5.1.$$

|                          | Test scores of students (out of 10) |   |   |   |   |   |   |    |   |   | Mean |
|--------------------------|-------------------------------------|---|---|---|---|---|---|----|---|---|------|
| Class A (constant sound) | 7                                   | 9 | 5 | 8 | 6 | 8 | 6 | 10 | 7 | 4 | 7    |
| Class B (variable sound) | 4                                   | 3 | 6 | 2 | 7 | 5 | 5 | 4  | 1 | 3 | 4    |
| Class C (no sound)       | 6                                   | 1 | 3 | 5 | 3 | 4 | 6 | 5  | 7 | 3 | 4.3  |
| Grand mean ->            |                                     |   |   |   |   |   |   |    |   |   | 5.1  |

Looking at the above table, we might assume that the mean score of students from Group A is definitely greater than the other two groups, so the treatment must be helpful. Maybe it's true, but there is also a slight chance that we happened to select the best students from class A, which resulted in better test scores (remember, the selection was made at random). This leads to a few questions, like:

1. How do we decide that these three groups performed differently because of the different situations and not merely by chance?
2. In a statistical sense, how different are these three samples from each other?
3. What is the probability of group A students performing differently than the other two groups?

To answer all these questions, first, we will calculate the F-statistic, which can be expressed as the ratio Between Group variability and Within Group Variability.

$$\mathbf{SS_{between} = 54.6}$$

$$\mathbf{SS_{within} = 90.1}$$

$$\mathbf{df_{between} = 2}$$

$$\mathbf{df_{within} = 27}$$

$$\mathbf{MS_{between} = 27.3}$$

$$\mathbf{MS_{within} = 3.33}$$

$$\mathbf{F = 8.18}$$

Let's complete the ANOVA test for our example with

= 0.05.

## Limitations of One-Way ANOVA

A one-way ANOVA tells us that at least two groups are different from each other. *But it won't tell us which groups are different.* If our test returns a significant f-statistic, we may need to run a post-hoc test to tell us exactly which groups differ in means. Below I have mentioned the steps to perform one-way ANOVA in Microsoft Excel along with a post-hoc test.

## Step-by-Step to Perform One-Way ANOVA With Post-hoc Test

**Step 1:** Input your data into columns or rows in Excel. For example, if three groups of students for music treatment are being tested, spread the data into three columns.

**Step 2:** Click the "Data" tab and then click "Data Analysis." If you don't see Data Analysis, load the 'Data Analysis Toolpak' add-in.

**Step 3:** Click "ANOVA Single Factor" and then click "OK."

**Step 4:** Type an input range into the Input Range box. For example, if the data is in cells A1 to C10, type "A1:C10" into the box. Check the "Labels in the first row" if we have column headers, and select the Rows radio button if the data is in rows.

**Step 5:** Select an output range. For example, click the "New Worksheet" radio button.

**Step 6:** Choose an alpha level. For most hypothesis tests, 0.05 is standard.

**Step 7:** Click "OK." The results from ANOVA will appear in the worksheet.

The results for our example look like this:

| Class A | Class B | Class C |
|---------|---------|---------|
| 7       | 4       | 6       |
| 9       | 3       | 1       |
| 5       | 6       | 3       |
| 8       | 2       | 5       |
| 6       | 7       | 3       |
| 8       | 5       | 4       |
| 6       | 5       | 6       |
| 10      | 4       | 5       |
| 7       | 1       | 7       |
| 4       | 3       | 3       |

| Anova: Single Factor |       |     |          |          |          |          |
|----------------------|-------|-----|----------|----------|----------|----------|
| SUMMARY              |       |     |          |          |          |          |
| Groups               | Count | Sum | Average  | Variance |          |          |
| Class A              | 10    | 70  | 7        | 3.333333 |          |          |
| Class B              | 10    | 40  | 4        | 3.333333 |          |          |
| Class C              | 10    | 43  | 4.3      | 3.344444 |          |          |
| ANOVA                |       |     |          |          |          |          |
| Source of Variation  | SS    | df  | MS       | F        | P-value  | F crit   |
| Between Groups       | 54.6  | 2   | 27.3     | 8.18091  | 0.001669 | 3.354131 |
| Within Groups        | 90.1  | 27  | 3.337037 |          |          |          |
| Total                | 144.7 | 29  |          |          |          |          |

Here, we can see that the F-value is greater than the F-critical value for the alpha level selected (0.05). Therefore, we have evidence to reject the null hypothesis and say that at least one of the three samples have significantly different means and thus belongs to an entirely different population.

**Another measure for ANOVA is the p-value.** If the p-value is less than the alpha level selected (which it is, in our case), we reject the Null Hypothesis.

## Methods for Finding Two Different Populations

There are various methods for finding out which samples represent two different populations. I'll list some for you:

1. [Bonferroni approach](#)
2. [Least significant difference test](#)
3. [Tukey's HSD](#)

We won't be covering all of these here in this article, but I suggest you go through them.

Now to check which samples had different means, we will take the Bonferroni approach and perform the post hoc test in Excel through following steps:

**Step 8:** Again, click on “Data Analysis” in the “Data” tab and select “t-Test: Two-Sample Assuming Equal Variances,” and click “OK.”

**Step 9:** Input the range of Class A column in Variable 1 Range box and range of Class B column in Variable 2 Range box. Check the “Labels” if you have column headers in the first row.

**Step 10:** Select an output range. For example, click the “New Worksheet” radio button.

**Step 11:** Perform the same steps (Step 8 to step 10) for Columns of Class B – Class C and Class A – Class C.

The results will look like this:

| t-Test: Two-Sample Assuming Equal Variances |            |            | t-Test: Two-Sample Assuming Equal Variances |            |            | t-Test: Two-Sample Assuming Equal Variances |            |            |
|---|------------|------------|---|------------|------------|---|------------|------------|
|   | Class A    | Class B    |   | Class B    | Class C    |   | Class A    | Class C    |
| Mean  | 7          | 4          | Mean  | 4          | 3.5        | Mean  | 7          | 3.5        |
| Variance                                    | 3.33333333 | 3.33333333 | Variance                                    | 3.33333333 | 3.34444444 | Variance                                    | 3.33333333 | 3.34444444 |
| Observations                                | 10         | 10         | Observations                                | 10         | 10         | Observations                                | 10         | 10         |
| ooled Variance                              | 3.33333333 |            | ooled Variance                              | 3.33888889 |            | ooled Variance                              | 3.33888889 |            |
| Hypothesized Mean Difference                | 0          |            | Hypothesized Mean Difference                | 0          |            | Hypothesized Mean Difference                | 0          |            |
| df  | 18         |            | df  | 18         |            | df  | 18         |            |
| t Stat                                      | 5.67024634 |            | t Stat                                      | 0.36711707 |            | t Stat                                      | 5.30405893 |            |
| P(T<=t) one-tail                            | 0.00080715 |            | P(T<=t) one-tail                            | 0.35050550 |            | P(T<=t) one-tail                            | 0.00157295 |            |
| t Critical one-tail                         | 1.73406387 |            | t Critical one-tail                         | 1.73406387 |            | t Critical one-tail                         | 1.73406387 |            |
| P(T<=t) two-tail                            | 0.0017143  |            | P(T<=t) two-tail                            | 0.71751178 |            | P(T<=t) two-tail                            | 0.00384579 |            |
| t Critical two-tail                         | 2.10795988 |            | t Critical two-tail                         | 2.10795988 |            | t Critical two-tail                         | 2.10795988 |            |

Here, we can see that the p-value of (A vs B) and (A vs C) is less than the alpha level selected (alpha = 0.05). This means that groups A and B & groups A and C have less than a 5% chance of belonging to the same population. Whereas for (B vs C), it is much greater than the significance level. This means that B and C belong to the same population. So, it is clear that A (constant music group) belongs to an entirely different population. Or we can say that the constant music had a significant effect on students' performance.

Voila! The music experiment actually helped in improving the results of the students.

## One-Way ANOVA Calculation

Another effect size measure for one-way ANOVA is called Eta squared. It works in the same way as R2 for t-tests. It is used to calculate how much proportion of the variability between the samples is due to the between-group difference. It is calculated as:

$$\eta^2 = \frac{SS_{Between}}{SS_{Total}}$$

For the above example:

$$\eta^2 = 54.6/90 = 0.60$$

Hence 60% of the difference between the scores is because of the approach that was used. Rest 40% is unknown. Hence, the Eta square helps us conclude whether the independent variable really impacts the dependent variable or whether the difference is due to chance or any other factor.

There are commonly two types of ANOVA tests for univariate analysis – One-Way ANOVA and Two-Way ANOVA. One-way ANOVA is used when we are interested in studying the effect of one independent variable (IDV)/factor on a population. In contrast, Two-way ANOVA is used for studying the effects of two factors on a population simultaneously. For multivariate analysis, such a technique is called MANOVA or Multi-variate ANOVA.

## Two-Way ANOVA

Using one-way ANOVA, we found out that the music treatment helped improve the test results of our students. But this treatment was conducted on students of the same age. What if the treatment was to affect different age groups of students in different ways? Or maybe the treatment had varying effects depending upon the teacher who taught the class.

Moreover, how can we be sure which factor(s) is affecting the students’ results more? Maybe the age group is a more dominant factor responsible for a student’s performance than the music treatment.

For such cases, when the outcome or dependent variable (in our case, the test scores) is affected by two independent variables/factors, we use a slightly modified technique called two-way ANOVA.

In the one-way [ANOVA test](#), we found that the group subjected to ‘variable music’ and ‘no music at all’ performed more or less equally. It means that the variable music treatment did not have any significant effect on the students.

So, while performing two-way ANOVA, we will not consider the “variable music” treatment for simplicity of calculation. Rather a new factor, age, will be introduced to find out how the treatment performs when applied to students of different age groups. This time our dataset looks like this:

|                    | Age Group(yrs) |          |           |
|--------------------|----------------|----------|-----------|
| Class Group        | 4-8 yrs        | 8-13 yrs | 13-17 yrs |
| A (Constant sound) | 6              | 4        | 7         |
|                    | 5              | 5        | 6         |
|                    | 5              | 6        | 10        |
|                    | 2              | 9        | 8         |
|                    | 4              | 8        | 9         |
| Sub-total A        | 22             | 32       | 40        |
| B (No sound)       | 1              | 4        | 6         |
|                    | 3              | 5        | 8         |
|                    | 2              | 6        | 4         |
|                    | 1              | 7        | 7         |
|                    | 2              | 3        | 5         |
| Sub-total B        | 9              | 25       | 30        |
| Grand Total        | 31             | 57       | 70        |

## Two Factors of ANOVA

Here, there are two factors – class and age groups with two and three levels, respectively. So we now have six different groups of students based on different permutations of class groups and age groups, and each different group has a sample size of 5 students.

A few questions that two-way [ANOVA](#) can answer about this dataset are:

1. Is music treatment the main factor affecting performance? In other words, do groups subjected to different music differ significantly in their test performance?

2. Is age the main factor affecting performance? In other words, do students of different ages differ significantly in their test performance?
3. Is there a significant interaction between the factors? In other words, how do age and music interact with regard to a student's test performance? For example, it might be that younger and older students reacted differently to such a music treatment.
4. Can any differences in one factor be found within another factor? In other words, can any differences in music and test performance be found in different age groups?

Two-way ANOVA tells us about the main effect and the interaction effect. The main effect is similar to a one-way ANOVA where the effect of music and age would be measured separately. In comparison, the interaction effect is the one where both music and age are considered at the same time.

two-way ANOVA can have up to three hypotheses, which are as follows:  
 Two null hypotheses will be tested if we have placed only one observation in each cell. For this example, those hypotheses will be:  
**H1:** All the music treatment groups have an equal mean score.  
**H2:** All the age groups have an equal mean score.

For multiple cell observations, we would also test a third hypothesis:  
**H3:** The factors are independent, or the interaction effect does not exist.

## F-Statistic for Each Hypothesis

An F-statistic is computed for each hypothesis we are testing.  
 Before proceeding with the calculation, look at the image below. It will help us better understand the terms used in the formulas.

|          |                    | factor 2                 |                   |                    |       |                     |
|----------|--------------------|--------------------------|-------------------|--------------------|-------|---------------------|
|          |                    | Age Group(yrs)           |                   |                    |       |                     |
| factor 1 | Class Group        | 4-8 yrs                  | 8-13 yrs          | 13-17 yrs          | $A_i$ | factor 1 total      |
|          | A (Constant sound) | 6,5,5,2,4<br>(22)        | 4,5,6,9,8<br>(32) | 7,6,10,8,9<br>(40) | 94    |                     |
|          | B (No sound)       | 1,3,2,1,2<br>(9)         | 4,5,6,7,3<br>(25) | 6,8,4,7,5<br>(30)  | 64    |                     |
|          | $A_m$              | 31                       | 57                | 70                 | 316   | Grand Total ( $G$ ) |
|          |                    | subtotal<br>( $A_{im}$ ) |                   |                    |       |                     |
|          |                    | factor 2 total           |                   |                    |       |                     |

The table shown above is known as a **contingency table**. Here,

$$A_i$$

it represents the total of the samples based only on factor 1 and

$$\text{similarly } A_m$$

represents the total of samples based only on factor 2. We will see in some time that these two are responsible for the main effect produced. Also, a term

is introduced representing the subtotal of factor 1 and factor 2. This term will be responsible for the interaction effect produced when both the factors are considered simultaneously. And we are already familiar with the

Grand total  $G$ ,

, which is the sum of all the observations (test scores), irrespective of the factors.

| SUMMARY                   | 4-8 yrs    | 8-13 yrs   | 13-17 yrs  | Total      |
|---------------------------|------------|------------|------------|------------|
| <i>A (Constant sound)</i> |            |            |            |            |
| Count                     | 5          | 5          | 5          | 15         |
| Sum                       | 22         | 32         | 40         | 94         |
| Average                   | 4.4        | 6.4        | 8          | 6.26666667 |
| Variance                  | 2.3        | 4.3        | 2.5        | 4.92380952 |
| <i>B (No sound)</i>       |            |            |            |            |
| Count                     | 5          | 5          | 5          | 15         |
| Sum                       | 9          | 25         | 30         | 64         |
| Average                   | 1.8        | 5          | 6          | 4.26666667 |
| Variance                  | 0.7        | 2.5        | 2.5        | 5.06666667 |
| <i>Total</i>              |            |            |            |            |
| Count                     | 10         | 10         | 10         |            |
| Sum                       | 31         | 57         | 70         |            |
| Average                   | 3.1        | 5.7        | 7          |            |
| Variance                  | 3.21111111 | 3.56666667 | 3.33333333 |            |

### Calculation of all the Means

**We have calculated all the means – sound class mean, age group mean, and mean of every group combination in the above table.**

Now, calculate the sum of squares (SS) and degrees of freedom (df) for sound class, age group, and interaction between factor and levels.

We already know how to calculate SS (within)/df (within) in our one-way ANOVA section, but in two-way ANOVA, the formula is different. Let's look at the calculation of two-way ANOVA:

$$SS_{within} = \sum_{i=1}^2 \sum_{j=1}^3 \sum_{k=1}^5 (Y_{ijk} - \bar{Y}_{ij.})^2 = 59.2$$

$$df_{within} = (r - 1) * a * b = 4 * 2 * 3 = 24$$



$$MS_{\text{within}} = SS_{\text{within}} / df_{\text{within}} = 59.2/24 = 2.46$$

$$SS_{\text{sound}} = r \cdot b \cdot \sum_{i=1}^2 (\bar{Y}_i - \bar{Y})^2 = 30$$

$$df_{\text{sound}} = 2-1 = 1$$

$$SS_{\text{age}} = r \cdot a \cdot \sum_{i=1}^3 (\bar{Y}_i - \bar{Y})^2 = 78.86$$

$$df_{\text{age}} = 3-1 = 2$$

In two-way [ANOVA](#), we also calculate **SS<sub>interaction</sub>** and **df<sub>interaction</sub>**, which defines the combined effect of the two factors.

$$SS_{\text{interaction}} = r \times \sum_{i=1}^2 \sum_{j=1}^3 \left( Y_{ij\cdot} - \bar{Y}_{i..} - \bar{Y}_{.j\cdot} + \bar{Y}_{...} \right)^2 = 1.8$$

$$df_{\text{interaction}} = (a - 1) \times (b - 1) = 2$$

$$MS_{\text{interaction}} = SS_{\text{interaction}} / df_{\text{interaction}} = 0.9$$

Since we have more than one source of variation (main effects and interaction effects), it is obvious that we will have more than one F-statistic also.

Using these variances, we compute the value of F-statistic for the main and interaction effect. So, the values of f-statistic are,

**F1** = 12.16

**F2** = 15.98

**F12** = 0.36

We can see the critical values from the table

**Fcrit1** = 4.25

**Fcrit2** = 3.40

**Fcrit12** = 3.40

Suppose for a particular effect, its F value is greater than its respective F-critical value (calculated using the F-Table). In that case, we reject the null hypothesis for that particular effect.

## Steps to Perform Two-Way ANOVA in Excel 2013

**Step 1:** Click the “Data” tab and then click “Data Analysis.” If you don’t see the Data analysis option, install the Data Analysis Toolpak.

**Step 2:** Click “ANOVA two factor with replication” and then click “OK.” The two-way ANOVA window will open.

**Step 3:** Type an Input Range into the Input Range box. For example, if your data is in cells A1 to A25, type “A1:A25” into the Input Range box. Ensure you include all of your data, including headers and group names.

**Step 4:** Type a number in the “Rows per sample” box. Rows per sample is actually a bit misleading. What this is asking you is how many individuals are in each group. For example, if you have 5 individuals in each age group, you would type “5” into the Rows per Sample box.

**Step 5:** Select an Output Range. For example, click the “new worksheet” radio button to display the data in a new worksheet.

**Step 6:** Select an alpha level. In most cases, an alpha level of 0.05 (5 percent) works for most tests.

**Step 7:** Click “OK” to run the two-way ANOVA. The data will be returned in your specified output range.

**Step 8:** Read the results. To figure out if you are going to reject the null hypothesis or not, you’ll basically be looking at two factors:

1. If the F-value ( $F$ ) is larger than the f critical value ( $F_{crit}$ )
2. If the p-value is smaller than your chosen alpha level.

And you are done!

**Note:** We don’t only have to have two variables to run a two-way ANOVA in Excel 2013. We can also use the same function for three, four, five, or more variables.

## Results for DataSet Test

The results for two-way ANOVA test on our example look like this:

|                     |  | Age Group(yrs) |          |           | SUMMARY            |            |            |            |             |
|---------------------|--|----------------|----------|-----------|--------------------|------------|------------|------------|-------------|
| Class Group         |  | 4-8 yrs        | 8-12 yrs | 13-17 yrs | 4-8 yrs            | 8-12 yrs   | 13-17 yrs  | Total      |             |
| A (Constant sound)  |  | 6              | 4        | 7         | A (Constant sound) |            |            |            |             |
|                     |  | 5              | 5        | 5         | Count              | 5          | 5          | 5          | 15          |
|                     |  | 5              | 6        | 10        | Sum                | 22         | 32         | 40         | 94          |
|                     |  | 2              | 9        | 8         | Average            | 4.4        | 6.4        | 8          | 5.20000007  |
|                     |  | 4              | 8        | 9         | Variance           | 2.3        | 4.3        | 2.5        | 4.92380952  |
| B (No sound)        |  | 1              | 4        | 5         | B (No sound)       |            |            |            |             |
|                     |  | 3              | 5        | 8         | Count              | 5          | 5          | 5          | 15          |
|                     |  | 2              | 6        | 4         | Sum                | 0          | 25         | 30         | 64          |
|                     |  | 1              | 7        | 7         | Average            | 1.8        | 5          | 6          | 4.26666667  |
|                     |  | 2              | 3        | 5         | Variance           | 0.7        | 2.5        | 2.5        | 5.00000007  |
|                     |  |                |          |           | Total              |            |            |            |             |
|                     |  |                |          |           | Count              | 10         | 10         | 10         |             |
|                     |  |                |          |           | Sum                | 31         | 57         | 70         |             |
|                     |  |                |          |           | Average            | 3.1        | 5.7        | 7          |             |
|                     |  |                |          |           | Variance           | 8.21111111 | 3.56666667 | 3.83333333 |             |
| ANOVA               |  |                |          |           |                    |            |            |            |             |
| Source of Variation |  |                |          |           | SS                 | df         | MS         | F          | P-value     |
| Sample              |  |                |          |           | 30                 | 1          | 30         | 11.1621622 | 0.001200531 |
| Columns             |  |                |          |           | 78.8666667         | 2          | 39.4333333 | 15.0864965 | 0.000088618 |
| Interaction         |  |                |          |           | 1.8                | 2          | 0.9        | 0.35485485 | 0.698075004 |
| Within              |  |                |          |           | 59.2               | 24         | 2.46666667 |            |             |
| Total               |  |                |          |           | 169.866667         | 29         |            |            |             |

As you can see in the highlighted cells in the image above, the F-value for sample and column, i.e., factor 1 (music) and factor 2 (age), respectively, are higher than their F-critical values. This means that the factors significantly affect the students’ results, and thus we can reject the null hypothesis for the factors. Also, the F-value for interaction effect is quite less than its F-critical value, so we can conclude that music and age did not have any combined effect on the population.

## Multi-Variate ANOVA (MANOVA)

Until now, we have been making conclusions about the performance of students based on just one test. Could there be a possibility that the music treatment helped improve the results of a subject like mathematics but would affect the results adversely for a theoretical subject like history?

How can we ensure the treatment won’t be biased in such a case? So again, we take two groups of randomly selected students from a class and subject each group to one kind of music environment, i.e., constant music and no music. But now we thought of conducting two tests (maths and history), instead of just one. This way, we can be sure how the treatment would work for different subjects.

We can say that one IDV/factor (music) will be affecting two dependent variables (maths scores and history scores) now. This kind of problem comes under a multivariate case; the technique we will use to solve it is known as MANOVA. Here, we will work on a specific case called one-factor MANOVA. Let us now see how our data looks:

|         | History | Maths |
|---------|---------|-------|
| Group A | 7       | 3     |
| Group A | 9       | 1     |
| Group A | 5       | 6     |
| Group A | 8       | 7     |
| Group A | 6       | 8     |
| Group A | 8       | 3     |
| Group A | 6       | 6     |
| Group A | 10      | 5     |
| Group A | 7       | 9     |
| Group A | 4       | 6     |
| Group B | 4       | 8     |
| Group B | 3       | 5     |
| Group B | 6       | 1     |
| Group B | 2       | 8     |
| Group B | 7       | 1     |
| Group B | 5       | 3     |
| Group B | 5       | 5     |
| Group B | 4       | 6     |
| Group B | 1       | 4     |
| Group B | 3       | 7     |

Here we have one factor, music, with 2 levels. This factor will affect our two dependent variables, i.e., the test scores in maths and history. Denoting this information in terms of variables, we can say that we have  $L = 2$  (2 different music treatment groups) and  $P = 2$  (maths and history scores). A MANOVA test also takes into consideration a null hypothesis and an alternate hypothesis.:

$$H_o : \mu_1 = \mu_2 = \cdots = \mu_L \qquad \text{Null hypothesis}$$

$$H_1 : \mu_i \neq \mu_m \qquad \text{Alternate hypothesis}$$

The Calculations of MANOVA are too complex for this article, so if you want to further read about it, check [this](#) paper. We will implement MANOVA in Excel using the ‘RealStats’ Add-ins. It can be downloaded from [here](#).

## Steps to Perform MANOVA in Excel 2013

- Step 1:** Download the ‘RealStats’ add-in from the link mentioned above
- Step 2:** Press “control+m” to open RealStats window
- Step 3:** Select “Analysis of variance”
- Step 4:** Select “MANOVA: single factor”
- Step 5:** Type an Input Range into the Input Range box. For example, if your data is in cells A1 to A25, type “A1:A25” into the Input Range box. Make sure you include all of your data, including headers and group names.
- Step 6:** Select “Significance analysis”, “Group Means” and “Multiple Anova”.
- Step 7:** Select an Output Range.
- Step 8:** Select an alpha level. In most cases, an alpha level of 0.05 (5 percent) works for most tests.
- Step 9:** Click “OK” to run. The data will be returned in your specified output range.
- Step 10:** Read the results. To figure out if you are going to reject the null hypothesis or not, you’ll basically be looking at two factors:

1. If the F-value ( $F$ )is larger than the f critical value ( $F_{crit}$ )
2. If the p-value is smaller than your chosen alpha level.

And you are done!

|                 |             |            |            |            |                |               |
|-----------------|-------------|------------|------------|------------|----------------|---------------|
| MANOVA          |             |            |            |            |                |               |
|                 |             |            |            |            |                |               |
|                 | <i>stat</i> | <i>F</i>   | <i>df1</i> | <i>df2</i> | <i>p-value</i> | <i>eta-sq</i> |
| Pillai Trace    | 0.55149076  | 10.4516721 | 2          | 17         | 0.00109662     | 0.55149076    |
| Wilk's Lambda   | 0.44850924  | 10.4516721 | 2          | 17         | 0.00109662     | 0.55149076    |
| Hotelling Trace | 1.22960848  | 10.4516721 | 2          | 17         | 0.00109662     | 0.55149076    |
| Roy's Lg Root   | 1.22960848  |            |            |            |                |               |

## RealStats Different Methods

RealStats add-on shows us the results by different methods. Each one of them denotes the same p-value. We will reject the null hypothesis because the p-value is less than the alpha value. Or in simpler terms, it means that the music treatment did have a significant effect on students’ test results.

But we still cannot tell which subject was affected by the treatment and which was not. This is one of the limitations of MANOVA; even if it tells us whether the effect of a factor on a population was significant, it does not tell us which dependent variable was actually affected by the factor introduced.

For this purpose, we will see the “Multiple ANOVA” table to generate a helpful summary. The result will look like this:

|                      |       |     |            |            |            |            |            |            |  |
|----------------------|-------|-----|------------|------------|------------|------------|------------|------------|--|
| Multiple ANOVA       |       |     |            |            |            |            |            |            |  |
| ANOVA: Single Factor |       |     | History    |            |            |            |            |            |  |
| DESCRIPTION          |       |     |            |            | Alpha      |            | 0.025      |            |  |
| Groups               | Count | Sum | Mean       | Variance   | SS         | Std Err    | Lower      | Upper      |  |
| Group A              | 10    | 70  | 7          | 3.33333333 | 30         | 0.57735027 | 5.44980826 | 8.55019174 |  |
| Group B              | 10    | 40  | 4          | 3.33333333 | 30         | 0.57735027 | 2.44980826 | 5.55019174 |  |
| ANOVA                |       |     |            |            |            |            |            |            |  |
| Sources              | SS    | df  | MS         | F          | P value    | F crit     | RMSSE      | Omega Sq   |  |
| Between Groups       | 45    | 1   | 45         | 13.5       | 0.00173543 | 5.97805246 | 1.161895   | 0.38461538 |  |
| Within Groups        | 60    | 18  | 3.33333333 |            |            |            |            |            |  |
| Total                | 105   | 19  | 5.52631579 |            |            |            |            |            |  |
| ANOVA: Single Factor |       |     | Maths      |            |            |            |            |            |  |
| DESCRIPTION          |       |     |            |            | Alpha      |            | 0.025      |            |  |
| Groups               | Count | Sum | Mean       | Variance   | SS         | Std Err    | Lower      | Upper      |  |
| Group A              | 10    | 54  | 5.4        | 6.04444444 | 54.4       | 0.79582243 | 3.26320815 | 7.53679185 |  |
| Group B              | 10    | 48  | 4.8        | 6.62222222 | 59.6       | 0.79582243 | 2.66320815 | 6.93679185 |  |
| ANOVA                |       |     |            |            |            |            |            |            |  |
| Sources              | SS    | df  | MS         | F          | P value    | F crit     | RMSSE      | Omega Sq   |  |
| Between Groups       | 1.8   | 1   | 1.8        | 0.28421053 | 0.60047575 | 5.97805246 | 0.16858545 | -0.0371179 |  |
| Within Groups        | 114   | 18  | 6.33333333 |            |            |            |            |            |  |
| Total                | 115.8 | 19  | 6.09473684 |            |            |            |            |            |  |

Here, we can see that the P value for history lies in a significant region (since P value is less than 0.025) while for maths, it does not. This means that the music treatment had a significant effect in improving the performance of students in history but did not have any significant effect in improving their performance in maths.

Based on this, we might consider picking and choosing subjects where this music approach can be used.

## Conclusion

I hope this article was helpful, and now you’d be comfortable solving similar problems using Analysis of Variance ([ANOVA](#)). I suggest you take different kinds of problem statements and take your time to solve them using the above-mentioned techniques.

You should also check out the below two resources to give your data science journey a huge boost:

- [Introduction to Data Science Course](#)
- [Certified Program: Data Science for Beginners \(with Interviews\)](#)

### Key Takeaways

- ANOVA is a statistical formula used to compare variances across the means (or average) of different groups.
- There are two types of commonly used [ANOVA](#): one-way ANOVA and two-way ANOVA.
- To analyze variance (ANOVA), statisticians or analysts use the f-test to compute the feasibility of variability amongst two groups more than the variations observed within the said groups under study.

## Frequently Asked Questions

### Q1. What is ANOVA used for in Excel?

A. In Excel, ANOVA is a built-in statistical test used to analyze the variances. For instance, we usually compare the available alternatives when buying a new item, which eventually helps us choose the best from all the available options.

### Q2. What is ANOVA? Explain its uses and applications.

A. One can use ANOVA to test for statistical differences between two or more groups to check if there is any significant difference between the means of those groups.

### Q3. What is the name of the Excel function for ANOVA?

A. ANOVA is not a function in Excel. In Microsoft Excel, ANOVA is part of Excel's "Data Analysis" tool.

---

Article Url - <https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/>



### **Gurchetan1000 Singh**

Budding Data Scientist from MAIT who loves implementing data analytical and statistical machine learning models in Python. I also understand big data technology like Hadoop and Alteryx.