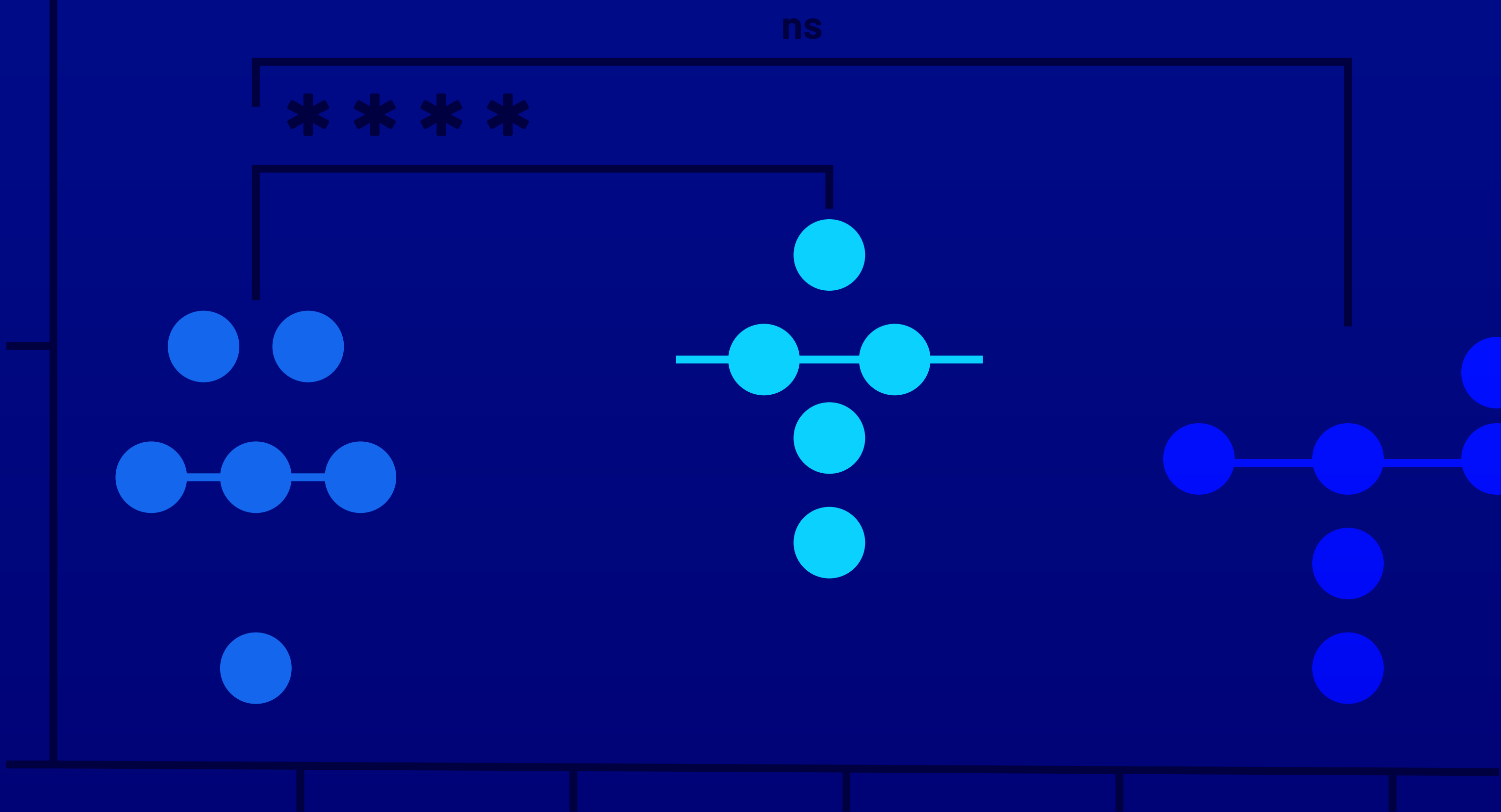
Statistics with GraphPad:

How and When to Use ANOVA



GraphPad

Analysis of variance (ANOVA) is a statistical tool commonly used to determine how different factors (independent variables) affect responses of interest (dependent variables). ANOVA tests the differences between two or more groups and helps determine if the means are significantly different from each other.

What is ANOVA and what is it used for?

ANOVA helps determine whether groups are different, overall, by comparing the differences in group means to a measure of noise.

Experimental designs use ANOVA to test for a relationship between categorical independent variables and a numeric response of interest, and can help determine if there are differences between two or more means.

ANOVA can be applied to many different types of data. It is a statistical tool that is very commonly used to analyze the results of biological experiments including in vitro treatments of cells with specific drugs and antagonists. For example, a one-way ANOVA test can be used to determine whether there are significant differences when treating cells with: no drug, drug, and drug + antagonist.

Variance

Variance is a central concept to ANOVA analysis. ANOVA compares the variation between groups to the variation within groups. What we are trying to find out with ANOVA is whether the differences in means of your groups (drug treatment vs. drug treatment + antagonist) are due to chance or due to the independent variable.

Assumptions in ANOVA

As with other statistical tools, ANOVA does make multiple assumptions. In a one-way ANOVA these are the assumptions:

- The samples are independent and randomly selected. The observations must not be correlated in any way, and if they are, a repeated measures ANOVA should be considered.
- Errors are normally-distributed. In other words, the data within each group are sampled from a Gaussian distribution.
- Homogeneity of variances. This means the group variances are the same—even if the means of the groups are different.

Null Hypothesis (H_0)

The null hypothesis (H_0) in the ANOVA F test generally states that there is no difference between the means. Any difference that you observe between groups is simply due to random sampling—chance.

Alternative Hypothesis (H_1)

The alternative hypothesis (H_1) states that at least two means—two groups— are statistically different. Usually, this is the effect you would like to observe after performing your experiment. Going back to the example where cells are treated with no drug, drug or drug + antagonist, generally, you are interested in observing that the drug and drug + antagonist treatments yield results that are significantly different from each other and/or from the control group.

P values and significance level

What do P values and statistically significant results mean? P values are calculated by ANOVA using the mean (M), standard deviations (SD or SEM) and sample size (n) but do not quantify the probability of the null hypothesis being true. There are many misconceptions of what P values mean. P values test the null hypothesis, and are calculated based on the assumption that the difference between groups is due to sampling error. A low p-value indicates that the differences are not due to sampling error and that a real effect exists.

P values determine if your experimental results (variance between means) are statistically significant. The P value is compared to a threshold called α. Commonly, this threshold (α) is set to a 0.05 ($P < 0.05$ is significant).

The following conclusions can be reached based on the P value results:

$P > \alpha$ Fail to reject null hypothesis. Any difference in means is due to random chance.

$P \leq \alpha$ Reject the null hypothesis. At least two of the means are different from each other.

There are two errors that can result when using ANOVA:

- Type I error – False rejection of the null hypothesis. The true means are not different, but the test says they are. This error rate is equal to α.
- Type II error – False acceptance of the null hypothesis. The true means are different, but the test did not detect this difference.

> P values determine if your experimental results (variance between means) are statistically significant.

Multiple Comparison Tests

As mentioned previously, ANOVA indicates that at least two means are different from each other when $P \leq 0.05$. This means that there is a difference overall, but ANOVA does not indicate which means are different.

In this case, you will want to determine what mean (means) are different. To do this, you can follow up by performing a multiple comparison test. Multiple comparison tests are post-hoc tests that help determine which groups are significantly different from other groups.

The type of post-hoc test you run depends on the type of ANOVA that you used to do your initial analysis. Prism can guide you through the process of choosing a multiple comparison test. These will be discussed in more detail below.

The ANOVA equations

These are the general ANOVA equations:

$$F = \frac{MST}{MSE}$$

$$MST = \frac{\sum_{i=1}^{K}\left(T_i^2/n_i\right) - G^2/n}{k-1}$$

$$MST = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k}(G^2/n)}{n-k}$$

F = variance ratio for the overall test, used to calculate the P value.

MST = mean square due to treatment/groups (between groups variability)

MSE = mean square due to error (within groups variability)

Yij = observation

Ti = group total

G = total of all observations

Ni = number in group I

N = total number of observations

The F ratio is important because it is used to calculate the P value. F is the ratio of two mean squares. In the case that the null hypothesis is true (means are the same), then F will be close to 1.0. If the F ratio is large then the variation between group means does not result from chance, and the null hypothesis is rejected.

> The F ratio is important because it is used to calculate the P value.

When you are first designing your experiment, you must define your hypothesis, sample size, and determine level of significance for statistical analysis.

As mentioned previously, the statistical significance (α) is generally set to a 0.05 ($P < 0.05$).

Then, you must choose the type of ANOVA that would work best for your experimental design and question.

What test do I use if my experiment has multiple factors?

If your experiment has two factors, then you should use two-way ANOVA. If your experiment has three factors, then you should use three-way ANOVA. These are discussed in more detail later in this article.

If your experiment has only one factor and there are three or more groups (levels), then use one-way ANOVA. For example, when a cellular response is measured after treatment with three different drugs. The independent factor in this case has three different levels (three different drugs). When the independent factor has two levels (male vs. female) then you should use a t test. Prism can answer your one-way ANOVA questions.

Learn how to perform a One-way ANOVA in Prism.

These are the possible one-way ANOVA tests you can use in your analysis:

- Classic ANOVA – Fisher's one-way ANOVA is the classic ANOVA that can be used to compare the means of three or more groups with one independent variable. For example, measuring enzyme levels in mice fed four different diets. The only independent variable is the type of diet and there are four different groups. There is fixed-effect (Type I) and random (Type II) ANOVA tests. Fixed-effect tests differences among means of the groups that data is collected from whereas random assumes that the groups in your experiment were randomly selected and the conclusions apply to all the groups, including those not in this experiment. Random ANOVA is not commonly used.
- Welch's ANOVA – This test assumes the standard deviations of groups are not equal (variances are not equal).
- Brown-Forsythe ANOVA – This test compares three or more sets of unpaired measurements (interval or ratio scale) and is recommended when data is skewed, in other words, the data do not have Gaussian distribution.
- Repeated Measures – This analysis compares the means of three or more matched groups. This test is not discussed in detail here.
- Kruskal-Wallis Test – This test is a nonparametric test used to compare three or more unmatched groups. It is an alternative to a classic one-way ANOVA. The dependent variable is continuous.
- Friedman's Test – This is a nonparametric test that compares three or more paired groups. It is an alternative to one-way ANOVA with repeated measures.

Choosing Post-Hoc Test – How do I know which groups are different?

Once you have run an ANOVA and determined that means are different (rejected the null hypothesis), you can choose the multiple comparisons test to determine what groups (means) are different.

These are the post-hoc tests you can choose from:

- Bonferroni – This analysis is sometimes called the Bonferroni-Dunn method. This method helps compare every mean to every other mean. Similar to the Šídák test. This test is commonly used and is recommended for comparison of selected pairs of columns (groups)—treated and control groups—allowing detection of differences that might not be detected when all the data are included.
- Tukey HSD – This method also lets you compare every mean to every other mean. These tests can be performed after a one-way or two-way ANOVA. This method also calculates the confidence interval for the difference between the two means.
- Dunnett – In prism, this method is used when comparing each of the treatment means[1] to the control group mean. Similar to the Tukey HSD test, Dunnett can be performed after a one-way or two-way ANOVA. This method also calculates the confidence interval for the difference between the two means. This is preferred to the Holm-Šídák.
- Sidak – This method allows you to compare every mean to every other mean and also gives the confidence interval for the difference between the two means. This test is similar to Bonferroni but it is more powerful. It is also more powerful than the Tukey test—can find a statistically significant difference that the Tukey cannot. Often called the Bonferroni-Šídák method.
- Holm-Šídák – This method is recommended when comparing every mean to every other mean. However, it does not give the confidence interval for the difference between the two means. This test can be performed after a one-way or a two-way ANOVA test. It is more powerful than the Holm test, especially in the case where you have to compare many groups.
- Newman-Keuls – This method can be used to compare all pairs of means and reports statistical significance. It does not report confidence intervals or multiplicity adjusted P values. However, it is too powerful, meaning that it does not maintain the family-wise error rate at the specified level. Thus, Type I error often results from using this test.
- Games-Howell – This method compares every mean with every other mean. This test is recommended for experiments with a large number of samples, and when you do not assume homoscedasticity—equal standard deviations.
- Fisher's Least Significant Difference (LSD) test – This test allows you to compare the mean of one group to the mean of another group after a one-way or two-way ANOVA. This test does not correct for multiple comparisons.

Enter your data and run the analysis - How do I organize my data?

In Prism, you can enter raw data or summary data (SD/SEM, mean, and n) for a one-way ANOVA.

However, note that entering summary data will allow one-way ANOVA but not repeated measures or nonparametric comparisons. Prism details the steps to enter data and graph your results.

For two-way ANOVA you can enter data different ways because your experiment has two factors. One factor is represented by the columns and the other factor is represented by the rows. You can enter raw data or summary data. Prism has a detailed guide for the application of two-way ANOVA.

In three-way ANOVA, you can also enter raw or summary data. However, with both two-way and three-way ANOVA, having a large number of replicates (>256 in version 8 and >512 in version 9) only allows you to enter summary data.

> The number of factors as well as whether you have raw or averaged will determine how you enter data for ANOVA analysis.

Two-way ANOVA

Two-way ANOVA identifies how a response (continuous dependent variable) in an experiment is influenced by two factors (variables).

In addition to testing the effect each factor has on the response individually, it answers the question: is there an interaction effect between the two independent variables?

For example, you can use two-way ANOVA after measuring the response of different cell types to serum starvation. The two independent factors are cell line and type of culture.

Table format: Grouped	Group A					Group B					Group C				
	Wild-type cells					GPP5 cell line					GPP7 cell line				
	A:Y1	A:Y2	A:Y3	A:Y4	A:Y5	B:Y1	B:Y2	B:Y3	B:Y4	B:Y5	C:Y1	C:Y2	C:Y3	C:Y4	C:Y5
1 Serum starved	34	36	41		43	98	87	95	99	88	77	89	97	66	76
2 Normal culture	23	19	26	29	25	32	29	26	33	30	33	45	35	46	54

Organization of data for two-way ANOVA

The two independent factors are cell line and type of culture.

Assumptions of two-way ANOVA

- Measurements of dependent variables have an interval or ratio scale.
- Samples are independent.
- Variance homogeneity – variance is the same for different groups
- Normally distributed errors.

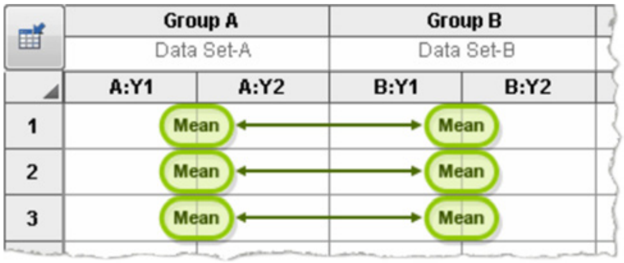
There are some things to consider when using two-way ANOVA for an experiment that involves a quantitative factor. For example, when comparing dose response curves one of the factors is dose. In this case, using a regression model would be a better approach. If ANOVA is used the different P values will tell you the following:

- First P value – Tests the null hypothesis that dose has no effect on the mean response. A small P value is expected.
- Second P value – Tests the null hypothesis that treatment has no effect on the mean response. Does not yield any valuable information.
- Third P value – Tests the null hypothesis that differences between treatment means are identical at all doses. This tests for interaction.

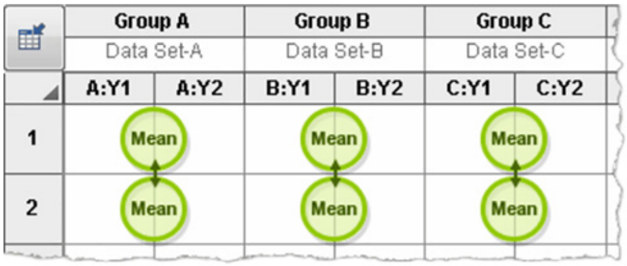
There are various alternatives to ANOVA when quantitative factors are involved.

Post-Hoc Tests - What comparisons do you want to make after two-way ANOVA?

There are many different types of comparisons you can make after performing a two-way ANOVA because your experiment involves two independent factors. To make this decision the main question you want to answer with this analysis must be clearly defined. For example, you can compare each cell mean with the other cell mean in that row as shown below:



You can also compare each cell mean with the other mean in that column as shown here:



Prism offers guidance as to what multiple comparison to perform with your data.

The tests that can be done after two-way ANOVA are Tukey HSD, Bonferroni, Holm, Newman Keuls and Dunnett. You should choose the test depending on the comparison you want to make, whether there are more than two rows and columns, and whether you would like to obtain a confidence interval. Tukey and Bonferroni are used when you want to compare every mean to every other mean and do give confidence interval value. Dunnett and Holm, however, are used to compare every mean to a control mean. Dunnett gives a confidence interval value, whereas Holm does not. Prism provides a summary of the multiple comparisons you may want to make.

Do you have additional questions about two-way ANOVA? Prism can answer your two-way ANOVA questions.

Three-way ANOVA

Analysis with three-way ANOVA identifies how a response (continuous dependent variable) in an experiment is influenced by three factors (variables).

This analysis characterizes the relationship between the independent variables and the response.

One example where you may want to perform a three-way ANOVA is to measure the effect of low fat or high fat diets in heavy smoker and light smoker females and males. Here, the independent variables are: sex, diet (low vs. high fat) and smoking (light smoker vs. heavy smoker) and you can enter raw as shown below.

Assumptions

- Measurements of dependent variables have an interval or ratio scale.
- Independent samples
- Variance homogeneity – variance is the same for different groups
- Normally distributed errors.

Analysis with three-way ANOVA can be very complicated and must be used with great caution. As with two-way ANOVA, the question you want to answer using statistical analysis must be clear to be able to use three-way ANOVA and obtain a useful answer. In three-way ANOVA, a non-significant independent variable may be removed, resulting in simplified analysis that may provide a better answer to your question.

Table format: Grouped		Group A Low fat Male			Group B Low fat Female			Group C High fat Male			Group D High fat Female		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3	C:Y1	C:Y2	C:Y3	D:Y1	D:Y2	D:Y3
1	Light smoker	24.1	29.2	24.6	20.0	21.9	17.6	14.6	15.3	12.3	16.1	9.3	10.8
2	Heavy smoker	17.6	18.8	23.2	14.8	10.3	11.3	14.9	20.4	12.8	10.1	14.4	6.1

Organization of data for three-way ANOVA

Because there are three different factors in this experiment, you can have seven different null hypotheses.

- Three test the main effect—the mean response is the same in females and males, for example.
- Three test two-way interactions—the effect of low fat vs high fat is the same for heavy smokers vs light smokers (pool the males and females), for example.
- One tests the three-way interaction.

There are a few alternatives to three-way ANOVA – Two-way ANOVA and linear regression analyses may provide better analysis and answers to your biological question.

The post-hoc tests you may want to use with three-way ANOVA are the Bonferroni, Tukey HSD, Dunnett, Šidák or Holm- Šidák test. Prism offers guidance to choose the right post-hoc test.

How do I read an ANOVA table and what do the results mean?

The results from an ANOVA test are given in the form of a table (three-way ANOVA table) as shown below.

The values shown on the table are the following:

- **Sum of Squares (SS) –** Sum of squares of deviations from the mean.
- **Degrees of Freedom (DF) –** Number of independent measures of variability for each component. Total degrees of freedom is defined as N-1.
- **Mean Squares (MF) –** This is the average variation.
- **F ratio (F) –** As stated previously, this ratio is used to compute the P value. This is the variance ratio.
- **P value –** Helps determine if your experimental results (variance between means) are statistically significant.

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Gender	70.38	1	70.38	F (1, 1) = 7.539	P=0.0144
Treatment	242.6	1	242.6	F (1, 1) = 25.98	P=0.0001
Dose	176.6	1	176.6	F (1, 1) = 18.92	P=0.0005
Gender x Treatment	72.45	1	72.45	F (1, 1) = 7.761	P=0.0132
Gender x Dose	11.07	1	11.07	F (1, 1) = 1.186	P=0.2923
Treatment x Dose	13.65	1	13.65	F (1, 1) = 1.462	P=0.2441
Gender x Treatment x Dose	1.87	1	1.87	F (1, 1) = 0.2004	P=0.6604
Residual	149.4	16	9.335		

Three-way ANOVA table

What do ANOVA results mean?

As you can see in the above table, using ANOVA can yield multiple P values that may or may not be useful in answering the question you are asking. This type of analysis requires careful consideration of the different factors and type of post-hoc test you want to use. Making sure that the question you want answered with ANOVA is clearly defined will help you decide what P value to focus on.

The different ANOVA tests divide the variability into different number of components, depending on the number of factors.

- **One-way ANOVA –** This test reports one P value. Variability is divided into two components; one component due to variability among group means (due to treatment, the factor) and a component that describes the variability within groups, which is noted as "Residual variation, within columns" in the one-way ANOVA results table. This analysis checklist for one-way ANOVA helps you evaluate your one-way ANOVA results.
- **Two-way ANOVA –** This test reports three P values. Variability is divided into three components and a residual error—interaction between rows and columns, among columns, among rows, and variation among replicates. Prism provides an analysis checklist that may help evaluate your results.
- **Three-way ANOVA –** This test reports seven P values. Variability is divided into seven components and a residual error: main effects, two-way interactions, three-way interaction, and variation among replicates. Prism offers guidance to take you through interpreting each of these components and provides an analysis checklist to evaluate your results.

SUMMARY

ANOVA is a very powerful statistical tool that can lead to novel findings, a greater understanding of the system being studied, and help improve experimental strategies. Our GraphPad Prism Guides will guide you through choosing the appropriate ANOVA test for your experiment.