*Ministry of Higher Education and Scientific Research*

*University of Babylon*

*College of Education for pure science*

*Department of Mathematics*


# *The Analysis of Variance*

# *(ANOVA)*

*A graduation research submitted to the Council of the Department of Mathematics in the College of Education for Pure Sciences / University of Babylon, as part of the requirements obtaining a bachelor's degree in mathematics*

**BY**

***Atyaf Zuhair***

***Supervised By***

***Rawasy Adnan Hameed***

**2023-2024**

**بِسْمِ اللهِ الرَّحْمنِ الرَّحِيم**

***("وَآخِرُ دَعْوَاهُمْ أَنِ الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِين")***

**(صدق الله العلي العظيم )**

**.[يونس: 10]**

# إهداء:

من قال أنا لها "نالها".

لم تكن الرحلة قصيرة ولا ينبغي لها أن تكون لم يكن الحلم قريبا و لا الطريق كان محفوفا بالتسهيلات

لكني فعلتها و نلتها .

الحمد لله حباً و شكراً و امتناناً ، الذي بفضله ها أنا اليوم أنظر

إلى حلماً طال انتظاره و قد أصبح واقعا أفتخر به.

إلى ملاكي الطاهر و قوتي بعد الله داعمتي الأولى و الأبدية "امي"

أهديك هذا الإنجاز الذي لولا تضحياتك لما كان له وجود، ممتنة لإن

الله قد اصطفاك لي من البشر أماً يا خير سند و عوض.

إلى من دعمني بلا حدود و أعطاني بلا مقابل

" أبي"

إلى من قيل فيهم:

سَنَشُدُّ عَضُدَكَ بِأَخِيكَ)

إلى من مدا يدهما دون كلل ولا ملل وقت ضعفي

"أخواني "أدامكم الله ضلعاً ثابتاً لي .

الى من تذكرني بقوتي وتقف خلفي كظلي "اختي الصغرى" .

## الشكر والتقدير

طُويت صفحة من التعب بفضل الله وتوفيقه ورعايته، وإنّني إذ أتقدم بالشكر لله أولًا، ولأساتذتي الأفاضل ثانيًا الذين لم يبخلوا عليّ بأيّ معلومة طوال مسيرتي الدراسية، وإلى كلّ من ساعدني لأُنجز هذا البحث العلمي الذي وضعت به خلاصة تعبي ومعرفتي ليكون بين يديّ أساتذتي وطلبة العلم، لهذا فإنّ ردّ الفضل لأهله هو أبسط ما يمكن أن أقدّمه في هذا المقام لأقول : شكرًا للدكتوره رواسي عدنان المشرفه على البحث وما ابدته من رعايه علميه متواصله وتوجيهات وملاحضات دقيقه ساعدتني على انجاز البحث .

شكرا لكلّ من بذل جهدًا ولو صغيرًا معي ليخرج هذا البحث إلى النور، كما أقدّم عظيم امتناني لأبي وأمي فقد كان منهم الدعم والتشجيع والتقدير، وكانوا لي منارة تنير لي الطريق المعتمة لأصل إلى ما أُريد، وأشكر إخواني وأخواتي وزملائي الذين لم يبخلوا عليّ بأي شيء .

الباحثة

## Abstract

ANOVAs are used to make comparisons across three or more groups of a dependent variable(s) with one or more independent variables. ANOVA is appropriate whenever you want to test differences between the means of an interval-ratio level dependent variable across three or more categories of an independent variable. There are two techniques for doing ANOVA: one-way ANOVA and two-way ANOVA  Also three-way ANOVA .

# Table of contents

# Introduction

In this research, I discussed the definition of ANOVA and what ANOVA is in statistics. I also discussed the types of ANOVA and factorial analysis of variance, how to do the ANOVA test, and also what happens when a second factor is added. I discussed the simple, main, and interaction effects in analysis of variance. I discussed the difference between ANOVA and the t-test. I discussed the limits of ANOVA and also learned about ANCOVA and multivariate analysis of variance (MANOVA). I compared ANCOVA and ANOVA, as well as compared ANOVA and MANO, how to perform F-tests in ANOVA, and is analysis of variance considered a similar and equal variance, as well as what does analysis of variance reveal? I also touched on the uses of ANIVA. It also touched on the uses of ANOVA in biomedical research .

## 1- What is the definition of ANOVA?

ANOVA stands for analysis of variance, and, true to its name, it is a statistical technique that analyzes how experimental factors influence the variance in the response variable from an experiment ANOVA, or Analysis of Variance, is a test used to determine differences between research results from three or more unrelated samples or groups.

You might use ANOVA when you want to test a particular hypothesis between groups, determining – in using one-way ANOVA – the relationship between an independent variable and one quantitative dependent variable.

An example could be examining how the level of employee training impacts customer satisfaction ratings. Here the independent variable is the level of employee training; the quantitative dependent variable is customer satisfaction.

You would use ANOVA to help you understand how employees of different training levels – for example, beginner, intermediate and advanced – with the null hypothesis for the test being that they have the same customer satisfaction ratings. If there is a statistically significant result, it means the null hypothesis is rejected – meaning the employee groups performed differently

The key word in 'Analysis of Variance' is the last one. 'Variance' represents the degree to which numerical values of a particular variable deviate from its overall mean. You could think of the dispersion of those values plotted on a graph, with the average being at the centre of that graph. The variance provides a measure of how scattered the data points

are from this central value Even though ANOVA involves complex statistical steps, it is a beneficial technique for businesses via use of AI. Organizations use ANOVA to make decisions about which alternative to choose among many possible options. For example, ANOVA can help to:

Compare the yield of two different wheat varieties under three different fertilizer brands.

Compare the effectiveness of various social media advertisements on the sales of a particular product

**Example 1:**

Three types of fertilizers are used on three groups of plants for 5 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply a one way ANOVA test at 0.05 significant level

| Fertilizer 1 | Fertilizer 2 | Fertilizer 3 |
|---|---|---|
| 6 | 8 | 13 |
| 8 | 12 | 9 |
| 4 | 9 | 11 |
| 5 | 11 | 8 |
| 3 | 6 | 7 |
| 4 | 8 | 12 |

**Solution:**

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1$: The means are not equal

| Fertilizer 1 | Fertilizer 2 | Fertilizer 3 |
|---|---|---|
| 6 | 8 | 13 |
| 8 | 12 | 9 |
| 4 | 9 | 11 |
| 5 | 11 | 8 |
| 3 | 6 | 7 |
| 4 | 8 | 12 |
| $\overline{X}_1 = 5$ | $\overline{X}_1 = 9$ | $\overline{X}_1 = 10$ |

Total mean, $\overline{X} = 8$

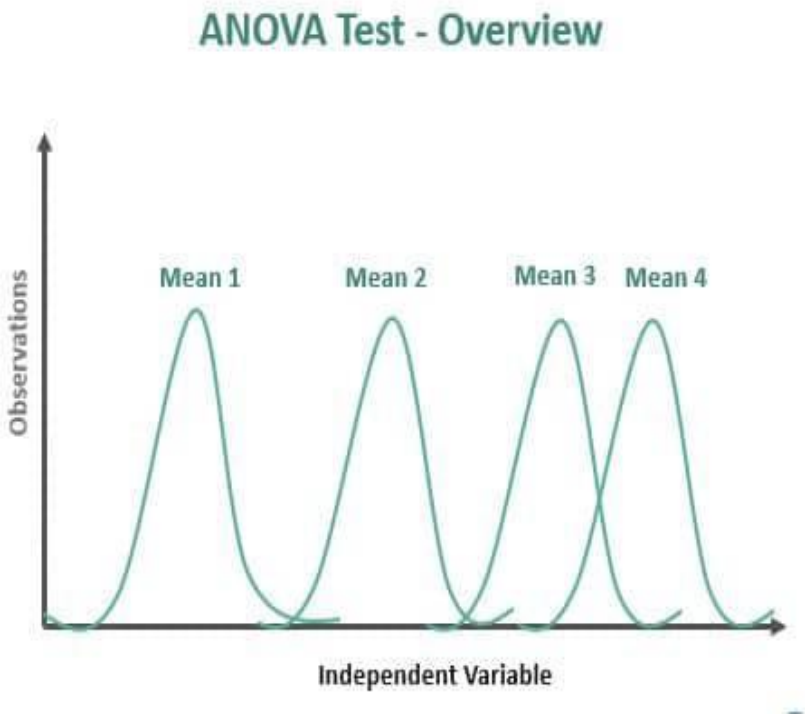
$n_1 = n_2 = n_3 = 6, k = 3$

$SSB = 6(5 - 8)^2 + 6(9 - 8)^2 + 6(10 - 8)^2$

$= 84$

$df1 = k - 1 = 2$

| Fertilizer 1 | $(X - 5)^2$ | Fertilizer 2 | $(X - 9)^2$ | Fertilizer 3 | $(X - 10)^2$ |
|---|---|---|---|---|---|
| 6 | 1 | 8 | 1 | 13 | 9 |
| 8 | 9 | 12 | 9 | 9 | 1 |
| 4 | 1 | 9 | 0 | 11 | 1 |
| 5 | 0 | 11 | 4 | 8 | 4 |
| 3 | 4 | 6 | 9 | 7 | 9 |
| 4 | 1 | 8 | 1 | 12 | 4 |
| $\overline{X}_1 = 5$ | Total = 16 | $\overline{X}_1 = 9$ | Total = 24 | $\overline{X}_1 = 10$ | Total = 28 |




## 2- What is ANOVA in statistics?

For a one-way ANOVA test, the overall ANOVA null hypothesis is that the mean responses are equal for all treatments. The ANOVA p-value comes from an F-test .

## 3- Types of ANOVA

There are various approaches to using ANOVA for your data analysis. Here's an introduction to some of the most common ones

## 3-1 One-way ANOVA

One-way ANOVA is its most simple form – testing differences between three or more groups based on one independent variable. For example, comparing the sales performance of different stores in a retail chain .

## a- When to use a one-way ANOVA?

Use a one-way ANOVA when you have collected data about one categorical independent variable and one quantitative dependent variable. The independent variable should have at least three levels ( i.e. at least three different groups or categorie.

ANOVA tells you if the dependent variable changes according to the level of the independent variable .

## For example:

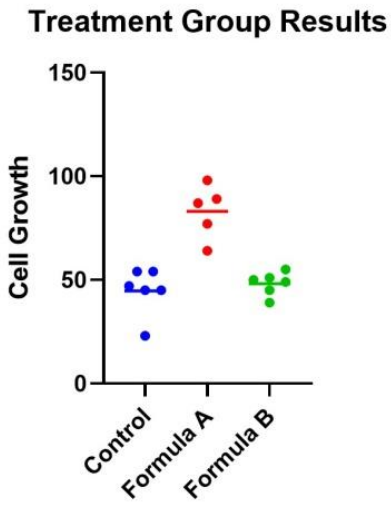
Your independent variable is social media use, and you assign groups to low, medium, and high levels of social media use to find out if there is a difference in hours of sleep per night.

Your independent variable is brand of soda, and you collect data on Coke, Pepsi, Sprite, and Fanta to find out if there is a difference in the price per 100ml.

You independent variable is type of fertilizer, and you treat crop fields with mixtures 1, 2 and 3 to find out if there is a difference in crop yield.

## b- Graphing one-way ANOVA

Below you can see a graph of the cell growth levels for each data point in each treatment group, along with a line to represent their mean. This can help give credence to any significant differences found, as well as show how closely groups overlap




## .. Determining statistical significance between groups

In addition to the graphic, what we really want to know is which treatment means are statistically different from each other. Because we are performing multiple tests, we'll use a multiple comparison correction. For our example, we'll use Tukey's correction (although if we were only interested in the difference between each formula to the control, we could use Dunnett's correction instead ) .

In this case, the mean cell growth for Formula A is significantly higher than the control (p<.0001) and Formula B (p=0.002), but there's no significant difference between Formula B and the control .

| Tukey's multiple comparisons test | Mean Diff. | 95.00% CI of diff. | Below threshold? | Summary | Adjusted P Value |
|---|---|---|---|---|---|
| Control vs. Formula A | -38.3 | -54.6 to -22.1 | Yes | **** | <0.0001 |
| Control vs. Formula B | -3.50 | -19.0 to 12.0 | No | ns | 0.8274 |
| Formula A vs. Formula B | 34.8 | 18.6 to 51.1 | Yes | *** | 0.0002 |

## 3-2 Two-way ANOVA

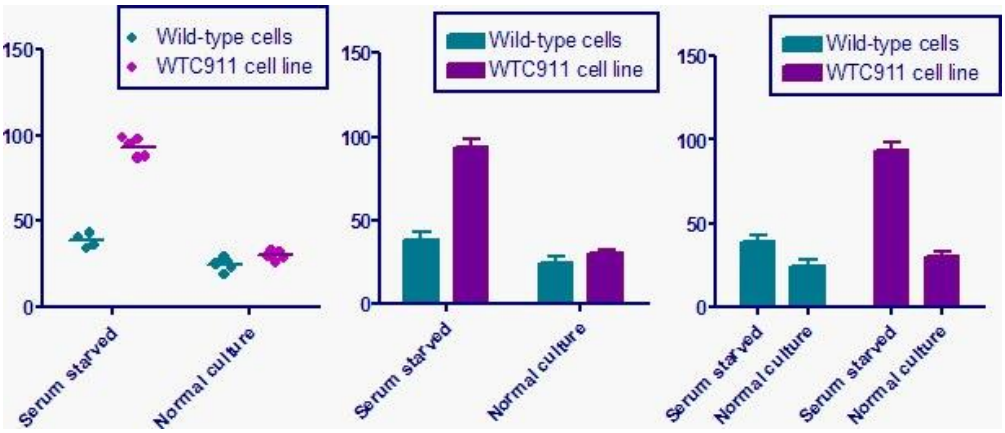
Used when there are two independent variables, two-way ANOVA allows for the evaluation of the individual and joint effects of the variables. For example, it could be used to understand the impact of both advertising spend and product placement on sales revenue .

### a- Two-way ANOVA example :

For two-way ANOVA, there are two factors involved. Our example will focus on a case of cell lines. Suppose we have a 2x2 design (four total groupings). There are two different treatments (serum-starved and normal culture) and two different fields. There are 19 total cell line "experimental units" being evaluated, up to 5 in each group (note that with 4 groups and 19 observational units, this study isn't balanced). Although there are multiple units in each group, they are all completely different replicates and therefore not repeated measures of the same unit

## b- Graphing two-way ANOVA

There are many options here. Like our one-way example, we recommend a similar graphing approach .




The graph above shows three ways to plot the sample data for two-way ANOVA. The graphs on the left and middle interleave the data sets. This is set on the second tab of the Format Graphs dialog. In this case, the data sets are defined by the figure legend, and the groups (rows) are defined by the labels on the X axis.

The graph on the right has the data sets grouped. In this graph, the labels on the X axis show the row title -- one per bar. Use the "number format" choice in the Format Axes dialog to change this to Column titles -- one per set of bars. With this choice, there wouldn't be much point in also having the legend shown in the box, and you would need to define the side by side bars ("serum starved" vs "normal culture" for this example) in the figure legend.

The graph on the left has the appearance set as a column dot plot. The other two graphs have the appearance set as bars with error bars plotted from the mean and SD. I prefer the column dot plot as it shows all the data, without taking up more space and without being harder to interpret. If you have a huge number of replicates, use a violin plot instead.

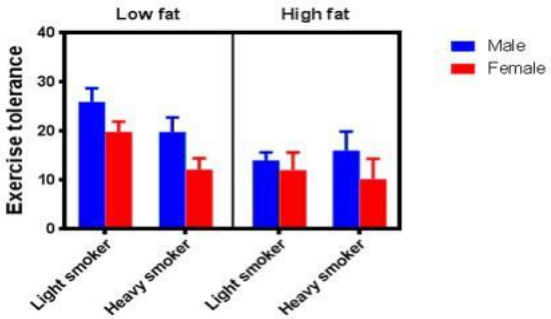
## c- What's the difference between one-way and two-way ANOVA tests?

This is defined by how many independent variables are included in the ANOVA test. One-way means the analysis of variance has one independent variable, two-way means the test has two independent variables

## 3-3 Three-way ANOVA

also called three-factor ANOVA, determines how a response is affected by three factors. For example, you might compare a response to drug vs. placebo in both men and women at two time points. Drug treatment is one factor .

## a- Graphing three_way ANOVA




## b- What is the difference between one-way, two-way and three-way ANOVA?

The number of "ways" in ANOVA (e.g., one-way, two-way, …) is simply the number of factors in your experiment

Although the difference in names sounds trivial, the complexity of ANOVA increases greatly with each added factor. To use an example from agriculture, let's say we have designed an experiment to research how different factors influence the yield of a crop

## 4- Factorial ANOVA

This variant of ANOVA is used when there are more than two independent variables. For example, a business might use a factorial ANOVA to examine the combined effects of age, income and education level on consumer purchasing habits..

## 5- How does an ANOVA test work?

ANOVA determines whether the groups created by the levels of the independent variable are statistically different by calculating whether the means of the treatment levels are different from the overall mean of the dependent variable.

If any of the group means is significantly different from the overall mean, then the null hypothesis is rejected.

## 6- What happens when you add a second factor?

If we have two different fields, we might want to add a second factor to see if the field itself influences growth. Within each field, we apply all three fertilizers which is still the main interest).

## 7- What are simple, main, and interaction effects in ANOVA?

Consider the two-way ANOVA model setup that contains two different kinds of effects to evaluate:

The $\alpha$ and $\beta$ factors are "main" effects, which are the isolated effect of a given factor. "Main effect" is used interchangeably with "simple effect" in some textbooks.

The interaction term is denoted as "$\alpha\beta$", and it allows for the effect of a factor to depend on the level of another factor. It can only be tested when you have replicates in your study. Otherwise, the error term is assumed to be the interaction term.

## 8- What is the difference between ANOVA and a t-test?

ANOVA is an extension of the t-test. If you only have two group means to compare, use a t-test. Anything more requires ANOVA.

## 9-Limitations of ANOVA

ANOVA can only tell if there is a significant difference between the means of at least two groups, but it can't explain which pair differs in their means. If there is a requirement for granular data, deploying further follow up statistical processes will assist in finding out which groups differ in mean value. Typically, ANOVA is used in combination with other statistical methods

## 10- Covariance (ANCOVA).

ANCOVA models include factors and covariates. Covariates are continuous independent variables that have a relationship with the dependent variable. Typically, covariates are nuisance variables that researchers cannot control during an experiment. Consequently, analysts include covariates in the model to control them statistically.

## 11- Repeated measures ANOVA

Repeated measures designs allow researchers to assess participants multiple times in a study. Frequently, the subjects serve as their own controls and experience several treatment conditions.

## 12- Multivariate analysis of variance (MANOVA)

MANOVA extends the capabilities of ANOVA by assessing multiple dependent variables simultaneously. The factors in MANOVA can influence the relationship between dependent variables instead of influencing a single dependent variable.

There's even a MANCOVA, which is MANOVA plus ANCOVA! It allows you to include covariates when modeling multiple independent variables

## 13- What is ANCOVA

ANCOVA, or the analysis of covariance, is a powerful statistical method that analyzes the differences between three or more group means while controlling for the effects of at least one continuous covariate

## 14- ANOVA vs ANCOVA

ANCOVA, which stands for Analysis of Covariance, is similar to ANOVA, but includes an additional variable that is used to adjust for differences in the groupsbeing compared. ANCOVA is useful when there are variables that can affect the outcome of

the experiment but are not of primary interest. ANOVA, on the other hand, assumes that the groups being compared are independent and does not account for any confounding variables.

## 15- ANOVA vs MANOVA

MANOVA, which stands for Multivariate Analysis of Variance, is a statistical method used to analyze data that has multiple dependent variables. ANOVA, on the other hand, is used to analyze data with a single dependent variable. MANOVA is useful when multiple outcome measures are being considered, such as in a clinical trial where multiple parameters are being measured to evaluate the efficacy of a drug.

## 16- How F-tests work in Analysis of Variance (ANOVA)?

Analysis of variance (ANOVA) uses F-tests to statistically assess the equality of means when you have three or more groups. In this post, I'll answer several common questions about the F-test.

## 17- Does ANOVA assume equal variance?

Yes, for conducting an ANOVA test, the data sets need to have 'means' that have similar or equal variance .

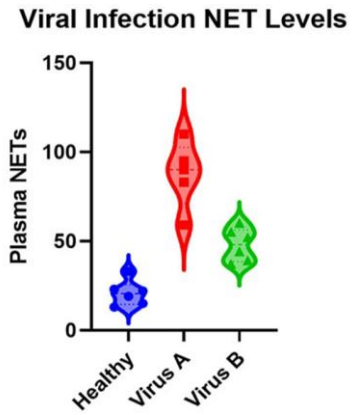
## 18- What ANOVA Reveals?

ANOVA splits an observed aggregate variability inside a data set into two parts: systematic factors and random factors. The systematic factors influence the given data set, while the random factors do not .

## 19- What is ANOVA used for?

ANOVA, or (Fisher's) analysis of variance, is a critical analytical technique for evaluating differences between three or more sample means from an experiment. As the name implies, it partitions out the variance in the response variable based on one or more explanatory factors.as you will see there are many types of ANOVA such as one-, two-, and three-way ANOVA as well as nested and repeated measures ANOVA. The graphic below shows a simple example of an experiment that requires ANOVA in which researchers measured

the levels of neutrophil extracellular traps (NETs) in plasma across patients with



Many researchers may not realize that, for the majority of experiments, the characteristics of the experiment that you run dictate the ANOVA that you need to use to test the results. While it's a massive topic (with professional training needed for some of the advanced techniques), this is a practical guide covering what most researchers need to know about ANOVA

. One should use the ANOVA test when one collects the data for one category of an independent variable having three different types and the data for contextual dependent variable too. Then, analysts use it to know the effect on the dependent variable concerning the change in the independent variable.

.For instance, if one has to use the Analysis of Variance test to find the effect of social media use on the users' sleep, then one has to assign three types – low usage, medium usage, and high usage to the social media variable. Only then is it possible to find contrast in the sleeping pattern of the users

## 20- Uses of ANOVA in Biomedical Research:

ANOVA is a useful tool for biomedical researchers to compare means across multiple groups. It can be used to analyze data from experiments that involve testing the effectiveness of a new drug, comparing the survival rates of different treatments for a disease, or analyzing the efficacy of a new medical device

## References


1- https://www.investopedia.com/terms/a/anova.asp
2- https://www.graphpad.com/guides/the-ultimate-guide-to-anova
3- https://www.scribbr.com/statistics/one-way-anova/
4- https://www.spotfire.com/glossary/what-is-analysis-of-variance-anova
5- https://datatab.net/tutorial/anova
6- https://statisticsbyjim.com/anova/
7- https://www.editage.com/blog/anova-types-uses-assumptions-a-quick-guide-for-biomedical-researchers/
8- https://surveysparrow.com/blog/anova/
9- https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/
10- https://www.wallstreetmojo.com/anova/
11- https://www.sciencedirect.com/topics/social-sciences/analysis-of-variance