Web of Statistics

Turning Data Into Knowledge

Procedures Manual

www.webofstatistics.com

Created by Prof. Dr. Silvio David Rodríguez based on material-dashboard template, licensed under MIT – Creative Tim 2019 (https://www.creative-tim.com/product/material-dashboard).

December 2024

Table of contents.

1. Features and Responsibilities.

This website provides a variety of tools used in the field of statistics, including descriptive statistics, probability table data, statistical inference, process control charts, analysis of variance (ANOVA) for both single-factor and multi-factor cases, regression analysis, and multivariate statistical tools.

The website was developed under Django environment as a collaborative effort by a group of professors and students to offer free, accessible statistical tools for educational purposes. Its primary aim is to assist in solving academic problems and to provide valuable resources for those learning and applying statistical concepts.

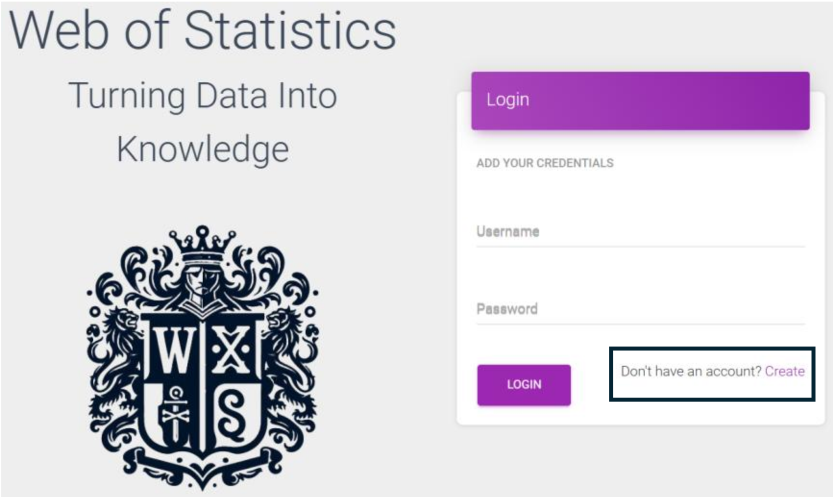
The use of this website or any of its components for personal or professional analyses is at the discretion of each user. The creators assume no responsibility for any potential errors that may arise from its use.

We kindly ask users to consider making a donation to help cover the hosting costs and support the ongoing maintenance of this free and ad-free platform. Your contribution, no matter the size, will help ensure that these resources remain available to everyone without interruptions.

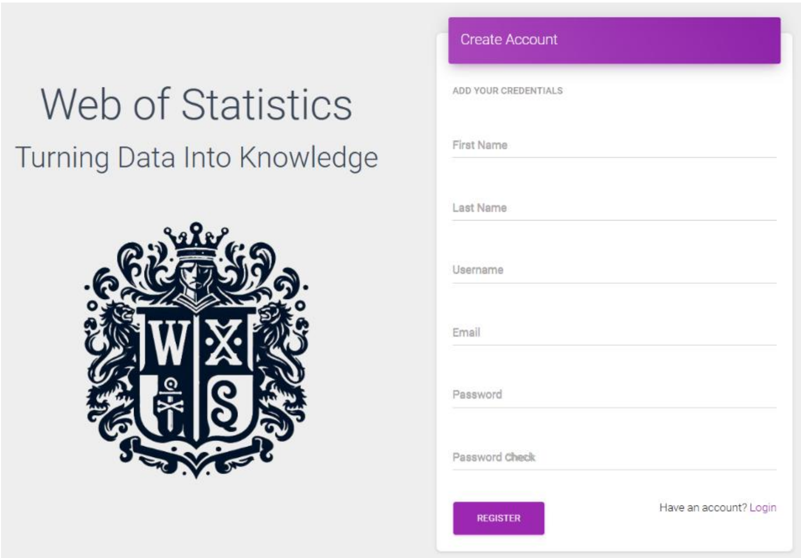
We also encourage users to report any issues they may encounter while using the website or to share suggestions for improvement. Your feedback is invaluable in helping us enhance the user experience and ensure the site continues to meet your needs effectively. Please feel free to contact us at webofstatistics@gmail.com for any comments or suggestions.

2. Login and Create an Account.

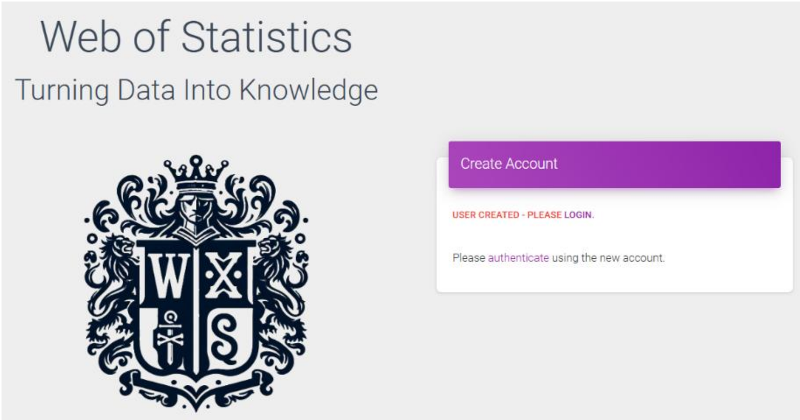
Once on the website www.webofstatistics.com, users will encounter the login page, where they can enter their credentials (username and password) and click 'Login' to gain access. If they do not have credentials, they must create an account by clicking the link associated with the word 'Create' next to the question 'Don't have an account?'



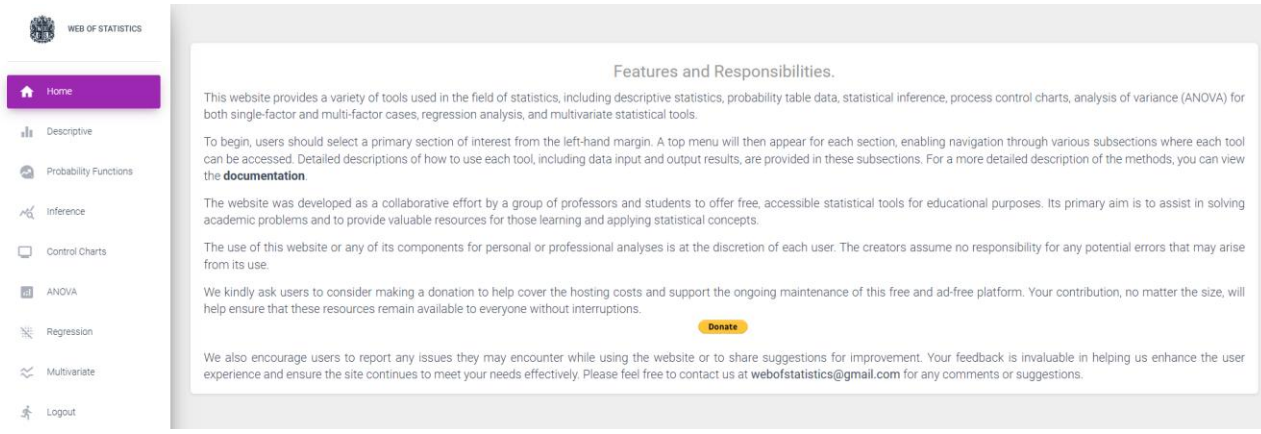
After clicking 'Create,' users will be redirected to the form they need to complete to create an account by entering all the required information and clicking the 'Register' button.



If the data is validated, upon clicking ‘Register,’ the message ‘User created – Please Login’ will appear on the screen, and the user will be able to log in to the site as a registered user.



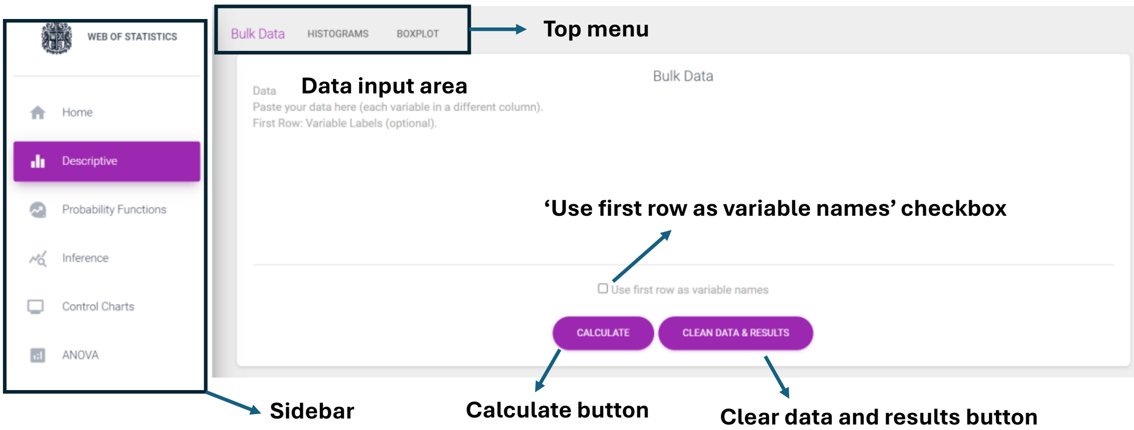
Once the credentials are completed and 'Login' is clicked, the main page will appear on the screen, which includes a right sidebar that allows the user to select the available options.



Each of the sections found in the sidebar will be discussed in detail in the following sections of this procedures manual. If users wish to log out, they should click the 'Logout' button.

3. Descriptive Statistics.

Clicking the 'Descriptive' button in the sidebar will display the options in a menu at the top. This section covers basic descriptive statistics calculations for bulk tabulated data. The three options available in the top menu are 'Bulk Data,' 'Histograms,' and 'Boxplot,' with the 'Bulk Data' section pre-selected by default.



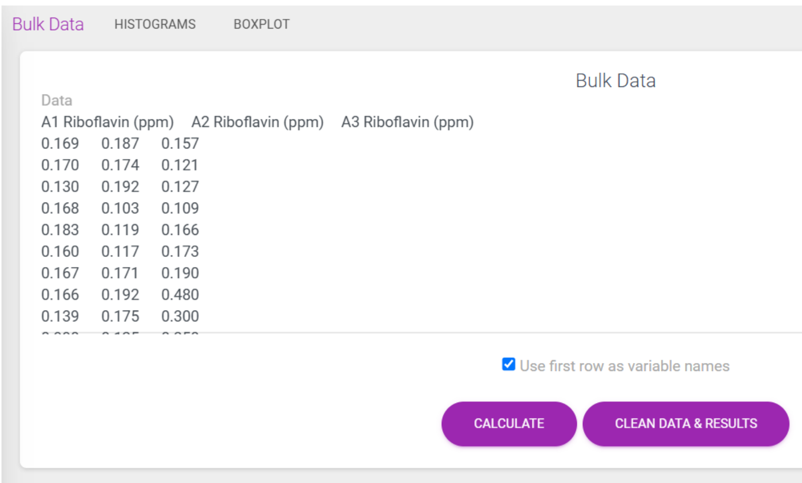
Once an option is selected from the top menu, it's time to enter the data in the 'data input area.' To simplify the use of the web platform, the data entry process is consistent across all sections and subsections. Simply follow the instructions on how the data should be organized and copy them from a spreadsheet.

For example, if 'Bulk Data' is selected, the data for one or more variables should be in different columns. It is optional to use the first row for the variable names, but if the variable names are provided, the checkbox that says 'Use first row as variable names' must be checked. If no names are provided in the first row and only data is entered, the mentioned checkbox should be unchecked, and automatic names will be generated for each variable entered.

Here is an example of how data can be entered from a spreadsheet, where there are three variables (it could be more or fewer) and some data. Simply copy and paste the data by right-clicking with the mouse or using the keyboard shortcuts (control + c and control + v).

A1 Riboflavin (ppm)	A2 Riboflavin (ppm)	A3 Riboflavin (ppm)
0.169	0.187	0.157
0.170	0.174	0.121
0.130	0.192	0.127
0.168	0.103	0.109
0.183	0.119	0.166
0.160	0.117	0.173
0.167	0.171	0.190
0.166	0.192	0.480
0.139	0.175	0.300
0.990	0.125	0.350
0.110	0.131	0.148
0.146	0.143	0.172
0.183	0.152	0.185
0.167	0.190	0.106
0.135	0.176	0.118
0.148	0.600	0.990

This is how the data will appear once pasted onto the page. It may seem like the values are misaligned, depending on the size of the names or the number of digits in the values, but since they are separated by tabs, there will be no issues with that. The values should use a period ('.') as the decimal symbol, and scientific notation can also be used for the values.

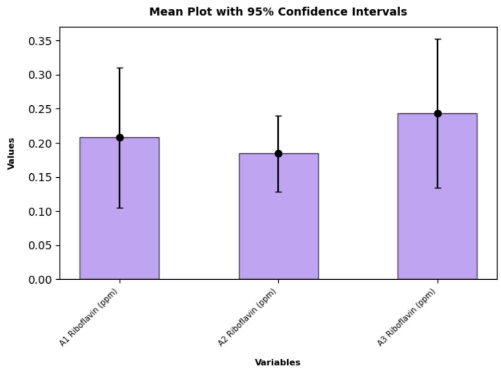


In this case, since the variable names were included in the first row, the checkbox that says 'Use first row as variable names' should be checked. Once this is done, you can click the 'Calculate' button. If incorrect values were entered, you can click the 'Clean data & results' button.

After selecting the 'Bulk Data' option, the results will include a table with all the most commonly used values in descriptive statistics and a bar chart, showing the average values for each variable and their 95% confidence intervals, as shown below.

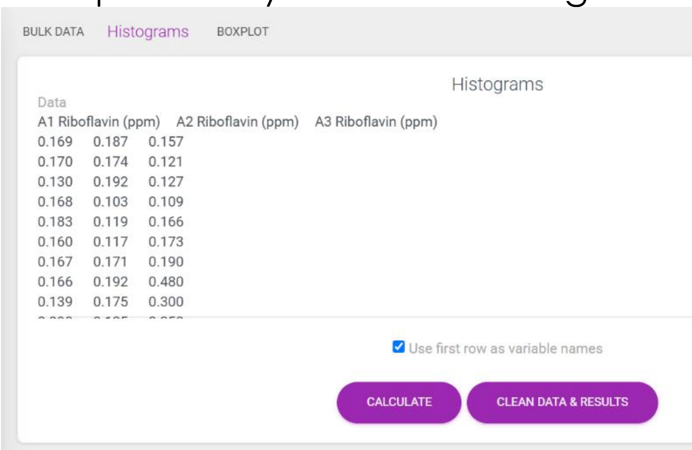
Results

Variable	Size	Min.	Max.	Range	First Quartile (25%)	Median	Third Quartile (75%)	Mean	Sample Variance	Sample Std. Dev.	Std. Error	Margin of Error (95% conf.)	C.I. of Mean (95% conf.)	Skewness	Kurtosis	Shapiro-Wilk W	Shapiro-Wilk p
A1 Riboflavin (ppm)	16	0.11	0.99	0.88	0.14425	0.1665	0.16925	0.20819	0.043866	0.20944	0.052361	± 0.10263	0.10556 ; 0.31081	3.937	15.65	0.3593	1.774e-07
A2 Riboflavin (ppm)	16	0.103	0.6	0.497	0.1295	0.1725	0.18775	0.18419	0.013194	0.11487	0.028716	± 0.056284	0.1279 ; 0.24047	3.535	13.46	0.5168	2.927e-06
A3 Riboflavin (ppm)	16	0.106	0.99	0.884	0.1255	0.169	0.2175	0.24325	0.049869	0.22331	0.055828	± 0.10942	0.13383 ; 0.35267	2.858	8.944	0.6111	2.065e-05



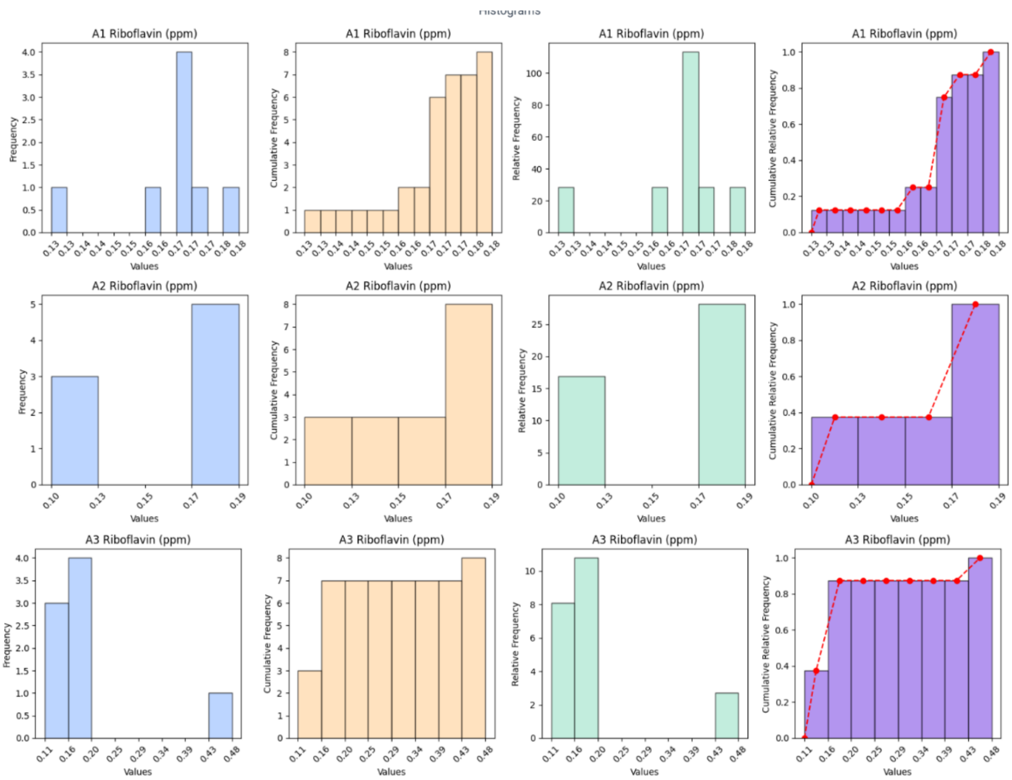
If the variables do not have the same size, the process can proceed in the same way. In the case of missing values in any of the variables, they are not considered and are removed, shortening the size of that particular variable.

Once a calculation is made, system retains the entered values. When another option is selected, such as 'Histograms,' the values will already be preloaded, allowing you to click the 'Calculate' button. For example, in the previous case, selecting the 'Histograms' option will yield the following:

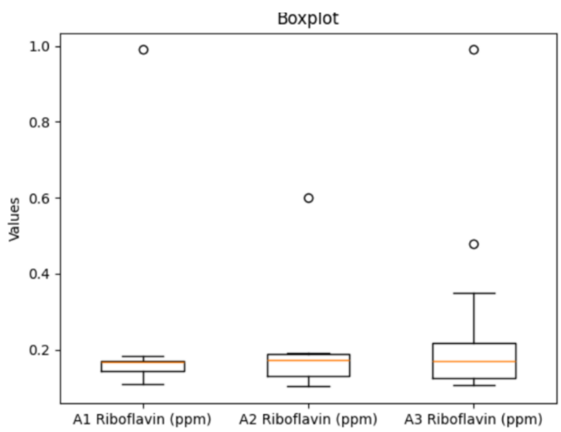


By clicking the 'Calculate' button, the results for that option will be displayed. However, it is important to note that the data can be deleted, and new data can be reloaded in any option.

The results include frequency histogram charts, relative frequency histograms, cumulative frequency histograms, and cumulative relative frequency histograms with frequency polygon.



If the last option, 'Boxplot,' is selected, the procedure is similar to the previous one. By clicking the 'Calculate' button, the boxplot chart for the variables that have been entered will be generated.



4. Probability Functions and Tables.

This section is dedicated to exploring the probability tables of four of the most commonly used probability distribution functions. By clicking ‘Probability’ in the sidebar, four options will appear in the top menu, each representing a static table for the most common functions (standard normal, Student's t, Chi-square, and Fisher-Snedecor F). Clicking any of these four options will simply display the different tables that relate the 'x' values to the corresponding probability 'p' values.

PROBABILITY FUNCTIONS Standard Normal Table STUDENT'S T TABLE CHI SQUARED TABLE F FISHER SNEDECOR TABLES

Standard Normal Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879

PROBABILITY FUNCTIONS STANDARD NORMAL TABLE Student's t Table CHI SQUARED TABLE F FISHER SNEDECOR TABLES

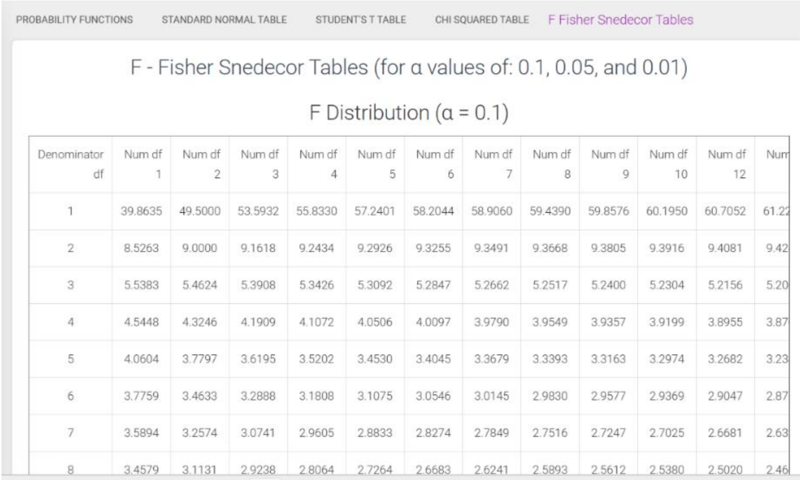
Student's t Table

df	0.8	0.85	0.9	0.95	0.975	0.99	0.995
1	3.0777	4.1653	6.3138	12.7062	25.4517	63.6567	127.3213
2	1.8856	2.2819	2.9200	4.3027	6.2053	9.9248	14.0890
3	1.6377	1.9243	2.3534	3.1824	4.1765	5.8409	7.4533
4	1.5332	1.7782	2.1318	2.7764	3.4954	4.6041	5.5976
5	1.4759	1.6994	2.0150	2.5706	3.1634	4.0321	4.7733
6	1.4398	1.6502	1.9432	2.4469	2.9687	3.7074	4.3168

PROBABILITY FUNCTIONS STANDARD NORMAL TABLE STUDENT'S T TABLE Chi Squared Table F FISHER SNEDECOR TABLES

χ^2 (chi squared) Table

df	0.005	0.010	0.025	0.050	0.100	0.500	0.900	0.950	0.975	0.990	0.995
1	0.0000	0.0002	0.0010	0.0039	0.0158	0.4549	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	1.3863	4.6052	5.9915	7.3778	9.2103	10.5966
3	0.0717	0.1148	0.2158	0.3518	0.5844	2.3660	6.2514	7.8147	9.3484	11.3449	12.8382
4	0.2070	0.2971	0.4844	0.7107	1.0636	3.3567	7.7794	9.4877	11.1433	13.2767	14.8603
5	0.4117	0.5543	0.8312	1.1455	1.6103	4.3515	9.2364	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	5.3481	10.6446	12.5916	14.4494	16.8119	18.5476

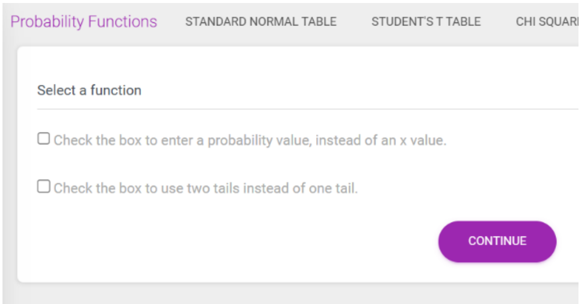


F - Fisher Snedecor Tables (for α values of: 0.1, 0.05, and 0.01)

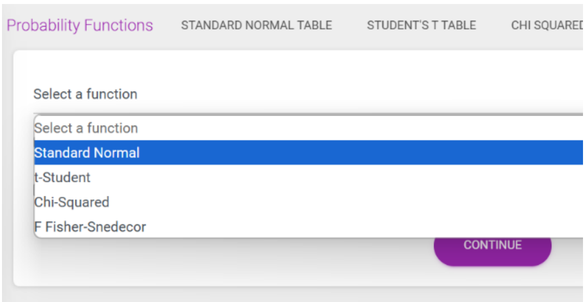
F Distribution (α = 0.1)

Denominator df	Num df 1	Num df 2	Num df 3	Num df 4	Num df 5	Num df 6	Num df 7	Num df 8	Num df 9	Num df 10	Num df 12	Num
1	39.8635	49.5000	53.5932	55.8330	57.2401	58.2044	58.9060	59.4390	59.8576	60.1950	60.7052	61.22
2	8.5263	9.0000	9.1618	9.2434	9.2926	9.3255	9.3491	9.3668	9.3805	9.3916	9.4081	9.42
3	5.5383	5.4624	5.3908	5.3426	5.3092	5.2847	5.2662	5.2517	5.2400	5.2304	5.2156	5.20
4	4.5448	4.3246	4.1909	4.1072	4.0506	4.0097	3.9790	3.9549	3.9357	3.9199	3.8955	3.87
5	4.0604	3.7797	3.6195	3.5202	3.4530	3.4045	3.3679	3.3393	3.3163	3.2974	3.2682	3.23
6	3.7759	3.4633	3.2888	3.1808	3.1075	3.0546	3.0145	2.9830	2.9577	2.9369	2.9047	2.87
7	3.5894	3.2574	3.0741	2.9605	2.8833	2.8274	2.7849	2.7516	2.7247	2.7025	2.6681	2.63
8	3.4579	3.1131	2.9238	2.8064	2.7264	2.6683	2.6241	2.5893	2.5612	2.5380	2.5020	2.46

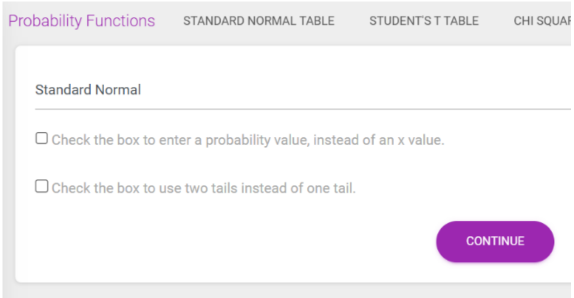
However, if you prefer not to work with tables, the first option, 'Probability Functions,' allows you to obtain the same values as those in the tables for each of the mentioned functions.



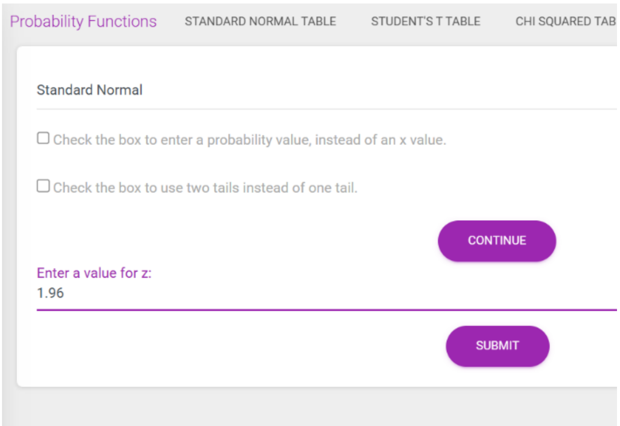
You simply need to select the desired function, for example, 'Standard Normal.' Upon selecting a function, the input fields below the selector will change accordingly.



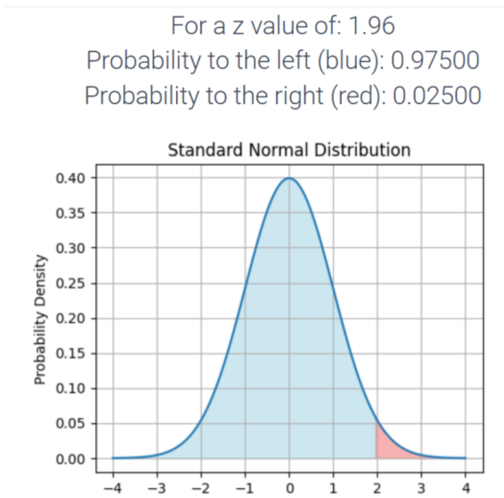
In the case of the standard normal function, several options are available. You can input either the value of ‘x’ to obtain ‘p’ or the value of ‘p’ to obtain ‘x.’ Additionally, you can select either a single value or two values to account for unilateral or bilateral probability cases.



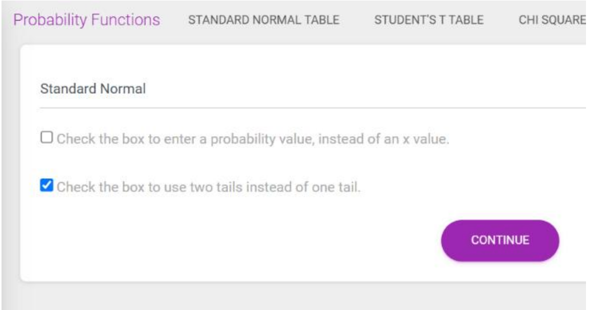
Suppose you want to find the probability value associated with x = 1.96 (z = 1.96). In this case, both checkboxes should be left unchecked, and you need to press 'Continue.' Doing this will reveal new fields below, which must be filled in with the necessary data in this example, the value 1.96.

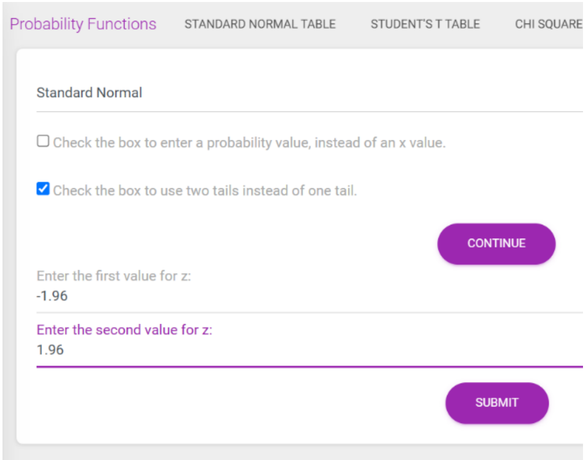


And the result after pressing 'Submit' will be as follows:

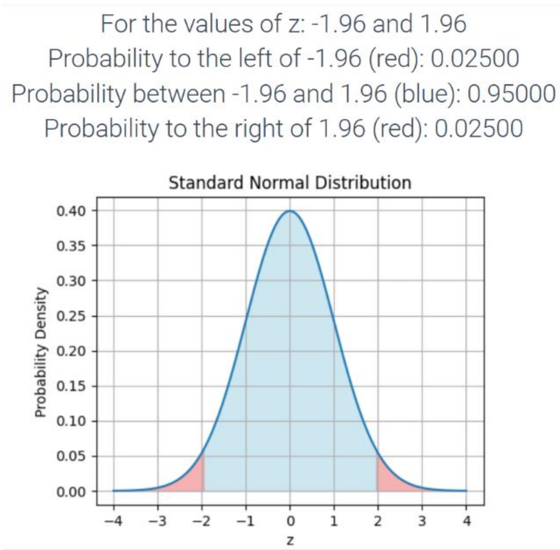


For example, in the case of a bilateral scenario, with 'x' values of -1.96 and 1.96:

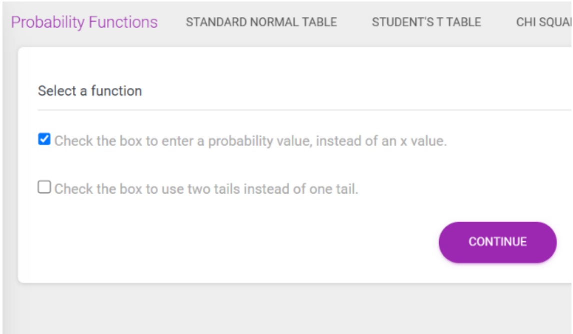


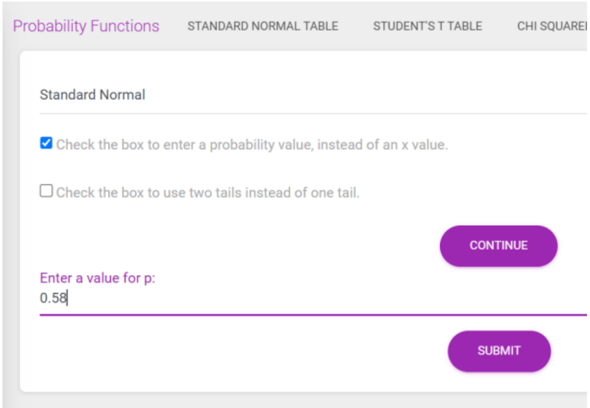


After entering the values and pressing 'Submit,' the result will be:

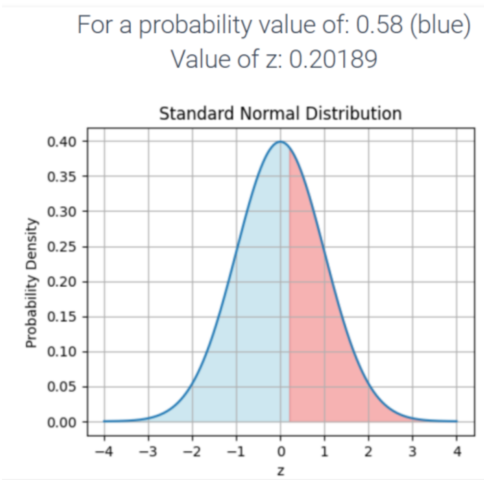


The same logic can be applied if, instead of entering the value of 'x,' you enter the value of 'p.' For example, in a unilateral case, by entering "p = 0.58" (remember that the values of 'p' can range between 0 and 1).

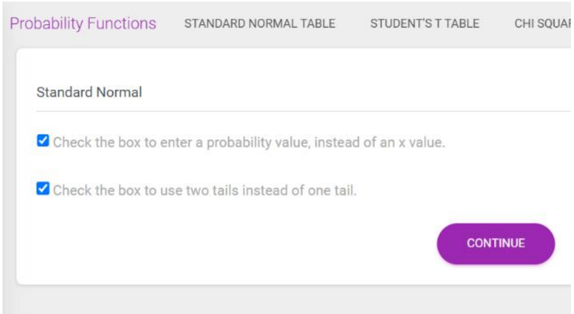


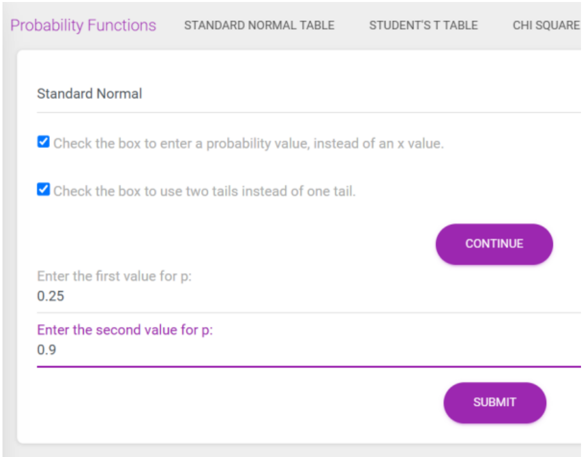


After entering the values and pressing 'Submit,' the result will be:

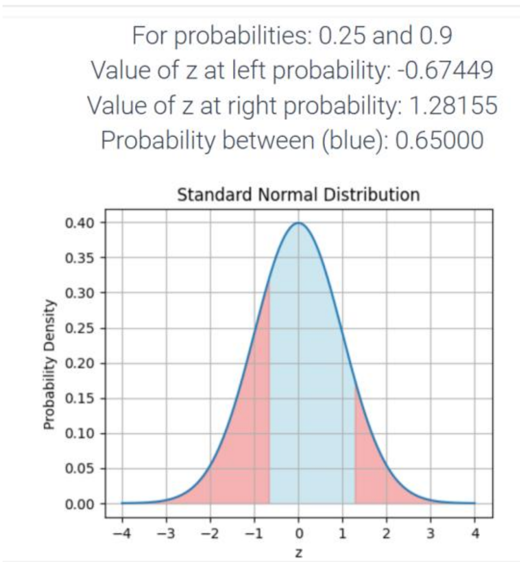


For example, in the case of a bilateral scenario, with 'p' values of 0.25 and 0.9:



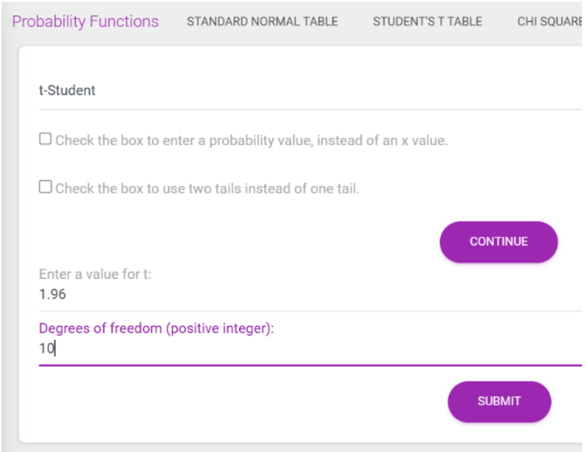


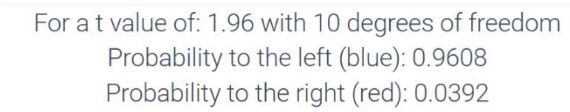
After entering the values and pressing 'Submit,' the result will be:

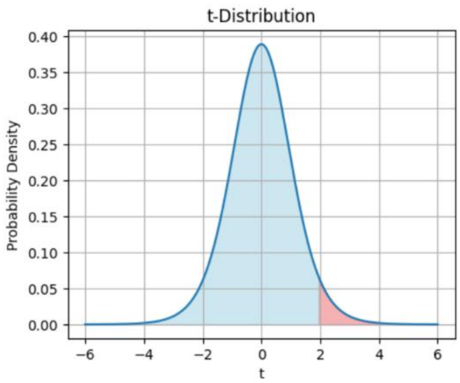


For the remaining functions, the method for preselecting the scenario is similar. After pressing the 'Continue' button, additional data will be requested, as these functions require the input of one or more degrees of freedom to generate the results.

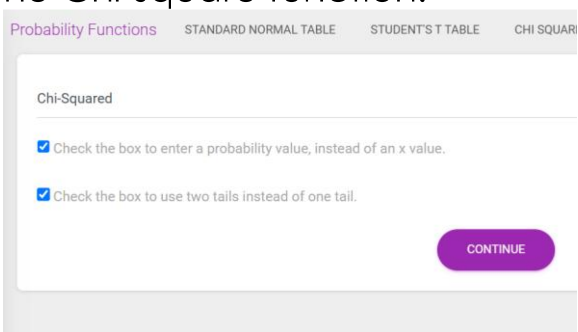
Example for Student's t: Obtain the value of 'p' with t = 1.96 and 10 degrees of freedom:

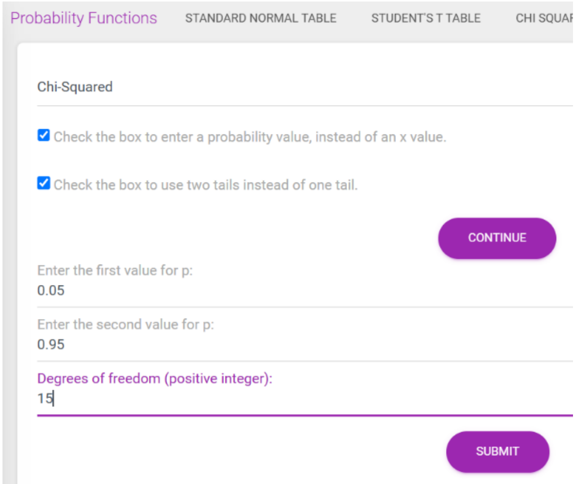


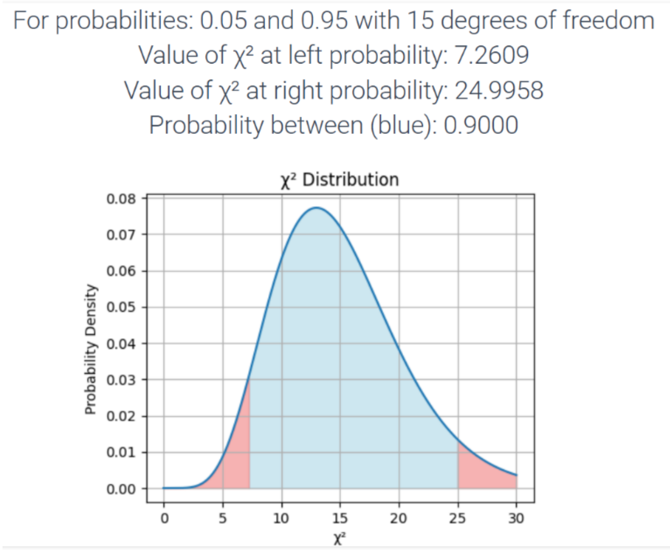




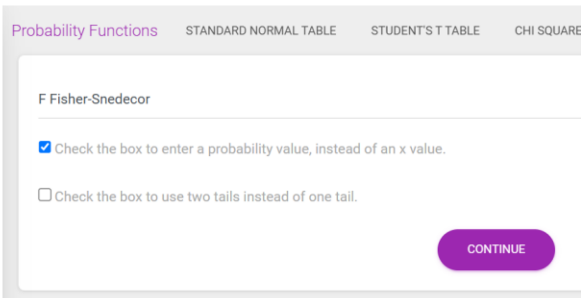
Example using the Chi-square function:

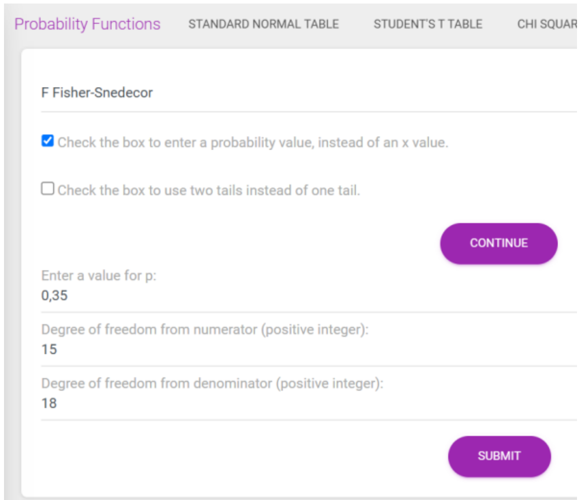




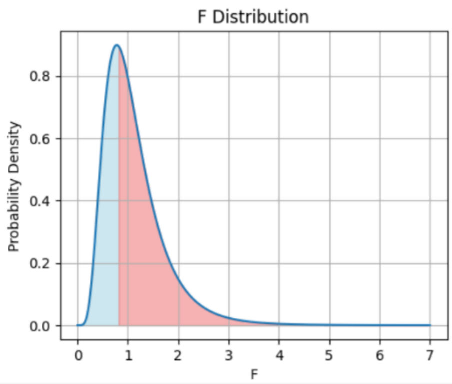


In the case of using the Fisher-Snedecor F function, two values for the degrees of freedom must be entered:





For a probability value of: 0.35 (blue) with 15 and 18 degrees of freedom
Value of F: 0.8173



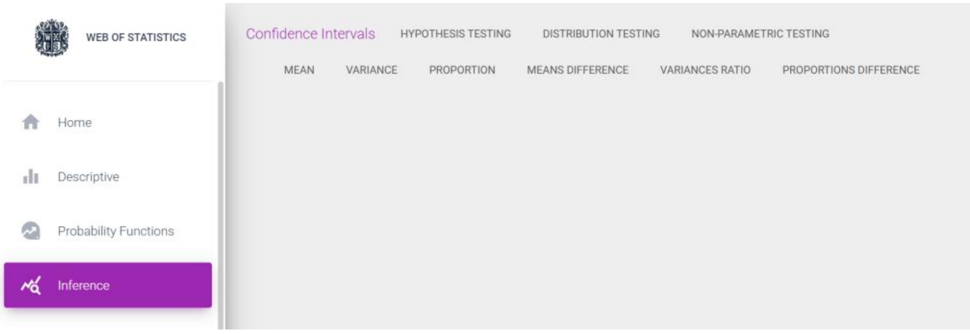
The possible combinations remain the same as those shown with the standard normal function; however, in the other cases, the degrees of freedom data is required. Not all possible cases are displayed here to avoid making this section too lengthy.

5. Inference.

This section includes several subsections defined in the top menu, where you can find 'Confidence Intervals,' 'Hypothesis Testing,' 'Distribution Testing,' and 'Non-Parametric Testing.' Each of these is further subdivided into different cases.

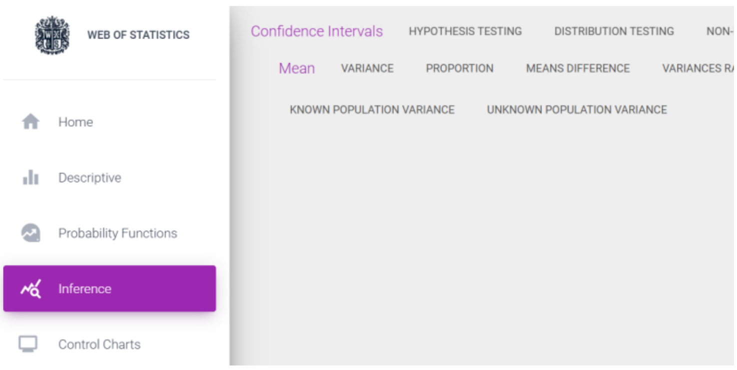
5.1 Confidence Intervals.

Let's start with 'Confidence Intervals.' This section, as the name suggests, is dedicated to the calculation of confidence intervals commonly used in various probability and statistics problems.

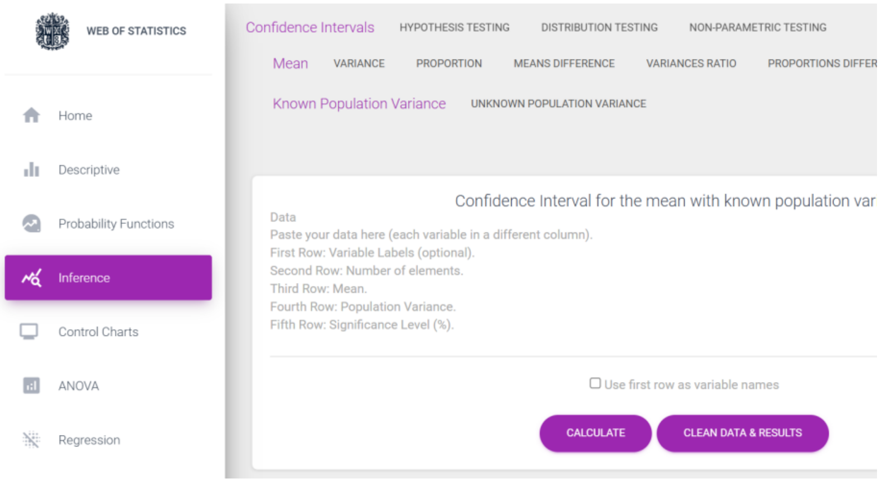


Within this section, you can navigate through the different cases, such as confidence intervals for the mean, variance, proportion, difference of means, ratio of variances, and difference of proportions. Upon selecting one of these cases, any available subcases will be displayed.

In the case of the confidence interval for the mean, there are two subcases: 'Population Variance Known' and 'Population Variance Unknown,' as shown in the following image:



By clicking on one of the subcases, a panel will appear where the required data can be entered, and the confidence interval calculation can be performed. Suppose we want to calculate the confidence interval for 2 variables with known variance; we select the subcase ‘Known Population Variance.’

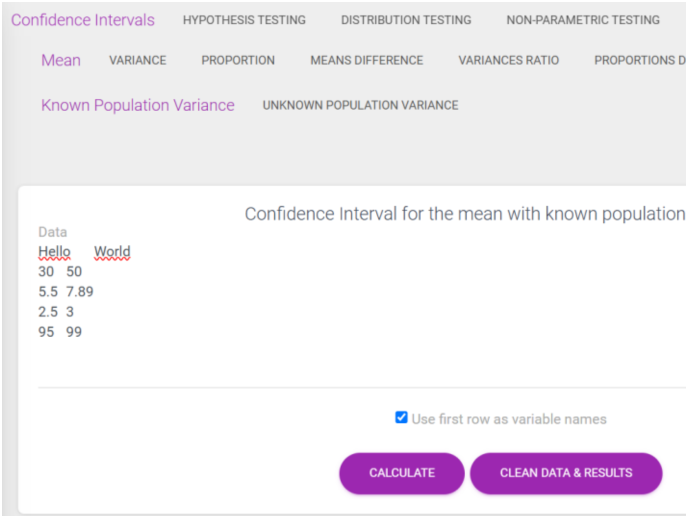


Here, the data from the spreadsheet must be pasted, following the specified guidelines: each column should contain one or more variables. For our example, we will consider 2 variables. If the variable names are included in the first row (optional), the checkbox ‘Use first row as variable names’ should be selected; otherwise, it should not. Finally, in the subsequent rows, the following values must be entered for each variable: the sample size (n), the mean, the population variance, and the confidence level expressed as a percentage (%).

The data in the spreadsheet will look like this:

Hello	World
30	50
5.5	7.89
2.5	3
95	99

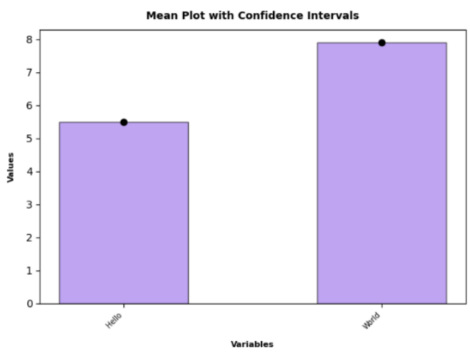
And once pasted onto the page, they will appear like this:



In this case, we have two variables, one called 'Hello' and the other 'World.' The first has 30 elements (n) and the second has 50. The means are 5.5 and 7.89, respectively, and the known population variances are 2.5 and 3.0. In the last row, the confidence level in percentage (%) should be entered. For example, for a confidence level of 95%, as in the case of the first variable, the significance level (α) is 0.05, and for the second variable, a confidence level of 99% corresponds to a significance level (α) of 0.01. After entering the required data and pressing the 'Calculate' button, the results will be displayed, as shown below.

Results

Variable	number of elements	Population Variance	Confidence Level	Mean	Confidence Interval
Hello	30.0	2.5	95.0	5.5	(5.4819, 5.5181)
World	50.0	3.0	99.0	7.89	(7.8869, 7.8931)

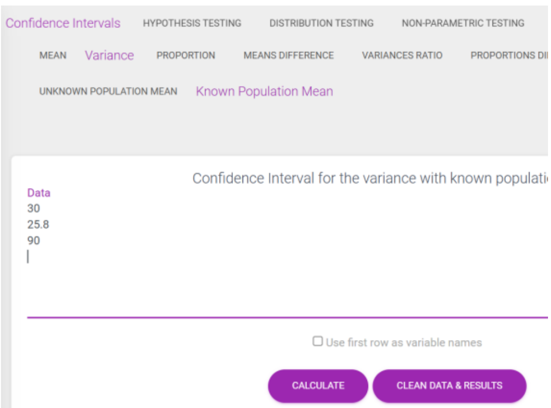


After pressing 'Calculate,' both the entered data, the checkbox selections, and the results will remain, even if you leave that section. However, it is recommended to press the 'Clear Data & Results' button to remove these values if they are no longer needed. If you want to enter new values, you can either rewrite the previously pasted values or enter new ones after pressing the 'Clear Data & Results' button.

The case of the confidence interval for the mean with unknown population variance is very similar to the previous case, except that instead of entering the population variance, you must enter the sample variance. If you have the values of a variable in bulk, you can use the 'Descriptive' section to obtain the estimators derived from the sample.

For the case of the confidence interval for variance, there are two different subcases. The first is based on not knowing the value of the population mean, and the second is when the population mean is known. In the first subcase, the sample size (n), sample variance, and confidence level (%) must be entered. In the second case, instead of the sample variance, the variance calculated with the population mean should be used.

Suppose we want to calculate the confidence interval for a sample of 30 data points, with a sample variance of 25.8 and a confidence level of 90%. The corresponding subcase is then selected, and the values are entered. In this case, we will enter data for only one variable, but more than one variable can also be entered.

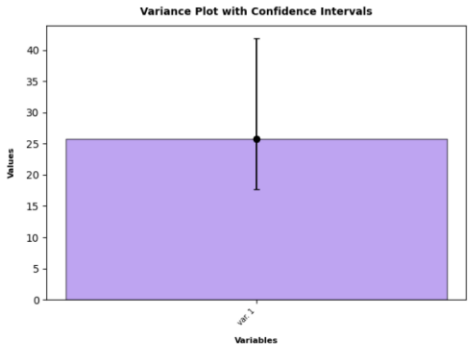


Since the variable name has not been entered, the checkbox at the bottom should not be selected.

The results obtained after pressing the 'Calculate' button are shown below.

Results

Variable	number of elements	Confidence Level	Sample Variance Using μ	Confidence Interval
var. 1	30.0	90.0	25.8	(17.682, 41.854)

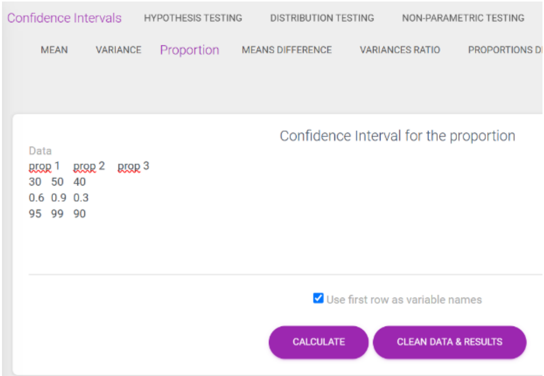


In a similar way, the procedure is followed for the case where the population mean is known, except that the variance entered in this case should be the variance calculated using that value.

For the subcase of the confidence interval for proportion, when selecting it, the proportion of the sample is requested. In the following example, there are 3 variables, and in the spreadsheet, they look like this:

prop 1	prop 2	prop 3
30	50	40
0.6	0.9	0.3
95	99	90

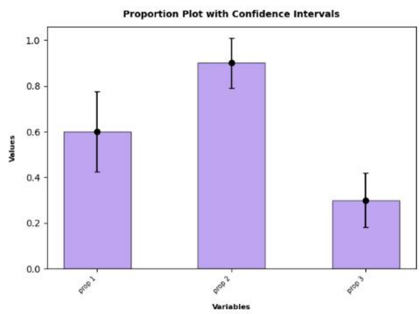
In this case, you should copy and paste the data and select the checkbox at the bottom to use the variable names.



Remember that proportions are values that must be between 0 and 1. In this case, the variables have proportions of 0.6, 0.9, and 0.3, respectively, with confidence levels of 95%, 99%, and 90%. The results are shown below:

Results

Variable	number of elements	Confidence Level	Sample Proportion	Confidence Interval
prop 1	30.0	95.0	0.6	(0.4247, 0.7753)
prop 2	50.0	99.0	0.9	(0.79072, 1.0093)
prop 3	40.0	90.0	0.3	(0.18082, 0.41918)

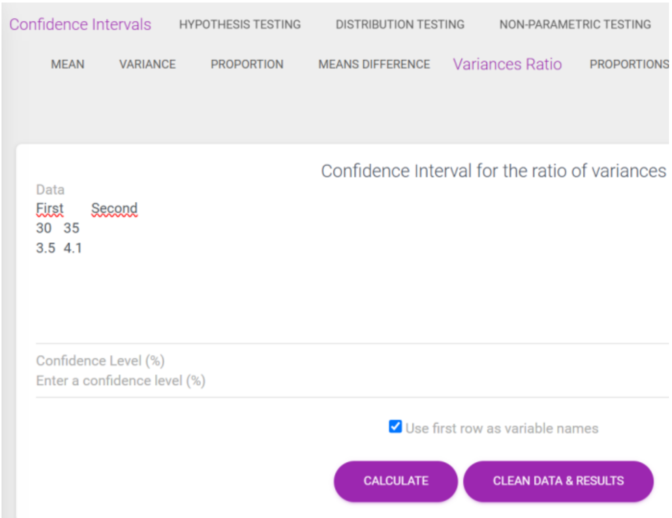


The remaining three subcases to analyze are based on the comparison of two variables, and for these, data must only be entered for two variables, meaning only two columns. The possible subcases are the confidence interval for the difference of means, the ratio of variances, and the difference of proportions.

Let us first analyze the case of the confidence interval for the ratio of variances. Here, the variable names are optionally provided in the first row, the number of elements in the second row, and the sample variance in the third row. The data in the spreadsheet will look like this:

First	Second
30	35
3.5	4.1

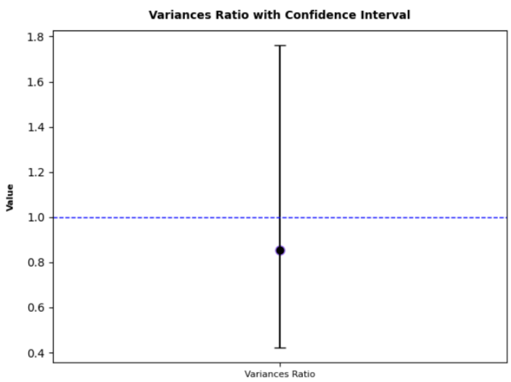
Once the corresponding data is pasted and the checkbox at the bottom is selected, the confidence level to work with must be manually entered. If no confidence level is provided, the default value of 95% will be used, and in that case, the results will include a warning about this.



Confidence Level not defined, defaulting to 95%.

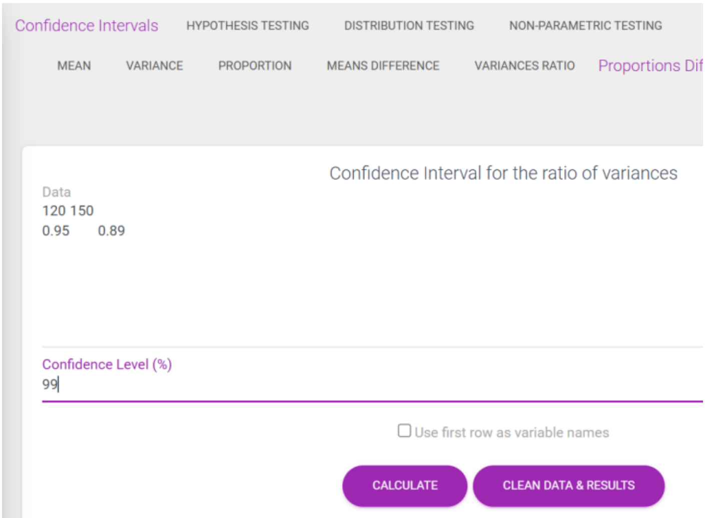
Results

Variable	number of elements	Sample Variances	Confidence Level	Variance Ratio	Confidence Interval
First, Second	30, 35	3.5, 4.1	95.0	0.8537	(0.4225, 1.7605)



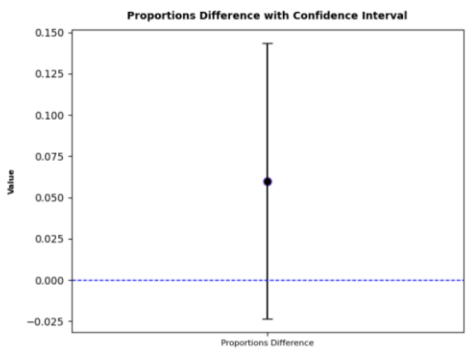
In the confidence interval graph, you can visualize the ratio of variances, the interval itself, and additionally, a dashed blue line at the value of 1. If the interval contains this value, it indicates that the variances are not significantly different under these conditions.

For the case of the confidence interval for the difference of proportions, the process is similar, but instead of entering variances, the sample proportions (values between 0 and 1) are entered. In this example, there are two samples of 120 and 150 elements each, with proportions of 0.95 and 0.89 respectively, and a confidence level of 99%.



Results

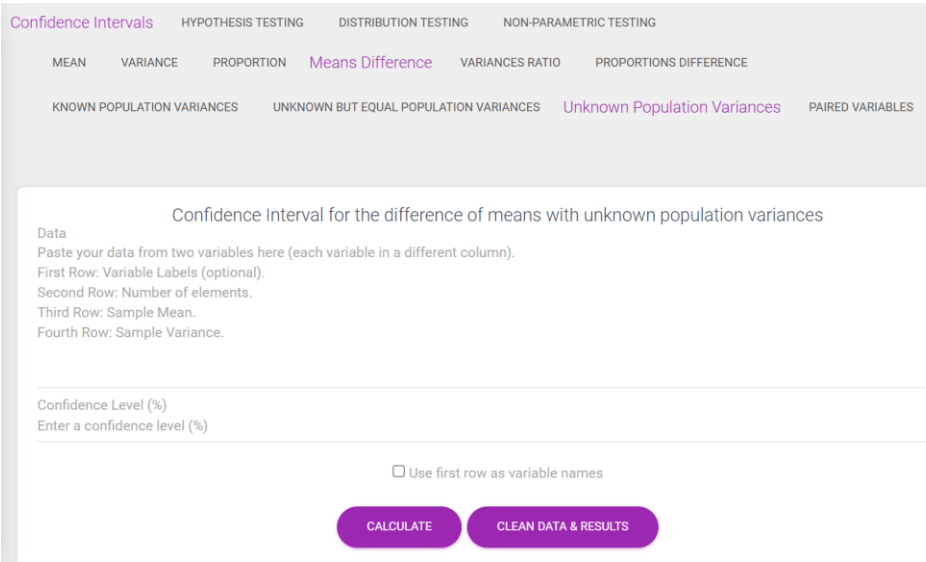
Variable	number of elements	Sample Variances	Confidence Level	Variance Ratio	Confidence Interval
var. 1, var. 2	120, 150	0.95, 0.89	99.0	0.0600	(-0.0234, 0.1434)



In this case, the results include a dashed blue line at the value of zero, as it pertains to a difference. If the interval contains zero, the samples will not be significantly different at the established confidence level.

To conclude the description of confidence intervals, the case of the confidence interval for the difference of means remains to be mentioned. This can be broken down into several subcases: known population variances, unknown but equal population variances, unknown and unequal population variances, and paired sample mean differences.

The first three subcases mentioned are very similar. Let us describe one of them: the subcase of unknown and unequal population variances.

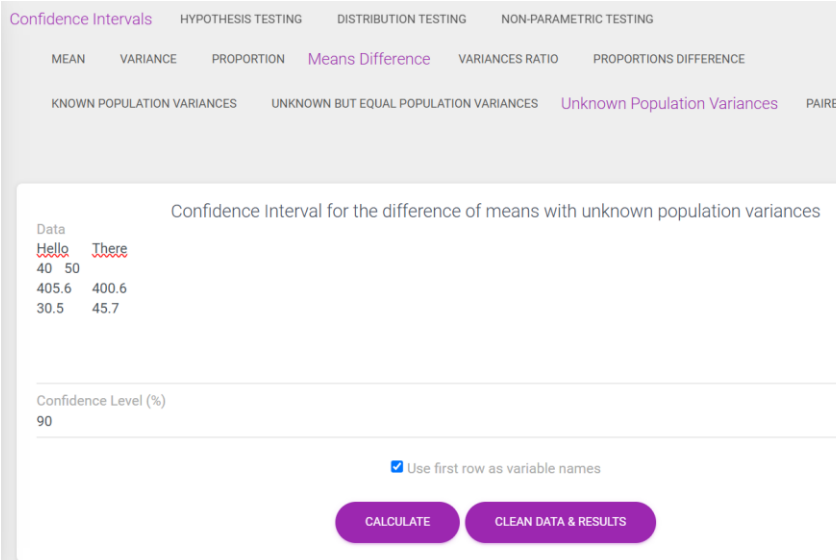


It can be observed that two columns need to be entered first, containing the two variables to be analyzed. Optionally, the names of the variables can be placed in the first row; if so, the corresponding checkbox must be selected, as we did in previous cases. Next, the second row requires the number of elements (n), followed by the mean value and the sample variance. Once this information is entered, the confidence level must be manually input. If none is provided, a default value of 95% will be applied.

For example, for the data shown in the following spreadsheet:

Hello	There
40	50
405.6	400.6
30.5	45.7

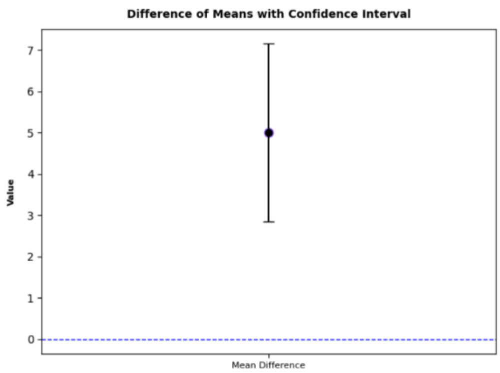
The entered data will be displayed as shown once pasted into the data panel, and the confidence level, for example 90%, is entered:



And the results obtained will be as follows:

Results

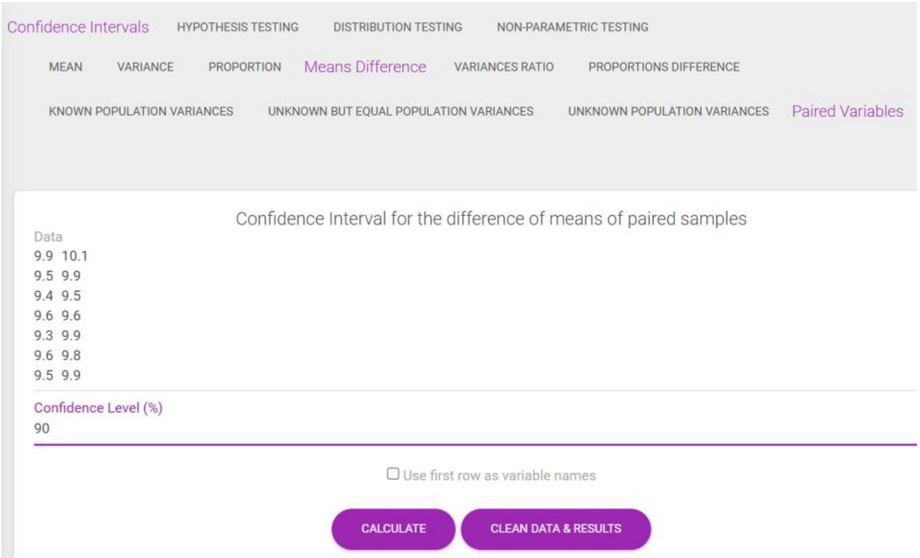
Variables	Number of elements	Population Variance	Means	Confidence Level	Mean Difference	Confidence Interval
Hello, There	40, 50	30.5, 45.7	405.6, 400.6	90.0	5.0000	(2.8476, 7.1524)



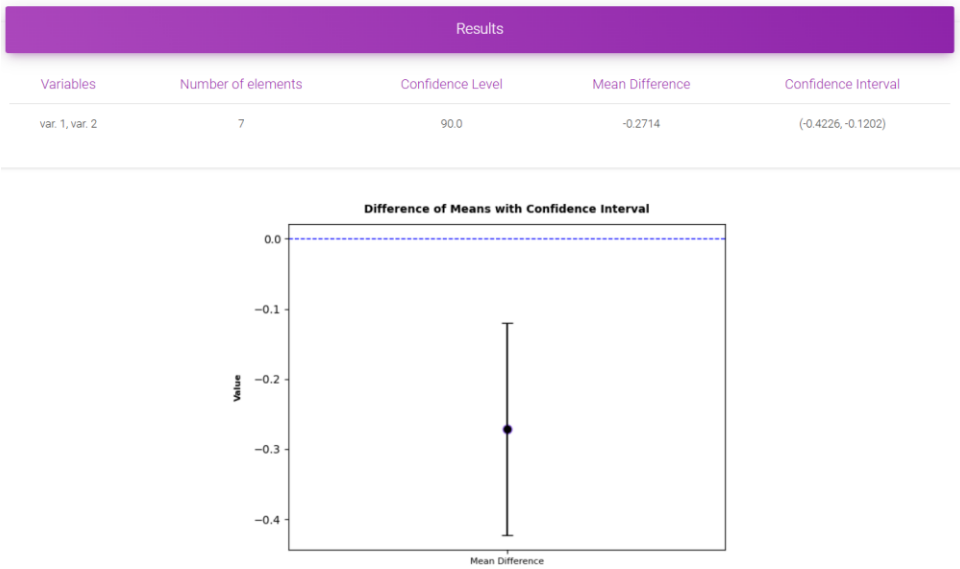
As can be seen, the generated confidence interval does not contain the value of zero, which indicates that it cannot be asserted that the samples have similar mean values at that confidence level.

The functionality of the subcases for confidence intervals with known population variances and unknown but equal variances is very similar, but in the first case, the required values for the population variances must be entered, while in the second case, the sample variances should be entered.

To conclude the description of all the cases and subcases for generating confidence intervals, the final subcase to describe is the paired sample mean difference. In this case, the data for both samples must be entered as bulk data, not parameters, because in this case, the calculations are based on the differences themselves. Consider the following example for bulk data values and a confidence level of 90%, without variable names:

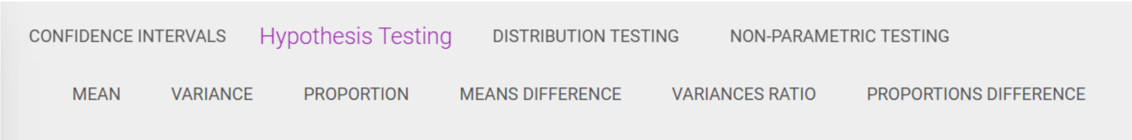


The results obtained will be as follows:

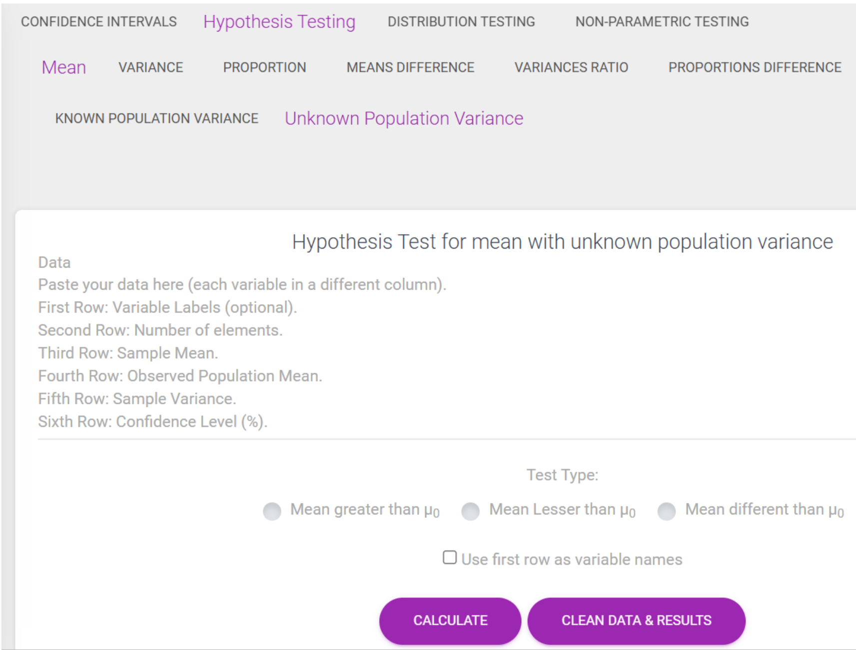


5.2 Hypothesis Testing.

In this section, the cases related to hypothesis tests under the 'Inference' tab will be described. It can be observed that these cases align with those discussed in section 5.1 (confidence intervals). Therefore, hypothesis tests can be conducted for the mean, variance, proportion, difference of means, ratio of variances, and difference of proportions.



For each displayed subcase, three types of tests can be performed: right-tailed, left-tailed, or two-tailed, which should be selected along with data input using radio buttons placed before the 'Calculate' button. For example, for the case 'Mean' and within the subcase 'Unknown Population Variance', the following will be observed:

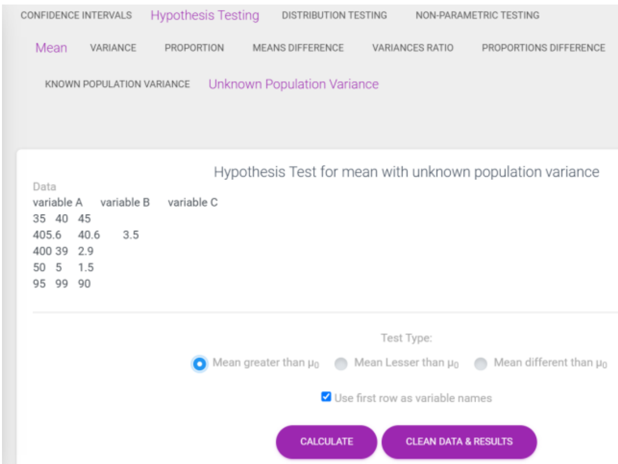


The data required to perform the test(s) are as follows: Initially, as many variables as needed can be entered across different columns. In the first row, optionally, the names of the variables can be provided. In the second row, the sample size (n) is entered, in the third row, the sample mean, in the fourth row, the observed population mean, in the fifth row, the sample variance (in this case), and finally, in the last row, the confidence level expressed as a percentage (%).

Let's consider the following example for 3 variables, where we want to test if the mean has increased in all three cases with confidence levels of 95%, 99%, and 90%. The data to be entered would be as follows:

variable A	variable B	variable C
35	40	45
405.6	40.6	3.5
400	39	2.9
50	5	1.5
95	99	90

And since the tests will be conducted to determine if the mean has increased, the right-tailed unilateral test (Mean greater than μ_0) should be selected.



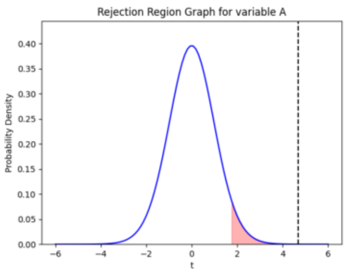
The results include a table, where each row describes the respective variable entered:

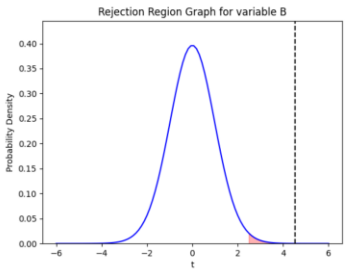
Results

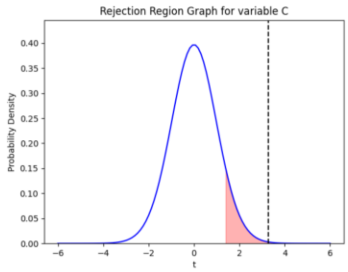
One-Tailed Right Test - Null Hypothesis (H_0): $\mu = \mu_0$ - Alternative Hypothesis (H_1): $\mu > \mu_0$

Variables	Number of elements	Sample Variance	Sample Mean	Observed Population Mean (μ_0)	Confidence Level	t value	Critical t value	P Value	Reject Null (H_0)
variable A	35	50.0	405.6	400.0	95.0	4.6853	1.6909	0.0	Yes
variable B	40	5.0	40.6	39.0	99.0	4.5255	2.4258	0.0	Yes
variable C	45	1.5	3.5	2.9	90.0	3.2863	1.3011	0.001	Yes

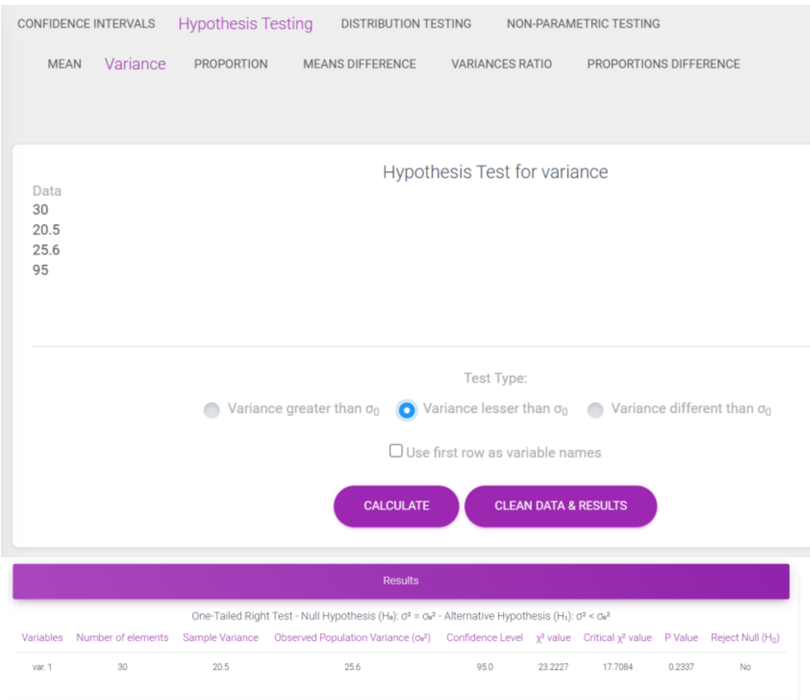
In addition, three graphs are presented (one for each variable), showing the test statistic value and the rejection region:

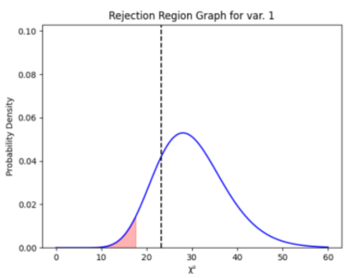




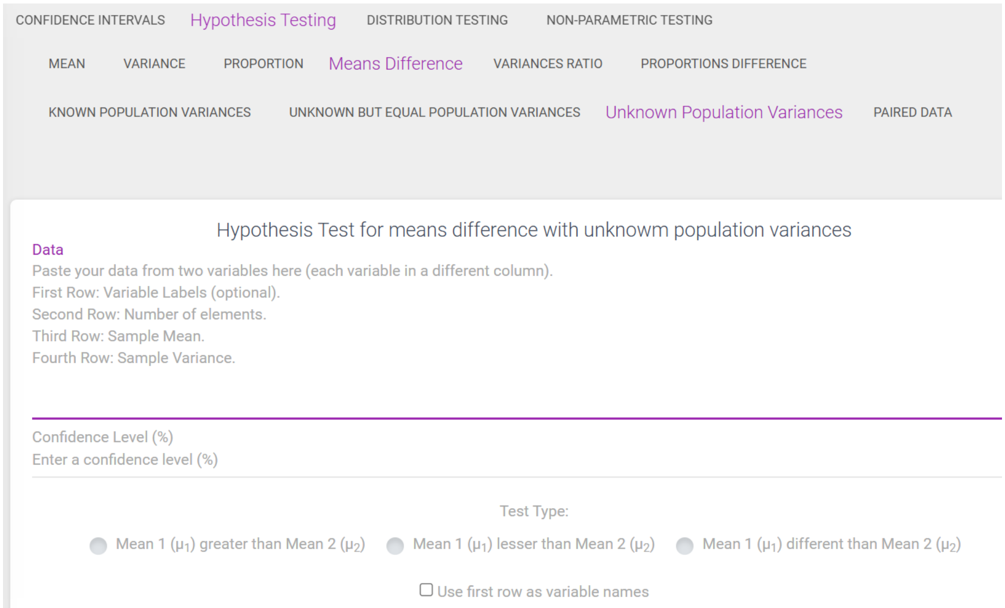


Another example of hypothesis testing can be performed, this time for variance. In this case, the test will be conducted on a single variable, without the need to enter a name, to determine whether the sample variance (20.5) has decreased compared to an observed population variance (25.6), with a confidence level of 95%. The type of test will then be "Variance lesser than σ_0'.





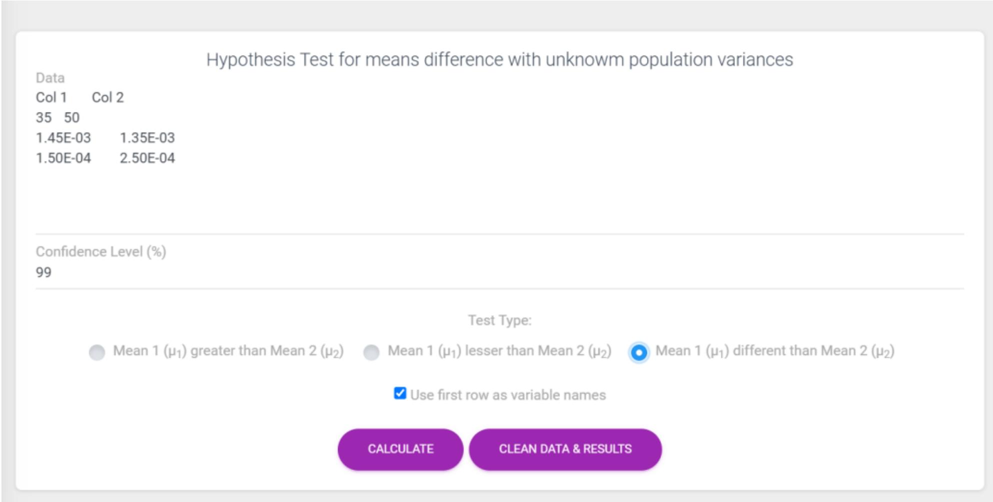
After providing two examples of one-tailed tests, a two-tailed or bilateral test will now be conducted. In this case, the level of significance is divided between both tails. Suppose we want to perform a test for the difference of means; in this case, only two variables need to be entered, one in each column:



The data to be entered should include the following: in the first row, the variable names (optionally). In the second row, the number of elements in each sample. In the third row, the sample mean value. Finally, in the last row, the sample variance, as we have selected a test for the difference of means with unknown population variances. For instance, the data in a spreadsheet would look like this:

Col 1	Col 2
35	50
1.45E-03	1.35E-03
1.50E-04	2.50E-04

And once pasted into the data input area, the data would appear as follows, with the bilateral test (Mean 1 different than Mean 2) selected, the confidence level percentage entered, and the checkbox enabled:



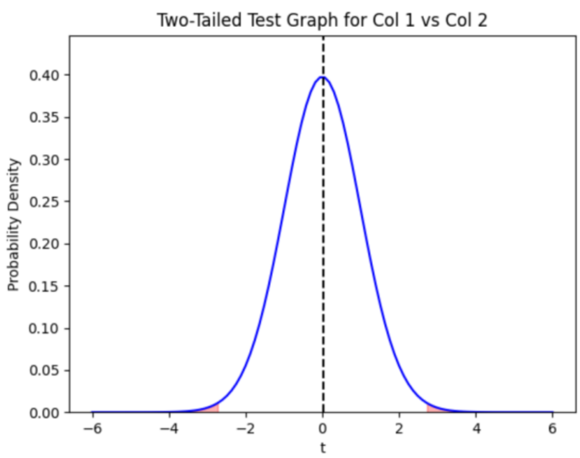
The result includes the following table, which provides an explanation of the test, the formulated hypotheses, and the obtained results:

Results

Two-Tailed Test - Null Hypothesis (H_0): $\mu_1 = \mu_2$ - Alternative Hypothesis (H_1): $\mu_1 \neq \mu_2$

Variables	Number of elements	Sample Variance	Sample Mean	Confidence Level	t value	Critical t value	Degrees of freedom	P Value	Reject Null (H_0)
Col 1, Col 2	35, 50	0.00015, 0.00025	0.00145, 0.00135	99.0	0.0328	-2.6371, 2.6371	82.08571428571429	0.9739	No

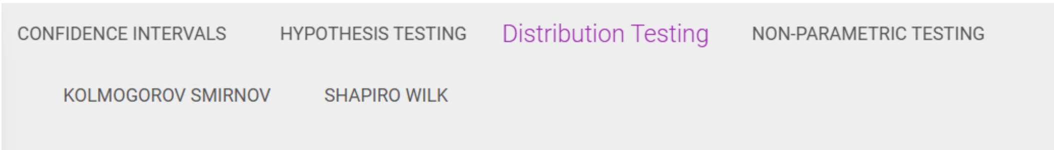
And finally, the graph is generated, showing the rejection regions and the value of the test statistic:



5.3 Distribution Testing

In the Distribution Tests section, there are two available options: the Kolmogorov-Smirnov Test and the Shapiro-Wilk Test.

The Kolmogorov-Smirnov Test can be used to determine whether the data from one or more samples aligns well with either a normal distribution or a uniform distribution, depending on the user's selection. The Shapiro-Wilk Test, however, is specifically designed to evaluate goodness-of-fit for a normal distribution.



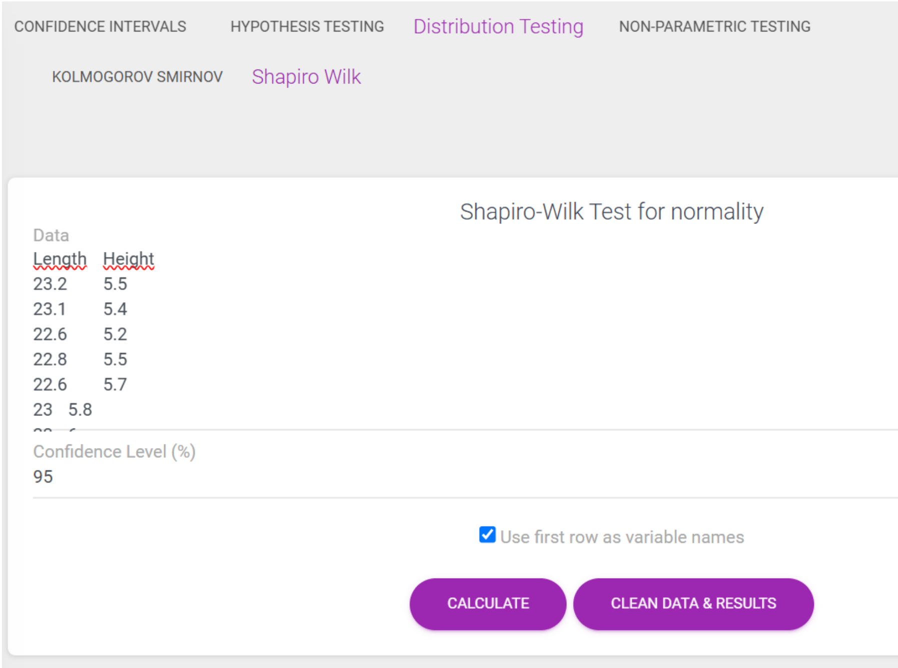
After selecting the desired test type, a panel will appear for data entry and result visualization. Here, you can input one or more variables (with or without names) as raw data. Additionally, you should specify the confidence level—if no value is provided, it defaults to 95%.

For the Kolmogorov-Smirnov Test, you must also select the target distribution (normal or uniform). Finally, click the 'Calculate' button to display the results.

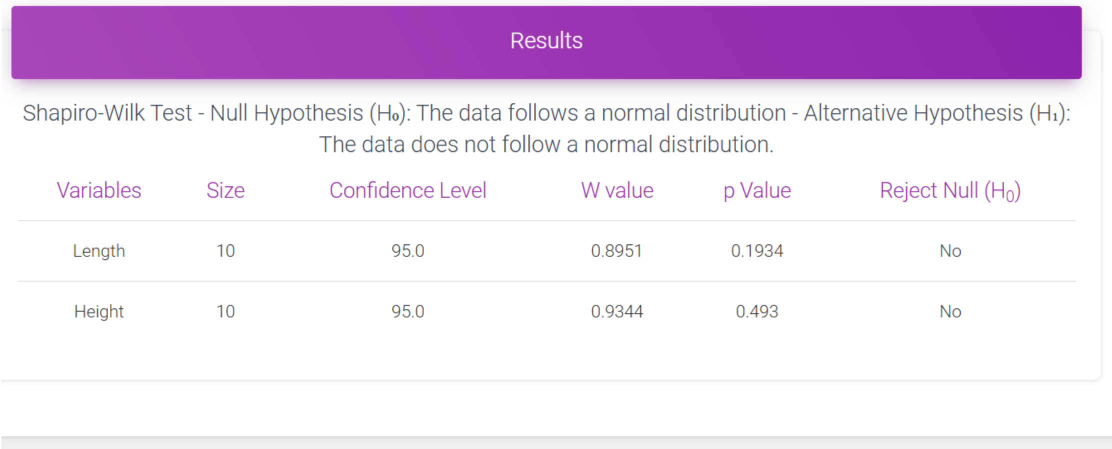
Here is an example dataset to copy and paste, featuring two variables for the Shapiro-Wilk test with a confidence level of 95%.

Length	Height
23.2	5.5
23.1	5.4
22.6	5.2
22.8	5.5
22.6	5.7
23	5.8
23	6
23.3	5.2
23.2	5.5
23.1	5.3

Once the data is copied and the options are selected, the panel will look as follows:



And after pressing the 'Calculate' button, the results are displayed as follows:



Results

Shapiro-Wilk Test - Null Hypothesis (H_0): The data follows a normal distribution - Alternative Hypothesis (H_1): The data does not follow a normal distribution.

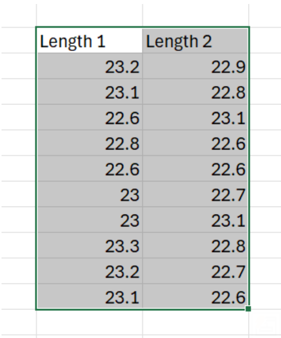
Variables	Size	Confidence Level	W value	p Value	Reject Null (H_0)
Length	10	95.0	0.8951	0.1934	No
Height	10	95.0	0.9344	0.493	No

5.4 Non-Parametric Testing

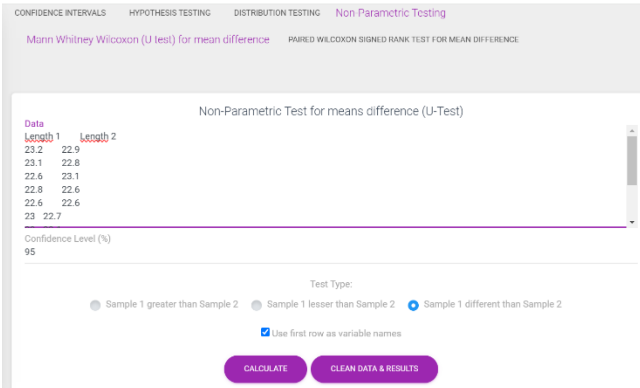
In cases where the assumptions for parametric tests are not valid for comparing means, the following non-parametric options are provided: the Mann-Whitney-Wilcoxon test (U test) for unpaired samples and the Wilcoxon Signed Rank Test for paired samples.

The process is very similar to what has been shown earlier in this manual. The values of two variables, whose means are to be compared, must be entered in bulk data format in two columns, with the option to include variable names in the first row.

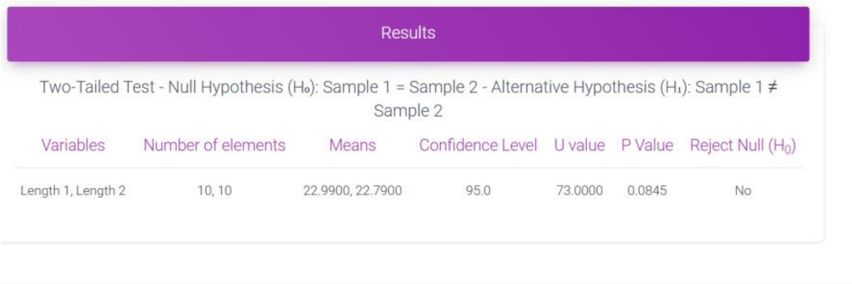
As an example, the following data will be entered:



Length 1	Length 2
23.2	22.9
23.1	22.8
22.6	23.1
22.8	22.6
22.6	22.6
23	22.7
23	23.1
23.3	22.8
23.2	22.7
23.1	22.6



After pressing the 'Calculate' button, the test results will be obtained.



Results

Two-Tailed Test - Null Hypothesis (H_0): Sample 1 = Sample 2 - Alternative Hypothesis (H_1): Sample 1 ≠ Sample 2

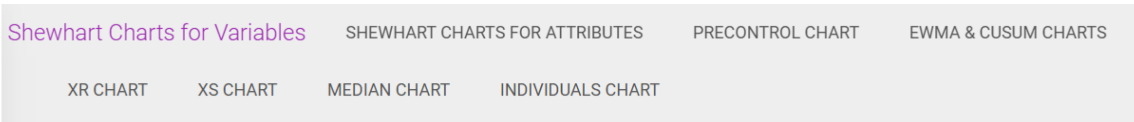
Variables	Number of elements	Means	Confidence Level	U value	P Value	Reject Null (H_0)
Length 1, Length 2	10, 10	22.9900, 22.7900	95.0	73.0000	0.0845	No

The data entry and operation are very similar in the case of the test for paired samples.

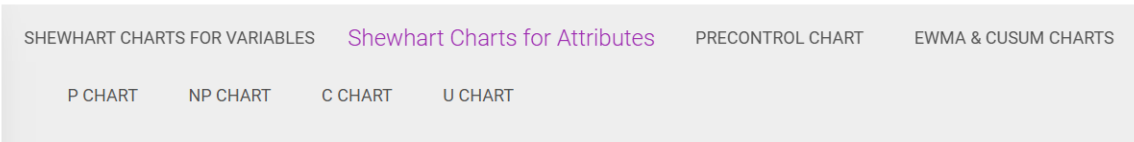
6. Statistical Process Control Charts.

In the field of statistical process control, different types of control charts can be constructed. Once the 'Control Charts' tab is selected, the various types of charts available in this application will be displayed. On one hand, there are Shewhart charts, both for variables and attributes. Additionally, pre-control charts can be selected, followed by EWMA and CUSUM charts.

The Shewhart charts for variables include XR, XS, Median, and Individual charts.



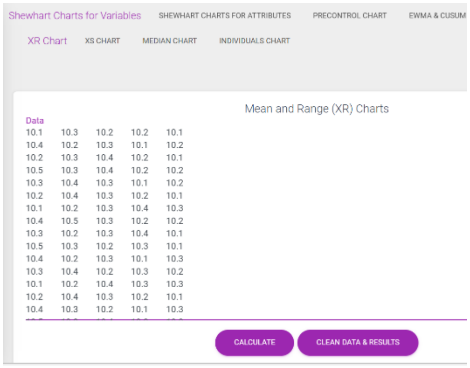
The options for Shewhart control charts for attributes include the following: p chart, np chart, c chart, and u chart.



The construction of Shewhart control charts utilizes parameter values derived from the Control Chart Constants Tables published by the Institute of Quality and Reliability (https://web.mit.edu/2.810/www/files/readings/ControlChartConstantsAndFormulae.pdf).

In the case of Shewhart charts for variables, XR, XS, and median charts, the procedure for data entry is very similar. You will need to copy and paste the sample data into the different columns and the subgroups into the different rows, using only the values without the names, as shown below:

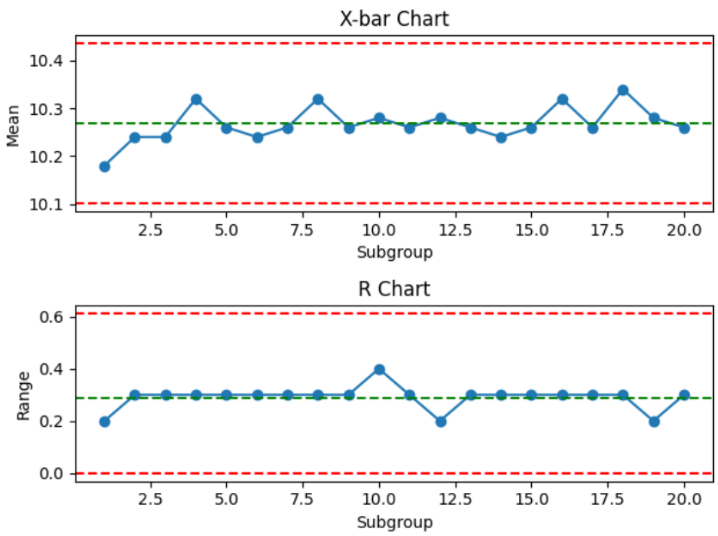
Subgroup	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	10.1	10.3	10.2	10.2	10.1
2	10.4	10.2	10.3	10.1	10.2
3	10.2	10.3	10.4	10.2	10.1
4	10.5	10.3	10.4	10.2	10.2
5	10.3	10.4	10.3	10.1	10.2
6	10.2	10.4	10.3	10.2	10.1
7	10.1	10.2	10.3	10.4	10.3
8	10.4	10.5	10.3	10.2	10.2
9	10.3	10.2	10.3	10.4	10.1
10	10.5	10.3	10.2	10.3	10.1
11	10.4	10.2	10.3	10.1	10.3
12	10.3	10.4	10.2	10.3	10.2
13	10.1	10.2	10.4	10.3	10.3
14	10.2	10.4	10.3	10.2	10.1
15	10.4	10.3	10.2	10.1	10.3
16	10.5	10.2	10.4	10.3	10.2
17	10.3	10.4	10.2	10.3	10.1
18	10.4	10.3	10.5	10.2	10.3
19	10.2	10.3	10.2	10.4	10.3
20	10.3	10.4	10.3	10.2	10.1



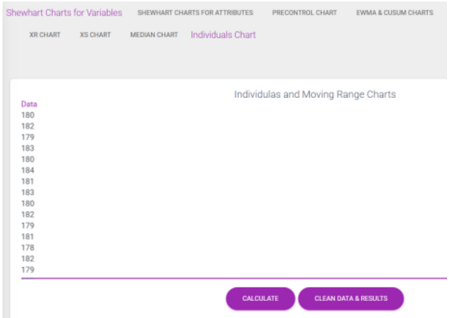
The results include a table with data and the requested control charts:

Results

Metric	Value
Overall Mean	10.27
Mean Range	0.29
UCL (X-bar)	10.44
LCL (X-bar)	10.1
UCL (R Chart)	0.6131
LCL (R Chart)	0
Subgroup Means	10.18, 10.24, 10.24, 10.32, 10.26, 10.24, 10.26, 10.32, 10.26, 10.28, 10.26, 10.28, 10.26, 10.24, 10.26, 10.32, 10.26, 10.34, 10.28, 10.26
Subgroup Ranges	0.2, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.4, 0.3, 0.2, 0.3, 0.3, 0.3, 0.3, 0.3, 0.3, 0.2, 0.3

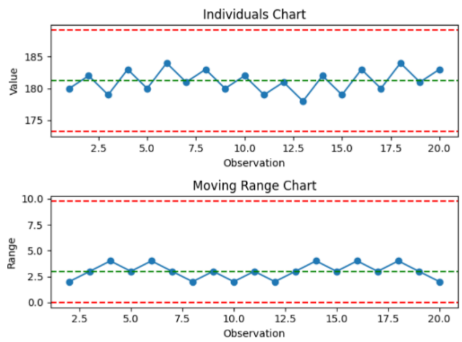


In the case of the individuals chart, only a single column of data needs to be entered. In this case, the number of subgroups for calculating the moving range will be 2.



Similarly, the results include a table with the data and the corresponding control charts.

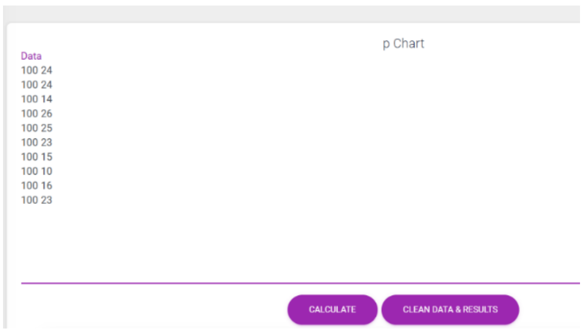
Results	
Metric	Value
Overall Mean	181.2
Mean Moving Range	3.0
UCL (I Chart)	189.2
LCL (I Chart)	173.2
UCL (MR Chart)	9.801
LCL (MR Chart)	0



For Shewhart attribute control charts, these are based on counting the number of defects or defective items in a series of samples. In all cases, only numerical values can be entered. If only the number of defects is required, a single column of data should be provided. However, if both the number of inspected samples and the number of defects are required, the data should be entered in two columns.

Let us assume that a p-chart needs to be constructed using the following data from a spreadsheet. In this case, the values to be copied would be as follows:

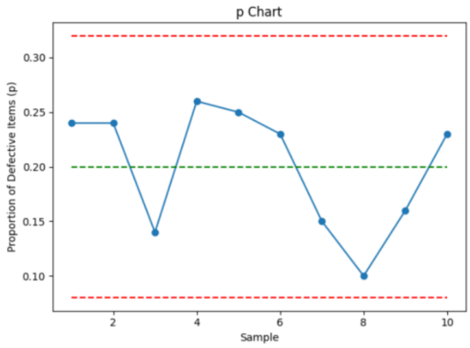
Quantity	Defects
100	24
100	24
100	14
100	26
100	25
100	23
100	15
100	10
100	16
100	23



And after pressing the 'Calculate' button, the results will be displayed.

Results

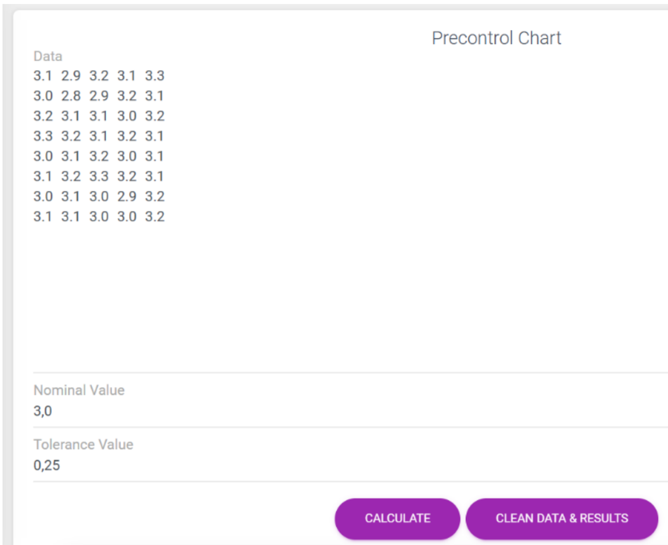
Metric	Value
Mean p	0.2



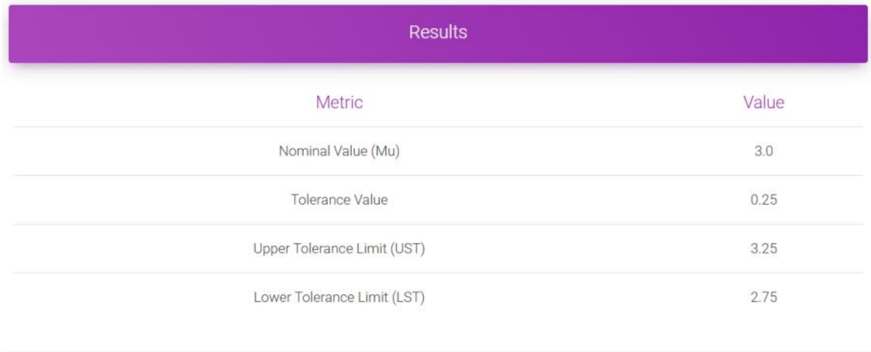
In the remaining cases of attribute control charts, the procedure is similar, with the exception that for the c-chart, only a single column with the number of defects is required.

Finally, the case of the remaining three charts needs to be analyzed, starting with the pre-control chart. In this chart, in addition to entering numerical data (with subgroups in rows and samples in columns), the nominal value (expected target value) and a tolerance value are also required.

Subgroup	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
1	3.1	2.9	3.2	3.1	3.3
2	3.0	2.8	2.9	3.2	3.1
3	3.2	3.1	3.1	3.0	3.2
4	3.3	3.2	3.1	3.2	3.1
5	3.0	3.1	3.2	3.0	3.1
6	3.1	3.2	3.3	3.2	3.1
7	3.0	3.1	3.0	2.9	3.2
8	3.1	3.1	3.0	3.0	3.2

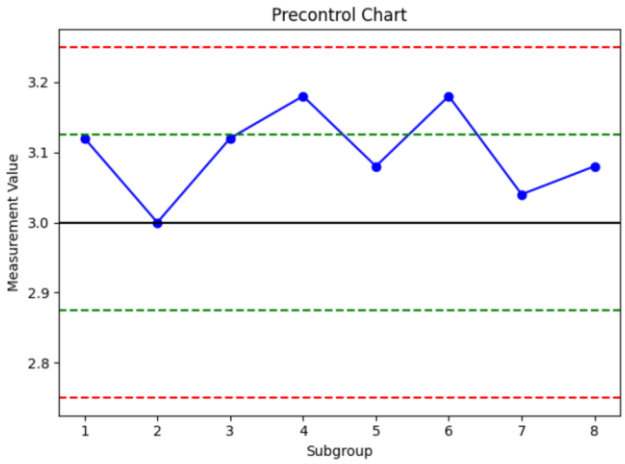


As always, the results are displayed after pressing the 'Calculate' button.

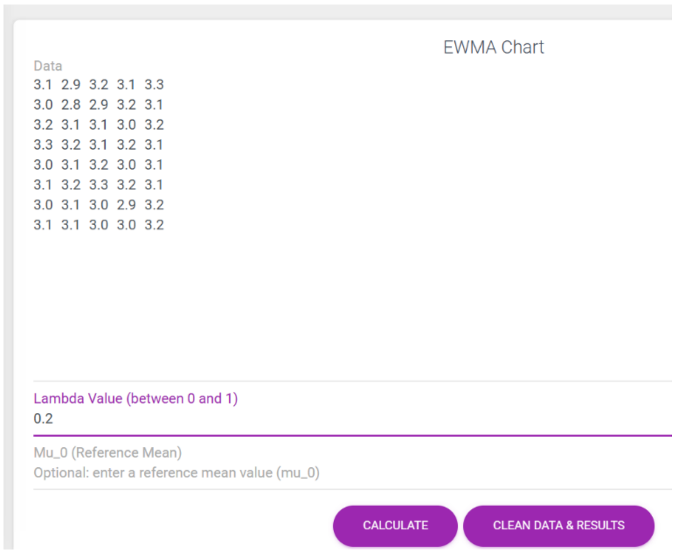


Results

Metric	Value
Nominal Value (Mu)	3.0
Tolerance Value	0.25
Upper Tolerance Limit (UST)	3.25
Lower Tolerance Limit (LST)	2.75



For the EWMA (Exponentially Weighted Moving Average) chart, the input parameters must include the lambda value (commonly 0.2 or 0.3) and a reference value for the mean. If no value is provided for the mean reference, the default is set to the global average of all the entered data.

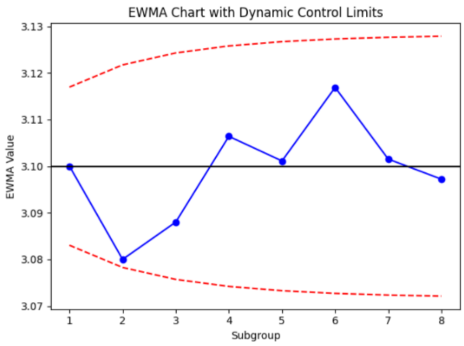


In this case, the results include an alert message indicating that the value for the parameter was not entered, and the global average of all the data was used as the default mean value.

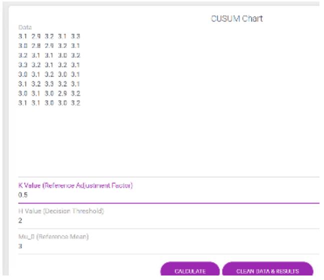
No target mean (mu_0) was provided. Using global mean of the data as default.

Results

Metric	Value
Overall Mean	3.1
Lambda Value	0.2
Reference Mean (Mu_0)	3.1

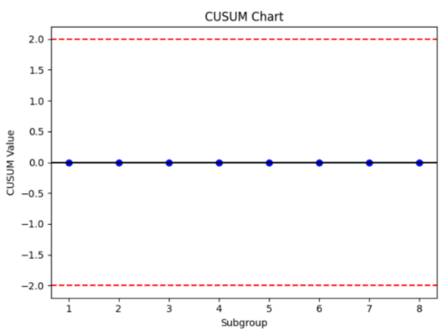


Finally, the case of the CUSUM (Cumulative Sum Control Chart) is considered, where, in addition to the data, the values for the parameters K and H must be entered, along with an optional estimated mean value, similar to the previous case. If the latter parameter is not entered, the global average of the data will be used as the default. Both the data entry process and the results are shown below.



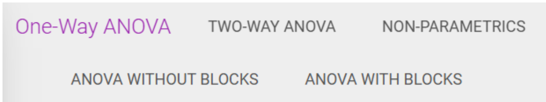
Results

Metric	Value
Reference Mean (Mu_0)	3.0
K Value	0.5
H Value	2.0

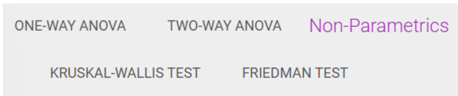


7. Analysis of Variance (ANOVA).

The web application presented offers the possibility of performing 3 different types of ANOVA. By clicking on the ANOVA tab in the sidebar, the available options can be viewed in the top menu. One-Way ANOVA displays two subcases: one without blocks and another where blocks are used in addition to the factor. Additionally, Two-Way ANOVA can also be performed.

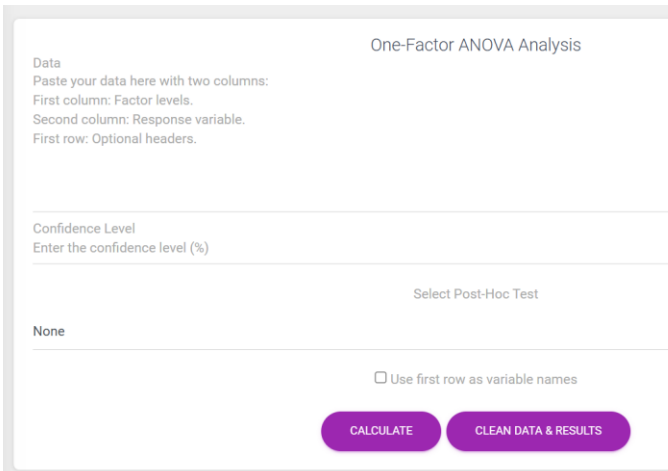


If the assumptions of ANOVA are not met, a third option will be available in the top menu, offering two non-parametric tests: the Kruskal-Wallis test, analogous to the one-way ANOVA without blocks, and the Friedman test, for the case of a one-way ANOVA with blocks.



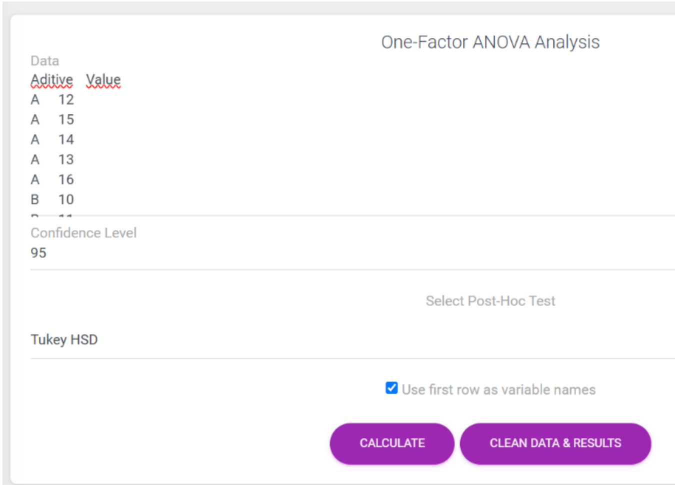
7.1 One-Way ANOVA

After selecting the One-Way ANOVA case from the top menu and the subcase of ANOVA without Blocks, the data entry panel, confidence level input, and post-hoc test selection will appear, as shown in the following figure.



The format in which the data should be copied and pasted must follow the structure shown below, where the first row optionally allows for the inclusion of the factor name and the response variable name, followed by the corresponding values in the columns.

Aditive	Value
A	12
A	15
A	14
A	13
A	16
B	10
B	11
B	9
B	12
B	11
C	14
C	13
C	15
C	16
C	17



In this case, a confidence level of 95% was entered, and if no value is provided, the test will be performed with this default value. Then, in the post-hoc test selector, Tukey-HSD was selected, but other options can also be chosen, such as Bonferroni, Scheffé, Dunnett, Fisher's LSD, Holm-Bonferroni, and Duncan. Finally, the checkbox is selected in this case because the first row contains the names of the variables.

The results obtained after pressing the 'Calculate' button are shown below and include a general table with the ANOVA results, as well as a table with the results of the post-hoc test if it was performed.

ANOVA Results

Null Hypothesis (H_0): Group means are equal - Alternative Hypothesis (H_1): At least one group mean differs.

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-statistic	P-value	Reject Null
Aditive	2	53.2	26.6	12.667	0.0011028	Yes
Residual	12	25.2	2.1			
Total	14	78.4				

Post Hoc Analysis - Tukey HSD

Group 1	Group 2	Mean Difference	P-Value	Lower	Upper	Reject Null
A	B	-3.4	0.0078	-5.845	-0.9549	Yes
A	C	1	0.5372	-1.445	3.445	No
B	C	4.4	0.0012	1.955	6.845	Yes

Additionally, the results of the Shapiro-Wilk and Levene tests on the residuals of the model are presented:

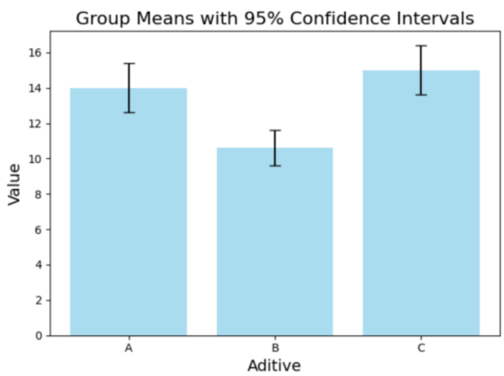
Shapiro-Wilk Test for Normality

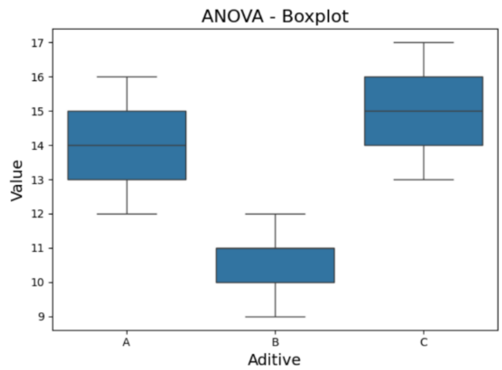
W Value: 0.9461 P Value: 0.465 Normality: Yes

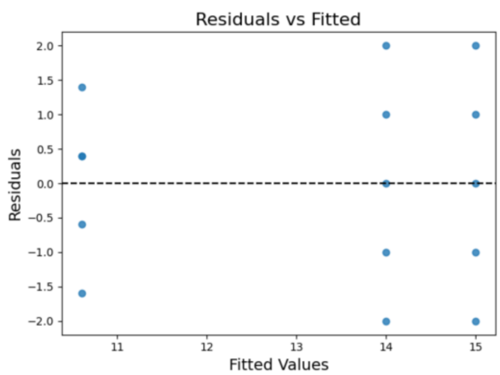
Levene's Test for Homogeneity of Variances

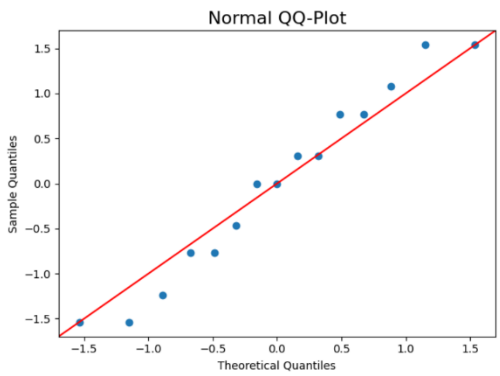
Statistic: 0.381 P Value: 0.6912 Homogeneity: Yes

Finally, a series of graphs are presented: the first shows the means with their 95% confidence interval, the second is a boxplot, and the last two graphs display the residuals vs. fitted values and the QQ-plot.







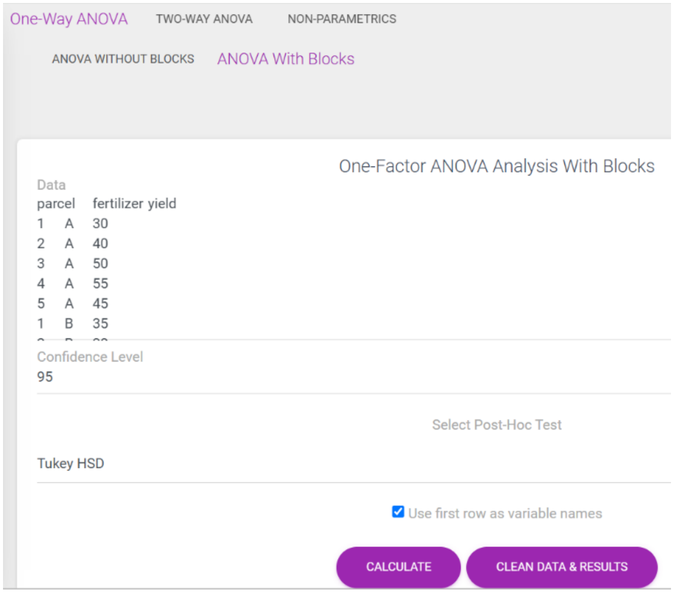


All the results provided form a comprehensive analysis that will allow for an excellent description of the model, which in turn will facilitate decision-making based on these results.

In the case that a one-way ANOVA with blocks is desired, the procedure is very similar to the previous one, with the exception that the first column of the data should contain the variable corresponding to the block. The accepted data type for describing the levels of both the factor and the blocks can be either text or numbers.

Here is an example for a one-way ANOVA with blocks, using a 95% confidence level and a Tukey HSD post-hoc test. First, this is how the data looks in the spreadsheet, and then after copying and pasting it into the panel.

parcel	fertilizer	yield
1	A	30
2	A	40
3	A	50
4	A	55
5	A	45
1	B	35
2	B	38
3	B	45
4	B	42
5	B	43
1	C	25
2	C	33
3	C	40
4	C	48
5	C	39



The results include both tables (ANOVA and post-hoc).

ANOVA Results

Null Hypothesis (H_0): Group means are equal - Alternative Hypothesis (H_1): At least one group mean differs.

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-statistic	P-value	Reject Null
fertilizer	2	122.53	61.267	4.5893	0.047034	Yes
parcel	4	622.4	155.6	11.655	0.0020312	
Residual	8	106.8	13.35			

Post Hoc Analysis - Tukey HSD

Group 1	Group 2	Mean Difference	P-Value	Lower	Upper	Reject Null
A	B	-3.4	0.7739	-16.55	9.753	No
A	C	-7	0.3621	-20.15	6.153	No
B	C	-3.6	0.7508	-16.75	9.553	No

Additionally, the results include the Shapiro-Wilk and Levene tests on the model's residuals.

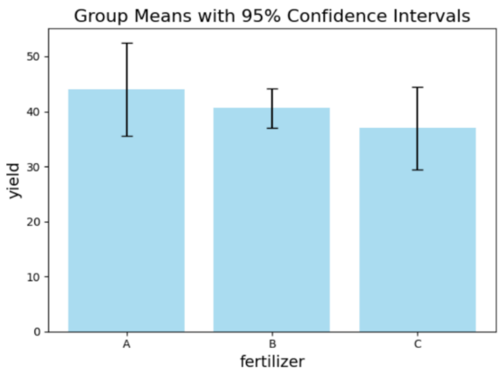
Shapiro-Wilk Test for Normality

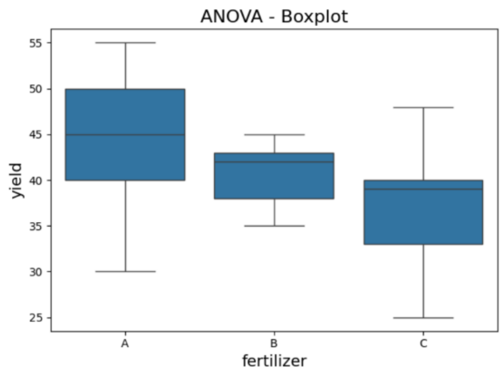
W Value: 0.96 P Value: 0.6928 Normality: Yes

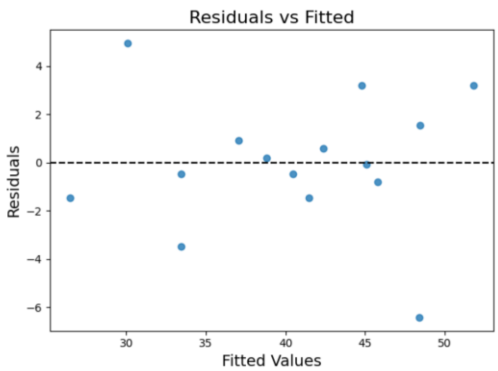
Levene's Test for Homogeneity of Variances

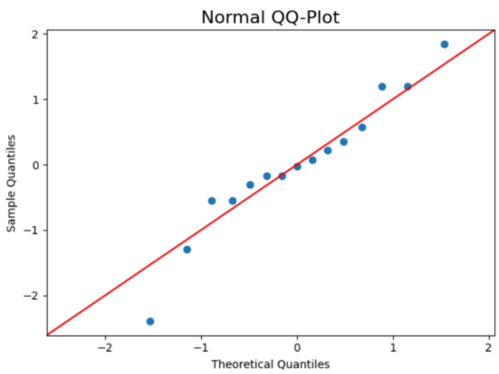
Statistic: 0.8844 P Value: 0.4383 Homogeneity: Yes

And all the graphs that have been previously described.







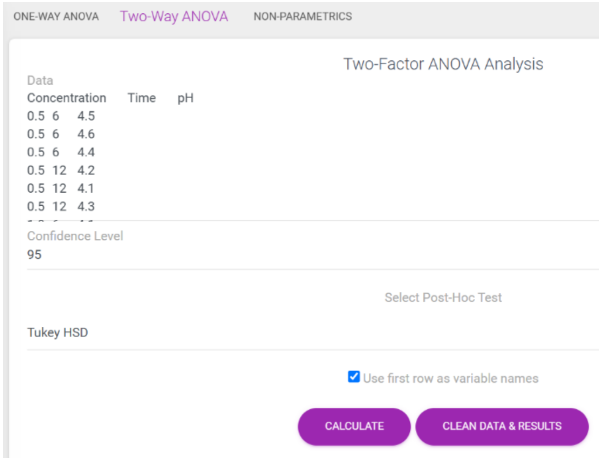


7.2 Two-Way ANOVA

To input data for a Two-Way ANOVA, the data formatting procedure is similar to the previous case, with the following distinctions: the first column should contain the values for the levels of the first factor, the second column should contain the levels of the second factor, and the third column should include the numerical values of the response variable.

This example illustrates a case where the first factor has two levels (0.5 and 1.0), and the second factor also has two levels (6 and 12). After copying the data, the confidence level (95%) is entered, the post-hoc test (Tukey-HSD) is selected, and the checkbox is checked to indicate that the variable names are included in the first row.

Concentration	Time	pH
0.5	6	4.5
0.5	6	4.6
0.5	6	4.4
0.5	12	4.2
0.5	12	4.1
0.5	12	4.3
1.0	6	4.1
1.0	6	4
1.0	6	4.2
1.0	12	3.8
1.0	12	3.9
1.0	12	3.7



The results include, as in the previous cases, a table with the ANOVA results and another table with the post-hoc test results (if selected).

ANOVA Results

Null Hypothesis (H_0): All group means are equal across factors and their interaction - Alternative Hypothesis (H_1): At least one group mean differs, or there is significant interaction between factors.

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F-statistic	P-value	Reject Null
Concentration	1	0.48	0.48	48	0.00012104	Yes
Time	1	0.27	0.27	27	0.00082628	Yes
Concentration:Time	1	4.4743e-30	4.4743e-30	4.4743e-28	1	No
Residual	8	0.08	0.01			

Post Hoc Analysis - Tukey HSD

Comparison Type	Group 1	Group 2	Mean Difference	P-Value	Lower	Upper	Reject Null
Concentration	0.5	1.0	-0.4	0.0041	-0.6407	-0.1593	Yes
Time	12	6	0.3	0.0528	-0.0044	0.6044	No
Interaction	0.5 x 12	0.5 x 6	0.3	0.0259	0.0385	0.5615	Yes
Interaction	0.5 x 12	1.0 x 12	-0.4	0.0052	-0.6615	-0.1385	Yes
Interaction	0.5 x 12	1.0 x 6	-0.1	0.6298	-0.3615	0.1615	No
Interaction	0.5 x 6	1.0 x 12	-0.7	0.0001	-0.9615	-0.4385	Yes
Interaction	0.5 x 6	1.0 x 6	-0.4	0.0052	-0.6615	-0.1385	Yes
Interaction	1.0 x 12	1.0 x 6	0.3	0.0259	0.0385	0.5615	Yes

Additionally, the results of the Shapiro-Wilk and Levene tests are displayed. If either of these tests fails, a warning message is displayed below the respective test result.

Shapiro-Wilk Test for Normality

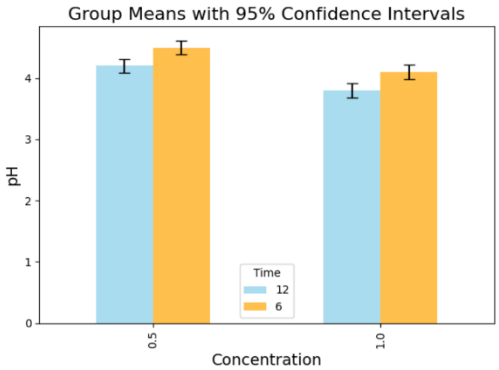
W Value: 0.8108 P Value: 0.01246 Normality: No

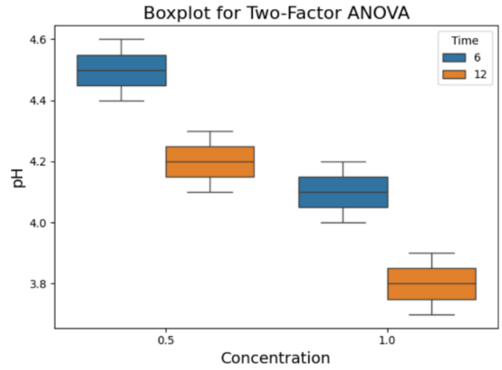
Shapiro-Wilk Test for normality failed. Data may not be normally distributed.

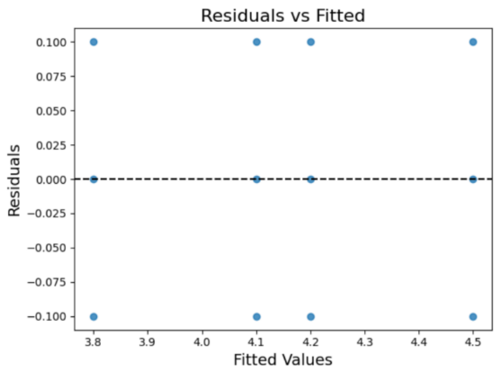
Levene's Test for Homogeneity of Variances

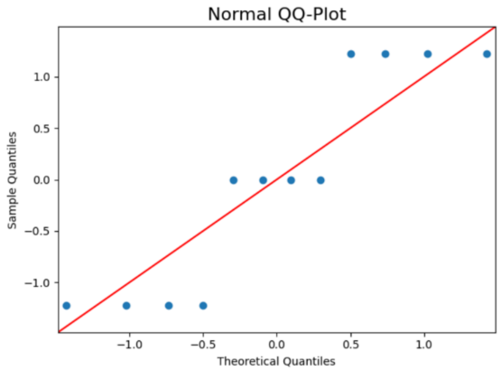
Statistic: 1.739e-29 P Value: 1 Homogeneity: Yes

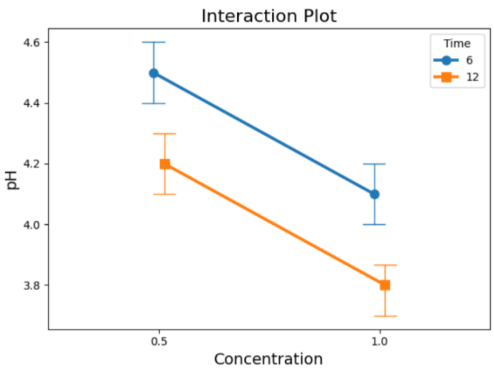
Then, all the graphs are displayed, similar to those described previously, with the addition of the interaction plot between the factors.











7.3 Non-Parametric Tests

If the assumptions of normality or homoscedasticity are not met in one-way ANOVA tests, whether without or with blocks, a non-parametric test can be used as an alternative.

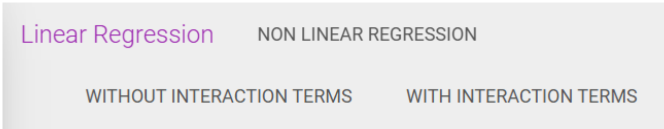
If a non-parametric analysis is desired for data from an experiment without blocks, the Kruskal-Wallis test should be selected under the ‘Non-Parametrics’ option. In this case, the data entry is similar to the One-Way ANOVA test, with the only available post-hoc test option being the Dunn test.

If the analysis requires the use of blocks, then the Friedman test should be selected. The data entry format is similar to the One-Way ANOVA with blocks. In this case, the only available option for the post-hoc test is the Nemenyi test.

The results of both tests are more limited compared to the parametric tests and do not include all the graphs as in the previous cases. Nevertheless, the primary functionality is to verify the differences between the sample treatments with a given level of confidence.

8. Regression.

In this application, the tab labeled 'Regression' in the sidebar enables the available options for regression analysis. The top menu displays two options: 'Linear' and 'Non-Linear'. For linear regression, it is further divided into two subcases: one that does not consider interaction terms between variables, and one that includes these terms.



8.1 Linear Regression

For the linear regression subcases, you can enter a Y variable in the first column and one or more X variables in the subsequent columns.

In this case, for example, if only one X variable is included in the subcase without considering the interaction term, we have the simple linear regression case, using the following model:

$$y = \beta_0 + \beta_1 x_1$$

Or, in the case of entering more X variable values in the copied data, the terms will increase as follows.

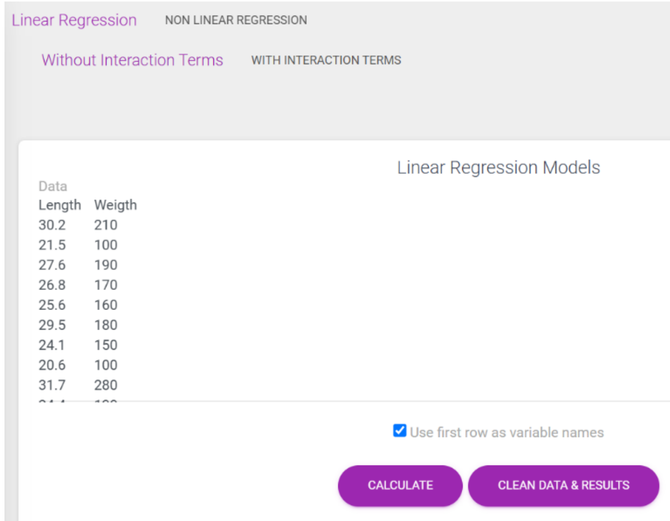
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Where the subscript 'k' represents the number of X variables entered.

For example, let's consider the case of simple linear regression, where we will enter the data in only two columns: the first column will contain the values of the Y variable, and the second column will contain the values of the X variable. Optionally, the variable names can be entered in the first row.

The data copied from a spreadsheet, for example, will look like this:

Length	Weigth
30.2	210
21.5	100
27.6	190
26.8	170
25.6	160
29.5	180
24.1	150
20.6	100
31.7	280
24.4	130
25.3	170
22.5	120
16.9	60
18.1	60
27.2	170
19.5	80
20.0	80



The results include a complete table with a full description of the model and the regression parameters.

Regression Results

Length = β0 + (β1.Weigth)

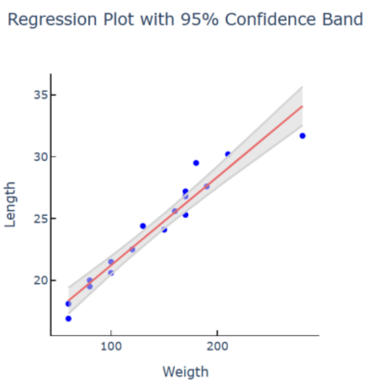
Model Summary:

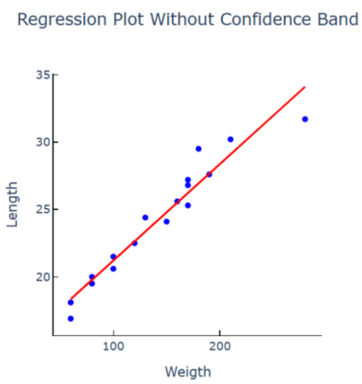
R^2: 0.934 Adjusted R^2: 0.9298 F-Statistic: 212.9209 F p-value: 0.0000

Coefficients:

Parameter	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
β0	14.0560	0.7499	18.7436	0.0000	12.4576	15.6544
β1	0.0716	0.0049	14.5918	0.0000	0.0611	0.0821

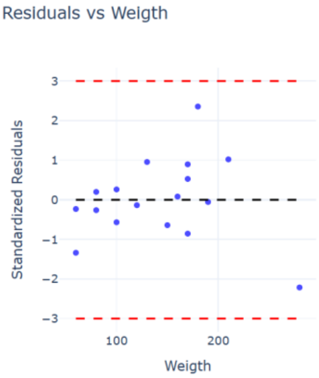
Next, a couple of scatter plots are displayed: on the left with the 95% confidence band and on the right without the confidence band, showing the data and the fitted model.





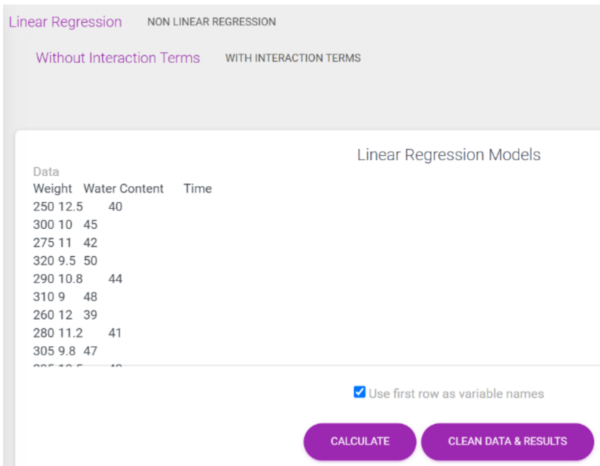
It is worth mentioning that these graphs provide the option to display the values when hovering the mouse over the data points.

Finally, a graph of the standardized residuals is displayed, also in an interactive format.



Next, we can observe, through the following example, what happens if two variables X are entered instead of just one.

Weight	Water Content	Time
250	12.5	40
300	10	45
275	11	42
320	9.5	50
290	10.8	44
310	9	48
260	12	39
280	11.2	41
305	9.8	47
295	10.5	43
315	9.2	49
270	11.5	40
285	10.2	42
325	8.8	51
265	12.3	38



Automatically, the application recognizes that another variable has been entered in this subcase without interaction, and the results will be displayed in a similar manner. First, a table will show the model and the regression parameters.

Regression Results

Weight = β0 + (β1.Water Content) + (β2.Time)

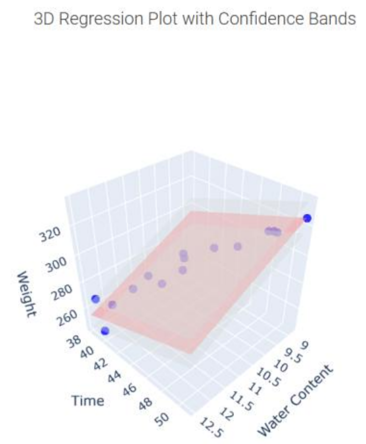
Model Summary:

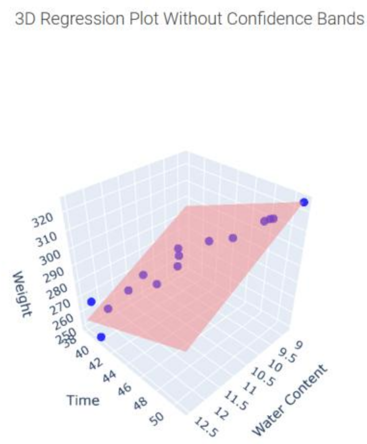
R^2: 0.954 Adjusted R^2: 0.9460 F-Statistic: 123.6616 F p-value: 0.0000

Coefficients:

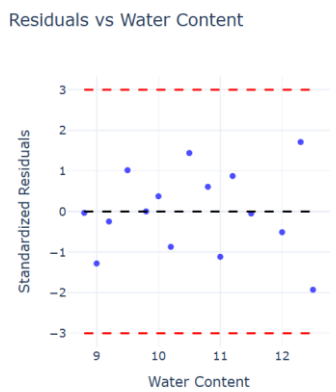
Parameter	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
β0	305.9141	77.5379	3.9453	0.0019	136.9735	474.8547
β1	-11.0423	3.4192	-3.2295	0.0072	-18.4920	-3.5926
β2	2.2827	0.9715	2.3495	0.0367	0.1659	4.3995

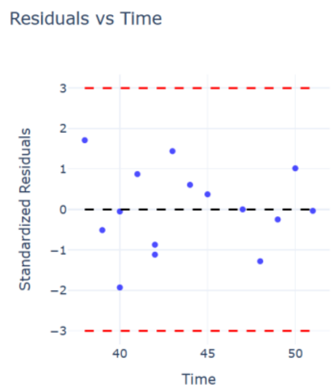
The regression graphs will be displayed, showing the data and a fitted plane. On the left side, the 95% confidence surfaces will be included, while on the right side, these surfaces will not be shown. These graphs are interactive and, being three-dimensional, they can be rotated by dragging with the mouse, allowing users to move them and find the best visualization angle.





Finally, interactive plots of the standardized residuals are also provided, in this case, against each of the variables.





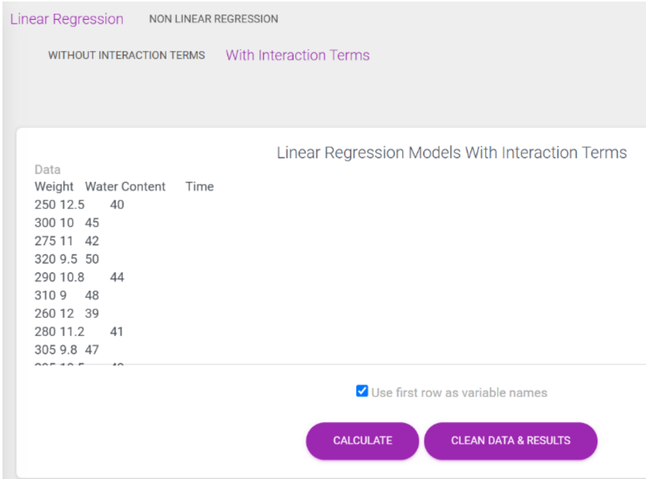
If the goal is to include interaction terms between the entered variables in the models, the option on the right within the 'Linear Regression' menu on the top bar should be selected.

For example, in the case of considering only two X variables, the model would be as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

This is how the data that needs to be copied from the spreadsheet to the panel would look:

Weight	Water Content	Time
250	12.5	40
300	10	45
275	11	42
320	9.5	50
290	10.8	44
310	9	48
260	12	39
280	11.2	41
305	9.8	47
295	10.5	43
315	9.2	49
270	11.5	40
285	10.2	42
325	8.8	51
265	12.3	38



Once you press the 'Calculate' button, the results will be displayed, starting with the table showing the model and the corresponding parameters.

Regression Results

Weight = β0 + (β1 * Water Content) + (β2 * Time) + (β3 * Water Content*Time)

Model Summary:

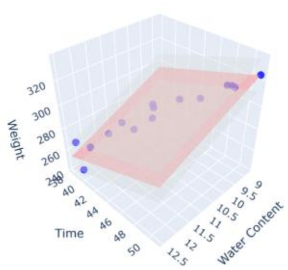
R^2: 0.954 Adjusted R^2: 0.9413 F-Statistic: 75.8678 F p-value: 0.0000

Coefficients:

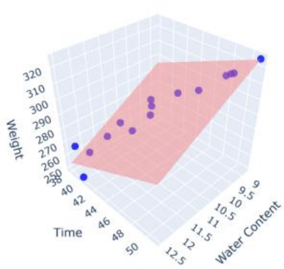
Parameter	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
β0	336.5070	171.0036	1.9678	0.0748	-39.8695	712.8834
β1	-14.4339	17.0822	-0.8450	0.4161	-52.0316	23.1637
β2	1.4953	4.0084	0.3730	0.7162	-7.3271	10.3178
β3	0.0866	0.4268	0.2030	0.8428	-0.8527	1.0260

Then, all the graphs will be displayed, similar to how we described earlier:

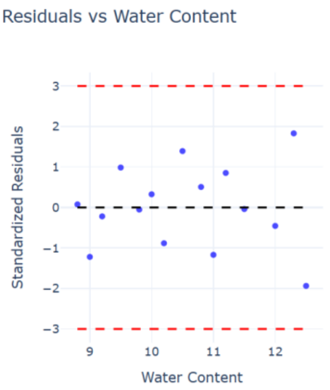
3D Regression Plot with Confidence Bands

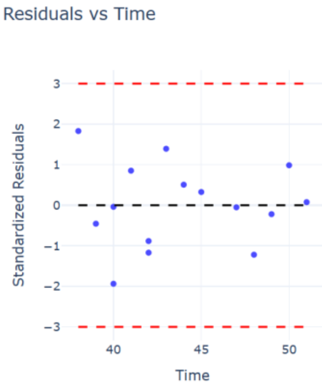


3D Regression Plot Without Confidence Bands



And finally, the standardized residuals plots will be displayed:





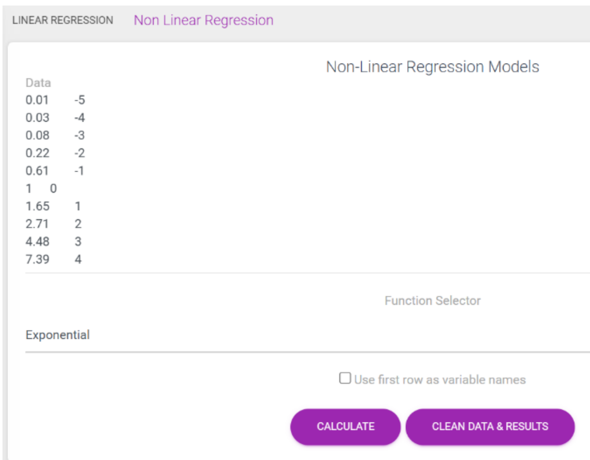
8.2 Non-Linear Regression

This case involves non-linear regression, using data from a Y variable and a single X variable. Therefore, the data should be copied into only two columns. Naturally, in the first row, the names of these variables can be optionally entered, and in that case, the checkbox must be selected.

Once the data is entered, the function should be selected from the dropdown menu, and then the 'Calculate' button should be pressed. Within the dropdown menu, the available options are quadratic, cubic, exponential, logarithmic, logistic, power, sinusoidal, Gaussian, hyperbolic, and sigmoidal.

Let's look at an example of how the data is copied into the panel and the results obtained when selecting the exponential function. For this example, the variable names are not included, so by default, they will be labeled as 'y' and 'x'.

0.01	-5
0.03	-4
0.08	-3
0.22	-2
0.61	-1
1	0
1.65	1
2.71	2
4.48	3
7.39	4



The results are similar to other regression types, starting with a table that provides the model description and the regression parameters.

Regression Results

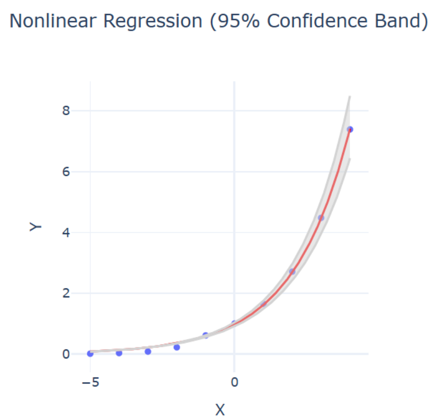
Model Summary:
R^2: 0.9990

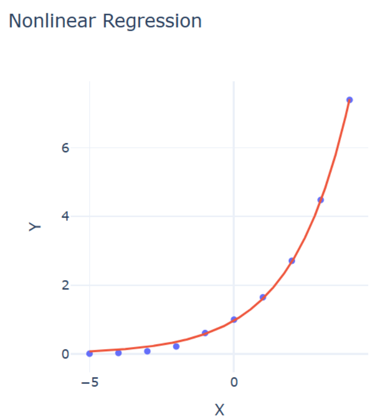
Equation:
y = 0.9702e^(0.5084x)

Coefficients:

Parameter	Coefficient	Std. Error	t-value	p-value	95% CI Lower	95% CI Upper
a	0.9702	0.0322	30.1467	0.0000	0.9071	1.0333
b	0.5084	0.0093	54.5645	0.0000	0.4902	0.5267

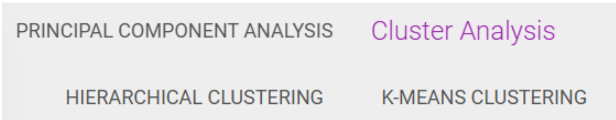
Additionally, the interactive scatter plots of the data with the fitted model, both with (left) and without (right) confidence bands, will be displayed.





9. Multivariate tools.

This section describes the multivariate exploratory methods available in the web application. Once the 'Multivariate' option is selected in the sidebar, the options in the top menu are: Principal Component Analysis (PCA) and Cluster Analysis (CA). In the case of CA, two subcases are presented: Hierarchical Clustering and non-hierarchical clustering, represented by the k-means method.



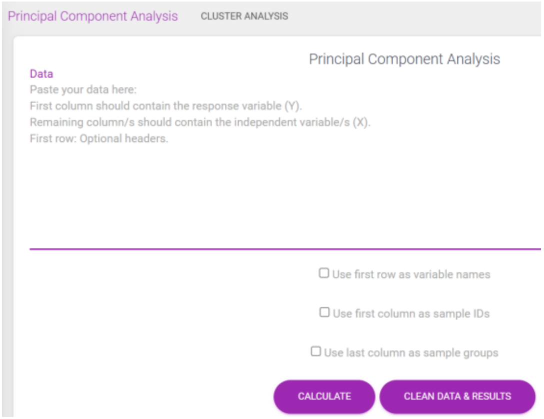
9.1 Principal Component Analysis.

PCA is based on the calculation of the eigenvalues and eigenvectors of the covariance matrix of the input data with multiple variables (expressed in different columns). This leads to the representation of the different samples (expressed in different rows) along new axes called principal components, which capture the greatest amount of variability in the data, from the first to the last component.

It is frequently used for dimensionality reduction and the identification of similar patterns in the samples through visual analysis, while also determining which original variables have the most influence on the most significant components.

In the data panel, the data should be pasted as follows: the sample names, variable names, and the factor (which will later be used by the system to color the points in the graphs) are optional. The numerical data values should have the variables in the columns and the different samples in the rows.

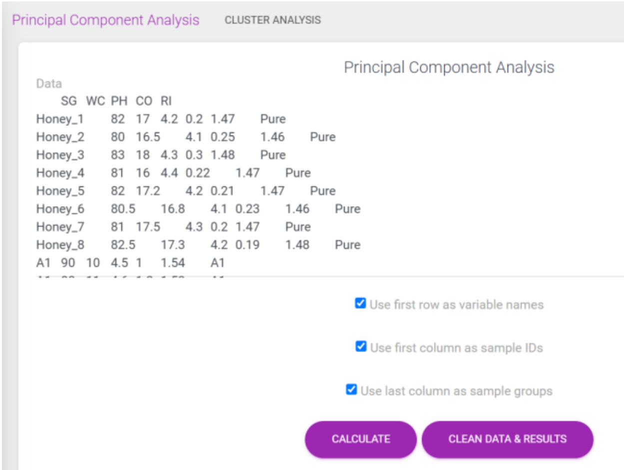
Here is an example of a data table collected from the determination of 5 variables (sugar content, water content, pH, conductivity, and refractive index) of a series of samples. The first 8 are measurements taken from pure honeys, and the remaining samples are artificially adulterated honeys. The goal is to determine whether these variables can help detect groupings of similar samples.



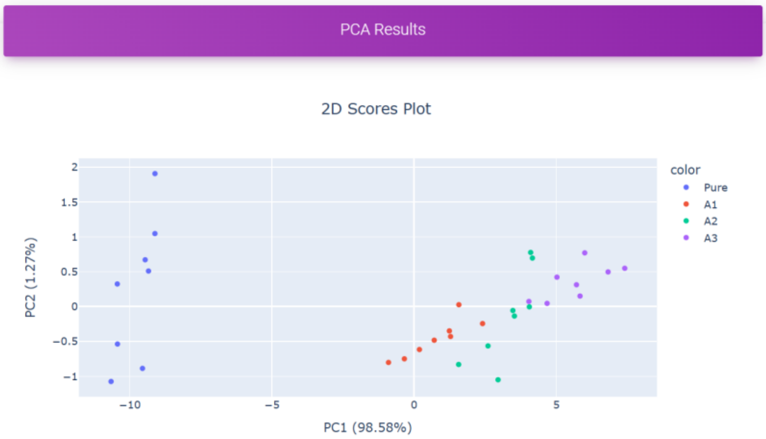
In this context, the data to be copied is shown below in a spreadsheet. As seen, the first column contains the names of each sample, the first row only contains the names of the variables, and the last column represents the pattern for coloring in the graphs.

	SG	WC	PH	CO	RI	
Honey_1	82	17	4	0.2	1.47	Pure
Honey_2	80	16.5	4	0.25	1.46	Pure
Honey_3	83	18	4	0.3	1.48	Pure
Honey_4	81	16	4	0.22	1.47	Pure
Honey_5	82	17.2	4	0.21	1.47	Pure
Honey_6	80.5	16.8	4	0.23	1.46	Pure
Honey_7	81	17.5	4	0.2	1.47	Pure
Honey_8	82.5	17.3	4	0.19	1.48	Pure
A1	90	10	5	1	1.54	A1
A1	88	11	5	1.2	1.53	A1
A1	89	10.5	4	1.1	1.55	A1
A1	91	9.5	5	1.3	1.56	A1
A1	90.5	10.2	5	1.1	1.54	A1
A1	88.5	10.7	4	1	1.53	A1
A1	89.5	10.3	4	1.1	1.55	A1
A1	90	10.1	5	1.2	1.54	A1
A2	92	9	5	0.9	1.58	A2
A2	91	8.5	4	0.8	1.57	A2
A2	93	9.2	5	0.85	1.59	A2
A2	92.5	8.7	5	0.9	1.58	A2
A2	90	9.5	5	0.8	1.56	A2
A2	91	9.1	5	0.75	1.57	A2
A2	92	8.9	5	0.8	1.58	A2
A2	93	9.3	5	0.85	1.59	A2
A3	94	8	5	1.4	1.6	A3
A3	93.5	8.5	5	1.5	1.59	A3
A3	95	7.5	5	1.6	1.61	A3
A3	94.5	8.2	5	1.4	1.6	A3
A3	92.5	8.8	5	1.5	1.58	A3
A3	93	8.4	5	1.5	1.59	A3
A3	95.5	7.2	5	1.7	1.61	A3
A3	94	7.8	5	1.5	1.6	A3

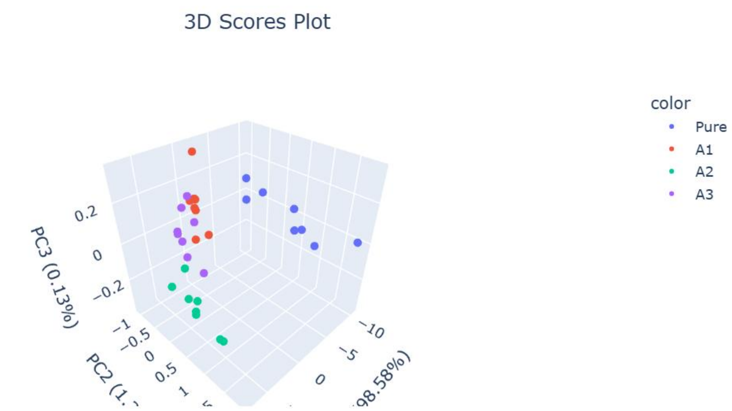
As previously mentioned, the sample names are optional. If they are not provided, values will be assigned from 1 to the number of samples entered. Similarly, if the variable names are not provided, default names will be assigned. If the last column (for pattern/coloring) is missing, all the points will be plotted in the same color. In this scenario, all three checkboxes must be selected for the system to recognize these values correctly.



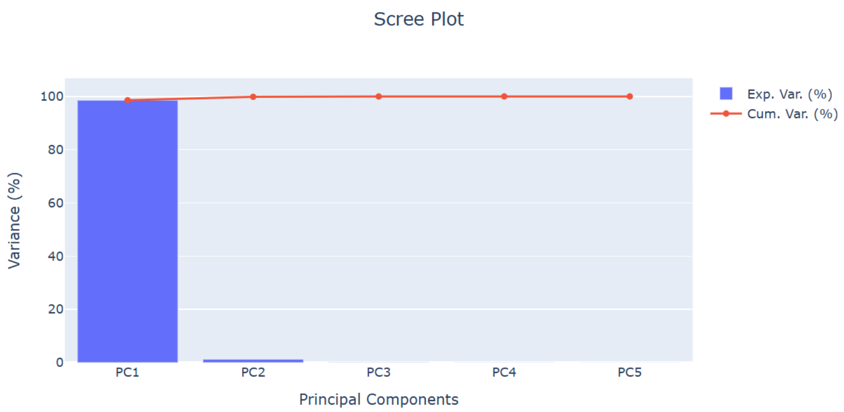
Upon pressing the 'Calculate' button, the PCA results will be displayed, starting with an interactive two-dimensional scatter plot of the first two principal components.



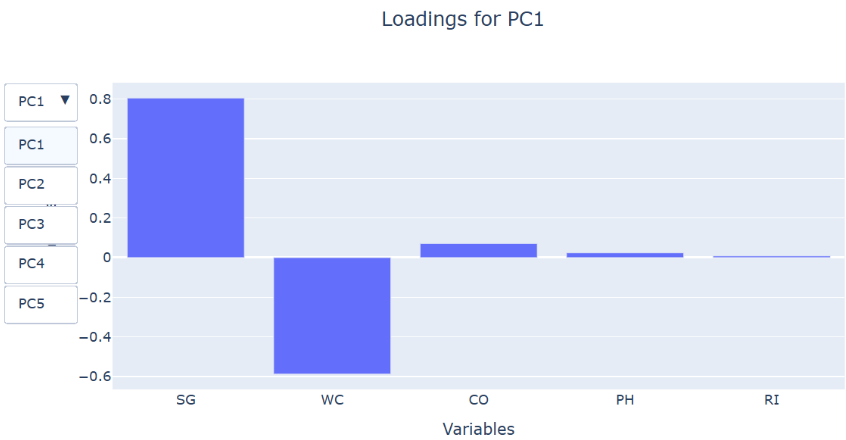
Next, an interactive three-dimensional plot will be displayed, which can be rotated by dragging with the mouse.



The scree plot, displaying the variance values for each of the components, will also be shown.



An interactive plot of the loading values will be displayed, where each component can be selected using the selector on the left side.



Additionally, if users require it, the values of the explained variance vector, the loading matrix, and the score matrix will be provided.

9.2 Cluster Analysis

This section presents the clustering methods, which include two types: the hierarchical method, which results in dendrogram-type graphs, and a non-hierarchical method called k-means, which partitions the samples into a predefined number of groups, resulting in a silhouette plot.

In both cases of CA (hierarchical and k-means), the data is entered similarly to the PCA case, but without the variable names and without the last column as a factor for coloring.

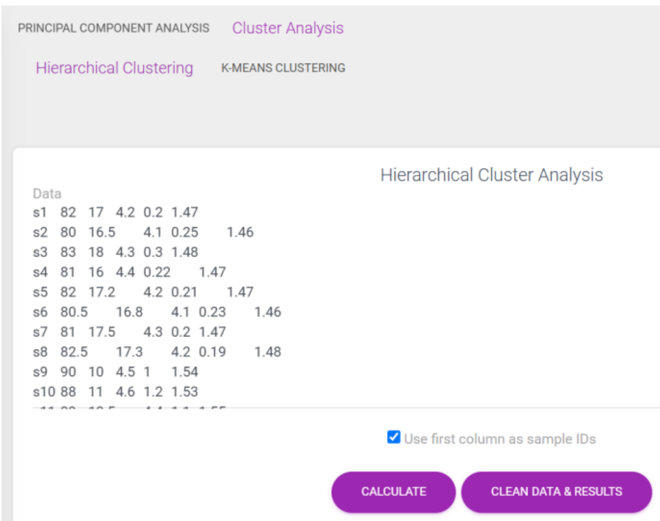
Only numeric values should be entered, with each column representing a variable and each row representing a sample. Optionally, the first column can include the sample names for easy identification in the graphs.

The k-means method requires an additional input: an integer specifying the number of clusters into which you wish to partition the data.

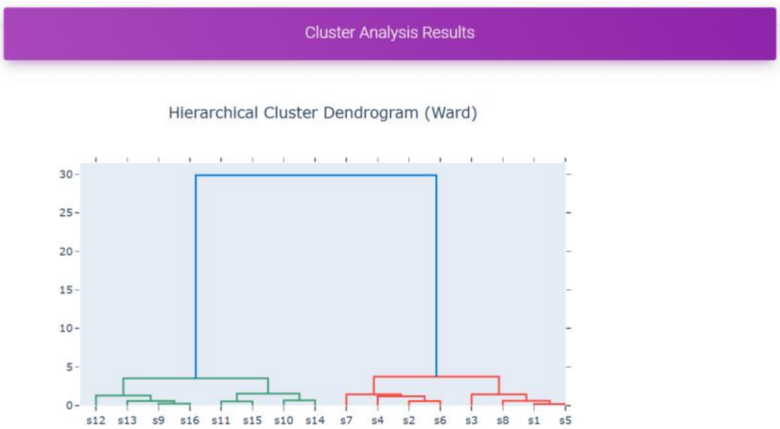
As an example, here is a table showing the required data format for performing cluster analysis (CA).

s1	82	17	4.2	0.2	1.47
s2	80	16.5	4.1	0.25	1.46
s3	83	18	4.3	0.3	1.48
s4	81	16	4.4	0.22	1.47
s5	82	17.2	4.2	0.21	1.47
s6	80.5	16.8	4.1	0.23	1.46
s7	81	17.5	4.3	0.2	1.47
s8	82.5	17.3	4.2	0.19	1.48
s9	90	10	4.5	1	1.54
s10	88	11	4.6	1.2	1.53
s11	89	10.5	4.4	1.1	1.55
s12	91	9.5	4.7	1.3	1.56
s13	90.5	10.2	4.5	1.1	1.54
s14	88.5	10.7	4.3	1	1.53
s15	89.5	10.3	4.4	1.1	1.55
s16	90	10.1	4.6	1.2	1.54

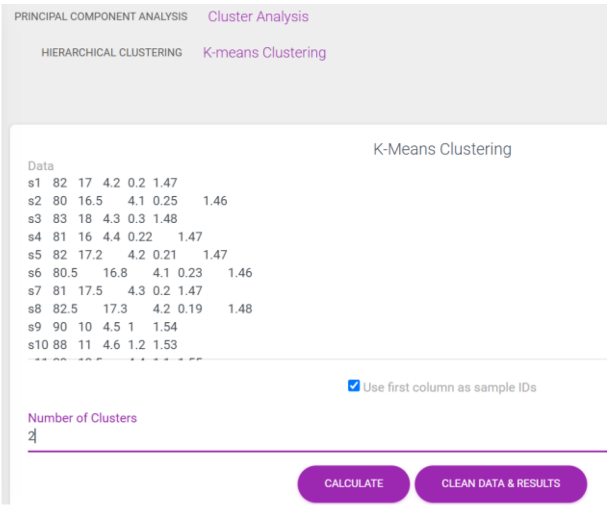
If the data is pasted into the hierarchical clustering panel, it would appear as follows:



As a result of this analysis, a dendrogram plot is generated using the Ward method.



In the case of performing a k-means cluster analysis with a cluster value of 2, the data would look as follows:



And a silhouette plot is obtained, as shown below:

At the bottom of the plot, the average silhouette width is also reported.

10. Final Considerations.

In summary, the web application provides a comprehensive and intuitive platform for performing various statistical analyses, including basic descriptive, inference, control charts, ANOVA, regression, and multivariate methods such as PCA and cluster analysis. Each method is designed to be user-friendly, with a straightforward process for inputting data and obtaining results. The application not only offers essential statistical outputs but also provides interactive visualizations, allowing users to explore their data and results more effectively.

The flexibility in handling different data formats and the inclusion of various statistical tests and post-hoc options make it a versatile tool for both beginners and experienced users.

By leveraging these features, users can gain valuable insights into their data, identify patterns or groupings, and make informed decisions based on robust statistical analysis. Whether for academic, research, or professional purposes, this application simplifies complex analyses, making them accessible and actionable.

The objective of this manual is to provide a quick and accessible reference for the functionality of the platform, allowing users to efficiently familiarize themselves with the available features. This manual is not intended to replace or serve as a comprehensive source of statistical theory. For a deeper and more detailed understanding of the concepts and statistical methods used, it is recommended to consult specialized textbooks in the field of statistics and scientific research.

Additionally, it is important to note that the platform is constantly evolving. New tools and features are being developed across different sections to gradually enhance the platform's functionality and user experience. These improvements are being implemented based on user feedback and needs, ensuring a complete and effective platform for conducting advanced statistical analyses.