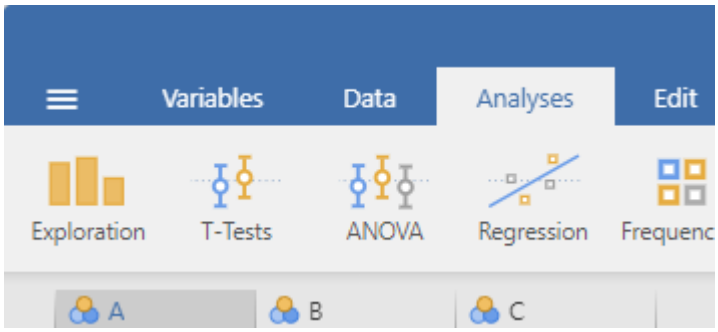
Instructions for computer-use in solving the proposed practical activities

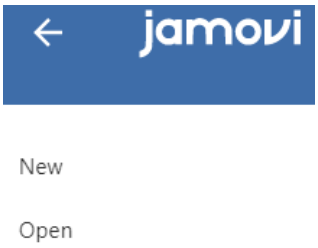
Instructions for using Jamovi for data analysis

Importing the database

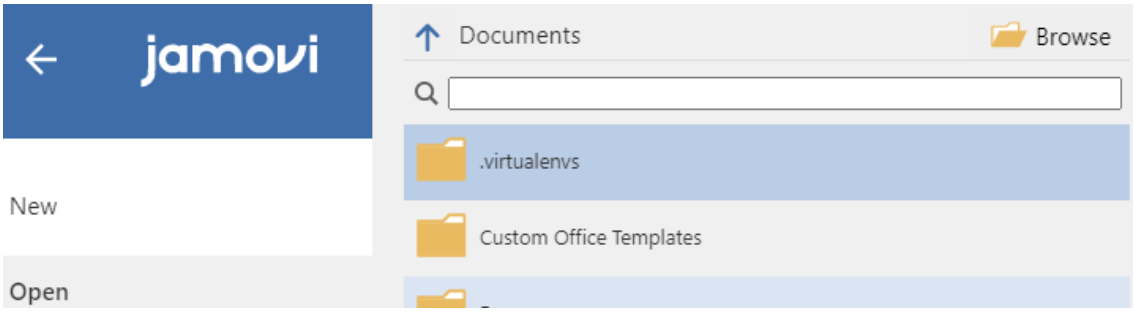
Use the **hamburger** menu icon .



Then select the **Open** option

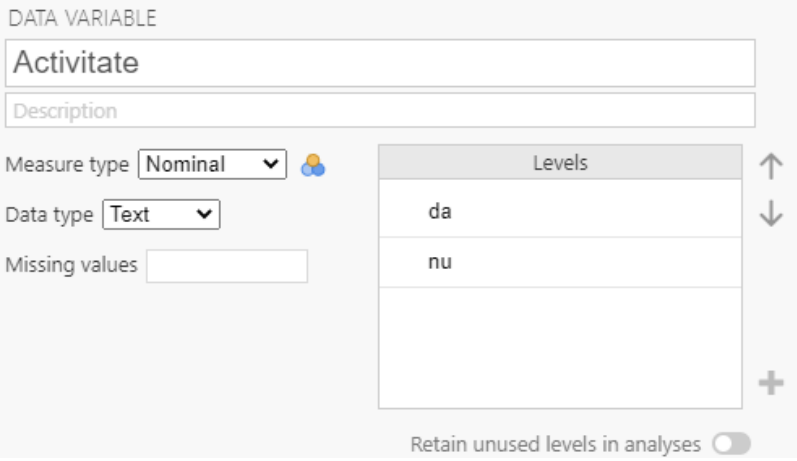


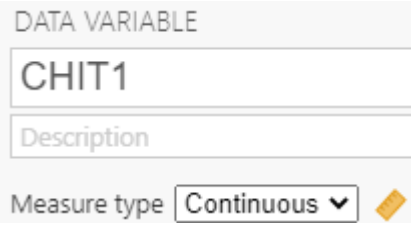
Afterwards, search for the database folder by pressing the **Browse button** , select the desired file and press the **Open button** to import it.



Data type checking

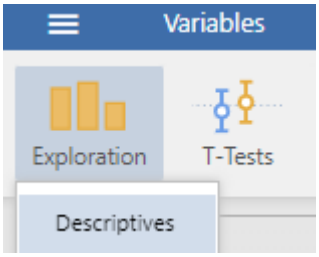
Before starting the analysis, it is necessary to check whether the program has correctly classified the variables in the database. The Activity variable is a qualitative variable, and the other 4 variables are quantitative. To check the type of a variable, **double-click** on **the column title** (variable name). In the case of the Activity variable we ensure that in the **Measure** selection area **type** option is selected **Nominally** . For the other quantitative variables we ensure that in the Measure selection area type option **Continuous** is selected .



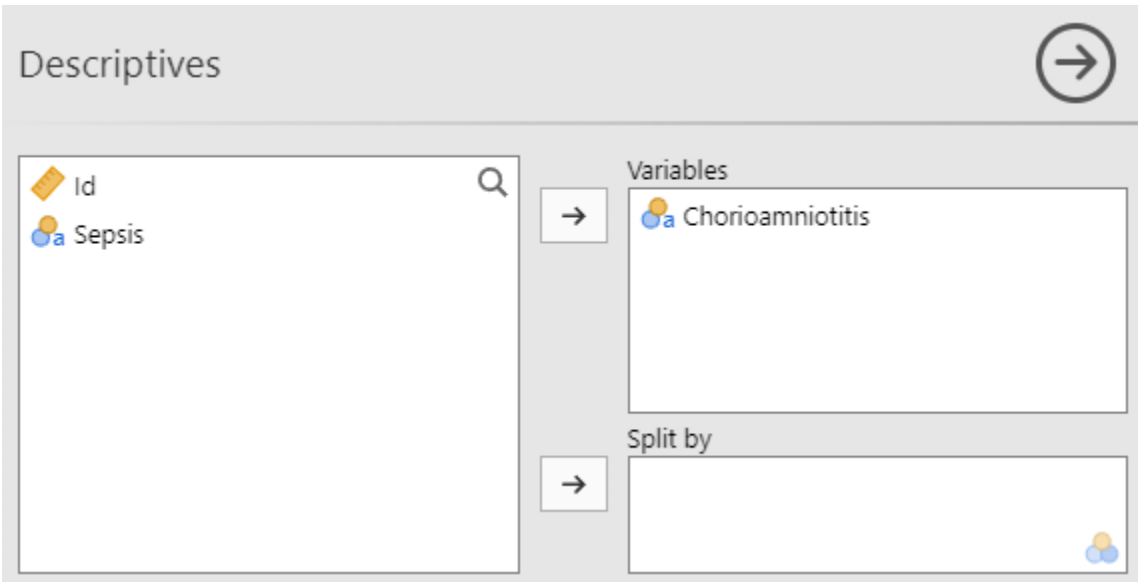


Description of a qualitative variable – frequency table, column graph

In the **Analyses** tab, the **Exploration** button, choose the **Descriptives** option.



The variable of interest (e.g. **disease**, or **risk factor**), should be moved by pressing the arrow next to the **Variables** field.



To display the frequency table, check the **Frequency tables** option.

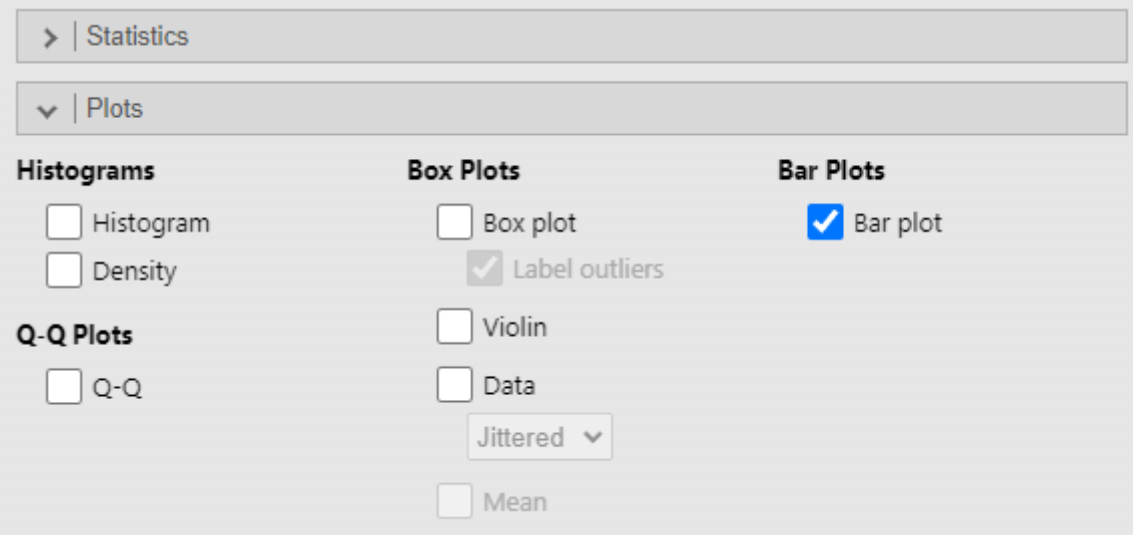
A frequency table similar to the one below is obtained:

Frequencies

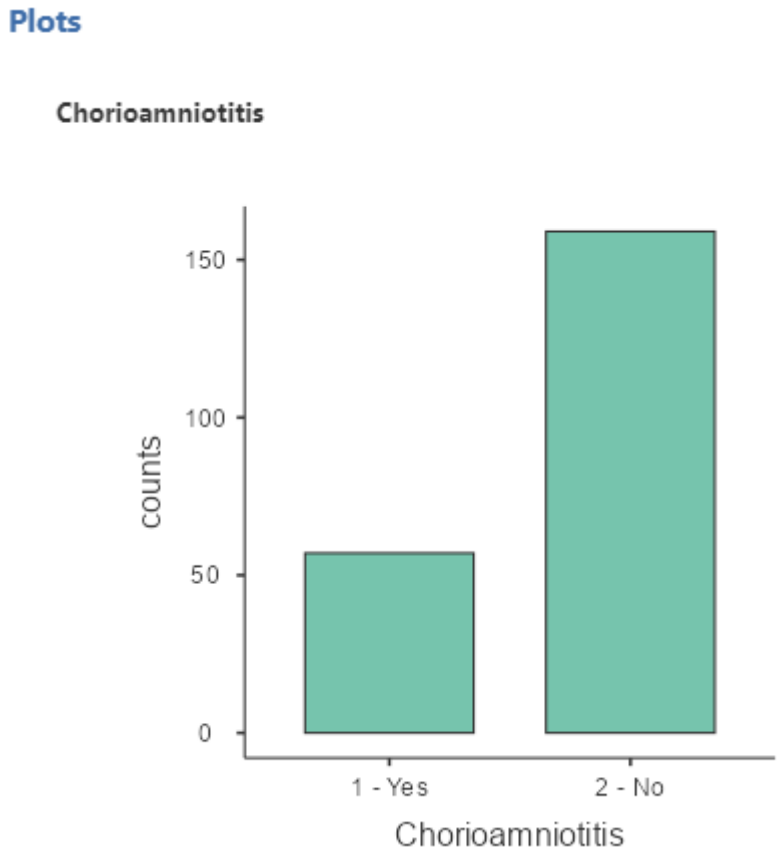
Frequencies of Chorioamniotitis

Chorioamniotitis	Counts	% of Total	Cumulative %
1 - Yes	57	26.4 %	26.4 %
2 - No	159	73.6 %	100.0 %

To display a column chart, check the **Bar plot** option.

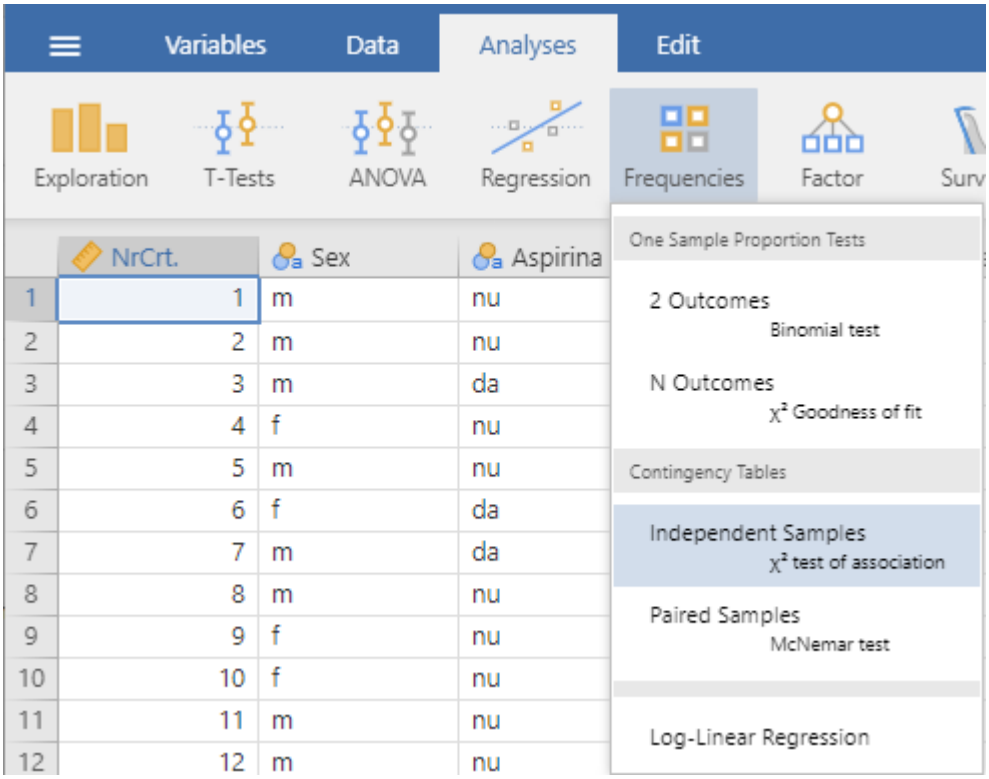


A column chart is obtained:



Analysis between 2 qualitative variables – Chi square test, Fisher test, contingency table, OR, RR, RA, column graph

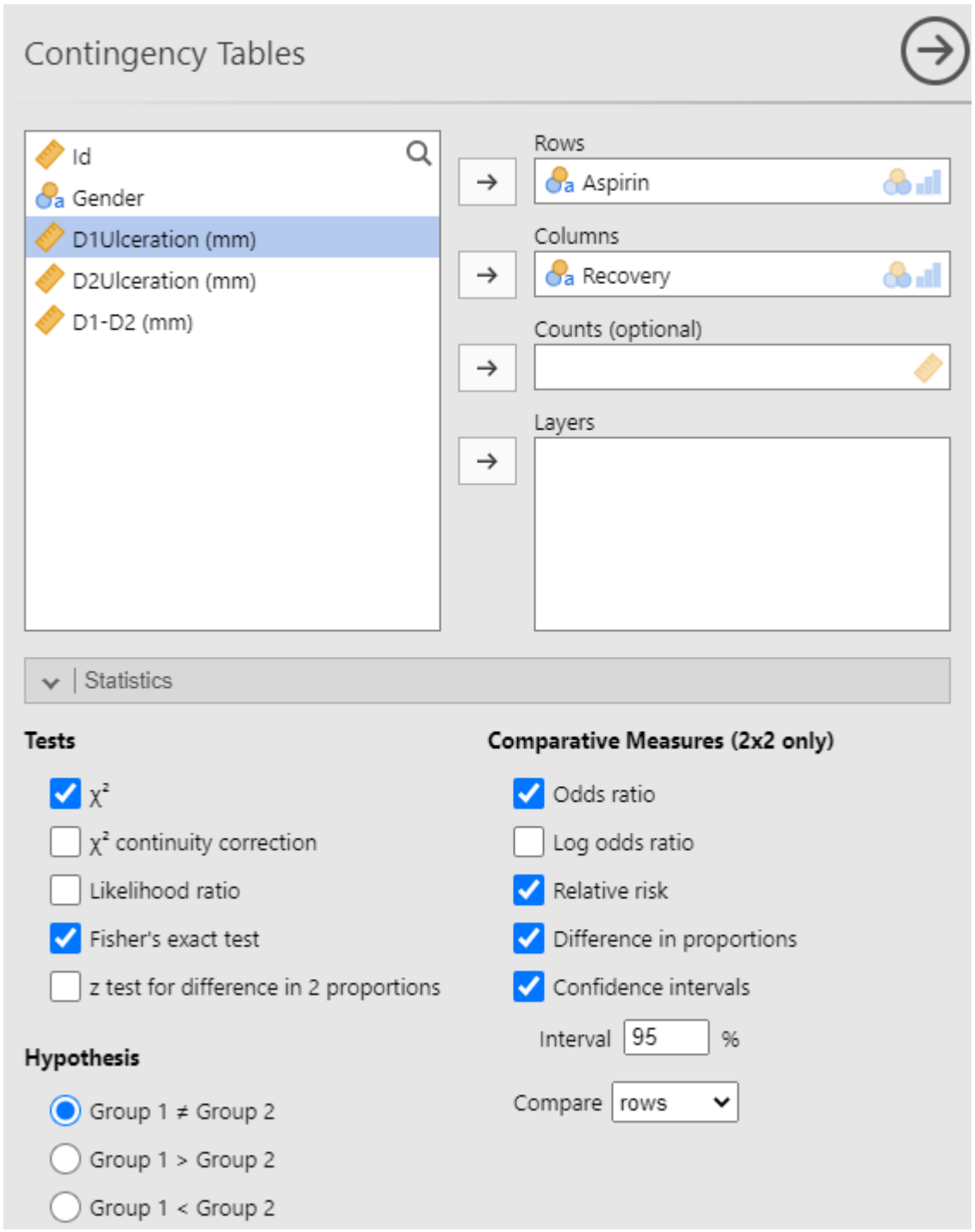
In the **Analyses** tab**,** the **Frequencies** module one can choos the option **Independent Samples (X^2 test of association).**

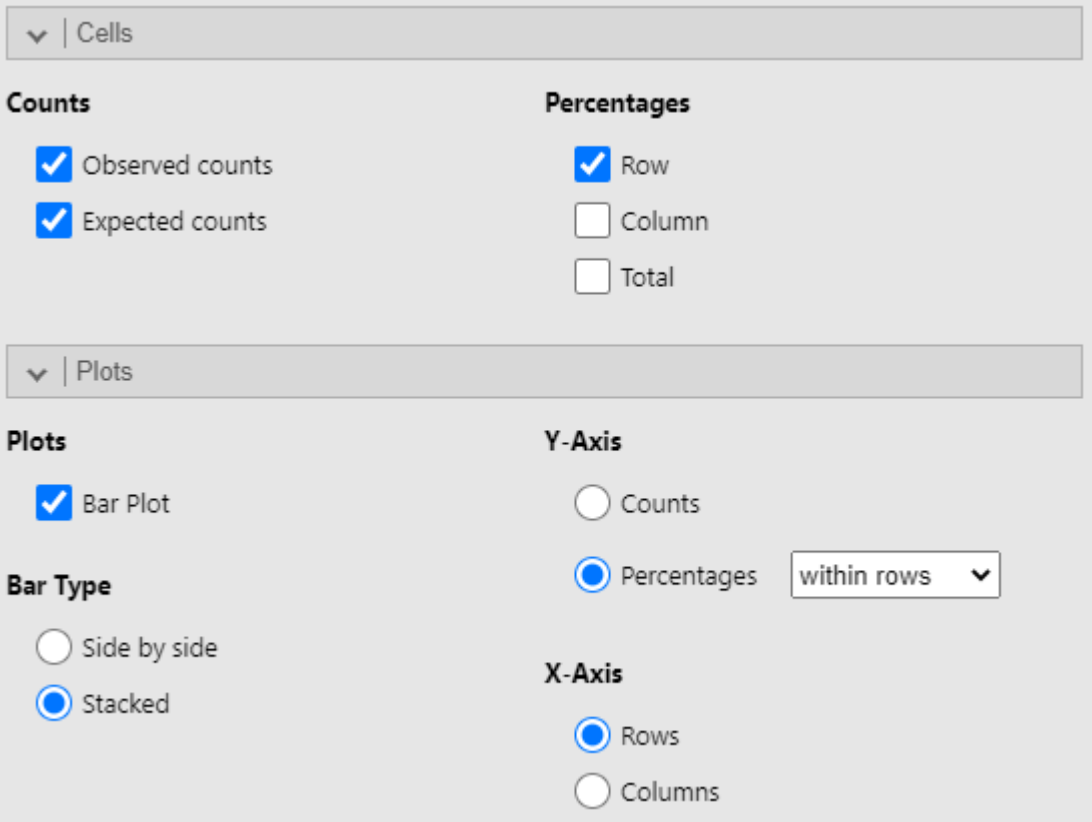


Grouping variable (e.g. **risk factor**, or **treatment**), should be moved by pressing the arrow next to the **Rows field**.

The variable that is of interest (e.g. **disease**, or **complication**, or **treatment result**), should be moved by pressing the arrow next to the **Columns field**.

As options we recommend the following choices: Tests: **X2**, **Fisher's exact test**; Comparative measures (2x2 only): **Odds ratio**, **Relative risk**, Attributable risk/absolute risk reduction (**Difference in proportion)**, **Confidence intervals**; Counts: **Observed counts**, **Expected counts**; Percentages: **Row**; Plots: column (**Bar Plot**), **Y-Axis**: percentages, **within rows**; Bar **Type**, Stacked; **X-Axis**: **Rows**.


Contingency Tables
Id
Gender
D1Ulceration (mm)
D2Ulceration (mm)
D1-D2 (mm)
Rows
Aspirin
Columns
Recovery
Counts (optional)
Layers
Statistics
Tests
χ²
χ² continuity correction
Likelihood ratio
Fisher's exact test
z test for difference in 2 proportions
Hypothesis
Group 1 ≠ Group 2
Group 1 > Group 2
Group 1 < Group 2
Comparative Measures (2x2 only)
Odds ratio
Log odds ratio
Relative risk
Difference in proportions
Confidence intervals
Interval 95 %
Compare rows



Contingency table

Contingency Tables

Aspirina		Vindecare da	Vindecare nu	Total
da	Observed	182	18	200
	Expected	162	38.0	200
	% within row	91.0 %	9.0 %	100.0 %
nu	Observed	142	58	200
	Expected	162	38.0	200
	% within row	71.0 %	29.0 %	100.0 %
Total	Observed	324	76	400
	Expected	324	76.0	400
	% within row	81.0 %	19.0 %	100.0 %

Chi Squared and Fisher exact test Results

The value of p, corresponding to the tests, is in the column p.

χ^2 Tests

	Value	df	p
χ^2	26.0	1	< .001
Fisher's exact test			< .001
N	400		

Medical indicators of association strength

Medical indicator results are presented in the **Comparative Measures** table, where the point estimator is in the **Value** column, and the ends of the 95% confidence interval are in the **Lower, Upper (95% Confidence Intervals) columns**.

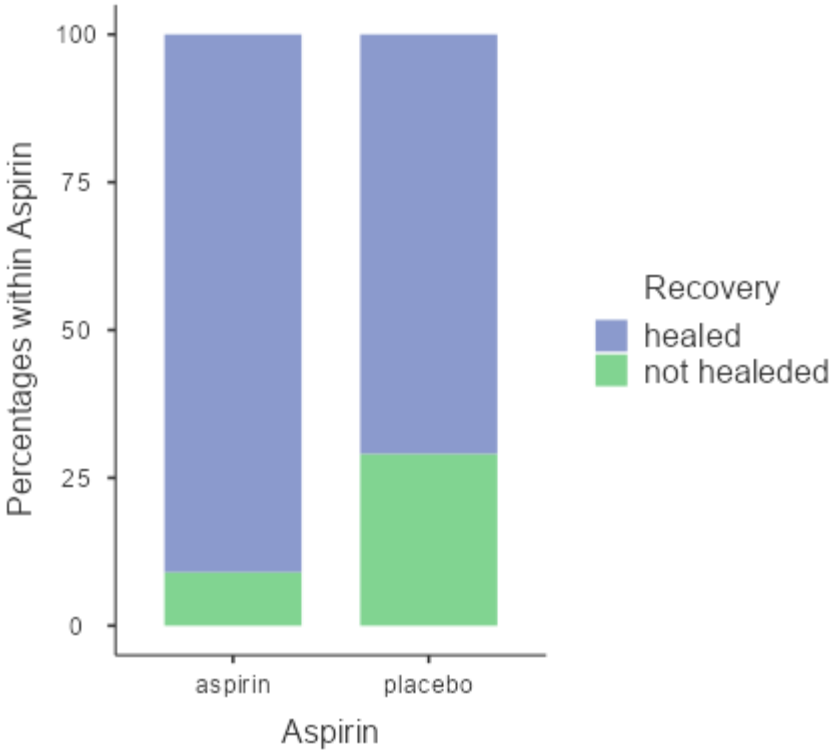
Comparative Measures

		95% Confidence Intervals	
	Value	Lower	Upper
Difference in 2 proportions	0.200[a]	0.126	0.274
Odds ratio	4.13	2.33	7.32
Relative risk	1.28[a]	1.16	1.41

[a] Rows compared

Column chart

Finally, we have the graphical representation of the association between the two qualitative variables.



Student test for two independent groups

Evaluation of the Student test application conditions – normality

The Student Test can only be **applied if the data of both compared groups** follow a **normal distribution**. Various methods can be used to assess normality: data distribution indicators: skewness - asymmetry coefficient, excess kurtosis; tests for normality; graphs (histogram, QQ plot).

Data distribution indicators can be used to learn about the normality of the data. If **both** the skewness **coefficient and** the kurtosis coefficient (**both coefficients**) are **within the range (-1; 1)** it is a suggestion that the data follow a **normal distribution**; if the skewness coefficient **or** the excess kurtosis (**any of the coefficients**) is **outside the range (-1; 1)** we have a suggestion that the data **does not follow a normal distribution.** Example in the table: for variable D2 ulceration (mm) – suggests that data do not follow a normal distribution (skewness: yes=4.30, no=2.54; kurtosis: yes=20.2; no=6.58); for variable D1-D2 (mm) – suggests that data follows a normal distribution (skewness: yes=-0.0718, no=0.119; kurtosis: yes=-0.480; no=-0.403).

One **option** for assessing the normality of data is with the help **of statistical tests to assess normality**. If the sample has **less than 50 observations** we look at the result of the **Shapiro-Wilk test**, if

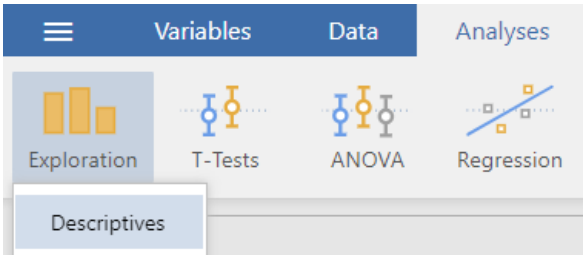
it has **more than 50 observations** we look at the result of the **Kolmogorov-Smirnov** test (in case it is available). The outcome of interest is **the p-value** of the test (Shapiro-Wilk p). If the p-value is **less than 0.05**, it is a **suggestion** that the data **does not follow a normal distribution**, **otherwise** if the p-value is **greater than 0.05**, it is a **suggestion** that the data follows a **normal distribution**. Example in the table: for variable D2 ulceration (mm) – suggests that data does not follow a normal distribution (yes: p<0.001, no: p<0.001); for variable D1-D2 (mm) – suggests that data follows a normal distribution (yes: p=0.547, no: p=0.512).

Another **way** to **evaluate the normality of the** data is with the help of graphs, such as the histogram, or the Q-Q-plot graph.

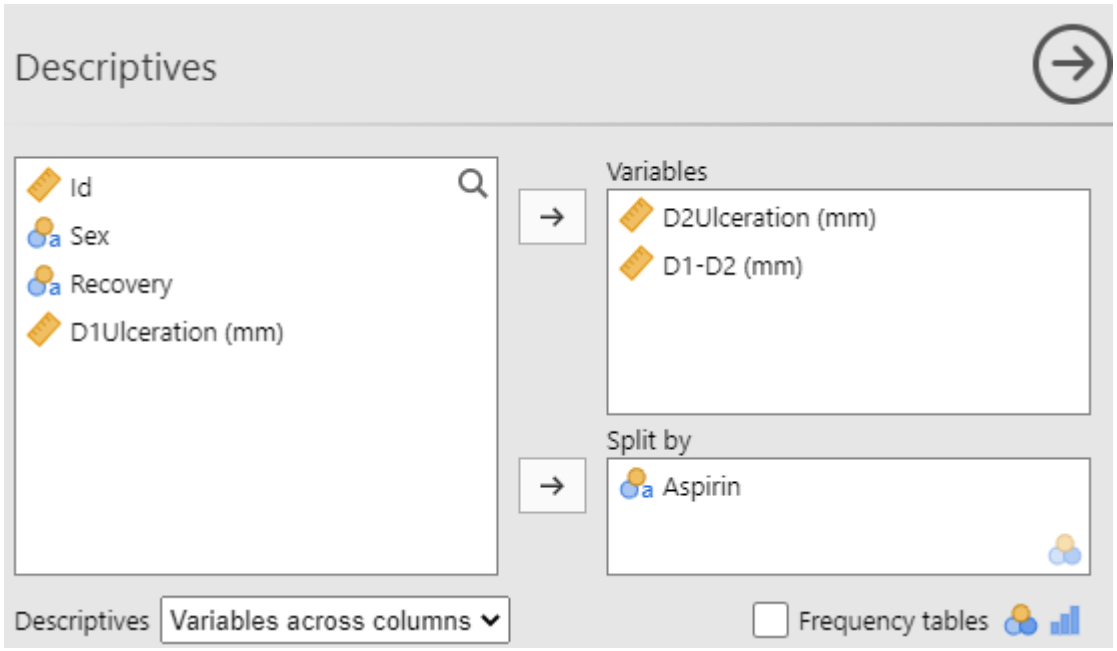
Histogram chart interpretation. If the graph follows an **approximately, relatively symmetrical bell shape**, there is a suggestion that the data follows a **normal distribution**, **otherwise** we have a suggestion that the data **does not follow a normal distribution.** Example in the image: for variable D2 Ulceration (mm) – suggests that data does not follow a normal distribution; for variable D1-D2 (mm) – suggests that data follows a normal distribution.

Interpretation of the Q-Q plot graph. If **the tendency** of the observations is to be arranged **far** from **the solid line** (which represents a normal distribution), then it is a **suggestion** that the data **does not follow a normal distribution**, **otherwise**, if the tendency of the observations is to be close to the solid line, it is a **suggestion** that the data follow a **normal distribution**.

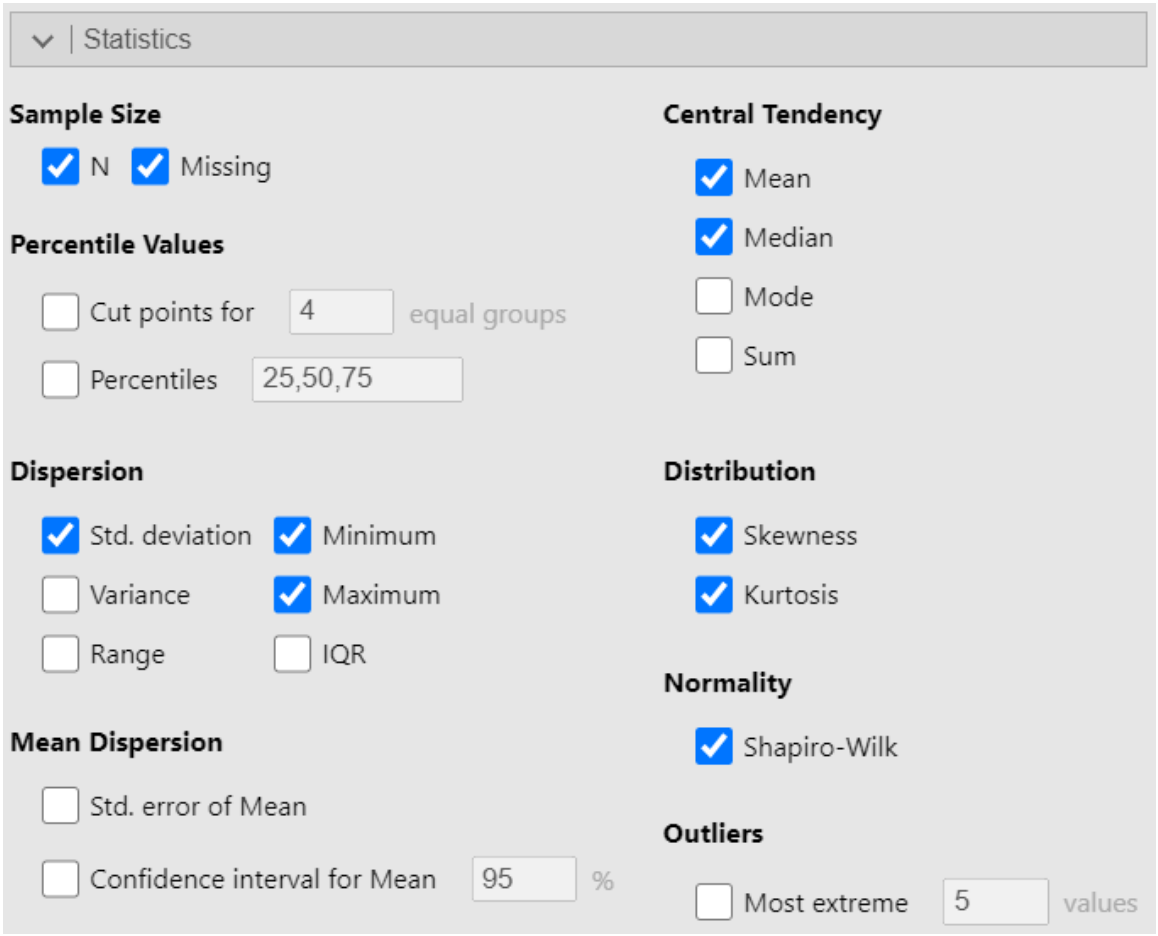
The normality evaluation will be done in the **Analysis** tab, the **Exploration** menu, the **Descriptives** command.



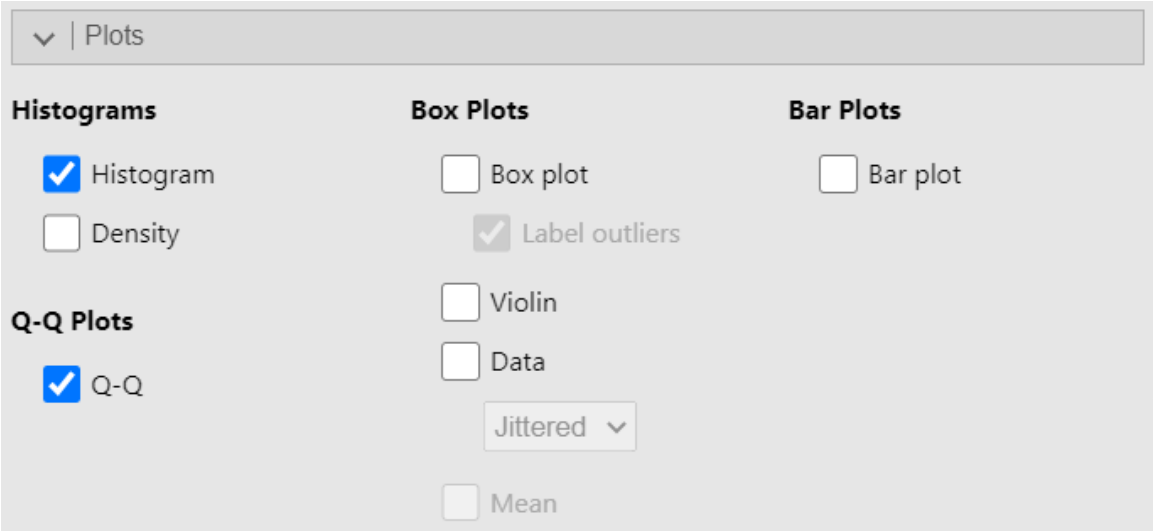
In the **Variables** section, the quantitative variables for which normality is to be evaluated (e.g. D2Ulceration (mm), D1-D2 (mm)) are moved. In the **Split by** section , the qualitative variable is moved, which identifies the groups to be compared (e.g. Aspirin).



In the **Statistics** section**, check Skewness, Kurtosis, Shapiro-Wilk**.



In the **Plots** section, **check Histogram** and **Q-Q**.

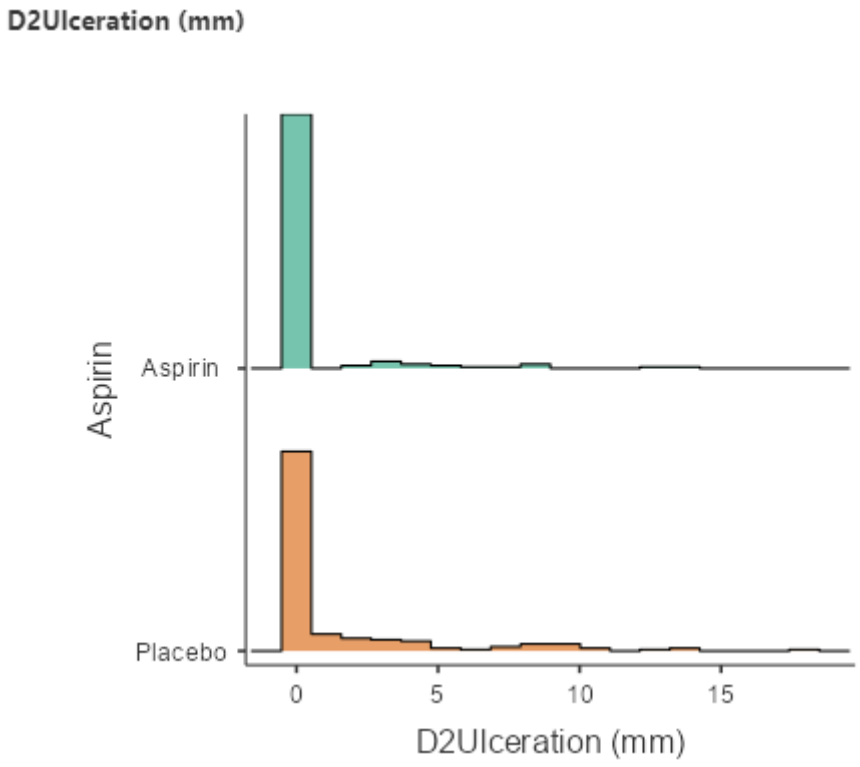


A table with statistics on each variable is obtained, for each group:

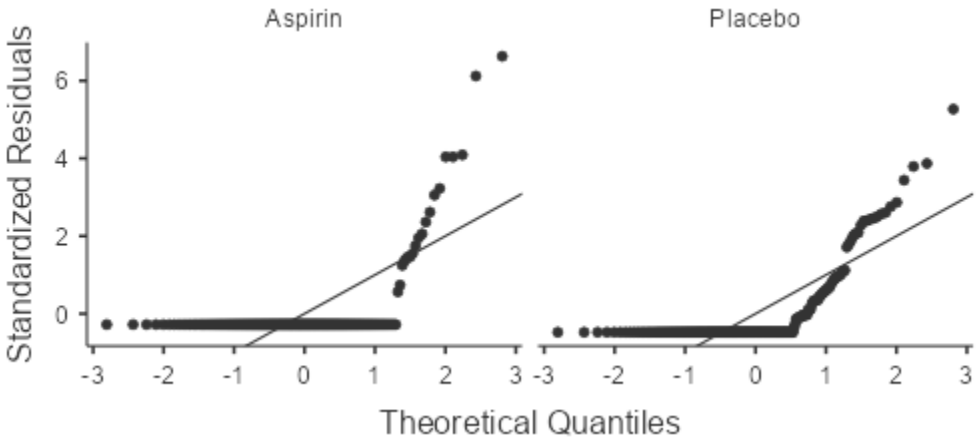
Descriptives

	Aspirin	D2Ulceration (mm)	D1-D2 (mm)
N	Aspirin	200	200
	Placebo	200	200
Missing	Aspirin	0	0
	Placebo	0	0
Mean	Aspirin	0.547	23.7
	Placebo	1.49	22.3
Median	Aspirin	0.00	23.6
	Placebo	0.00	21.9
Standard deviation	Aspirin	1.97	3.85
	Placebo	3.12	4.18
Minimum	Aspirin	0.00	13.8
	Placebo	0.00	12.6
Maximum	Aspirin	13.6	32.9
	Placebo	17.9	33.2
Skewness	Aspirin	4.30	−0.0718
	Placebo	2.54	0.119
Std. error skewness	Aspirin	0.172	0.172
	Placebo	0.172	0.172
Kurtosis	Aspirin	20.2	−0.480
	Placebo	6.58	−0.403
Std. error kurtosis	Aspirin	0.342	0.342
	Placebo	0.342	0.342
Shapiro-Wilk W	Aspirin	0.313	0.994
	Placebo	0.554	0.993
Shapiro-Wilk p	Aspirin	< .001	0.547
	Placebo	< .001	0.512

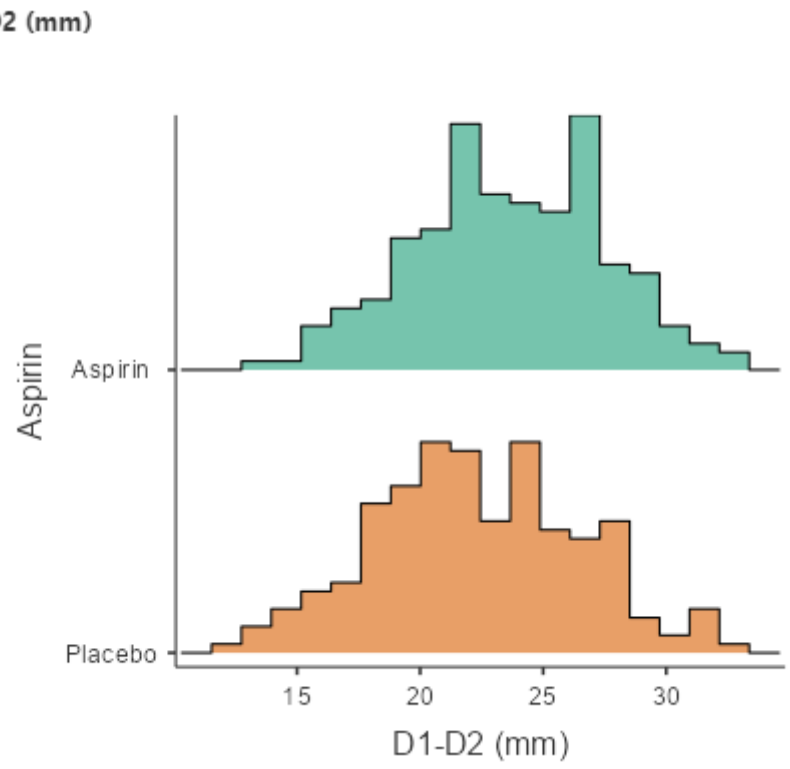
The **histogram for** a variable for which the data **does not follow a normal distribution (marked right- skewness, marked kurtosis):**



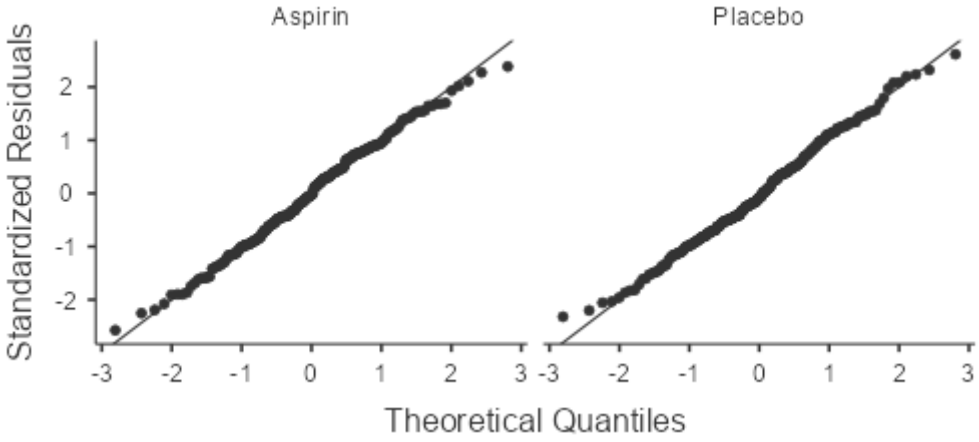
Q-Q plot for a variable for which the data **does not follow a normal distribution (observations are far from the normal line):**



The **histogram chart** for a variable for which the data **follows an approximately normal distribution (bell-shaped, relatively symmetrical chart):**



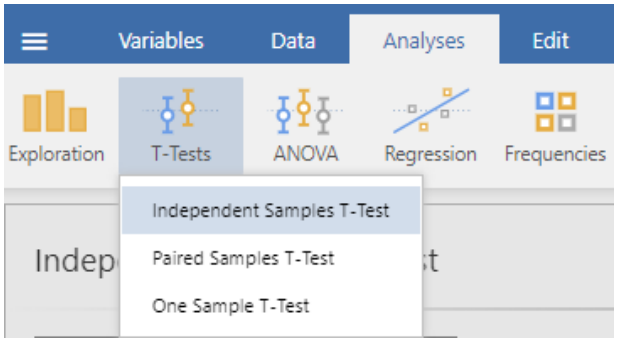
Q-Q plot for a variable for which the data **follows a normal distribution (the observations are close to the normal line):**



In the **Analyses** tab, choose the **Independent Samples T-Test option**.

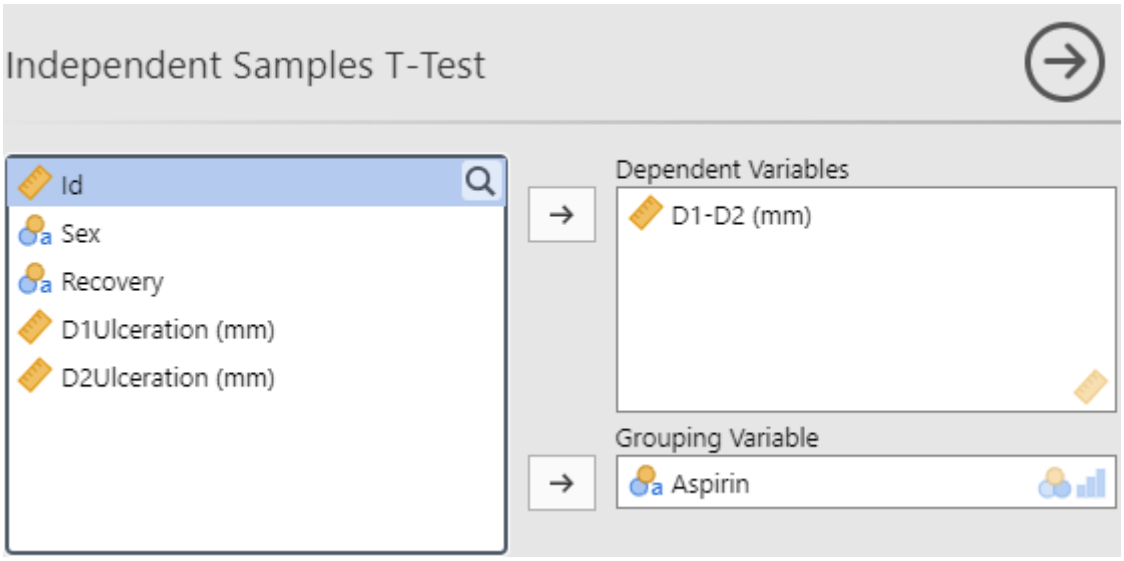
Conducting the Student Test for Independent Samples

In the **Analyses tab**, choose the Independent **Samples T-Test option**.



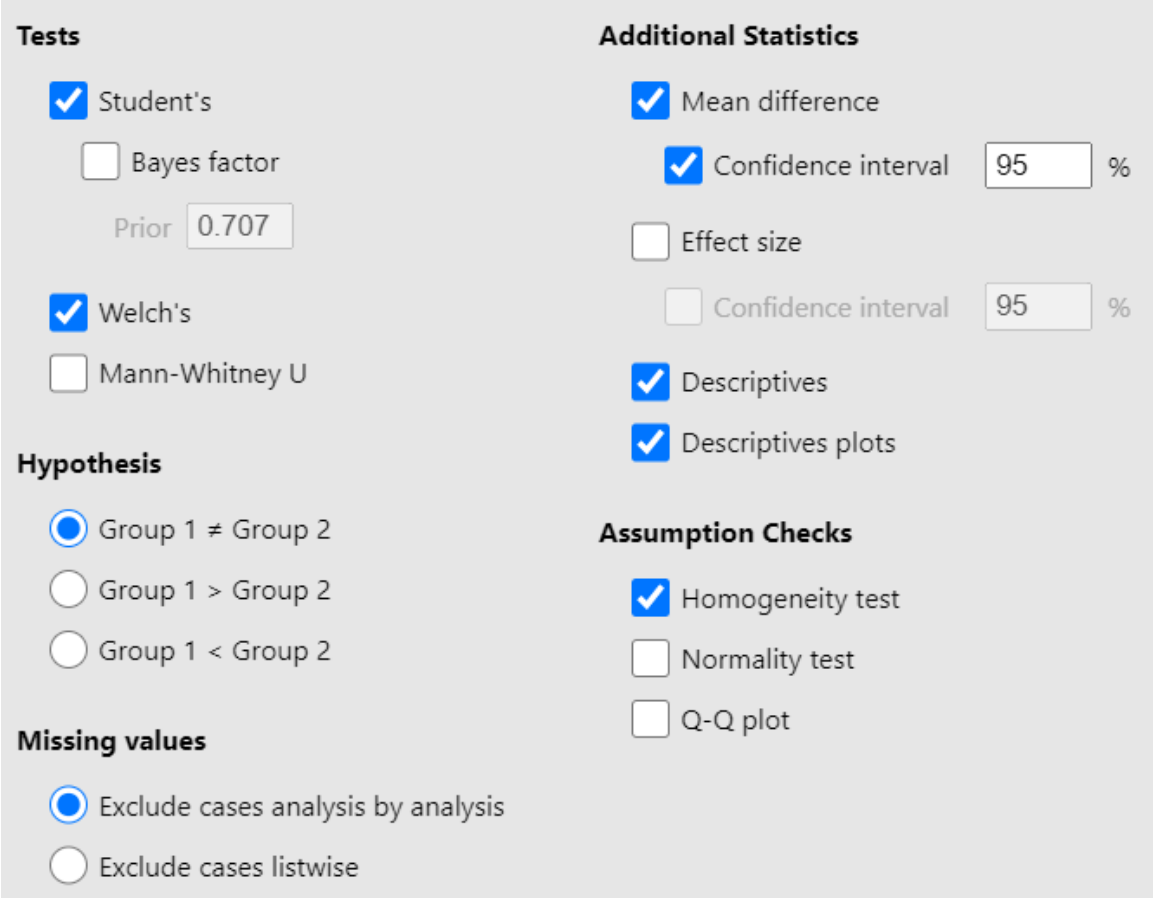
The qualitative variable that **identifies the compared groups** should be moved by pressing the arrow next to the **Grouping Variable field**.

The quantitative variable(s) that are of interest should be moved by pressing the arrow next to the **Dependent Variable field**.



Check the following options:

1. in the Tests section: **Student's**, **Welch's section**
2. in section **Additional statistics**: **Mean difference**, **Confidence interval**, **Descriptives**, **Descriptive plots**.
3. In section **Assumptions checks**: **Homogeneity test**.



Choosing between the Equal Variance Student Test and the Welch Test – Unequal Variance

In the case **of the Student test for independent groups**, there are **two variants**, for the situation where **equal variances** or **unequal variances** are assumed. To choose between the two, tests can be used **to compare variances**. In the **Homogeneity of Variances Tests table**, we look at **the p-value** obtained with **the Leven test**. If **the p-value** is **less than 0.05**, it is a **suggestion** that groups **have unequal variances**, **otherwise** it is a **suggestion** that groups have **equal variances**.

Assumptions

Homogeneity of Variances Test (Levene's)

	F	df	df2	p
D1-D2 (mm)	1.17	1	398	0.280

Note. A low p-value suggests a violation of the assumption of equal variances

[3]

If **the variances** are assumed to be **equal**, we choose the corresponding results for **the Student test**, otherwise we choose the corresponding ones for the **Welch test** in the **Independent Samples T-**

Test table. In column **P** we have **the p value** of the test, in the column **Mean difference**, we have the **difference between the averages of** the two groups, and in **Lower, Upper (95% Confidence Interval),** there is **the 95% confidence interval** associated with the difference between the averages of the groups.

Independent Samples T-Test

Independent Samples T-Test

							95% Confidence Interval	
		Statistic	df	p	Mean difference	SE difference	Lower	Upper
D1-D2 (mm)	Student's t	3.44	398	< .001	1.38	0.402	0.591	2.17

Note. H_a $\mu_{Aspirin} \neq \mu_{Placebo}$

Descriptive statistics

Numeric

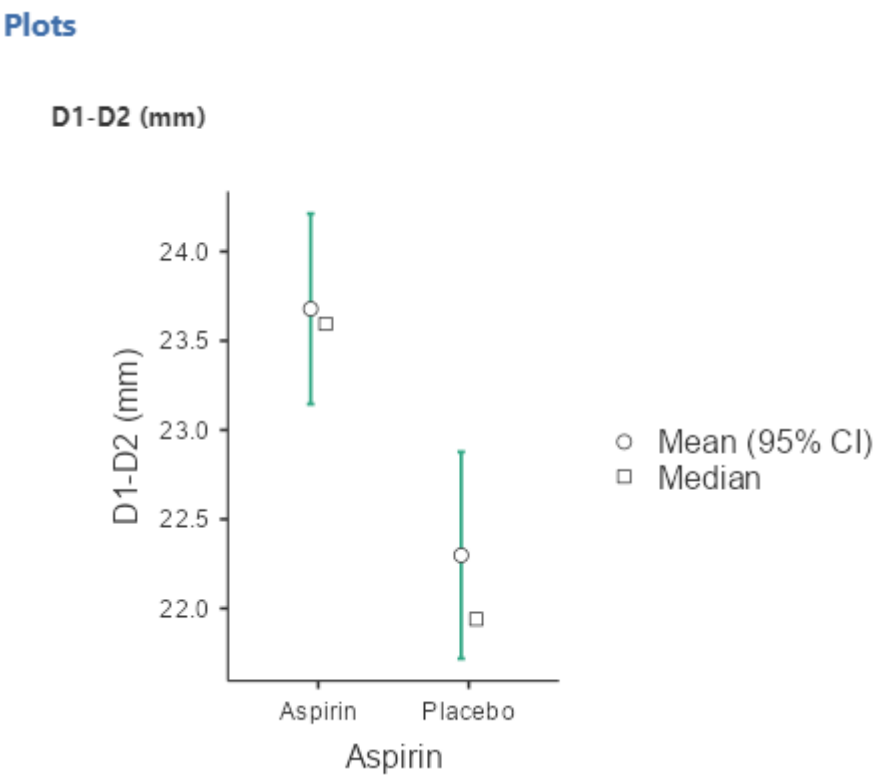
Descriptive statistics can be found numerically in the **Group Descriptives table**. **The number of subjects** is in column **N**, **mean** in column **Mean**, **standard deviation** in column **SD.**

Group Descriptives

	Group	N	Mean	Median	SD	SE
D1-D2 (mm)	Aspirin	200	23.7	23.6	3.85	0.272
	Placebo	200	22.3	21.9	4.18	0.296

Graph of means

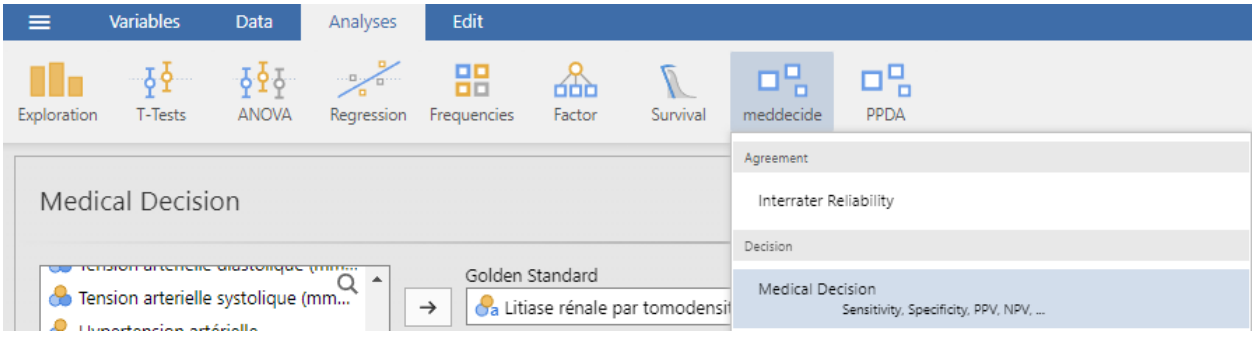
A graph of means **is also provided**, where the circle represents the mean, error bars represent the 95% confidence interval, and square represents the median.



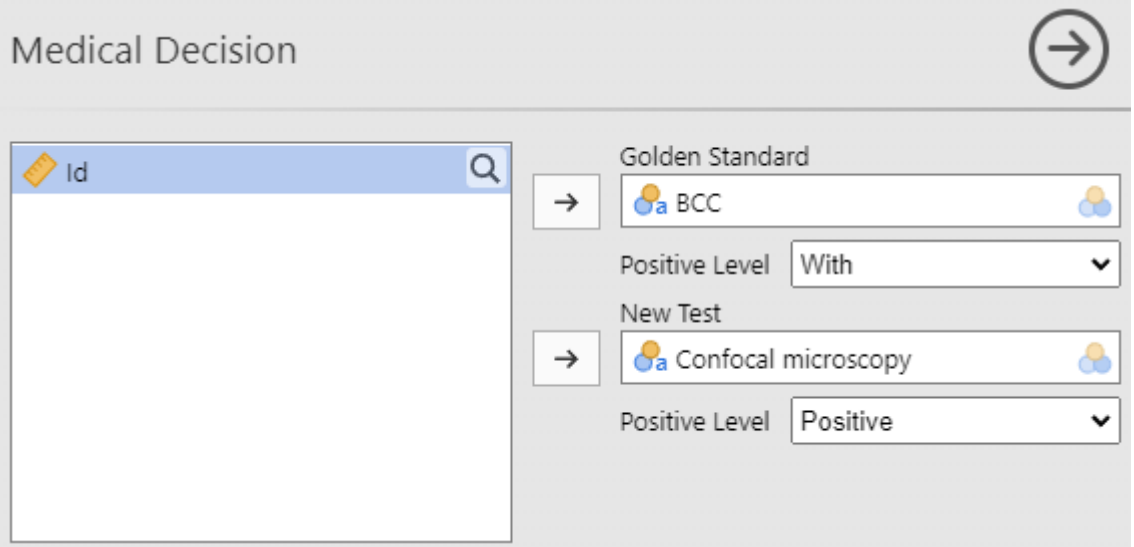
Diagnostic Tests

Make sure that the **Meddecide – Functions for medical decision in ClinicoPath** module is installed. See the chapter **Installing an additional module**.

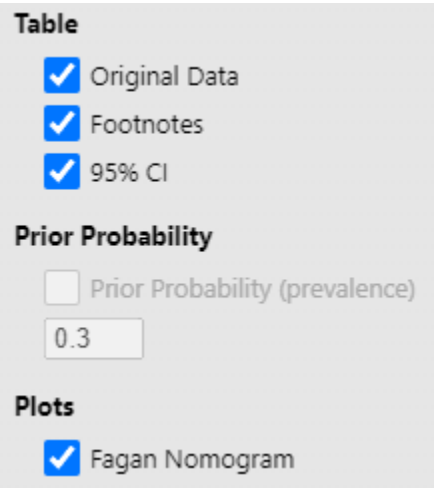
In the **Analyses** tab, press the **Meddecide** button, from the menu, select the Medical Decision option (Sensitivity, Specificity, PPV, NPV).



Select the variable that represents the **standard test** (for example, BCC) and press **the arrow button** next to the **Golden Standard** field. Specify which category of the variable represents the presence of the disease, as indicated by the standard test (here With), in the **Positive Level list**. **Select** the variable that represents the **new test** (for example, Confocal microscopy) and press **the arrow button** next to the **New Test** field. Specify which category of the variable represents a positive test result (here Positive), in the **Positive Level list**.



Select the following options: in the Table section: Original Data, Footnotes, 95% CI; in the Plots section: Fagan Nomogram.



Original Data

	BCC	
Confocal microscopy	With	Without
Negative	6	20
Positive	57	5

	Confocal microscopy	BCC	n
1	Negative	With	6
2	Negative	Without	20
3	Positive	With	57
4	Positive	Without	5

Check if the data is correctly placed in the table below, compared to the table at the above! If there are any problems, change the Positive Level option.

Recoded Data for Decision Test Statistics

	Gold Positive	Gold Negative	Total
Test Positive	57	5	62
Test Negative	6	20	26
Total	63	25	88

Below there are statistics of the totals of the participants, by different categories:

	n
Totalundefined	88[a]
Diseasedundefined	63[b]
Healthyundefined	25[d]
Positive Testsundefined	62[e]
Negative Testsundefined	26[f]
True Testundefined	77[g]
Wrong Testundefined	11[h]

[a] Total Number of Subjects

[b] Total Number of Subjects with Disease

[d] Total Number of Healthy Subjects

[e] Total Number of Positive Tests

[f] Total Number of Negative Tests

[g] Total Number of True Test Results

[h] Total Number of Wrong Test Results

Below there are the **classic statistics of diagnostic tests, with their 95% confidence interval**.

	n
Totalundefined	88[a]
Diseasedundefined	63[b]
Healthyundefined	25[d]

Positive Testsundefined	62[e]
Negative Testsundefined	26[f]
True Testundefined	77[g]
Wrong Testundefined	11[h]

[a] Total Number of Subjects

[b] Total Number of Subjects with Disease

[d] Total Number of Healthy Subjects

[e] Total Number of Positive Tests

[f] Total Number of Negative Tests

[g] Total Number of True Test Results

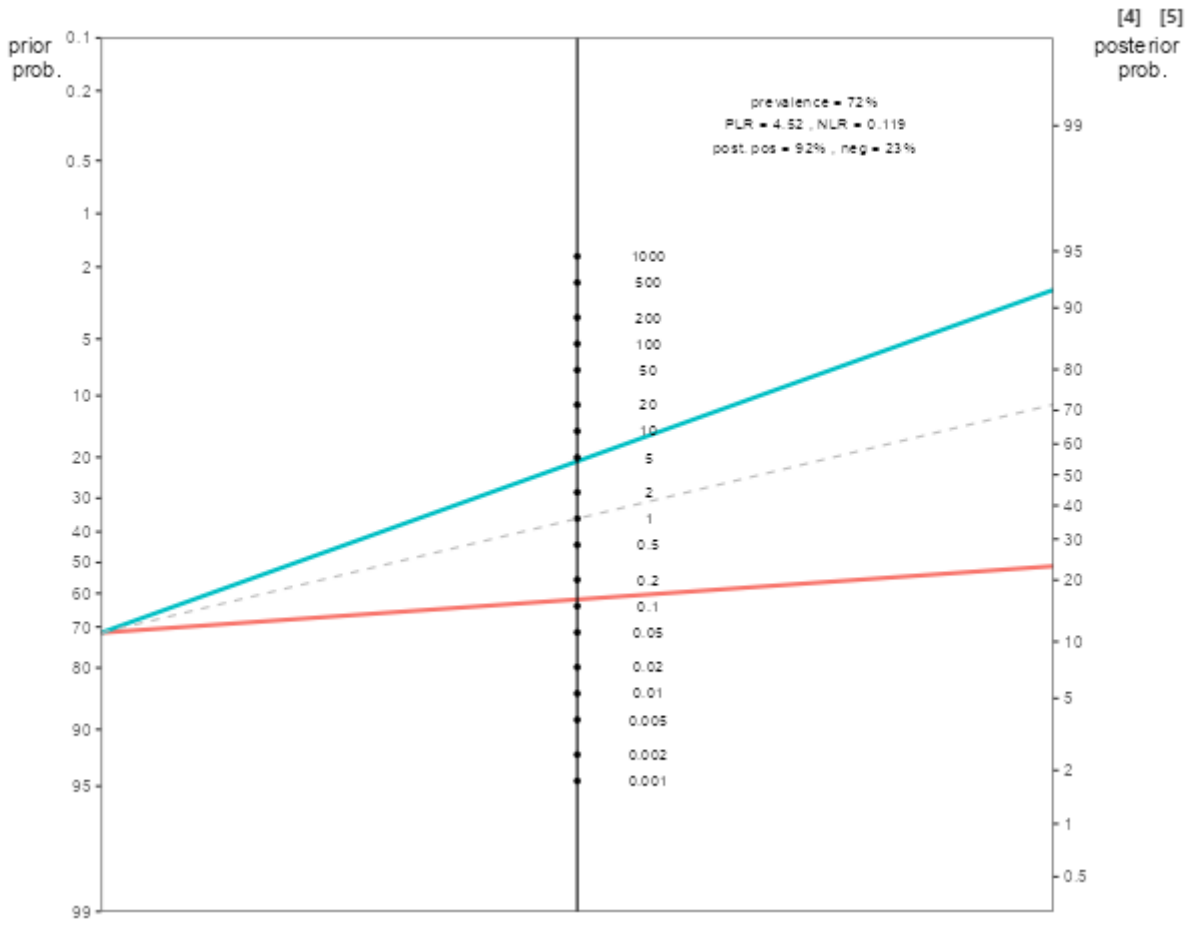
[h] Total Number of Wrong Test Results

		95% Confidence Interval	
Decision Statistics	**Estimate**	**Lower**	**Upper**
Apparent prevalence	70.5%	59.8%	79.7%
True prevalence	71.6%	61.0%	80.7%
Test sensitivity	90.5%	80.4%	96.4%
Test specificity	80.0%	59.3%	93.2%
Diagnostic accuracy	87.5%	78.7%	93.6%
Positive predictive value	91.9%	82.2%	97.3%
Negative predictive value	76.9%	56.4%	91.0%
Proportion of false positives	20.0%	6.8%	40.7%
Proportion of false negative	9.5%	3.6%	19.6%
False Discovery Rate	8.1%	2.7%	17.8%
False Omission Rate	23.1%	9.0%	43.6%

Decision Statistics	Estimate	95% Confidence Interval	
		Lower	Upper
Diagnostic odds ratio	38.000	10.4445	138.255
Number needed to diagnose	1.419	1.1162	2.518
Youden's index	0.705	0.3971	0.896
Likelihood ratio of a positive test	4.524	2.0571	9.949
Likelihood ratio of a negative test	0.119	0.0542	0.261

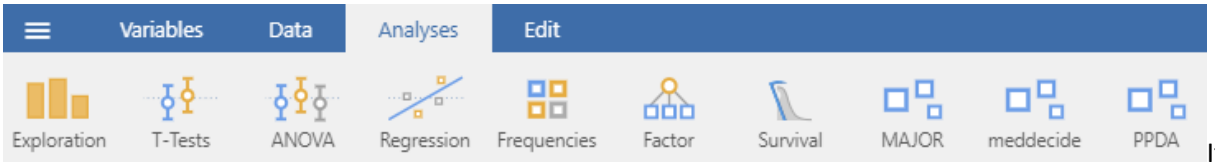
Below there is the **Fagan nomogram,** based on the prevalence observed in the sample, imagining a positive test result, and another negative.

Fagan nomogram

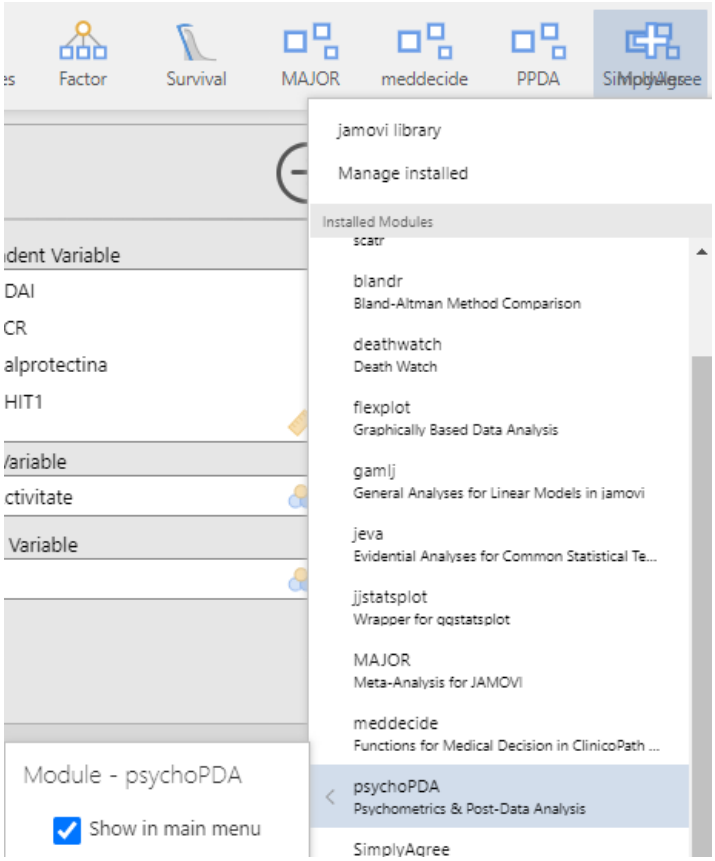


ROC analysis

In the table **Analyzes** , press the **PPDA button** .



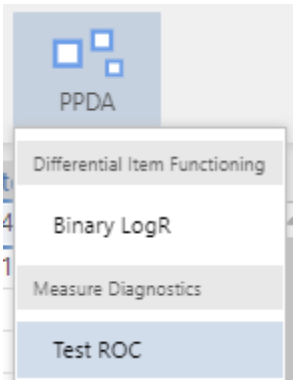
If you can't find it, it may not be displayed. In the table **Analyzes** , on the right, tap on + **Modules** and check if it is not present in the list of installed modules.



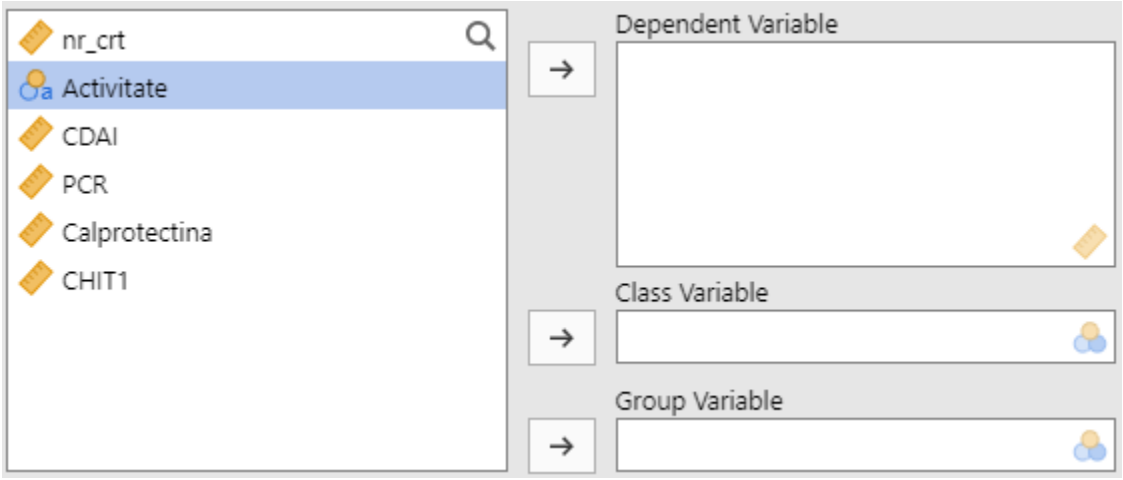
You can check that module to be visible in the table **Analyzes** by clicking the **Show in main option menus** .

If it is not installed, follow the steps indicated in the chapter **Installing the additional analysis module** .

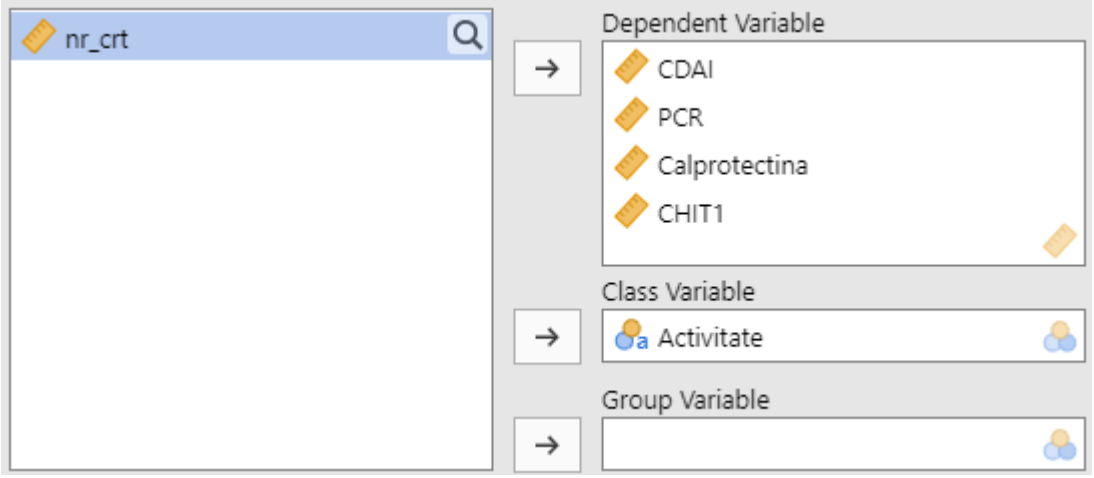
Clicking on the PPDA module, select the ROC Test option from the menu



Select the variable that represents **the standard test** (eg Activity), and press the **arrow button** next to the **Class field variables** . Select the quantitative **variables** that represent **the tests of interest** and press the arrow button next to the **Dependent variable field** .



In the end you will get something similar to the image below:



Tables of limit values and associated statistics

The program has already done the ROC analysis and presents you on the right side tables for different cutoff values (**Cutpoint**), with sensitivity (**Se**), specificity (**Sp**), positive predictive value (**PPV**) and negative (**NPV**), Youden 's index (**Youden's index**) , the area under the ROC curve (**AUC**).

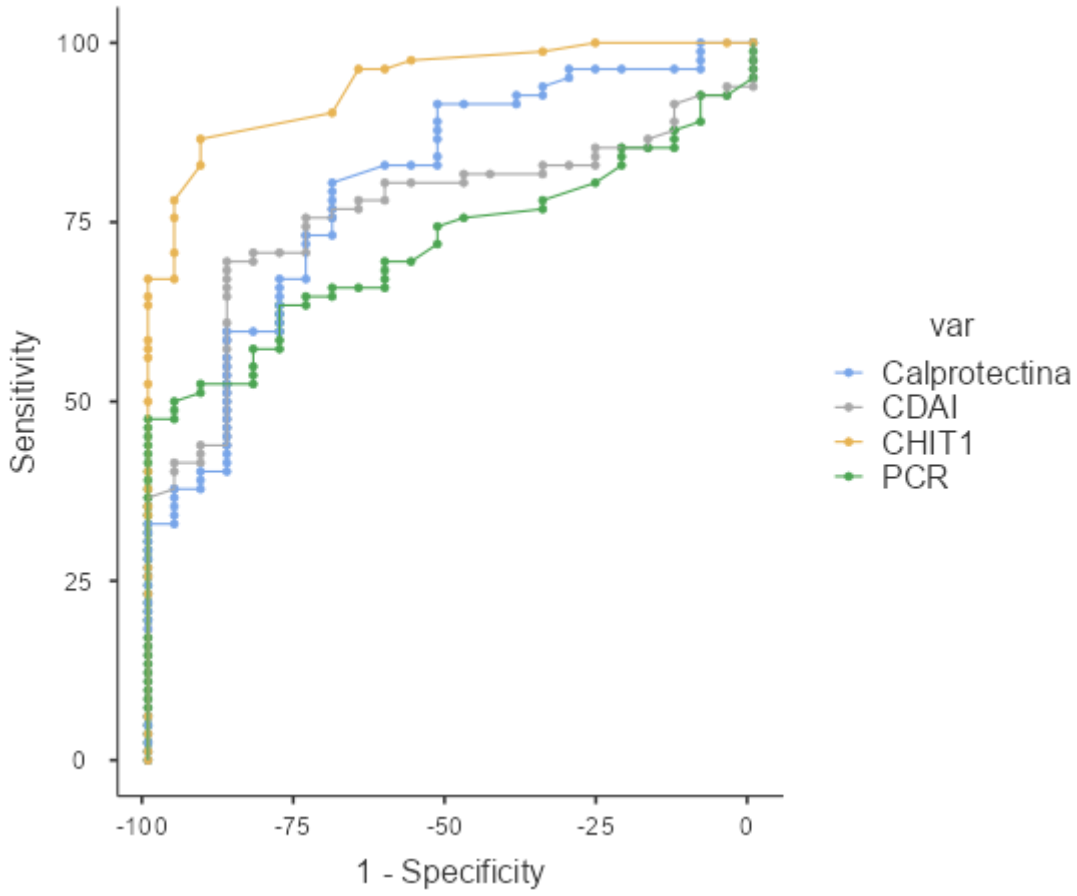
Thus **the best limit value** , which has the Youden index has the highest value, is 210, having the associated sensitivity and specificity of 80.49% and 69.57%, respectively.

Scale: Calprotectina

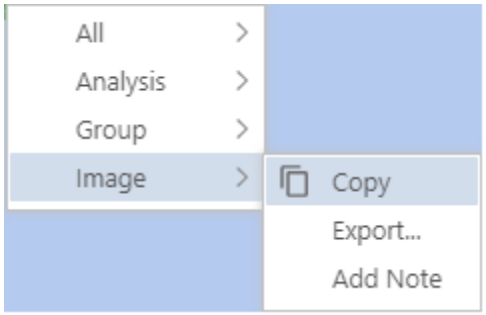
Cutpoint	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Youden's index	AUC	Metric Score
210	80.49%	69.57%	90.41%	50%	0.501	0.801	1.50
215	79.27%	69.57%	90.28%	48.48%	0.488	0.801	1.49
220	78.05%	69.57%	90.14%	47.06%	0.476	0.801	1.48
221.6	76.83%	69.57%	90%	45.71%	0.464	0.801	1.46
240	75.61%	69.57%	89.86%	44.44%	0.452	0.801	1.45
280	73.17%	73.91%	90.91%	43.59%	0.471	0.801	1.47
300	71.95%	73.91%	90.77%	42.5%	0.459	0.801	1.46
400	67.07%	78.26%	91.67%	40%	0.453	0.801	1.45
596	59.76%	86.96%	94.23%	37.74%	0.467	0.801	1.47
600	58.54%	86.96%	94.12%	37.04%	0.455	0.801	1.45

ROC curve graph

Below is the graph with the ROC curve (Receiver operating characteristic - the operating characteristic of the receiver).

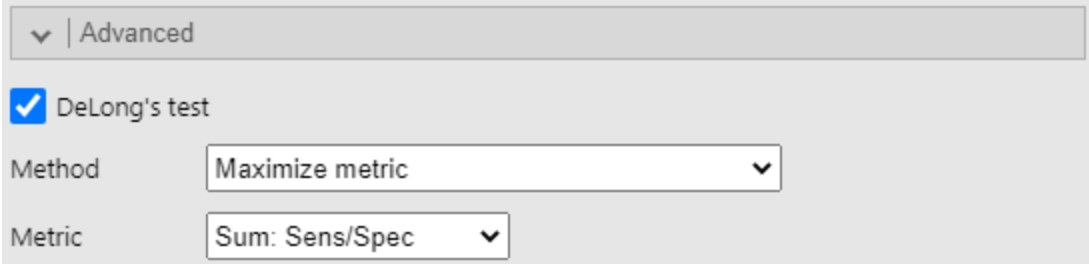


To copy the image press the right mouse button and select **Image Copy**



Comparison of ROC curves by statistical tests

To perform **statistical tests that compare the ROC curves** with each other, press the **> button** on the right **Advanced** , and select the **De Long's test option** .



The results look like the following:

DeLong Test of Difference between AUCs

Estimated AUC's:

	AUC	SD(Hanley)	P(H0: AUC=0.5)	SD(DeLong)	P(H0: AUC=0.5)
1	0.762	0.050	0.000	0.048	0.000
2	0.717	0.055	0.000	0.049	0.000
3	0.801	0.045	0.000	0.050	0.000
4	0.942	0.022	0.000	0.022	0.000

Pairwise comparisons:

	AUC Difference	CI(lower)	CI(upper)	P.Value	Correlation
1 vs. 2	0.045	-0.059	0.149	0.400	0.400
1 vs. 3	-0.039	-0.145	0.066	0.465	0.409
1 vs. 4	-0.179	-0.266	-0.092	0.000	0.406
2 vs. 3	-0.084	-0.187	0.019	0.110	0.437
2 vs. 4	-0.224	-0.318	-0.131	0.000	0.268
3 vs. 4	-0.140	-0.228	-0.052	0.002	0.455

Overall test:
p-value = 1.28e-06

The first table lists the estimates of **the areas under the ROC curve** for each diagnostic test, the standard deviation, as well as a **statistical test of significance for one ROC curve**, for each selected variable in the order of their selection. Thus for the CDAI clinical activity score, the area under the ROC curve is 0.717, and the result is statistically significant P (H0: AUC=0.5) being less than 0.05. (valerur du p<0.001). **The null hypothesis** of the test for an **ROC curve** is that the area under the ROC curve for CDAI, compared to histopathological examination, is not statistically significantly different from 0.5. **The alternative hypothesis** of the test is that the area under the ROC curve for CDAI compared to histopathological examination is statistically significantly different from 0.5.

In the second table are comparisons between tests taken two by two, presenting the difference between the surfaces of the ROC curves (AUC Difference), with 95% confidence interval, statistical test value (**P.Value**). For example, comparing CDAI clinical activity score (1) with C-reactive protein (2), the result is not statistically significant (p=0.400). **Null hypothesis** of **the two-ROC curve comparison** test**:** There is no statistically significant difference between the diagnostic accuracy of the CDAI clinical activity score and C-reactive protein, as measured by AUC, with histopathological examination as standard. **Alternative hypothesis** of the two-ROC curve comparison test: There is a statistically significant difference between the diagnostic accuracy of the CDAI clinical activity score and C-reactive protein, measured by AUC, with histopathological examination as standard.

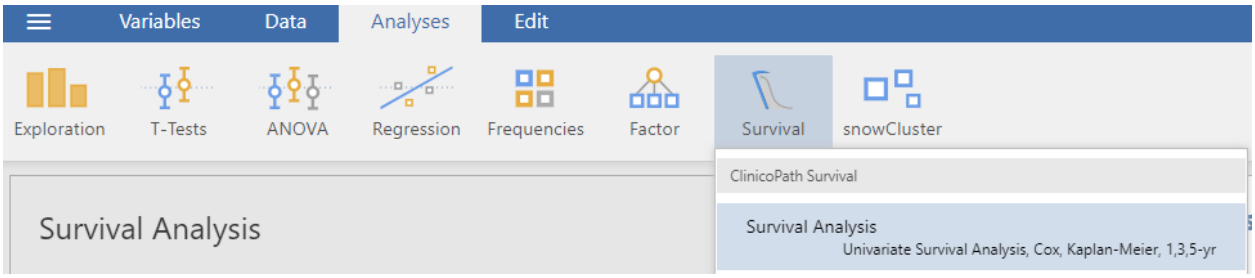
Finally, the result of a global statistical test comparing all ROC curves with each other is presented.

Survival analysis (Kaplan Meyer plot, log-rank test, Cox – HR regresion)

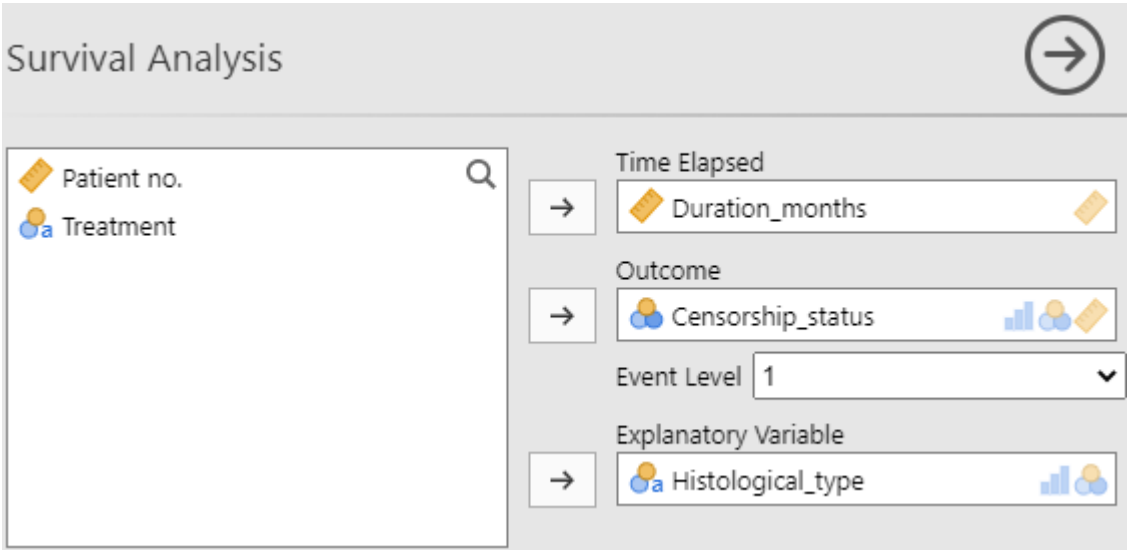
Make sure that the **jsurvival - Survival Module of ClinicoPath** for Jamovi 0.0.2 is installed (e.g. by verifying if the Survival button is present in the Analyses tab). See the chapter **Installation of additional analysis module (e.g. for ROC analysis, survival, diagnosis)**.

Check that the type of variables is well chosen in the **Data** tab, and that the order of the categories for each variable is as you prefer.

In the **Analyses** tab, clik **Survival** button, and choose in the menu the option **Survival Analysis (Univariate Survival Analysis, Cox, Kaplan-Meier, 1,3,5-yr).**

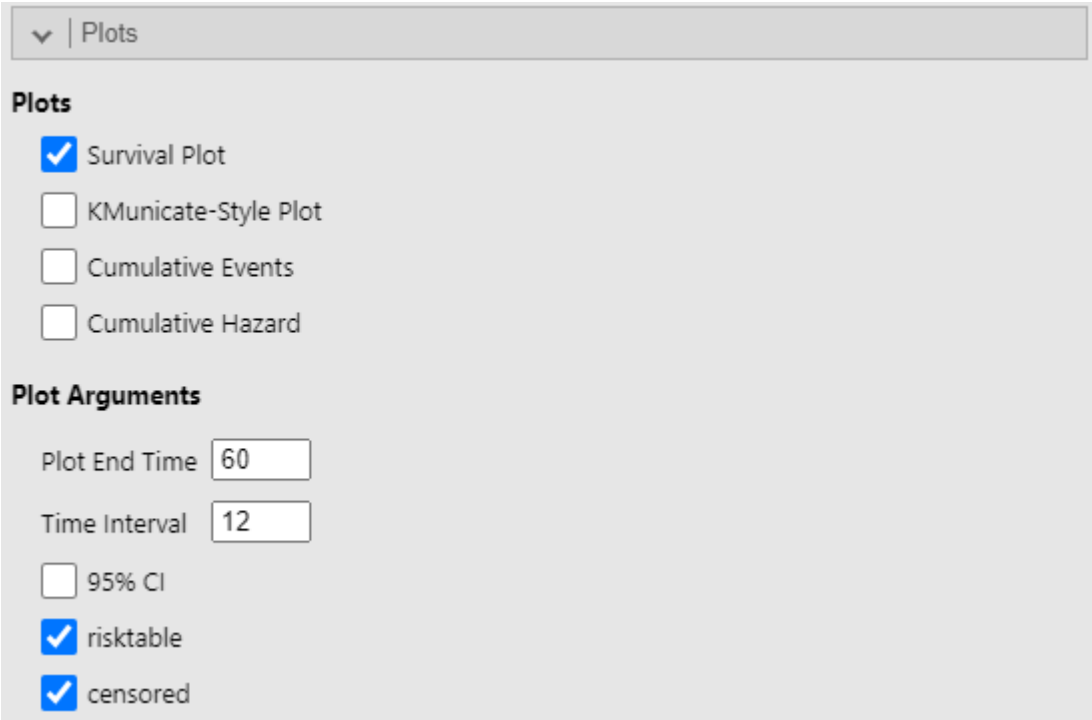


Move the variable that indicates the **survival time** with a click on the arrow next to the **Time Elapsed** text box. Move the variable that indicates the **censorship status/event** with a click on the arrow next to the **Outcome** text box. Make sure that the **category** that indicates **the event** is selected from the **Event Level** list (e.g. 1). Move the grouping variable with a click on the arrow next to the **Explanatory Variable** field (e.g. Type, Treatment, Risk Factor).

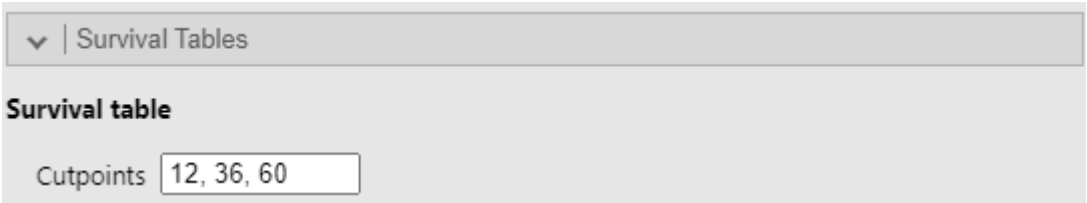


In the **Plots** section, check the **Survival Plot** option.

You can change the options of the graph: to show the p-value, the table with the number of patients at risk, the cases with censored observations. Please check **risktable** to make sure the legend is visible.



To find out the probability of survival at different times, you can enter these times in the **Survival table** section, in the **Cutpoints** field.



The median survival time

In the **Median Survival Summary and Table** section, the medians of survival time with 95% confidence intervals can be found in the text before the table. Also, in the table, in the **Median** column we find the median by group, and the 95% confidence interval in the columns **Lower** and **Upper 95% Confidence interval.** For example, for the group with small cell histology, the median survival is 2 months.

Median Survival Summary and Table - Histological_type

When Histological_type is small cell, median survival is 2 [1 - 2, 95% CI] months.
When Histological_type is squamous cell, median survival is 4 [3 - 10, 95% CI] months.

Median Survival Table: Levels for Histological_type

						95% Confidence Interval	
Levels	Records	Events	rmean	se_rmean	Median	Lower	Upper
Histological_type=small cell	48	45	2.72	0.497	2.00	1.00	2.00
Histological_type=squamous cell	35	31	7.76	1.601	4.00	3.00	10.00

Cox Regression– hazard ratio

In the table with the results of the Cox regression (**Cox Regression Summary and Table**), you can find the hazard ratio with the 95% confidence interval, in the **HR (Univariable) column**. To find out which group is **the reference group**, in the **HR (Univariable)** column you look for the group that shows the - **sign**. In the example below, it can be seen that the reference group is the histopathological small cell type. So, HR is obtained by dividing the hazard of participants with cancer of squamous histology, to the hazard of participants with cancer of small cell histology.

Cox Regression Summary and Table - Histological_type

When Histological_type is squamous cell, there is 0.43 (0.26-0.72, p=0.001) times risk than when Histological_type is small cell.

Cox Table- Histological_type

Explanatory	Levels	all	HR (Univariable)
Histological_type	small cell	48 (57.8)	-
	squamous cell	35 (42.2)	0.43 (0.26-0.72, p=0.001)

The survival probability at different times

The probability of survival at different points in time (e.g. in the **Time column**, at 12 months) can be found in the **Survival Summary and Table**, in the **Survival column**, and the 95% confidence interval in the **Lower, and Upper 95% confidence interval columns**), but also in text format, before the table. For example, the group with small cell histology has a probability of survival at 12 months of 5.9%.

1, 3, 5-yr Survival Summary and Table - Histological_type

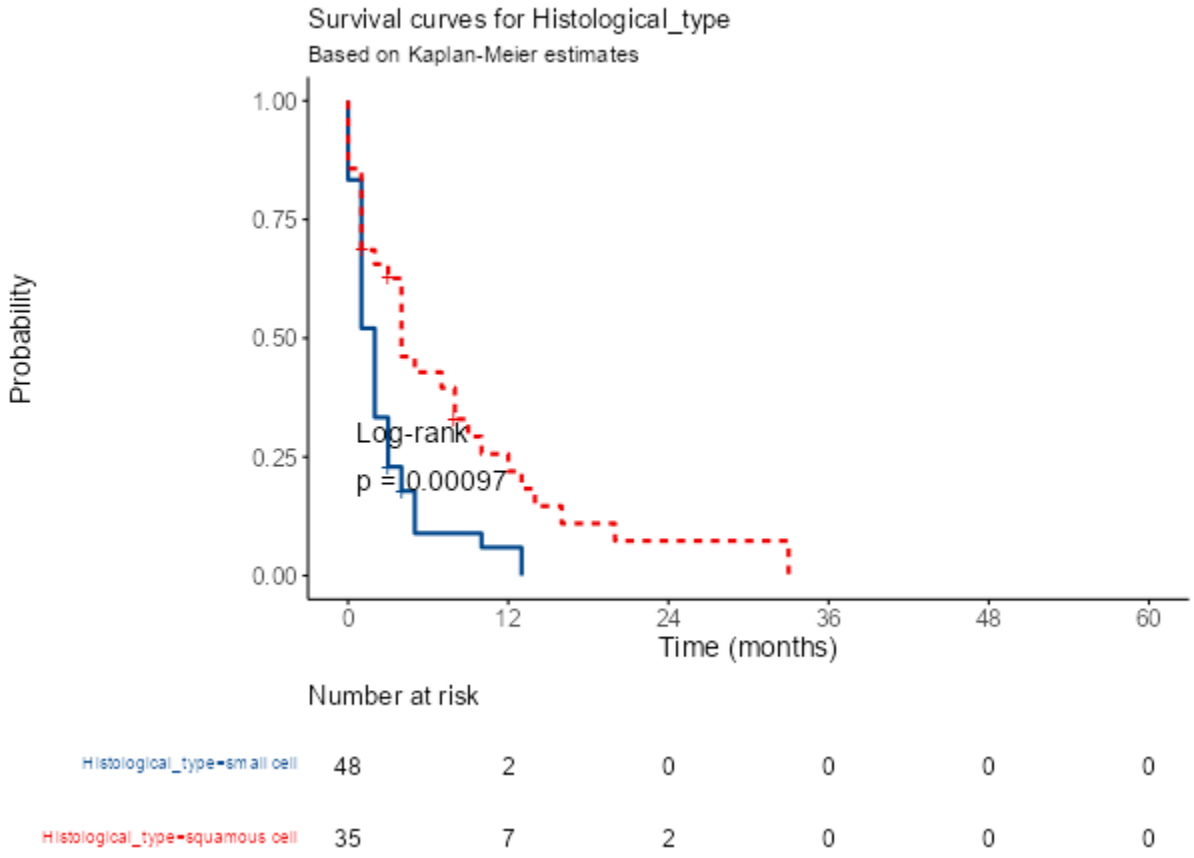
When Histological_type=small cell, 12 month survival is 6% [1.6%-22%, 95% CI].
When Histological_type=squamous cell, 12 month survival is 22% [11.1%-43%, 95% CI].

1, 3, 5 year Survival - Histological_type

					95% Confidence Interval	
Levels	time	Number at Risk	Number of Events	Survival	Lower	Upper
Histological_type=small cell	12	2	43	5.9 %	1.6 %	21.6 %
Histological_type=squamous cell	12	7	25	22.0 %	11.1 %	43.3 %

The Kaplan-Meier survival curve, with the log-rank test

The Kaplan-Meier survival curve, with the log-rank p-value, and the table with the number of subjects at risk, can be found in the Survival **Plot** section. In the table with the number of subjects at risk, at the beginning of the study there are 48 living subjects (at risk of having the event – death) in the group with small cell histology, and 35 subjects in the group with squamous histology. At 12 months there are only 2 living subjects (at risk of the event – death) left in the group with small cell histology, and 7 subjects in the group with squamous histology.



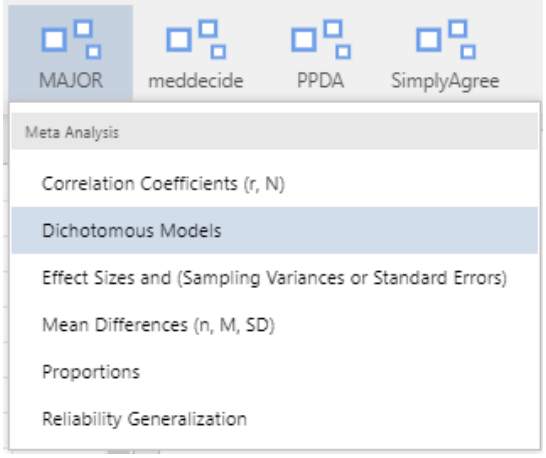
Meta-analysis

Meta-analysis to compare two groups in terms of dichotomous data

We make sure that **the MAJOR module** is **installed**. If it is not installed, follow the instructions in the **Installation of additional analysis modules** chapter.

We make sure that **the type of variables** representing the number of events, respectively the total number of subjects, is of the **Continuous type**.

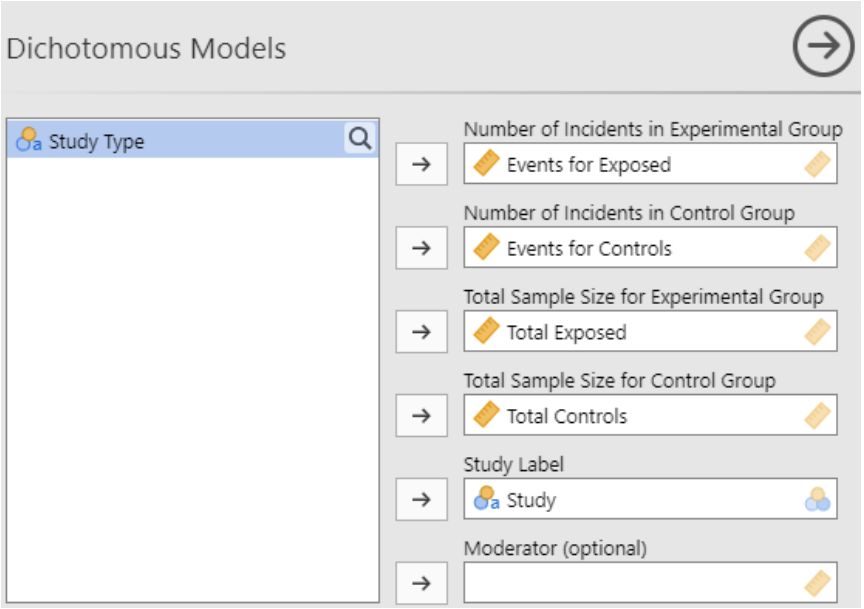
From the MAJOR module, **we choose** the **Dichotomous Models** option.

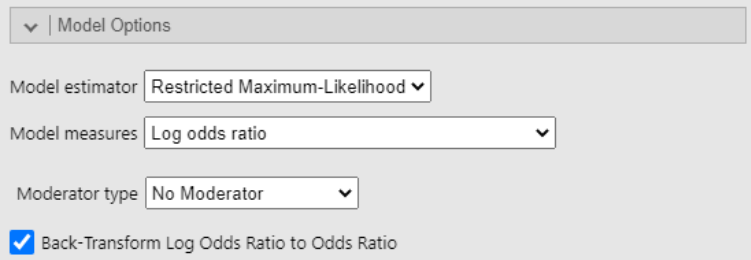


Move the variable indicating the number of events in the experimental/intervention/exposed group to the **Number of Incidents in Experimental Group** field ; the variable indicating the number of events in the control/non-exposed group to the **Number of Incidents in Control Group** field; the variable indicating the total number of participants in the experimental/intervention/exposed group in the **Total Sample Size for Experimental Group** field; the variable indicating the total number of participants in the control/non-exposed group in the **Total Sample Size for Control Group** field; the variable indicating the name of the study in the **Study Label** field.

In the **Model Options** section, choose for **Model Estimator**, **Restricted Maximum-Likelihood**; for **Model Measures**, Log odds ratio (for case-control studies), or risk difference (for cohort studies); check **Back-Transform Log Odds Ratio to Odds Ratio**.

The rest of the options can be left as they are. If desired, it can be reduced in case of reduced heterogeneity to the Fail-Safe N Rosenberg method.





Association estimator

If Log odds ratio has been selected as the estimator of the association, as well as if **Back-Transform Log Odds Ratio to Odds Ratio** has been checked, then in the table with the same name, the value of the OR point estimator of the can be found, as well as the 95% confidence interval.

Back-Transform Log Odds Ratio to Odds Ratio

Odds Ratio	CI Lower Bound	CI Upper Bound
1.542	1.135	2.096

In **the Random-Effects Model table** , the estimator value of the log odds ratio/log risk ratio/risk difference association is reported in the **Estimate** column with the confidence interval in the last two columns. The **p-value** for its statistical significance is in the **p**-column.

Random-Effects Model (k = 10)

	Estimate	se	Z	p	CI Lower Bound	CI Upper Bound
Intercept	0.433	0.157	2.77	0.006	0.126	0.740
	.	.	.	.	.	.

Note. Tau^2 Estimator: Restricted Maximum-Likelihood

Statistical heterogeneity of studies

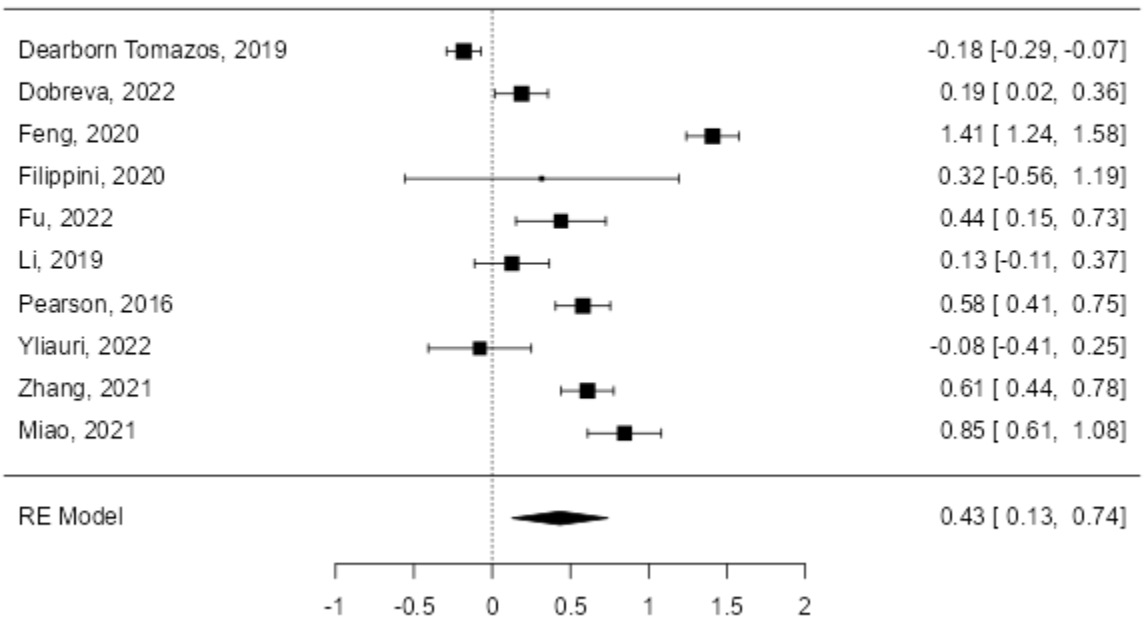
In the **Heterogeneity Statistics** table, in the column **I^2** there is the inconsistency index, and in column **p**, the p-value of the associated test.

Heterogeneity Statistics

Tau	Tau^2	I^2	H^2	R^2	df	Q	p
0.471	0.222 (SE= 0.1147)	95.71%	23.304	.	9.000	288.885	< .001

Forest plot

In the **Forest plot** section, you have the forest plot. Unfortunately, for the estimators of the log odds ratio, log risk ratio, the logarithm values of the OR or RR are written in the graph. The advantage is the symmetry of the confidence intervals. The downside is that it is more difficult to interpret. The value of 0 is equivalent to an OR or RR of 1. Values on the graph greater than 0 for OR/RR are greater than 1, and values less than 0 are less than 1.



Publication bias

The publication bias can be assessed with formal statistical tests and with plots, like the funnel plot.

Statistics

To evaluate the publication bias, in the **Publication Bias Assessment** section, one can use the result p-value (column p) of the statistical test for funnel plot asymmetry (**Regression Test for Funnel Plot Asymetry**).

Publication Bias Assessment

Fail-Safe N Analysis (File Drawer Analysis)

Fail-safe N	p
593.000	< .001

Note. Fail-safe N Calculation Using the Rosenthal Approach

Rank Correlation Test for Funnel Plot Asymmetry

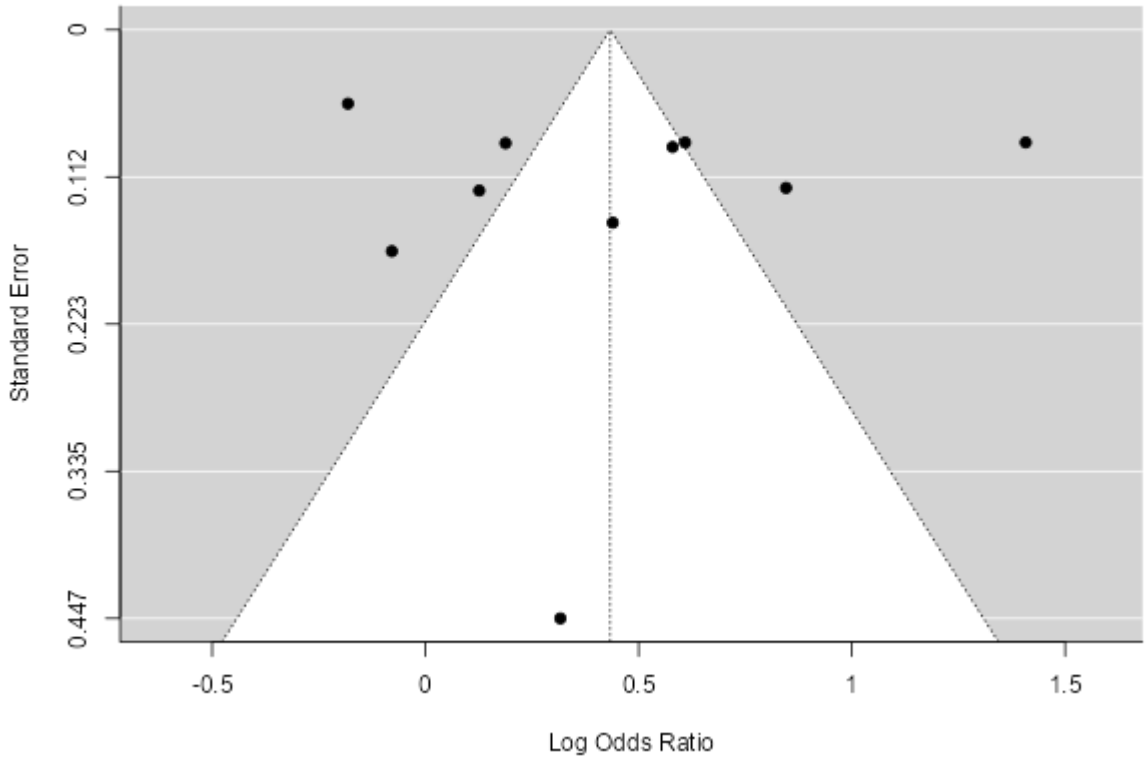
Kendall's Tau	p
−0.156	0.601

Regression Test for Funnel Plot Asymmetry

Z	p
−0.306	0.760

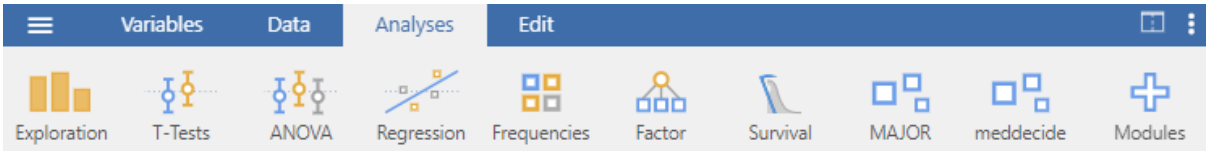
Funnel plot

The funnel plot can be found in the **Funnel plot section**.

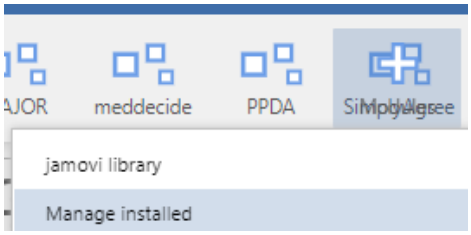


Installation of additional analysis modules (e.g. for ROC analysis, survival, diagnosis)

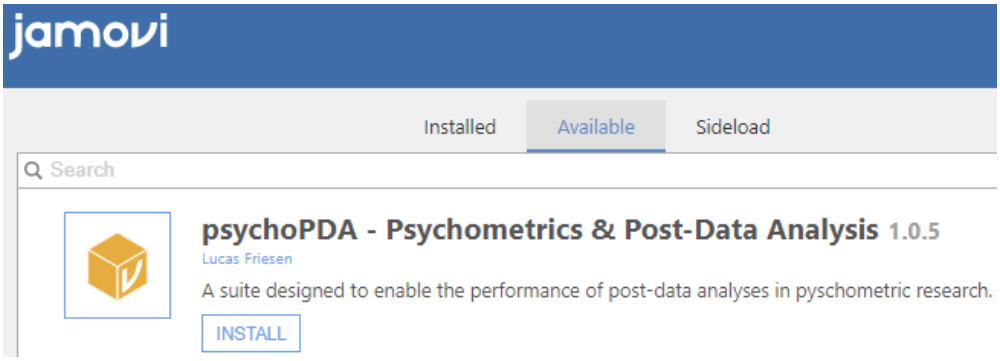
In the table Analyzes , on the right, tap on + Modules and check if it is not present in the list of installed modules.



To install analysis modules in addition to the default ones, select tab **Analyzes** , and on the right side press the + **Modules button** , then select **Manage installed** .



Select tab **Available** , scroll until you find the desired module (eg psychoPDA) and press the **INSTALL button** .



Return to the analysis window by pressing the arrow button: 

Instructions for using EpiInfo and Excel for data analysis

Downloading *EpiInfo* (for home use)

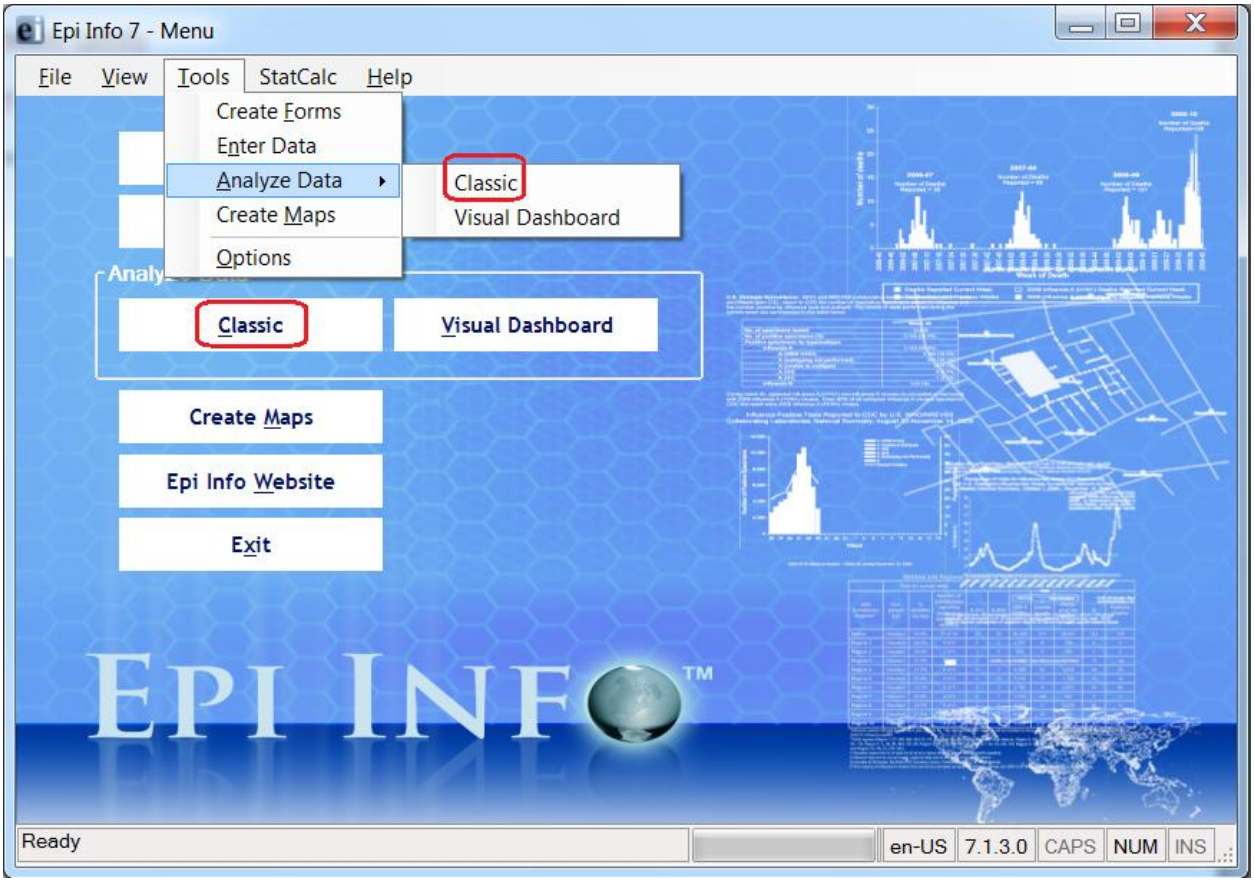
Use an Internet browser to navigate to: http://www.cdc.gov/epiinfo/installation.htm and follow the installation instructions.

Starting Epi Info

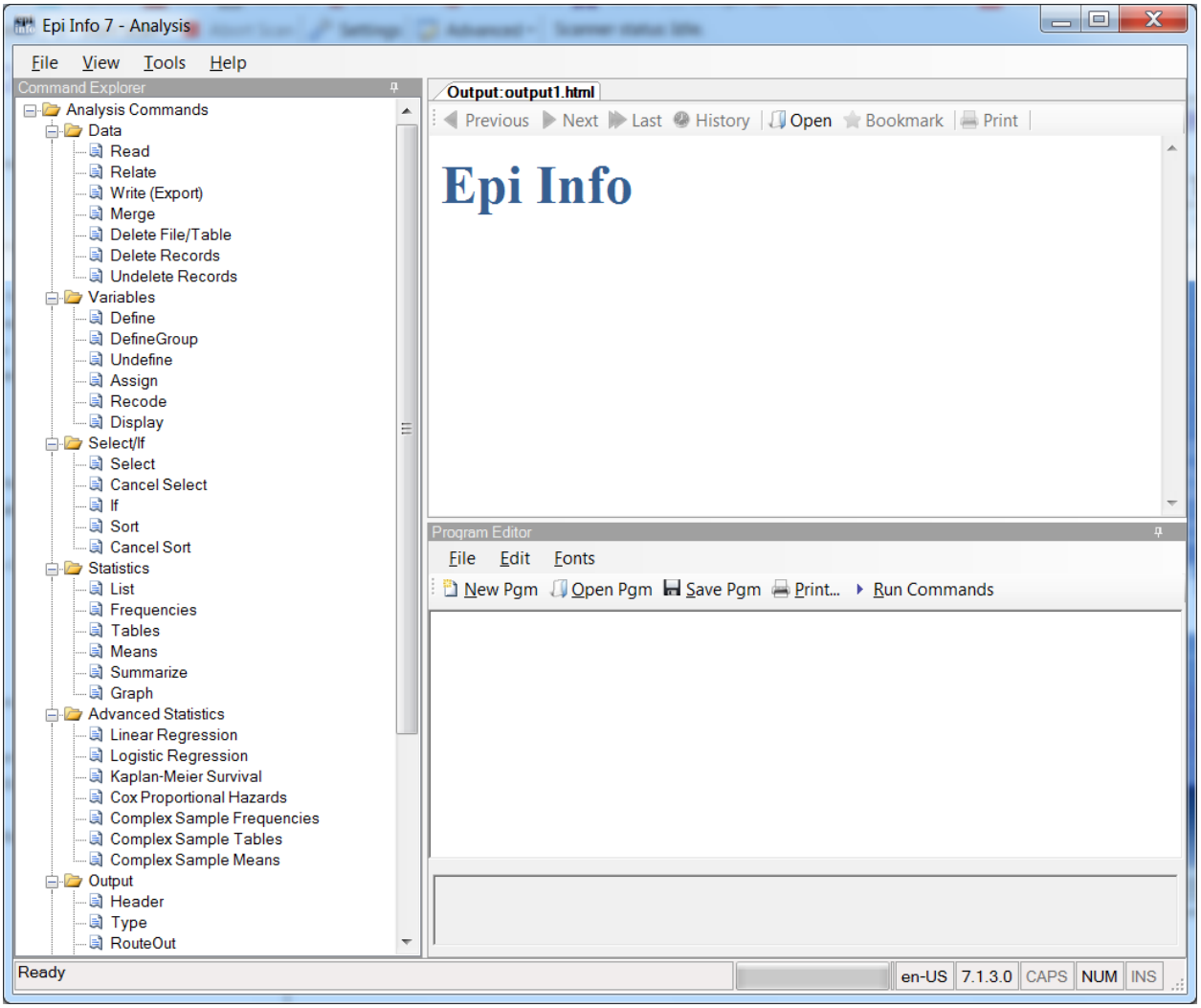
Start button, shortcut **Epi Info™ 7** or **Start button**, **All Programs**, **CDC**, **Epi Info 7**, **Epi Info™ 7**

Analyzing data in Epi Info

Click on **Classic** in the **Analyze Data section** or using the menu.

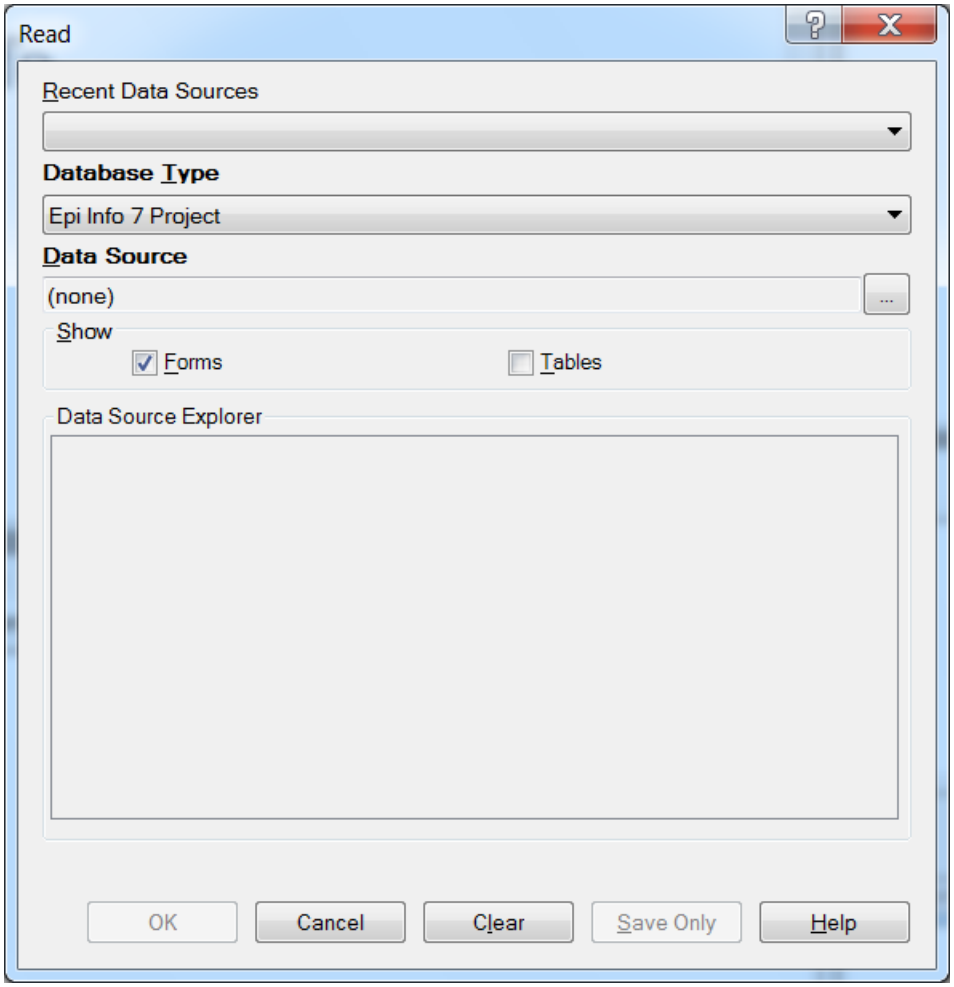


The data analysis window:

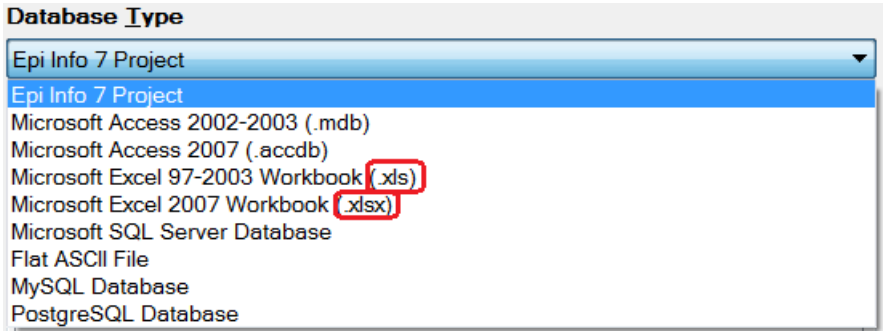


Importing an Excel file in Epi Info

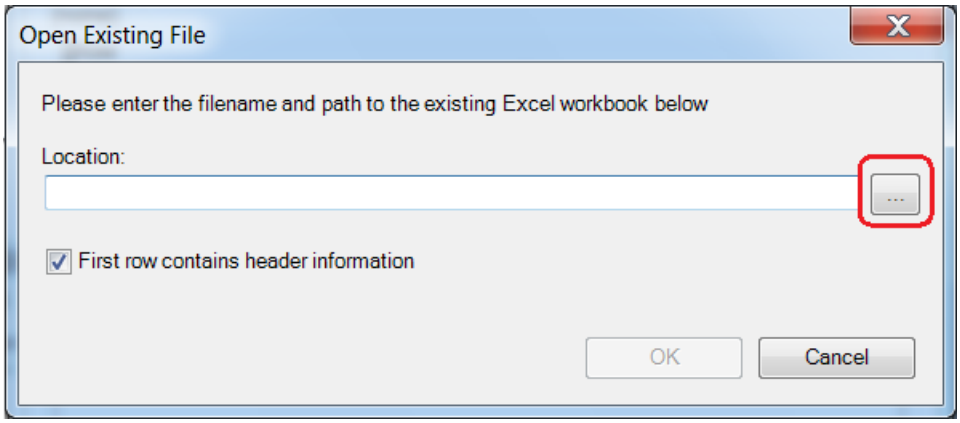
In the commands section - **Command Explorer –** left hand side of the window, **Analysis Commands**, in the section : **Data** – click on **Read**. The window for importing data will open. Select the type of file to import.



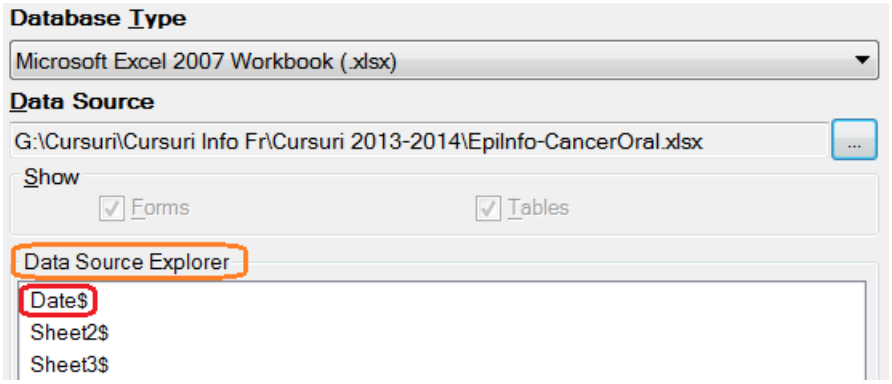
If the **Excel** file has the **.xlsx**, extension we choose **Microsoft Excel 2007 Workbook (.xlsx)**, if the file has the **.xls** extension, we choose **Microsoft Excel 97-2003 Workbook (.xls)**.



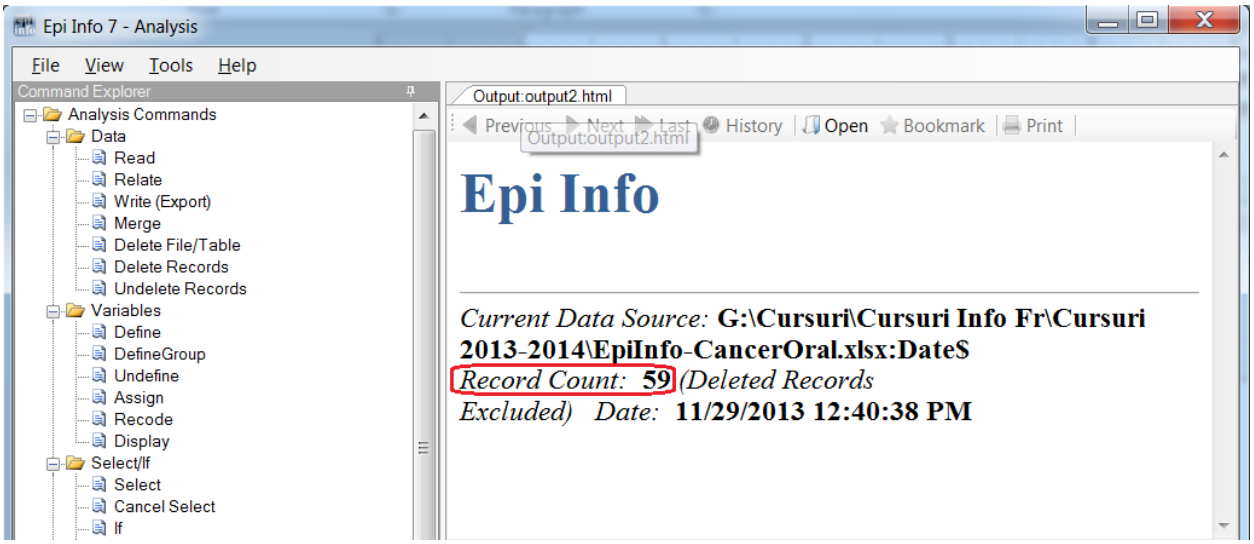
After this we search for the file to import after clicking on the button  corresponding to **Data Source**. In the window that will open we click the button  of the **Location** field, and we browse for the Excel file. We select the file and click **OK**, letting checked the option : **First row contains header information**.



After this EpiInfo shows all the worksheets in the file, in the **Data Source Explorer** section. Select the worksheet that contains the data and click the OK button to import it.



After importing Epi Info shows in the results section (**Output**), the imported file name and the number of records in the file (**Record Count**).



Activating the *Data Analysis* module in *Microsoft Excel*

Click on a void cell, then click on ***Add-Ins*** in the ***Tools*** menu. Check the box next to ***Analysis ToolPack*** and then click OK. Select another empty cell, then search for ***Data Analysis*** in the ***Tools*** menu.

If ***Data Analysis*** does not appear in the ***Tools*** menu, despite being checked in ***Add-Ins***, uncheck the box in ***Add-Ins*** and repeat the above procedure.

Descriptive statistics

Qualitative (categorical) data:

Frequency tables

Use the ***COUNTIF*** function in ***Microsoft Excel*** to count how many times each value taken by a variable appears in the database (its absolute frequency).

E.g. to find out how many female persons (coded with F in the file) are in the sample, a void cell at the future location of the frequency table should contain a similar formula to =COUNTIF(A2:A58, "F"), if data regarding gender was recorded in cells A2 to A58. A correct frequency table should look like this:

Table 1. Gender distribution in the studied sample

Gender	**Number of subjects**
Male	20
Female	37
Total	57

Note that any table has to be labeled on top of it, using a clear and precise title. Select the table, right click on this selection and choose ***Caption*** in order to label the table. Row and column labels should be visible and easily understandable by the reader, with no need to search for further explanations in order to understand the content of the table.

Pie charts

Follow the instructions above to create a frequency table using ***COUNTIF***.

Select only the cells containing the absolute frequencies and their labels (do not select the total or column labels). Use ***Insert - Graph*** and select ***Pie***. Click ***Next***. In the ***Chart Options*** window click on the tab ***Data Labels*** and tick ***Percentage***. Continue and finish the chart wizard. *A correct pie chart should look like this:*

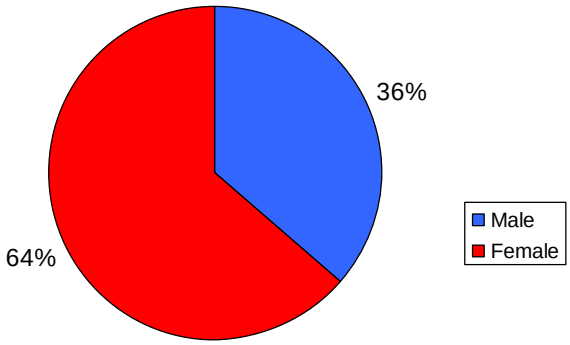


Figure 1. Gender distribution in the studied sample

Note that pie charts have to be labeled using visible percentages.

If you plan to use the chart in a ***PowerPoint*** presentation, make sure to label it on top, using a clear and precise chart title (what, how and for which subjects has been represented?).

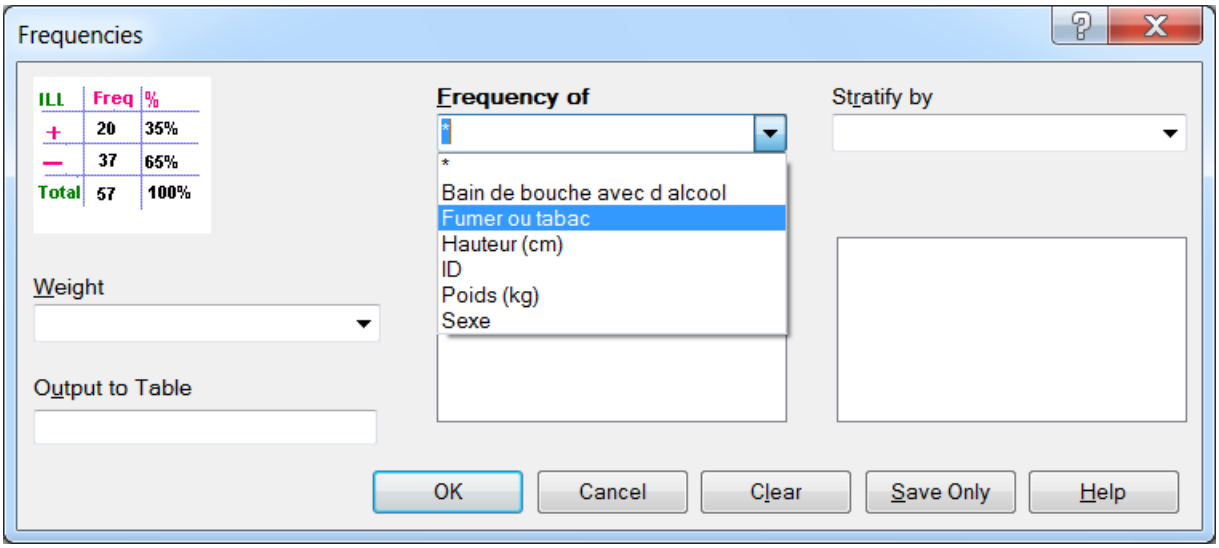
If you plan to use the chart as a figure in a ***Word*** document, erase the chart title in ***Excel*** but remember to label the chart in ***Microsoft Word***: select the figure, right click on this selection and use ***Caption***.

All labels and legend entries should be visible and easily understandable by the reader, with no need to search for further explanations in order to understand the content of the figure.

Frequency tables in Epi Info

In Command Explorer section **Statistics** we choose **Frequencies** .

In the new window we select the variable of interest in the list of **Frequency of** and we click the boutton **OK**.



In the Output we get the frequency table and the corresponding confidence intervals.

FREQ Gen

GEN	Frequency	Percent	Cum. Percent	
F	14	23.73%	23.73%	
M	45	76.27%	100.00%	
Total	59	100.00%	100.00%	

95% Conf Limits

F 13.62% 36.59%

M 63.41% 86.38%

Contingency tables

In ***Microsoft Excel***, select any cell containing data. Then, click in the menu bar ***Data*** – ***Pivot Table*** – ***Pivot Chart Report***. Work your way through the wizard and obtain a new worksheet containing an empty pivot table and a field list.

Drag and drop the field representing a prognostic factor (the risk factor, the new diagnostic test or the new treatment, depending on the given research scenario) to the area labeled ***Drop Row Fields Here.*** Drag and drop the field representing an outcome (the disease, the reference diagnostic test or the treatment response, depending on the given research scenario) to the area labeled ***Drop Column Fields Here***. Finally, drag and drop any of the formerly used fields to the area labeled ***Drop Data Fields Here***.

Rename row and column labels so that they are easily understandable by the reader, with no need to search for further explanations in order to understand the content of the table (*e.g. If male gender was coded as m, rename the corresponding row label: male*)

Right click on a row label and select order, to correct the row order in your contingency table. Right click on a column label and select order, to correct the column order in your contingency table.

After inserting the contingency table into your ***Word*** document, remember to label it using ***Caption*** and a correct title.

The column chart associated to a contingency table

After creating a pivot contingency table, select ***Insert - Chart*** from the menu bar.

To hide the chart buttons right click the button ***Count of*** and select ***Hide Pivot Chart Field Buttons***.

To show frequency labels, right click the empty chart area towards the upper left corner, select ***Chart Options*** and, in the ***Data Labels*** tab, tick ***Percentage*** or ***Value***.

Then, switch to the ***Titles*** tab and define clear and precise titles for your chart axes, including the units of measurement between brackets, where necessary.

After inserting the chart into your ***Word*** document, remember to label it using ***Caption*** and a correct title.

Quantitative data:

Individual description of quantitative variables

Mean, median, standard deviation, 95% confidence interval for means

In ***Microsoft Excel***, use ***Tools*** – ***Data Analysis*** – ***Descriptive Statistics*** to simultaneously compute the most important descriptive parameters for selected quantitative variables.

In the ***Descriptive Statistics*** window tick options ***Summary Statistics*** and ***Confidence Level for Mean***.

To find the lower limit of the 95% confidence interval, compute ***Mean*** minus ***Confidence Level (95%)***.

To find the upper limit of the 95% confidence interval, compute ***Mean*** plus ***Confidence Level (95%)***.

Frequency table and Histogram

In ***Microsoft Excel***, use ***Tools*** – ***Data Analysis*** – ***Descriptive Statistics*** to compute minimum, maximum and range for the desired quantitative variable.

Choose a convenient bin size for the variable of interest (a round-figure for which 7-10 non-overlapping intervals of that similar size will cover the whole variable range).

Label a void column as "***Bin*** *Variable name (units of measurement)*" on the same worksheet as the variable of interest.

Below this label, insert the value for minimum+ the chosen bin size.

Use ***Edit – Fill – Series*** (select options: in Columns, Step value=bin size, Stop value=maximum-bin size) to complete the column containing the bin values for your variable.

Now use ***Tools – Data Analysis – Histogram***:

For ***Input Range*** select the range of cells containing the quantitative variable for which you want to plot a frequency table and histogram. For ***Bin Range*** select the newly created column. In both cases, include the column labels in your selection and tick ***Labels***.

In ***New Worksheet Ply*** write a suggestive name for the worksheet that will contain the frequency table and histogram for your variable.

In order to display the histogram you need to select ***Chart Output***.

After pressing the OK button, both frequency table and histogram will appear in a raw, unfinished form.

In order to be comprehensible, both the frequency table and the histogram need adjustments:

1. replace the upper limit of each bin shown in the brute table with the corresponding bin interval
2. delete the chart legend since its information is redundant and only takes up space
3. delete the title ***Histogram***, since you will label the figure in ***Microsoft Word***, using ***Caption***
4. resize the chart area as needed, in order to have a clear view over your histogram
5. eliminate spaces between columns by right clicking any of the columns and using ***Format Data Series – Options*** and adjusting the ***Gap Width***
6. verify the content and font size of all labels, to make sure that your histogram is easily understandable by anyone who reads your work.

A correct frequency table and a correct histogram should look like this:

Table 2. Weight distribution in the studied sample

Weight intervals (kg)	*No. of subjects*
<=40	21
(40-50]	144
(50-60]	297
(60-70]	240
(70-80]	145
(80-90]	79
(90-100]	45
>100	29

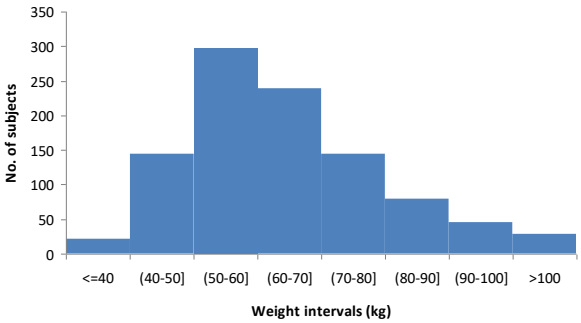


Figure 2. Histogram of weight in the studied sample

Description of a potential relation between two quantitative variables

Scatter chart

In ***Microsoft Excel***, select the columns containing the two quantitative variables, including their labels. Then select ***Insert – Chart*** and choose ***XY (Scatter)***.

Advance to step 3 of the chart wizard and define correct titles for both X and Y axes. Do not forget to specify after the title of each axis the corresponding units of measurement, between round brackets.

Then click on the ***Legend*** tab and uncheck the ***Show legend*** box, since no useful information derives from a legend when investigating only two variables at once.

In the ***Titles*** tab write the precise title of each axis, including the units of measurement in parentheses.

If you plan to use the chart in a ***PowerPoint*** presentation, make sure to label it on top, using a clear and precise chart title (what, how and for which subjects has been represented?).

If you plan to use the chart as a figure in a ***Word*** document, erase the chart title in ***Excel*** but remember to label the chart in ***Microsoft Word***: select the figure, right click on this selection and use ***Caption***.

All labels should be visible and easily understandable by the reader, with no need to search for further explanations in order to understand the content of the figure.

After finishing the chart wizard, right-click on any point from the data cloud and select ***Add Trendline***. The most common trend of data clouds is a linear one. In the ***Options*** tab check ***Display equation on chart*** and ***Display R-squared value on chart***.

To highlight the trendline using a contrasting color, right-click on the trendline and use ***Format Trendline***.

To highlight the trendline labels using a contrasting color, right-click on the label box and use ***Format Data Labels***.

A correct scatter chart should look like this:

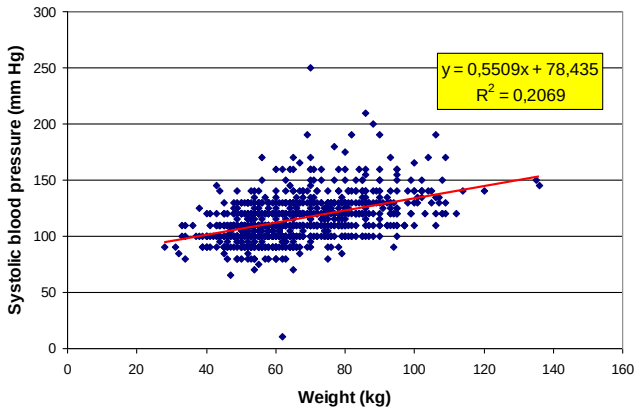


Figure 3. Relation between weight and systolic blood pressure for all subjects included in the studied sample

Survival data:

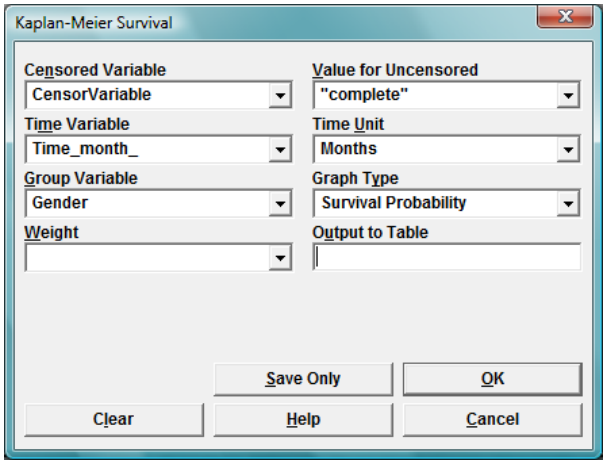
Median of survival time

In ***Microsoft Excel***, use ***Tools*** – ***Data Analysis*** – ***Descriptive Statistics***. In the ***Descriptive Statistics*** window check ***Summary Statistics***.

Survival probability chart

In the ***Analysis*** module of ***EpiInfo*** click ***Kaplan-Meyer Survival***, from the left panel.

Complete the dialog box as seen in the image below:



Change the ***Group Variable*** as needed for your comparison.

Data Analysis

Performing a Student test (t-test) in Excel

Before performing the test you need to sort your data according to the groups that you wish to compare. (e.g. if you wish to compare cholesterol values of males with cholesterol values of females, you need to sort your data by gender).

To sort your data, click on any cell inside your data range, then use ***Data*** – ***Sort***.

If the groups that you wish to compare are independent (e.g. comparing cholesterol values of women with those of men), use ***Tools*** – ***Data Analysis*** – ***t-Test: Two-Sample Assuming Unequal Variances.***

If the groups that you wish to compare are paired (e.g. comparing cholesterol values of the same subjects, before and after treatment), use ***Tools*** – ***Data Analysis*** – ***t-Test: Paired Two Sample for Means.***

In the test window, select for ***Variable 1 Range*** the cells containing the quantitative variable corresponding to the first group (e.g. initial cholesterol values for women) and for ***Variable 2 Range*** the cells containing the quantitative variable corresponding to the second group (e.g. initial cholesterol values for men), without selecting the column label. Pay attention not to select the grouping variable (e.g. *Gender*) instead of the corresponding quantitative variable that you wish to compare.

Since the null hypothesis (H_0) of your reasoning states the absence of difference between mean values of the compared variable, introduce 0 in the ***Hypothesized mean difference*** box.

Give a title to the future worksheet that will contain the test results, by entering a suggestive name in the ***New Worksheet Ply*** box.

Immediatly after pressing OK, rename the generic lables ***Variable 1*** and ***Variable 2*** using suggestive labels: include information regarding both the quantitative variable you have compared and the grouping variable that sets them apart. This will allow you to easily interpret your test results later on.

The two-tailed p-value rendered by the test shows the statistical significance of the investigated difference between the mean values of the compared groups.

If the rendered two-tailed p-value includes the letter E followed by a negative figure this means in fact a very low (i.e. significant) p-value (e.g. p = 3,22342E-6 = 3,22342 x 10^{-6} = 0,0000032234).

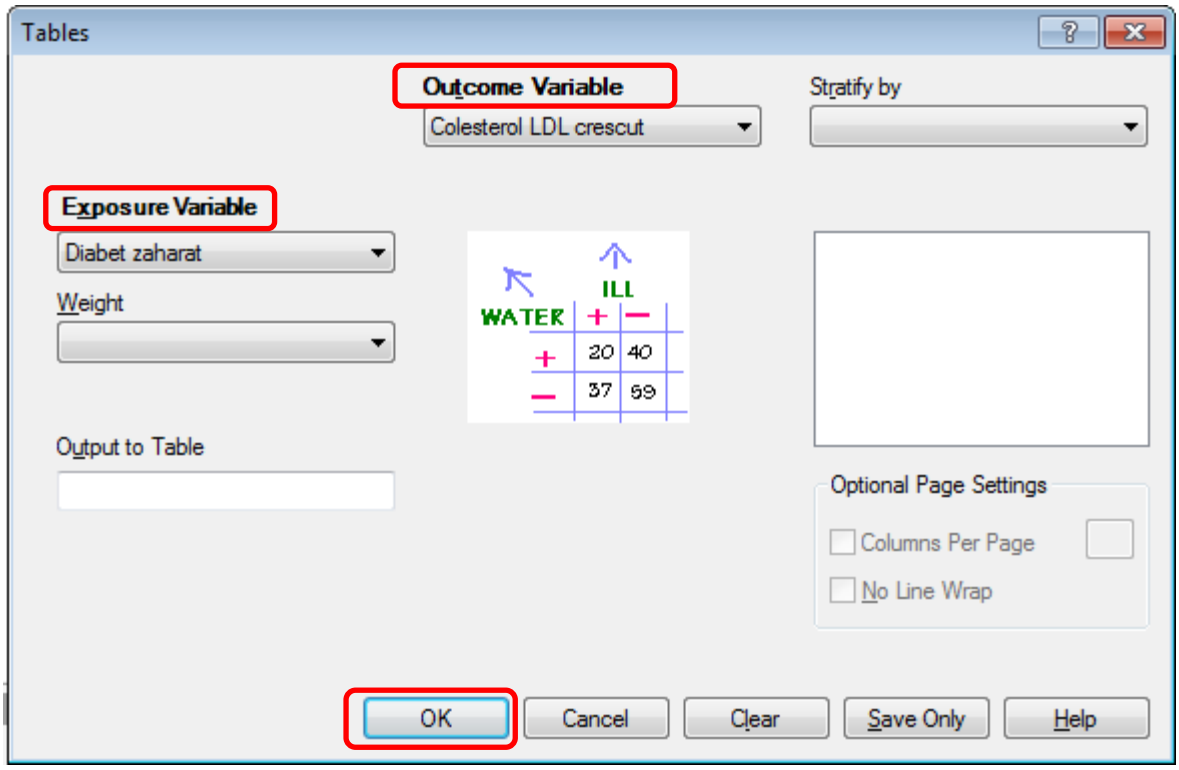
The test results also include the mean values of the compared variables. By subtracting them, you will be able to evaluate the difference between mean values, thus appraising the clinical significance of this difference.

Computing the contingency table, Risk Ratio (RR) and Odds Ratio (OR) in EpiInfo

Performing a Chi-square (X^2) test in EpiInfo

In the ***Analysis*** module of ***EpiInfo*** click ***Tables***, from the left panel.

In the dialog window that opens, select from the drop-down lists the ***Exposure Variable*** (e.g. the risk factor, treatment, etc.) and the ***Outcome Variable*** (e.g. the disease suspected to be an outcome of the risk factor, the improvement of health suspected to be an outcome of the treatment, etc.), then press OK.



	Colesterol LDL crescut		
DIABET ZAHARAT	**da**	**nu**	**Total**
da	40	42	82
Row%	48.78%	51.22%	100.00%
Col%	85.11%	56.00%	67.21%
nu	7	33	40
Row%	17.50%	82.50%	100.00%
Col%	14.89%	44.00%	32.79%
TOTAL	47	75	122
Row%	38.52%	61.48%	100.00%
Col%	100.00%	100.00%	100.00%

Depending on the type of data collection used in your research scenario, interpret only the appropriate indicator (RR / OR) and its ***95% Confidence Interval*** displayed to the right of the ***Point Estimate*** of each indicator.

Single Table Analysis

	Point Estimate	95% Confidence Interval Lower	Upper	
PARAMETERS: Odds-based				
Odds Ratio (cross product)	4.4898	1.7831	11.3049	(T)
Odds Ratio (MLE)	4.4367	1.8082	11.9542	(M)
		1.6804	13.2787	(F)
PARAMETERS: Risk-based				
Risk Ratio (RR)	2.7875	1.3725	5.6611	(T)
Risk Difference (RD%)	31.2805	15.2896	47.2714	(T)

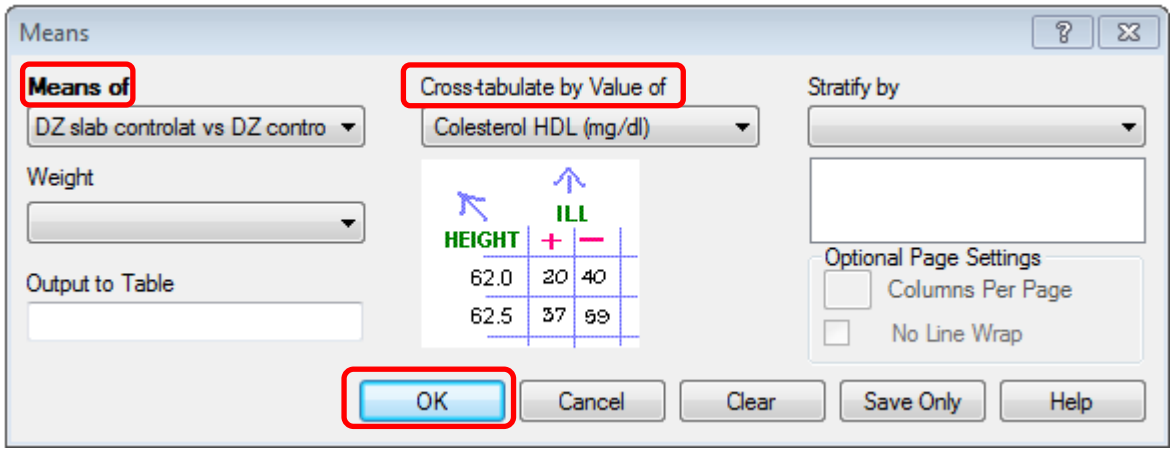
In most cases, the two-tailed p-value rendered by the ***Chi-square - uncorrected*** test shows the statistical significance of the investigated difference between the frequency distribution in the compared groups. Yet, sometimes, when one or more expected frequencies are lower than 5, a message will be displayed below the test results, telling you to interpret the p-value rendered by the ***Fisher exact*** test.

a.

STATISTICAL TESTS	Chi-square	1-tailed p	2-tailed p
Chi-square - uncorrected	11.1076		0.0008608989
Chi-square - Mantel-Haenszel	11.0166		0.0009041693
Chi-square - corrected (Yates)	9.8261		0.0017216883
Mid-p exact		0.0003773484	
Fisher exact		0.0006285038	0.0007946499

Comparing quantitative data (Test student/ANOVA/ ...) in Epi Info

In **Command Explorer, Statistics** section we choose the command **Means**.



In the opened window choose the quantitative variable from the list from **Means of** and the grouping variable in the list from **Cross-tabulate by Value of**, then press the OK button.

In the window with the results (output), we have:

1. **Descriptive statistics** for groups:

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev
DZ controlat	45.0000	2502.0000	55.6000	112.4727	10.6053
DZ slab controlat	37.0000	1932.0000	52.2162	77.8408	8.8227

	Minimum	25%	Median	75%	Maximum	Mode
DZ controlat	39.0000	48.0000	55.0000	62.0000	80.0000	54.0000
DZ slab controlat	36.0000	46.0000	52.0000	58.5000	76.0000	41.0000

1. **The result of the t test** (Student) to compare the means of two independent samples with equal variances (**Pooled**) or unequal (**Unequal**) variances

T-Test

	Method	Mean	95% CL Mean		Std Dev
Diff (Group 1 - Group 2)	Pooled	3.3838	-0.9633	7.7309	9.8432
Diff (Group 1 - Group 2)	Satterthwaite	3.3838	-0.8867	7.6543	

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	80	1.55	0.1253
Satterthwaite	Unequal	79.98	1.58	0.1188

2. **The result of ANOVA test** to compare the averages of three or more independent samples with equal variances:

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	232.49071	1	232.49071	2.39957
Within	7751.07027	80	96.88838	
Total	7983.56098	81		

P-value = 0.12532

3. The result of the Bartlett test to compare the variances of two independent samples:

Bartlett's Test for Inequality of Population Variances

Bartlett's chi square= 1.30103 df=1 P value=0.25403

A small p-value (e.g., less than 0.05 suggests that the variances are not homogeneous and that the ANOVA may not be appropriate.

4. The results of nonparametric tests for comparing two independent samples (**Mann-Whitney test / Wilcoxon Two-Sample**) or more than two independent samples (**Kruskal-Wallis test**):

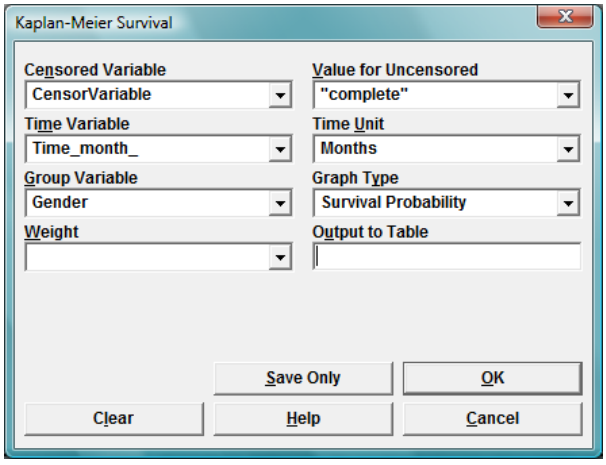
Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

Kruskal-Wallis H (equivalent to Chi square) = 1.4703
Degrees of freedom = 1
P value = 0.2253

Performing a Log-rank test for survival analysis in EpiInfo

In the ***Analysis*** module of ***EpiInfo*** click ***Kaplan-Meyer Survival***, from the left panel.

Complete the dialog box as seen in the image below:



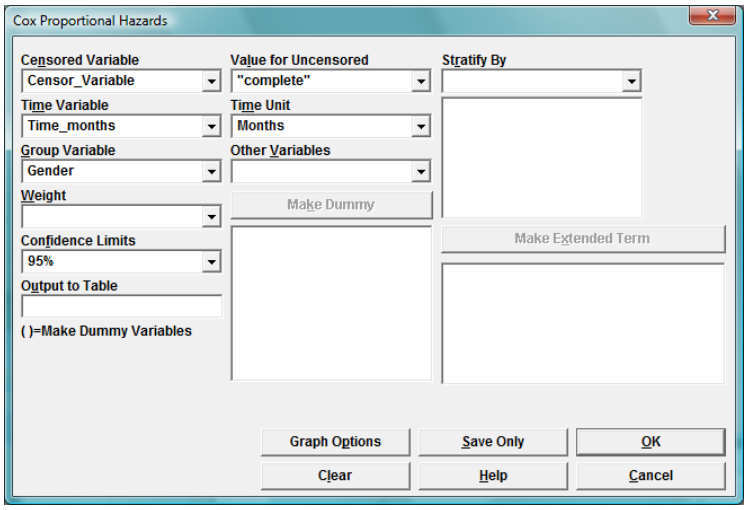
Change the ***Group Variable*** as needed for your comparison.

Below the survival chart you will find the result of the Log-rank test.

Performing a Cox Regression and computing the Hazard Ratio in EpiInfo

In the ***Analysis*** module of ***EpiInfo*** click ***Cox proportional hazard***, from the left panel.

Complete the dialog box as seen in the image below:



Change the ***Group Variable*** as needed for your comparison.

The Hazard Ratio (HR), its 95% confidence interval and the statistical significance of the Cox regression model will be listed below the regression model chart.

Performing a Simple Linear Regression in Microsoft Excel

Use ***Data Analysis*** – ***Regression*** from the ***Tools*** menu in ***Microsoft Excel***.

For ***Input Y Range*** select the cell range that contains the dependent variable (**y**) in your sample, the one you want to predict using a simple linear regression.

For ***Input X Range*** select the cell range that contains the independent variable (**x**) in your sample, the one you want to use in order to predict the dependent variable (**y**), using a simple linear regression.

Make sure to include in your selection the cells containing labels for both the dependent and the independent variable and check the ***Labels*** box. Also check the ***Confidence Level*** box for a 95% CI and enter a suggestive name for the new worksheet where your simple linear regression will be saved.

Performing a Multiple Linear Regression in Microsoft Excel

Use ***Data Analysis*** – ***Regression*** from the ***Tools*** menu in ***Microsoft Excel***.

For ***Input Y Range*** select the cell range that contains the dependent variable (**y**) in your sample, the one you want to predict using a multiple linear regression.

For ***Input X Range*** select the contiguous cell range that contains the independent variables (**x_1, x_2, … , x_n**) in your sample, the ones you want to use in order to predict the dependent variable (**y**), using a multiple linear regression. If the independent variables do not form a contiguous cell range, cut isolated variables before using ***Regression,*** and insert them into adjacent columns in order to form a contiguous cell range.

Make sure to include in your selection the cells containing labels, for all dependent and independent variables and check the ***Labels*** box. Also check the ***Confidence Level*** box for a 95% CI and enter a suggestive name for the new worksheet where your multiple linear regression will be saved.