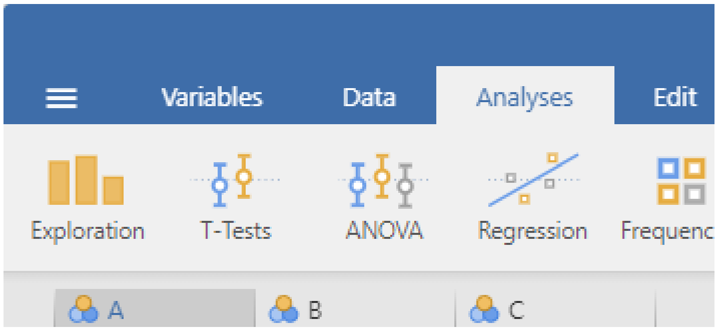
Instructions for computer-use in solving the proposed practical activities

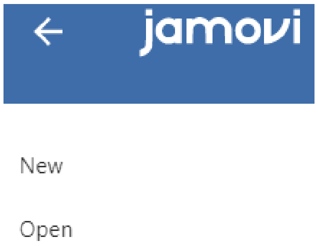
Instructions for using Jamovi for data analysis

Importing the database

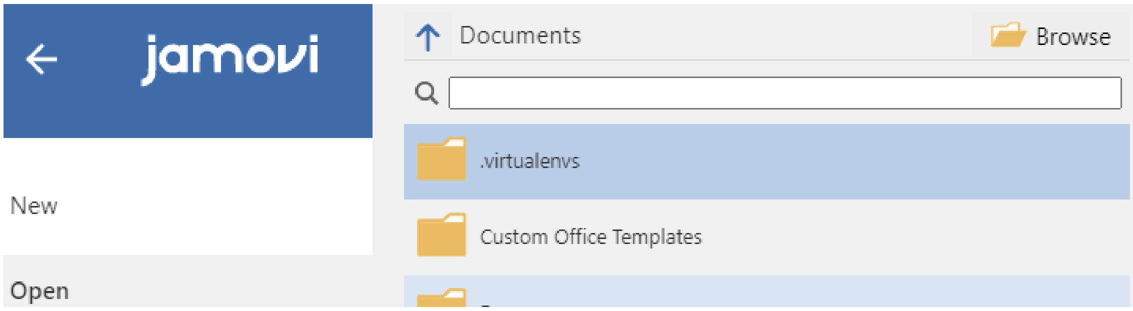
Use the **hamburger** menu icon .☰



Then select the **Open** option

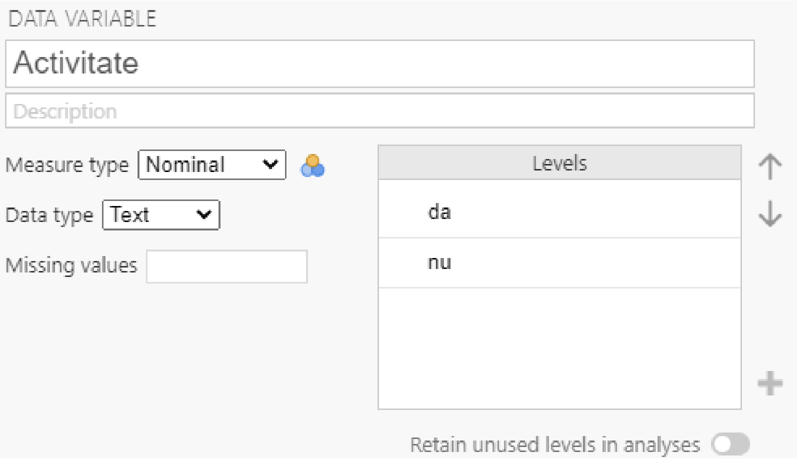


Afterwards, search for the database folder by pressing the **Browse button** , select the desired file and press the **Open button** to import it.



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 .



DATA VARIABLE

Activitate

Description

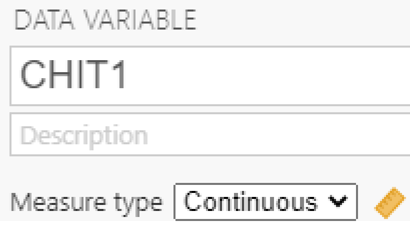
Measure type: Nominal

Data type: Text

Missing values:

Levels
da
nu

Retain unused levels in analyses



DATA VARIABLE

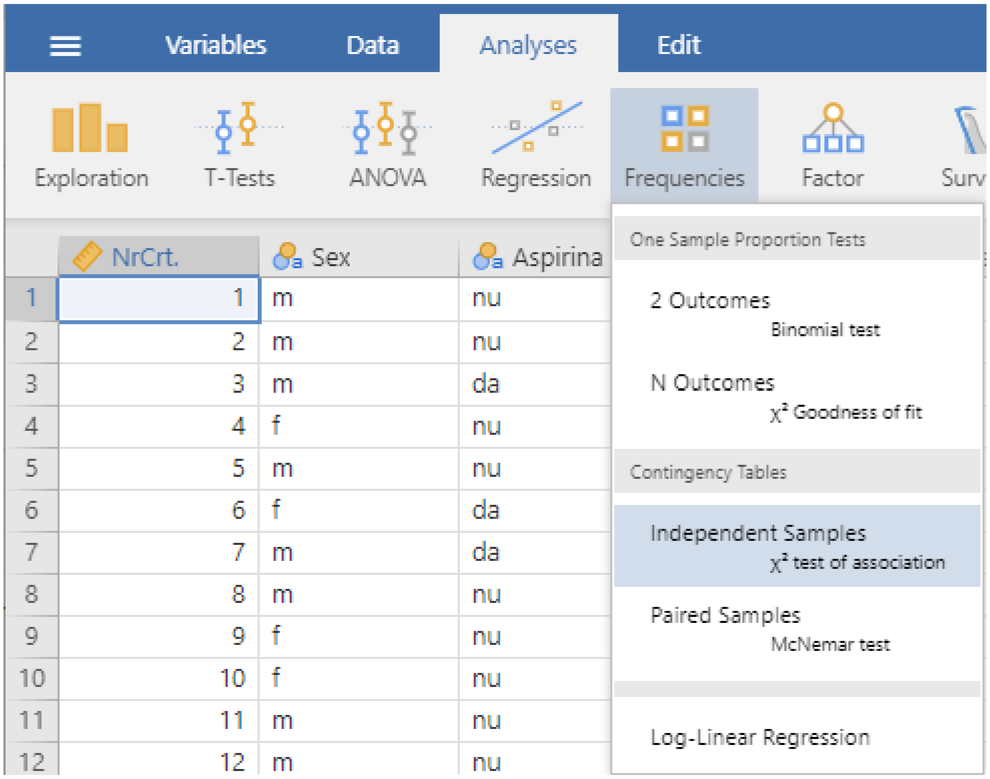
CHIT1

Description

Measure type: Continuous

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).**

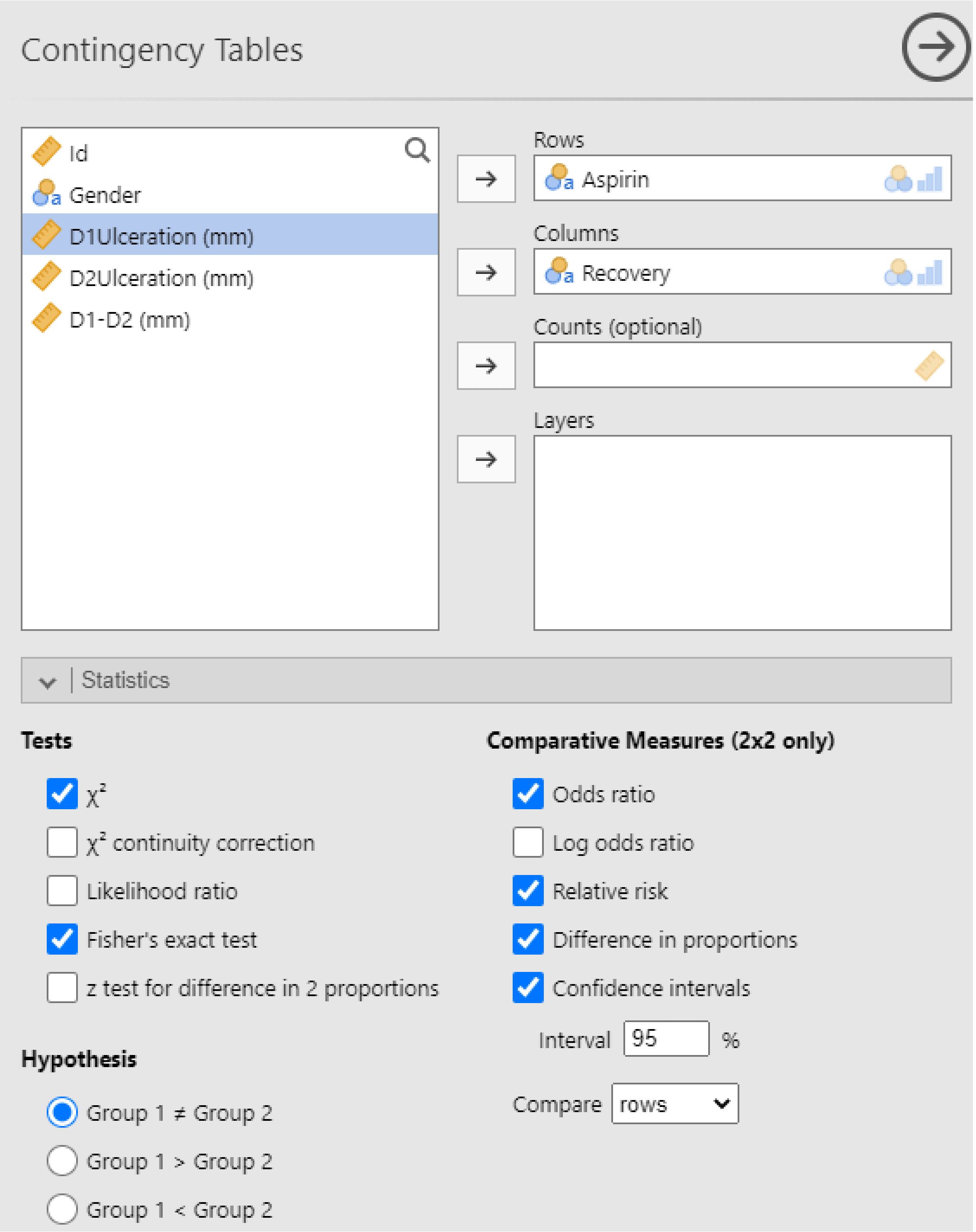


	NrCrt.	Sex	Aspirina
1	1	m	nu
2	2	m	nu
3	3	m	da
4	4	f	nu
5	5	m	nu
6	6	f	da
7	7	m	da
8	8	m	nu
9	9	f	nu
10	10	f	nu
11	11	m	nu
12	12	m	nu

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

☑ χ^2
□ χ^2 continuity correction
□ Likelihood ratio
☑ Fisher's exact test
□ z test for difference in 2 proportions

Hypothesis

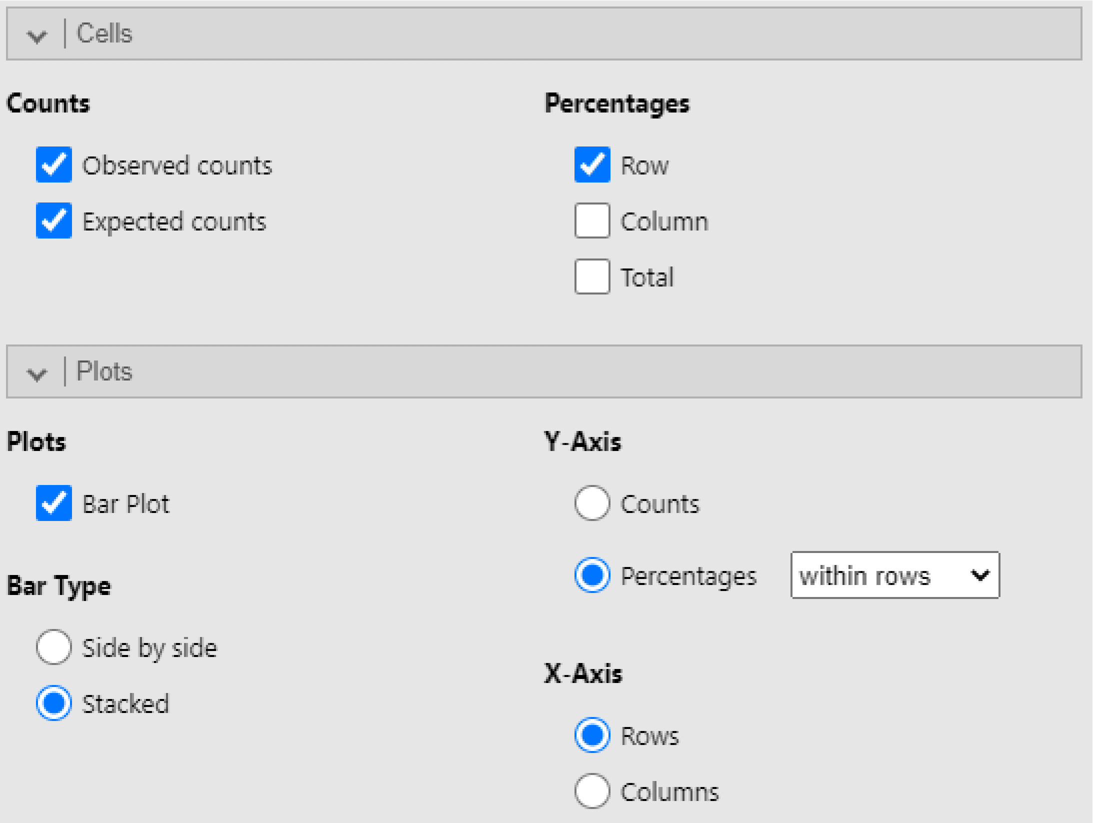
● Group 1 ≠ Group 2
○ Group 1 > Group 2
○ Group 1 < Group 2

Comparative Measures (2x2 only)

☑ Odds ratio
□ Log odds ratio
☑ Relative risk
☑ Difference in proportions
☑ Confidence intervals

Interval 95 %

Compare rows



Cells

Counts

☑ Observed counts

☑ Expected counts

Percentages

☑ Row

□ Column

□ Total

Plots

Plots

☑ Bar Plot

Bar Type

○ Side by side

● Stacked

Y-Axis

○ Counts

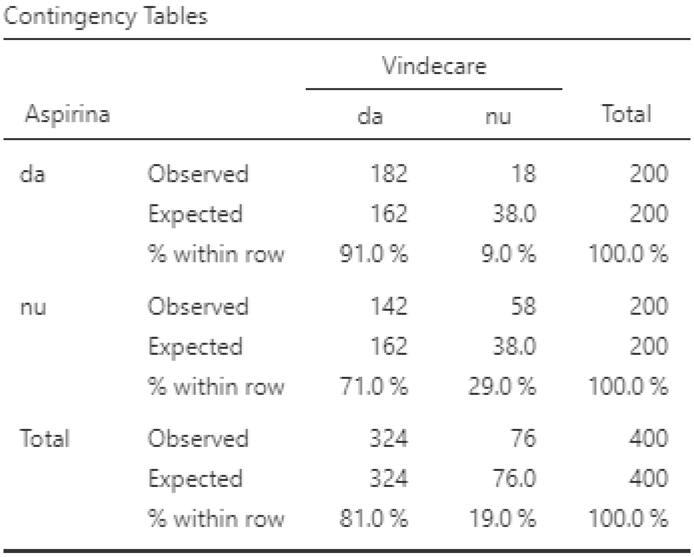
● Percentages within rows

X-Axis

● Rows

○ Columns

Contingency table

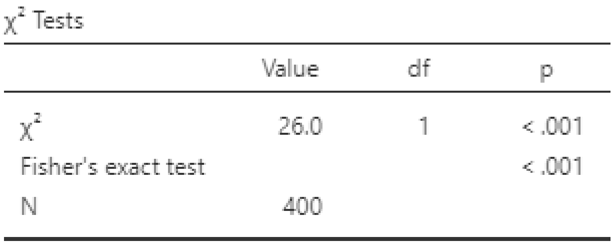


Contingency Tables

		Vindecare		
Aspirina		da	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.

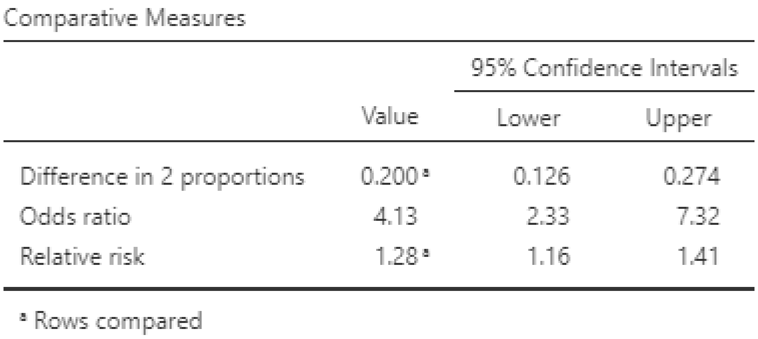


χ^2 Tests

	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**.



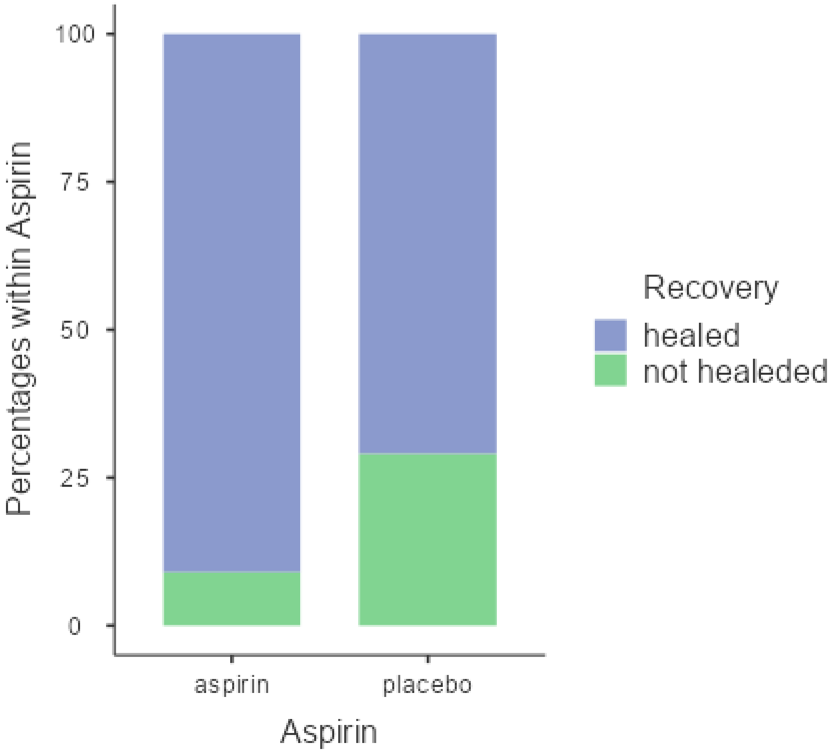
Comparative Measures

		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

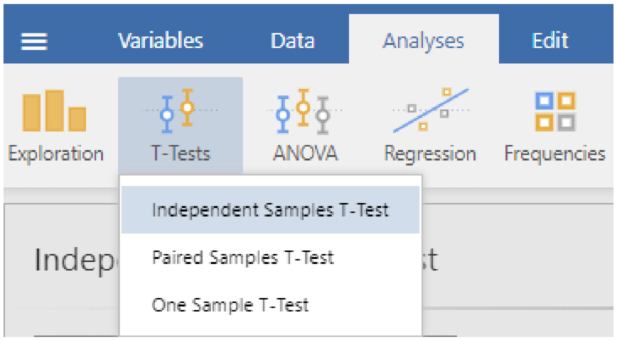
Column chart

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



Student test for two independent groups

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

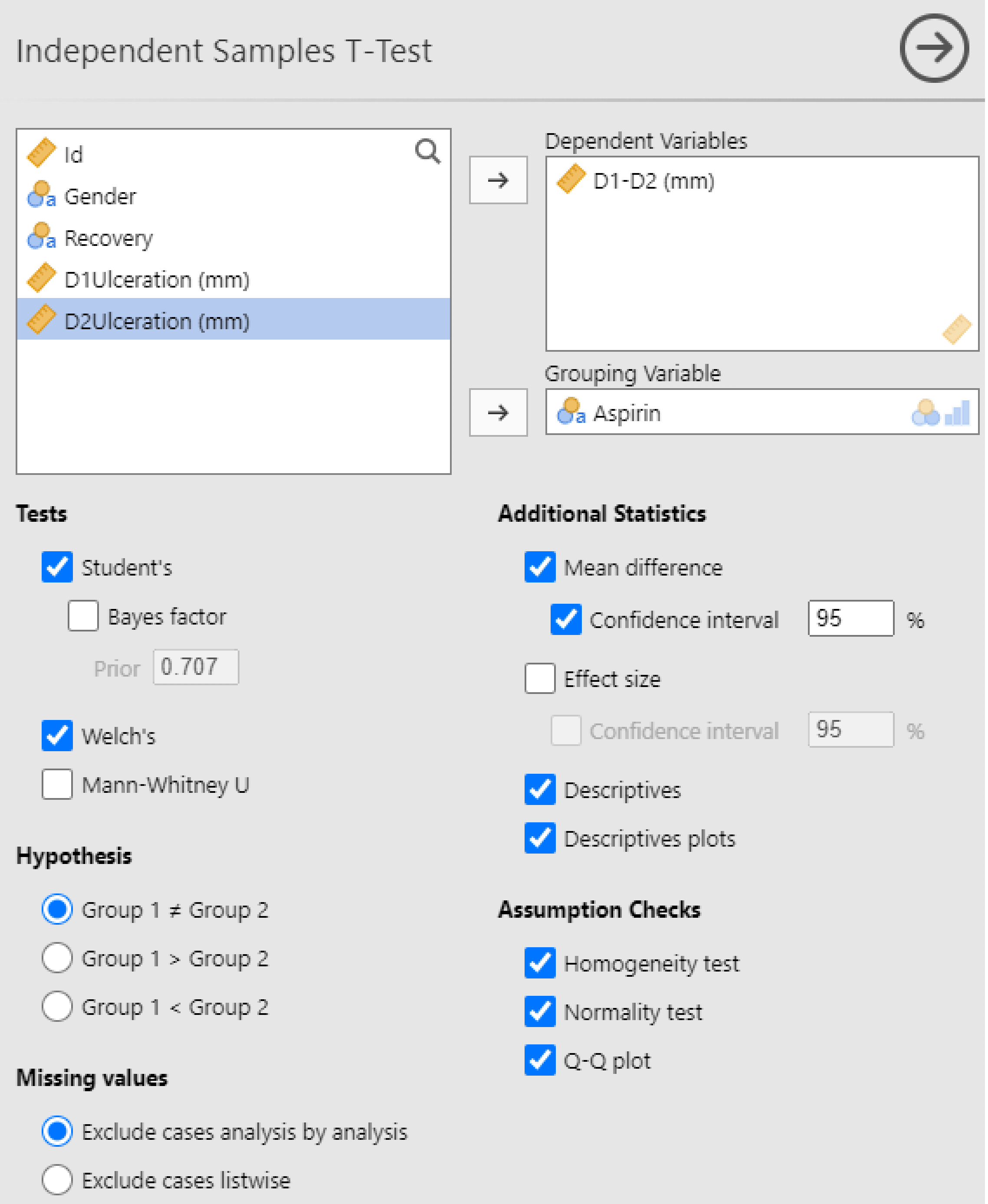


The qualitative variable that **identifies the compared groups** should be moved by pressing the arrow next to the **Grouping Variable field**.

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**, **Normality test**, **Q-Q plot**.



Independent Samples T-Test

Id
Gender
Recovery
D1Ulceration (mm)
D2Ulceration (mm)

Dependent Variables: D1-D2 (mm)

Grouping Variable: Aspirin

Tests

☑ Student's

□ Bayes factor

Prior 0.707

☑ Welch's

□ Mann-Whitney U

Hypothesis

● Group 1 ≠ Group 2

○ Group 1 > Group 2

○ Group 1 < Group 2

Missing values

● Exclude cases analysis by analysis

○ Exclude cases listwise

Additional Statistics

☑ Mean difference

☑ Confidence interval 95 %

□ Effect size

□ Confidence interval 95 %

☑ Descriptives

☑ Descriptives plots

Assumption Checks

☑ Homogeneity test

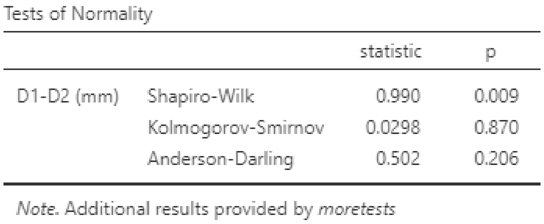
☑ Normality test

☑ Q-Q plot

Evaluation of Student test application conditions – normality

The Student Test can only be **applied if the data** follows a **normal distribution**. A **variant** of 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. The result of interest is **the p-value**, which is located in column p of the **Tests of Normality** table. If **the p-value** is **less than 0.05**, it is a **suggestion that the** data **does not follow a normal distribution**, **otherwise** it is a **suggestion** that the data follows a **normal distribution**.

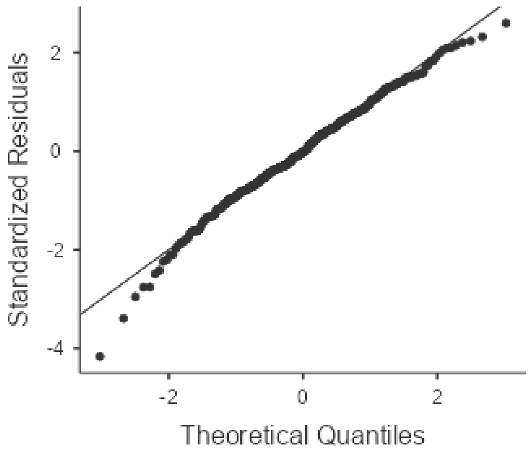


Tests of Normality

		statistic	p
D1-D2 (mm)	Shapiro-Wilk	0.990	0.009
	Kolmogorov-Smirnov	0.0298	0.870
	Anderson-Darling	0.502	0.206

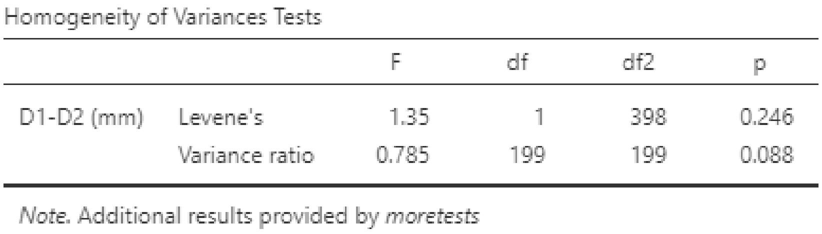
Note. Additional results provided by *moretests*

Another **way** to assess the normality **of** the data is with the help **of the Q-Q plot chart**. If the **tendency** of observations is for them to be arranged **away** from the **solid line** (which is a normal distribution), then it is a **suggestion** that the data **do not follow a normal distribution**, **otherwise**, if the trend is for observations to be close to the continuous line, it is a **suggestion** that the data follow a **normal distribution**.



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**.

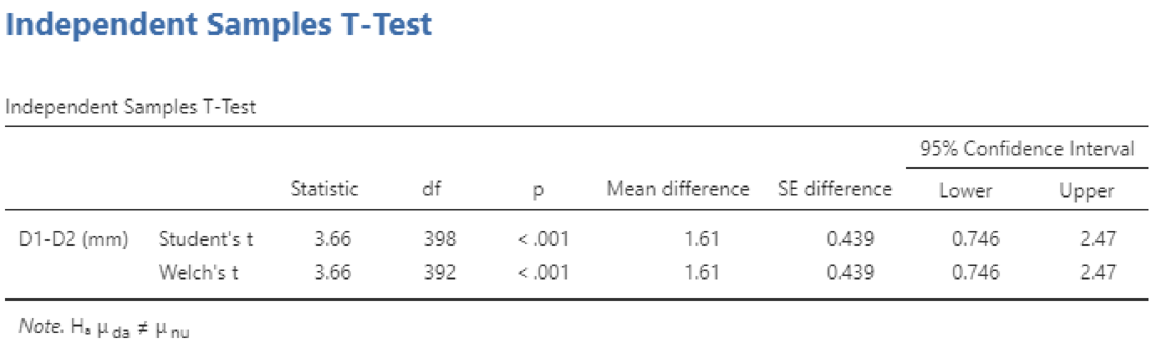


Homogeneity of Variances Tests

		F	df	df2	p
D1-D2 (mm)	Levene's	1.35	1	398	0.246
	Variance ratio	0.785	199	199	0.088

Note. Additional results provided by *moretests*

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

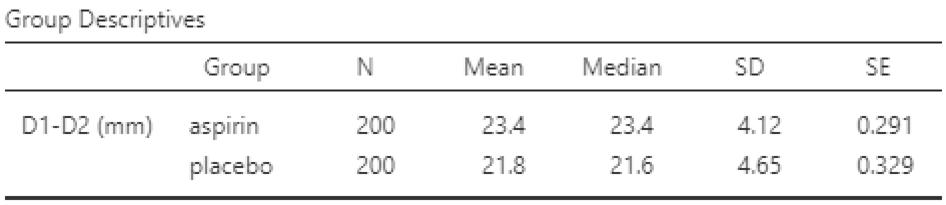
							95% Confidence Interval	
		Statistic	df	p	Mean difference	SE difference	Lower	Upper
D1-D2 (mm)	Student's t	3.66	398	< .001	1.61	0.439	0.746	2.47
	Welch's t	3.66	392	< .001	1.61	0.439	0.746	2.47

Note. H_a $\mu_{da} \neq \mu_{nu}$

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.**

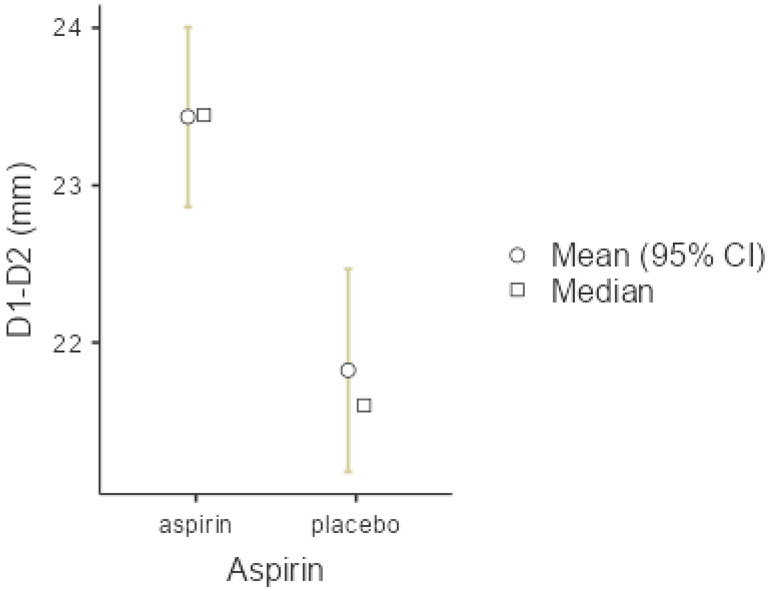


Group Descriptives

	Group	N	Mean	Median	SD	SE
D1-D2 (mm)	aspirin	200	23.4	23.4	4.12	0.291
	placebo	200	21.8	21.6	4.65	0.329

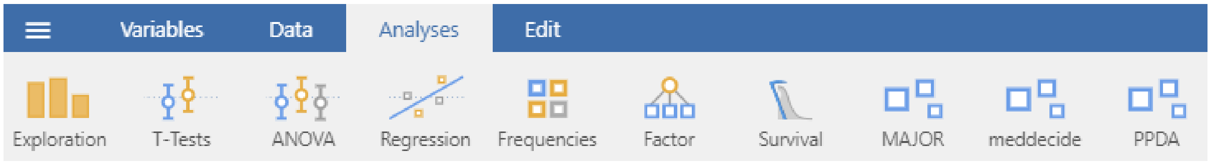
Graph of means

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.

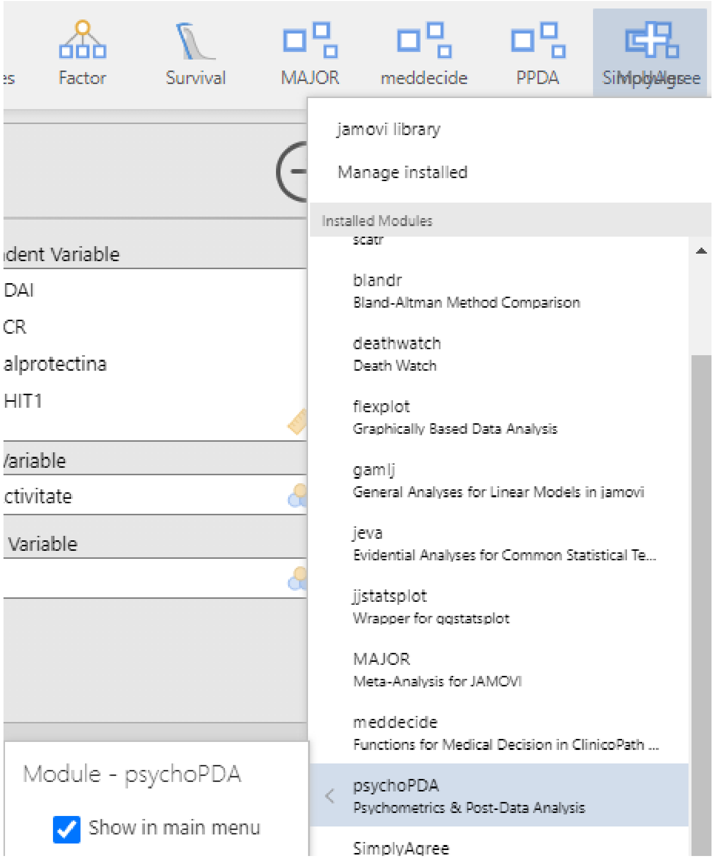


ROC analysis

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



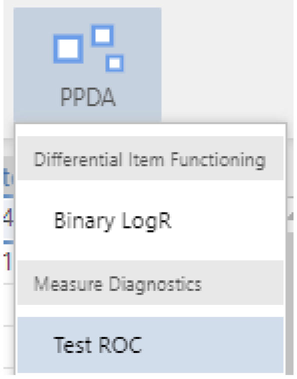
If you can't find it, it may not be displayed. In the table **Analyzes** , on the right, tap on + **Modules** and check if it is not present in the list of installed modules.



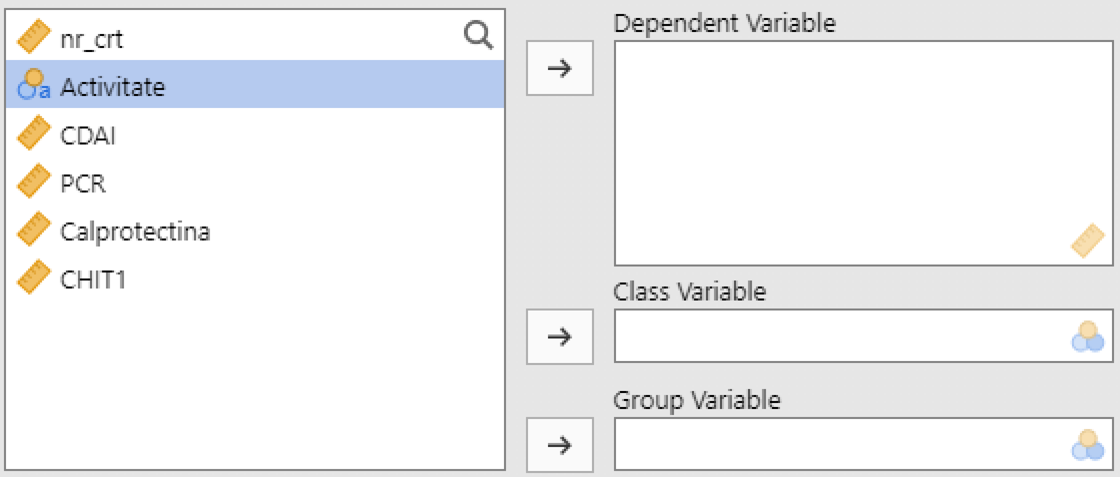
You can check that module to be visible in the table **Analyzes** by clicking the **Show in main option menus** .

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

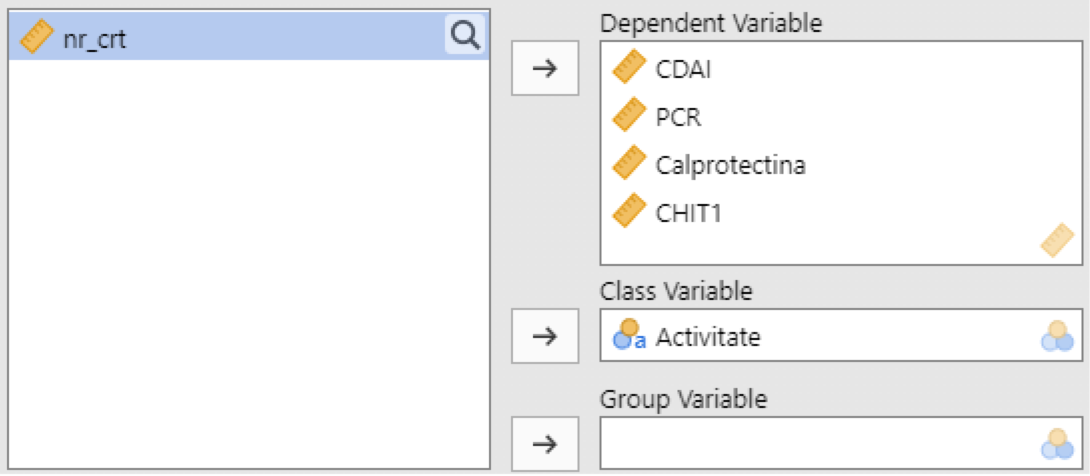
Clicking on the PPDA module, select the ROC Test option from the menu



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:



Dependent Variable
CDAI
PCR
Calprotectina
CHIT1

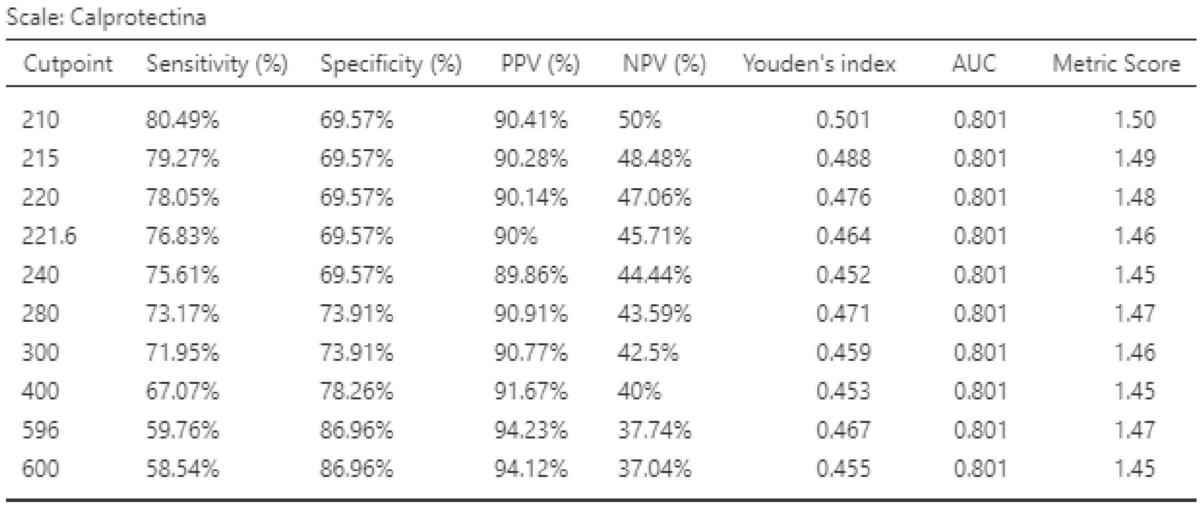
Class Variable
Activitate

Group Variable

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.

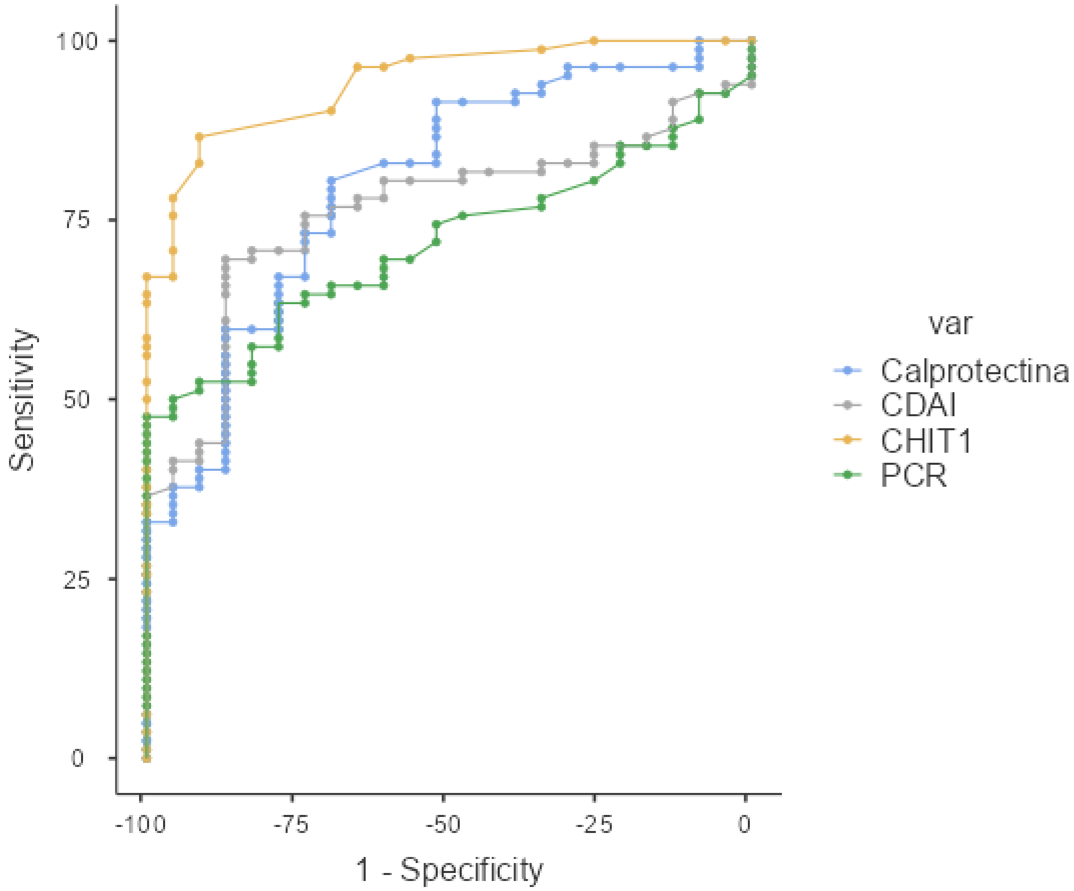


Scale: Calprotectina

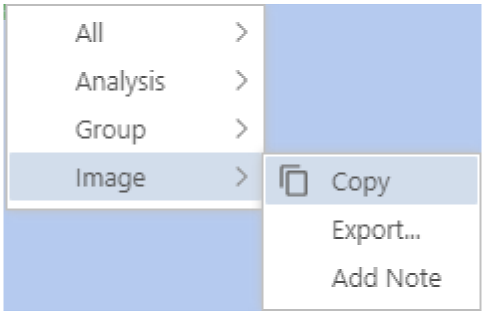
Cutpoint	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Youden's index	AUC	Metric Score
210	80.49%	69.57%	90.41%	50%	0.501	0.801	1.50
215	79.27%	69.57%	90.28%	48.48%	0.488	0.801	1.49
220	78.05%	69.57%	90.14%	47.06%	0.476	0.801	1.48
221.6	76.83%	69.57%	90%	45.71%	0.464	0.801	1.46
240	75.61%	69.57%	89.86%	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).

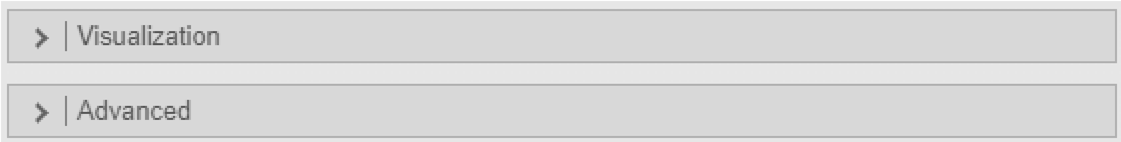


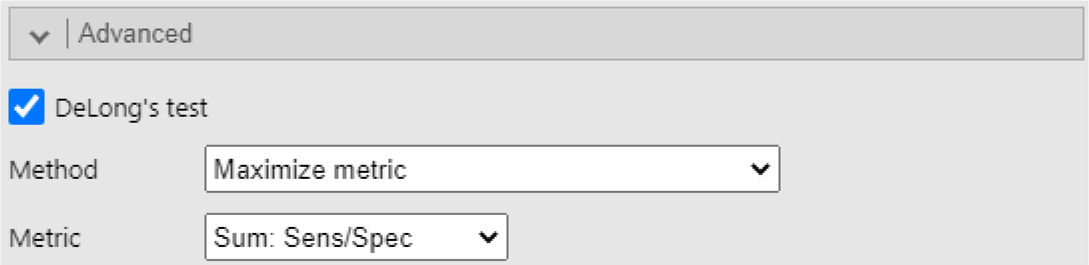
To copy the image press the right mouse button and select **Image Copy**



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** .



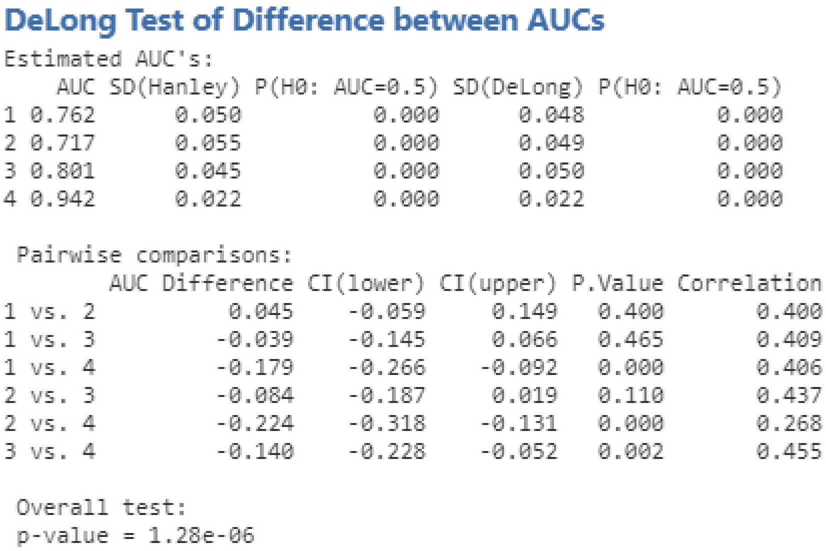


☑ DeLong's test

Method: Maximize metric

Metric: Sum: Sens/Spec

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
```

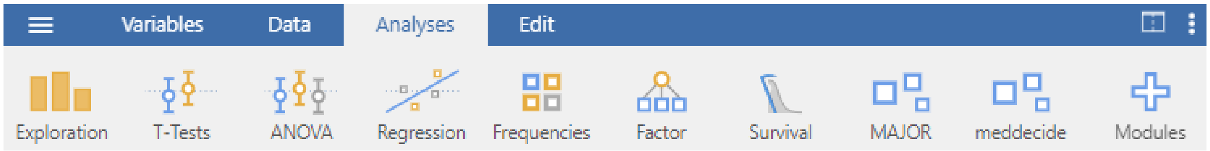
The first table lists the estimates of **the areas under the ROC curve** for each diagnostic test, the standard deviation, as well as a **statistical test of significance for one ROC curve**, for each selected variable in the order of their selection. Thus for the CDAI clinical activity score, the area 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.

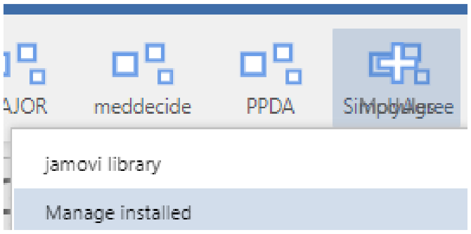
Finally, the result of a global statistical test comparing all ROC curves with each other is presented.

Installation of additional analysis module (e.g. for ROC analysis)

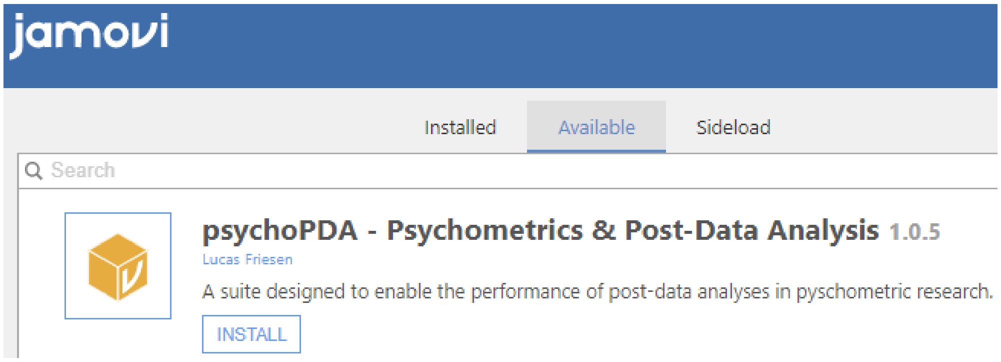
In the table Analyzes , on the right, tap on + Modules and check if it is not present in the list of installed modules.



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** .



psychoPDA - Psychometrics & Post-Data Analysis 1.0.5

Lucas Friesen

A suite designed to enable the performance of post-data analyses in pyschometric research.

Return to the analysis window by pressing the arrow button:

Instructions for using EpiInfo and Excel for data analysis

Downloading *EpiInfo* (for home use)

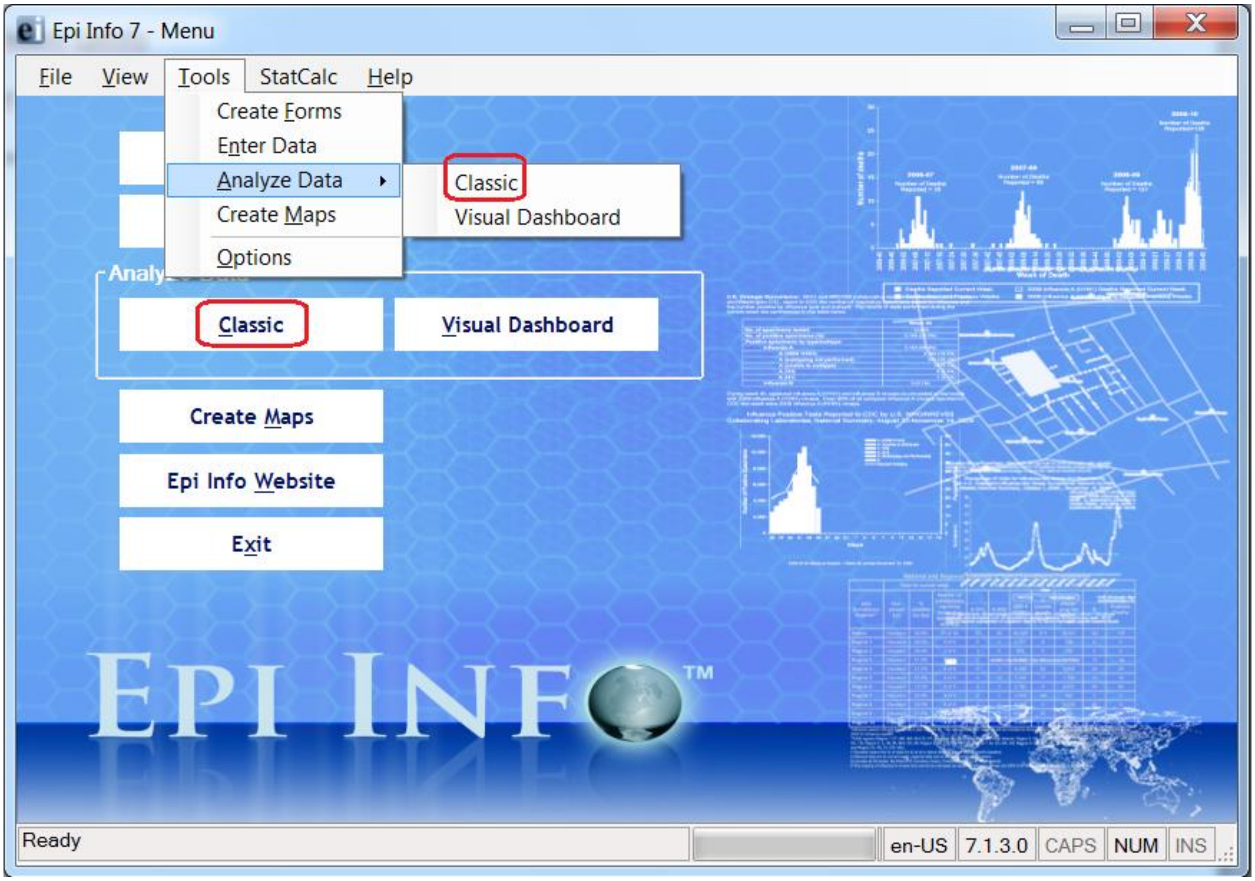
Use an Internet browser to navigate to: http://www.cdc.gov/epiinfo/installation.htm and follow the installation instructions.

Starting Epi Info

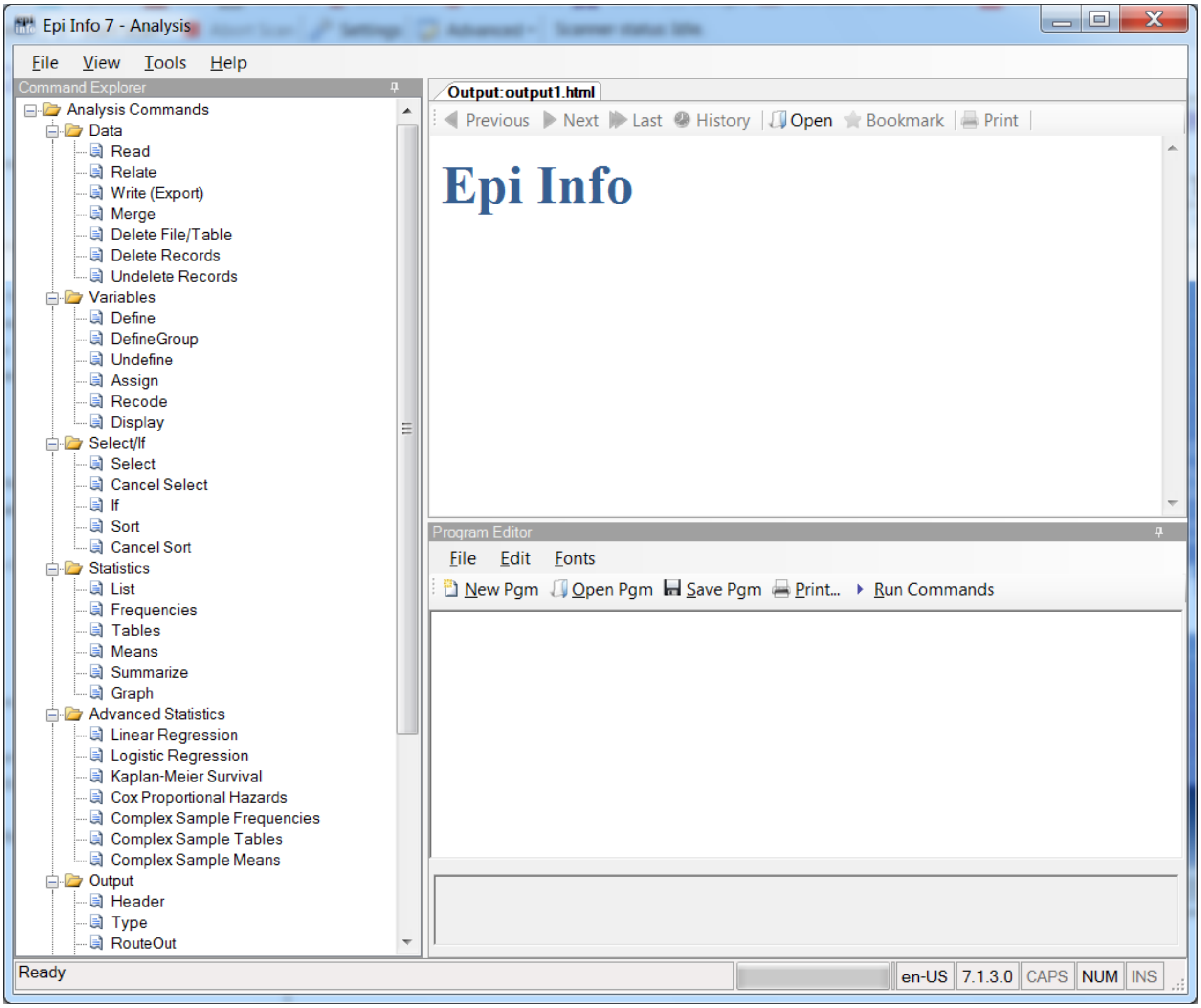
Start button, shortcut **Epi Info^{TM} 7** or **Start button**, **All Programs**, **CDC**, **Epi Info 7**, **Epi Info^{TM} 7**

Analyzing data in Epi Info

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

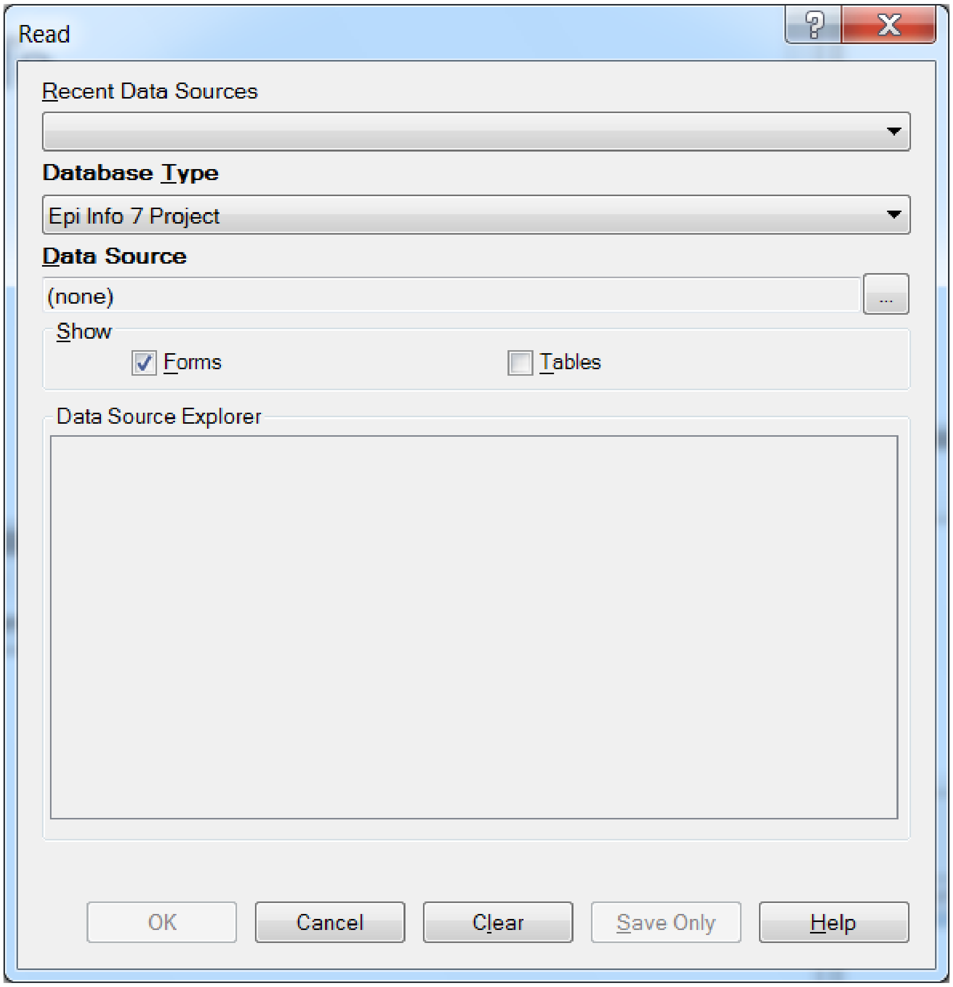


The data analysis window:



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.

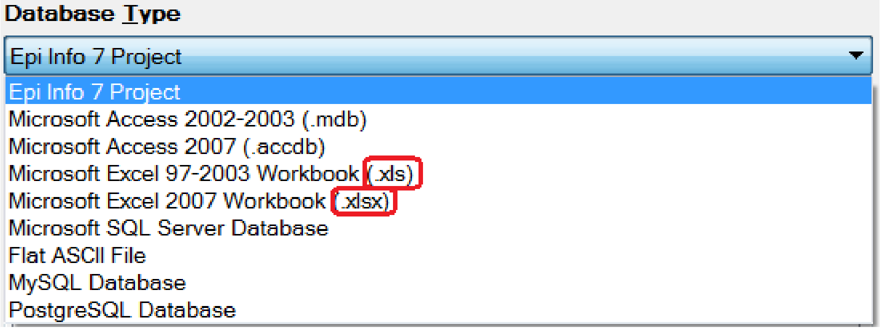


Database Type: Epi Info 7 Project

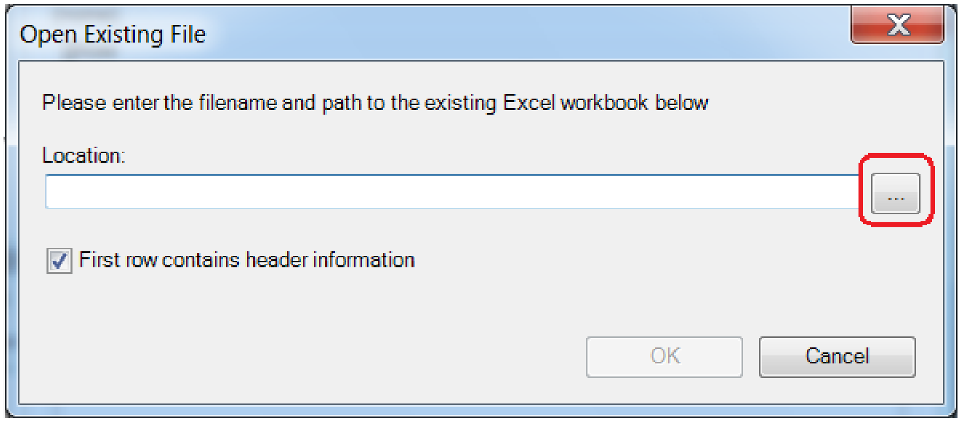
Data Source: (none)

Show: ☑ Forms ☐ Tables

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)**.


Database Type
Epi Info 7 Project
Epi Info 7 Project
Microsoft Access 2002-2003 (.mdb)
Microsoft Access 2007 (.accdb)
Microsoft Excel 97-2003 Workbook (.xls)
Microsoft Excel 2007 Workbook (.xlsx)
Microsoft SQL Server Database
Flat ASCII File
MySQL Database
PostgreSQL Database

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**.

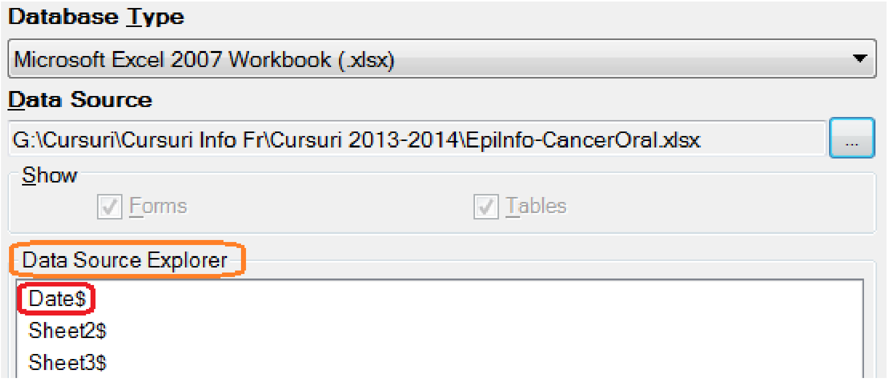

Open Existing File

Please enter the filename and path to the existing Excel workbook below

Location:

☑ 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.


Database Type: Microsoft Excel 2007 Workbook (.xlsx)

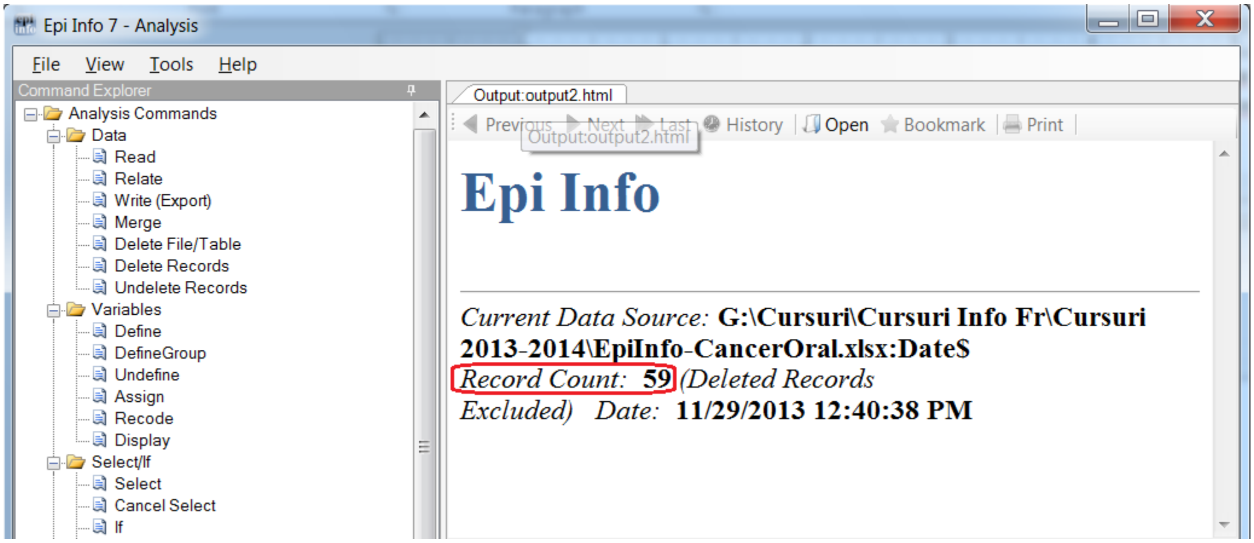
Data Source: G:\Cursuri\Cursuri Info Fr\Cursuri 2013-2014\EpiInfo-CancerOral.xlsx

Show: ☑ Forms ☑ Tables

Data Source Explorer

Date$
Sheet2$
Sheet3$

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


Epi Info 7 - Analysis

Current Data Source: **G:\Cursuri\Cursuri Info Fr\Cursuri 2013-2014\EpiInfo-CancerOral.xlsx:Date$**
Record Count: **59** *(Deleted Records Excluded)* *Date:* **11/29/2013 12:40:38 PM**

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:*

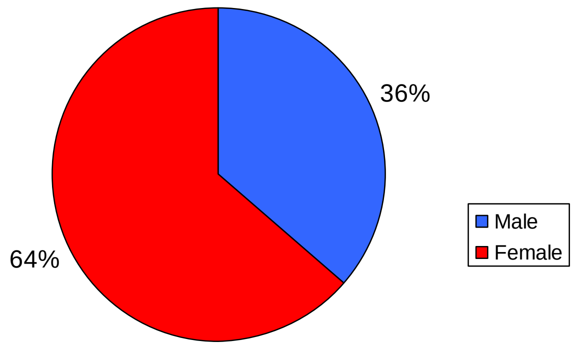


Figure 1. Gender distribution in the studied sample

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?).

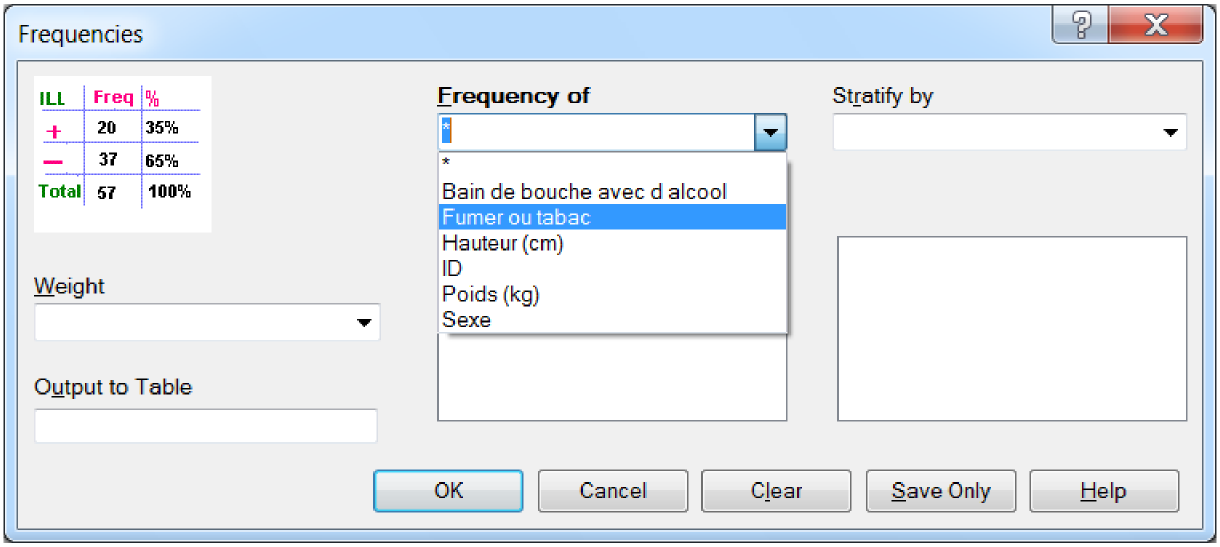
If you plan to use the chart as a figure in a ***Word*** document, erase the chart title in ***Excel*** but remember to label the chart in ***Microsoft Word***: select the figure, right click on this selection and use ***Caption***.

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.

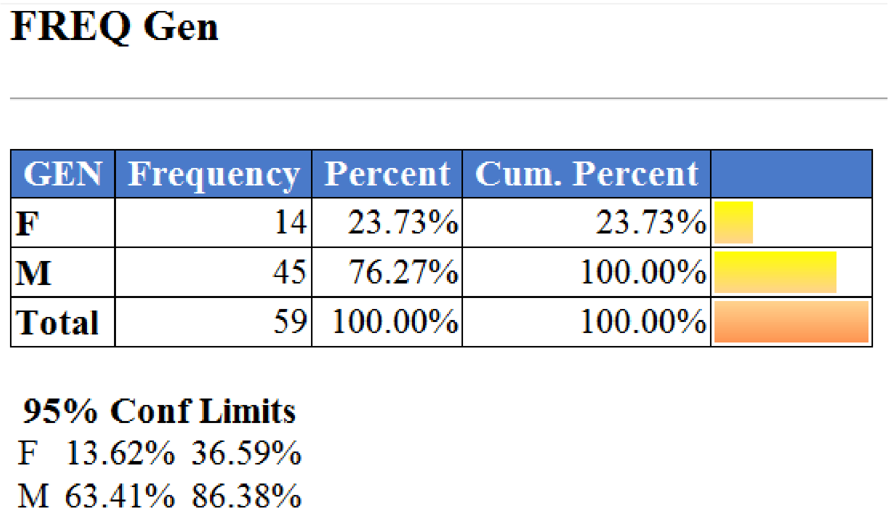
Frequency tables in Epi Info

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

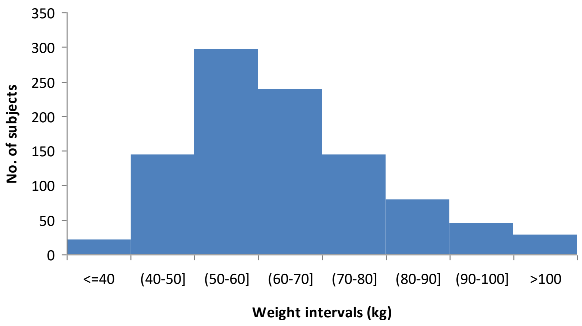


Figure 2. Histogram of weight in the studied sample

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:

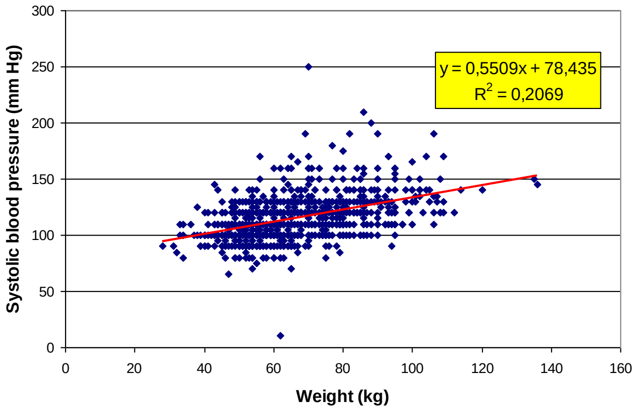


Figure 3. Relation between weight and systolic blood pressure for all subjects included in the studied sample

Survival data:

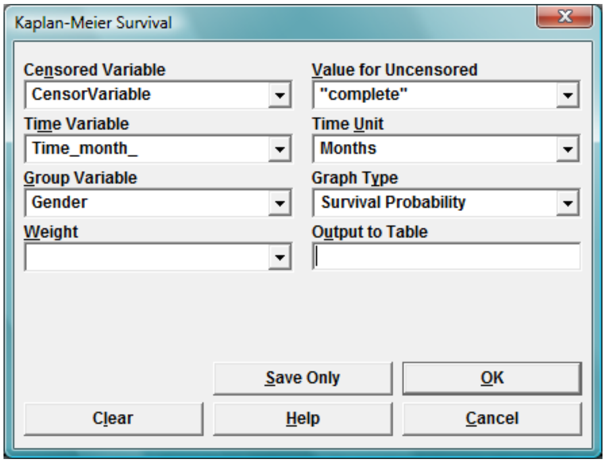
Median of survival time

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:



Censored Variable: CensorVariable

Value for Uncensored: "complete"

Time Variable: Time_month_

Time Unit: Months

Group Variable: Gender

Graph Type: Survival Probability

Weight:

Output to Table:

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. $p = 3{,}22342E\text{-}6 = 3{,}22342 \times 10^{-6} = 0{,}0000032234$).

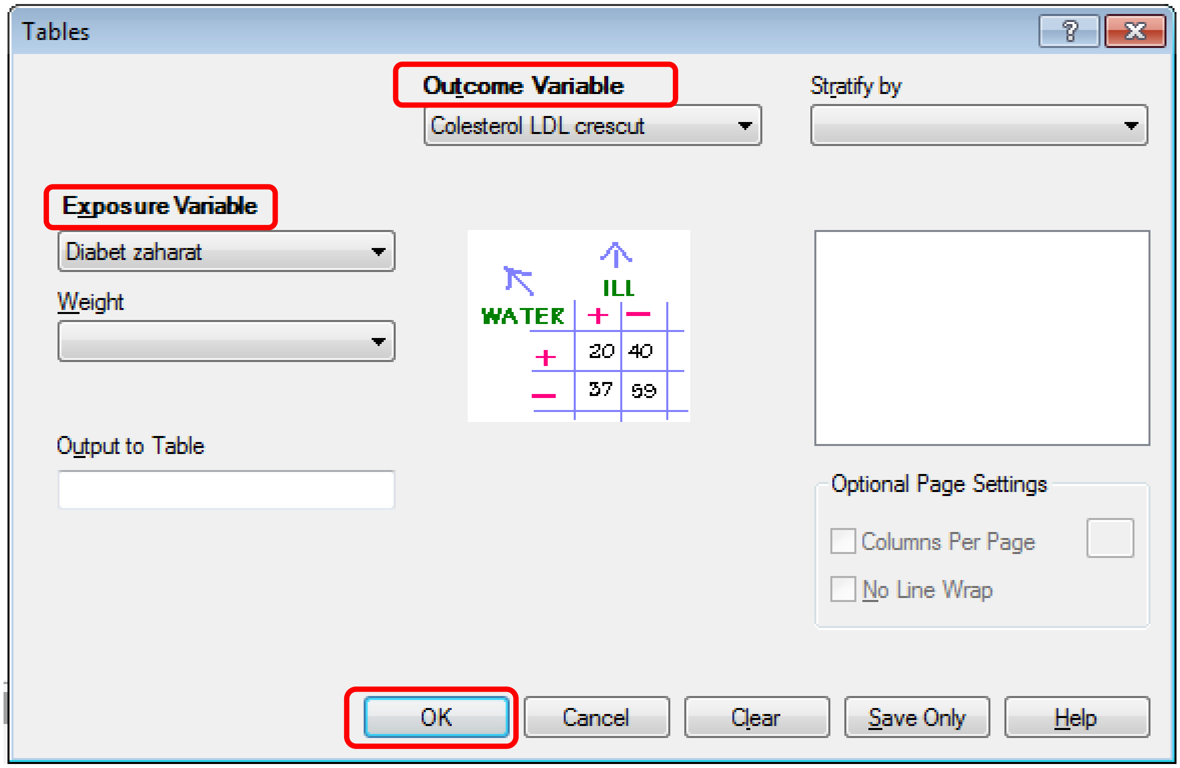
The test results also include the mean values of the compared variables. By subtracting them, you will be able to evaluate the difference between mean values, thus appraising the clinical significance of this difference.

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.



Tables

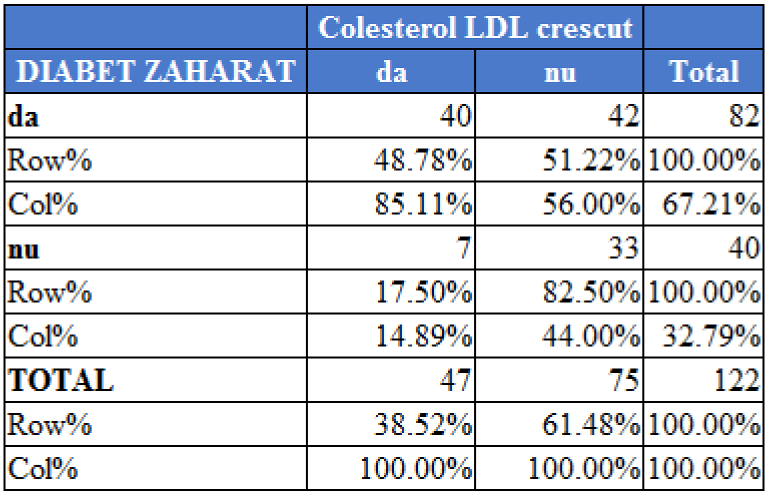
Outcome Variable: Colesterol LDL crescut

Stratify by:

Exposure Variable: Diabet zaharat

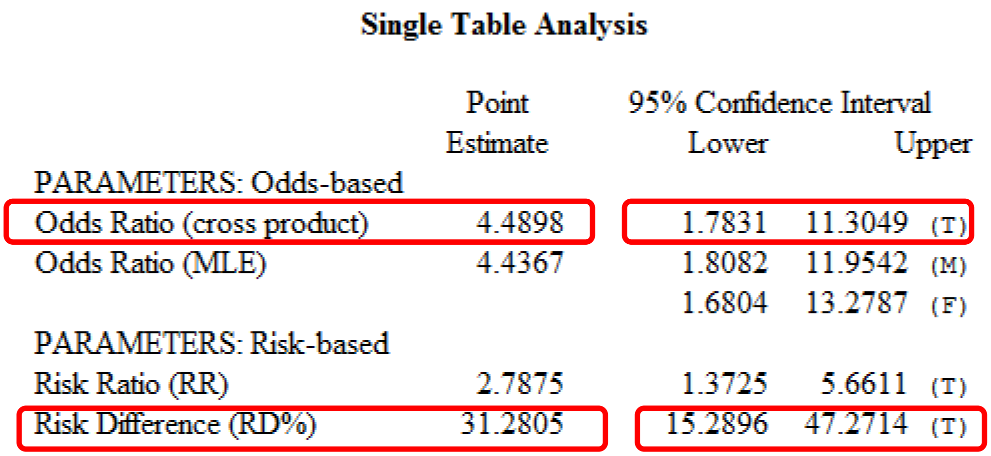
Weight:

Output to Table:



	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.

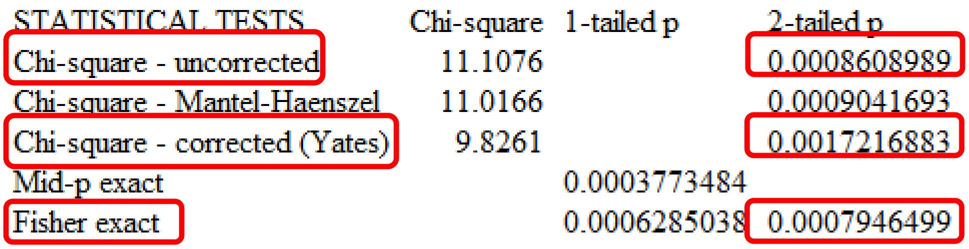


```
                    Single Table Analysis

                               Point       95% Confidence Interval
                              Estimate       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.

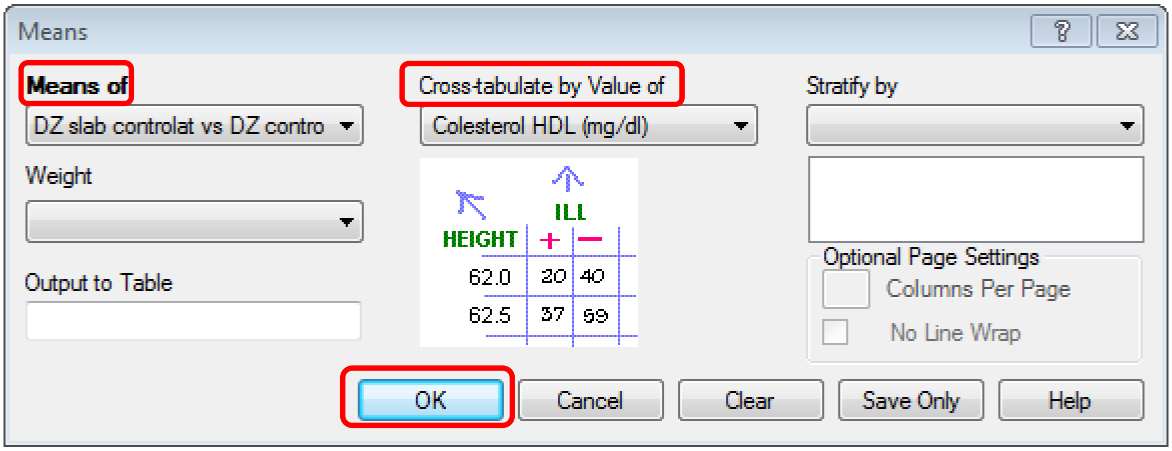
a.



```
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**.



Means

Means of: DZ slab controlat vs DZ contro

Cross-tabulate by Value of: Colesterol HDL (mg/dl)

Stratify by:

Weight:

Output to Table:

Optional Page Settings

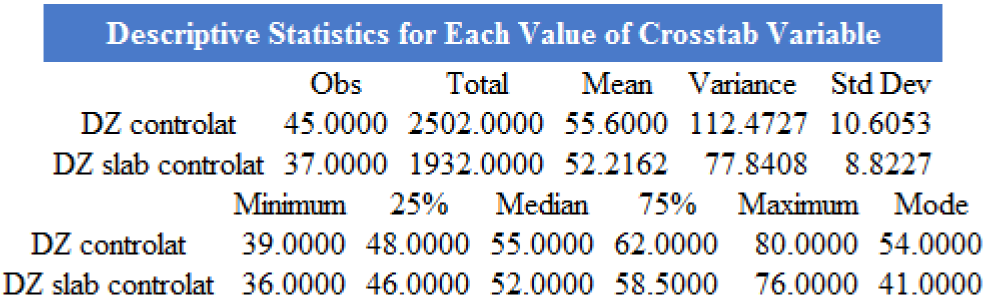
□ Columns Per Page

□ No Line Wrap

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.

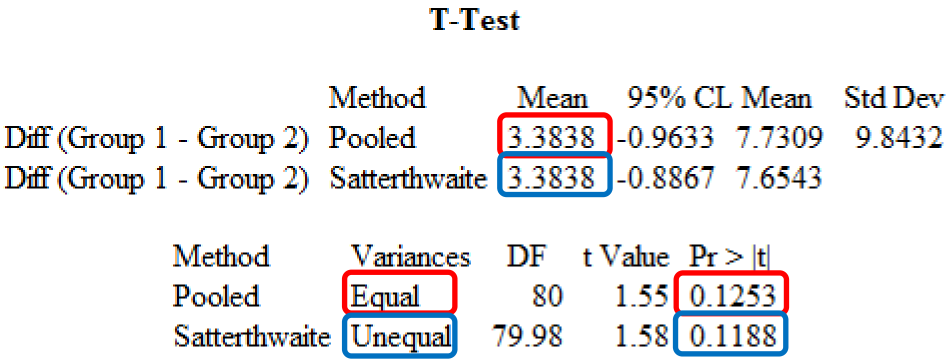
In the window with the results (output), we have:

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

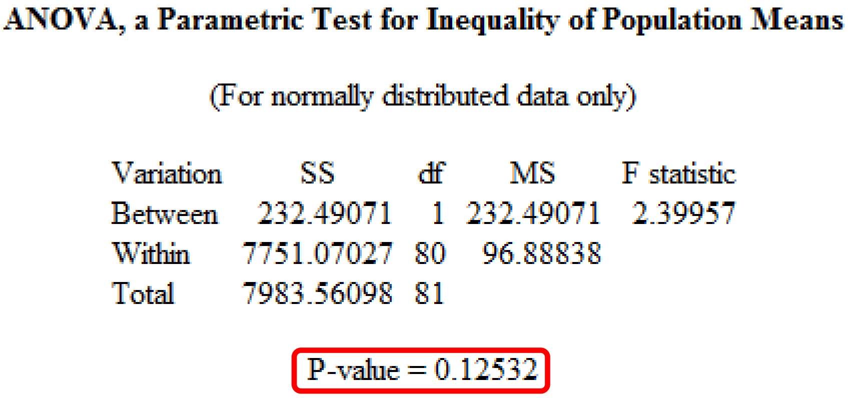


```
                                  T-Test

                          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:



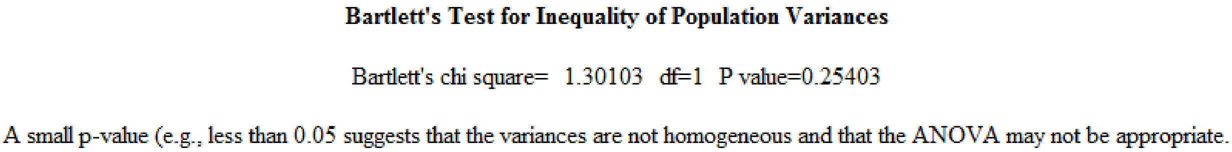
```
 ANOVA, a Parametric Test for Inequality of Population Means

              (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:

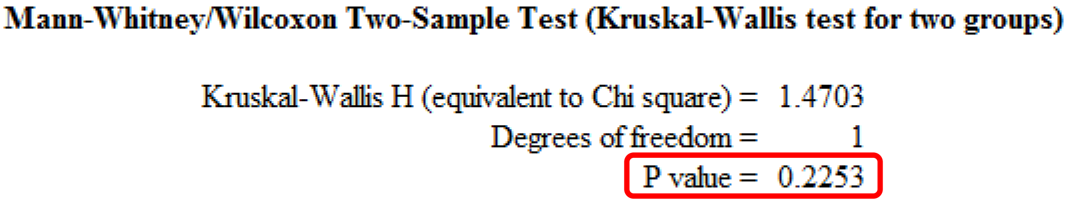


```
              Bartlett's Test for Inequality of Population Variances

              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**):



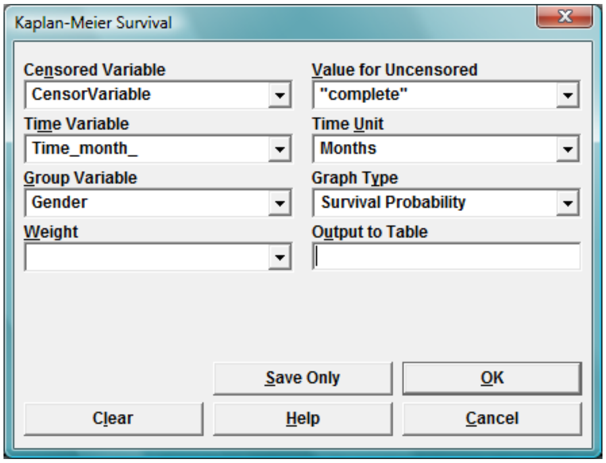
Mann-Whitney/Wilcoxon Two-Sample Test (Kruskal-Wallis test for two groups)

```
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:



Kaplan-Meier Survival

Censored Variable: CensorVariable

Value for Uncensored: "complete"

Time Variable: Time_month_

Time Unit: Months

Group Variable: Gender

Graph Type: Survival Probability

Weight:

Output to Table:

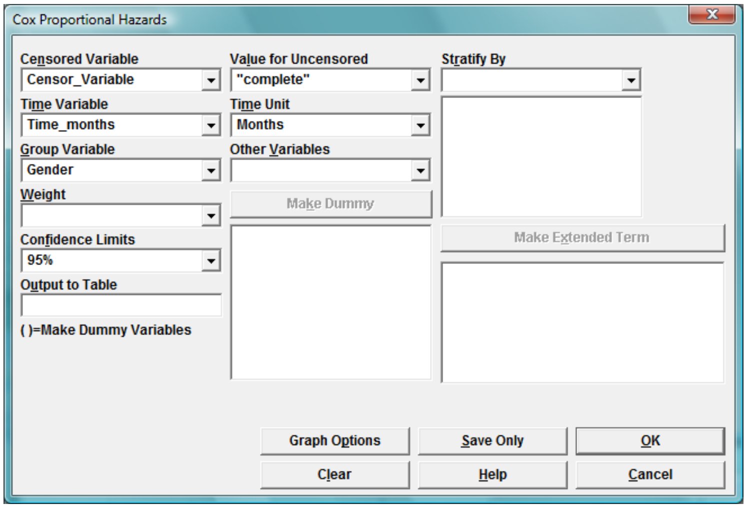
Change the ***Group Variable*** as needed for your comparison.

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:


Cox Proportional Hazards

Censored Variable: Censor_Variable

Value for Uncensored: "complete"

Time Variable: Time_months

Time Unit: Months

Group Variable: Gender

Confidence Limits: 95%

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.