Help Index

Main menu options (excluding this help file) include:-

File Menu
Edit Menu
Statistics Menu
Calculate
Options Menu
Spreadsheet
Graph Menu

Other topics:-

Data entry
Missing values
Altering data file size
Toolbar
Printing data
Data file structure

File Menu

The File menu contains commands that enable you to create new data files, load and save existing files, print the data, or exit from the program.   Some of these functions have buttons on the <u>toolbar</u> to allow rapid access.

New   Resets all variables and allows entry of a new data set.

Open   Allows you to load and use a previously saved data file.

Merge   Add variables stored in another file to the current data.

Delete   Erase a previously save data file.

Save   Write the data to a file.

Print   Output either variable statistics (mean, SE, n, etc) or data lists to a printer.

Exit   Quit the program.

Edit Menu

The Edit menu includes commands that enable you to edit and manipulate the information stored in the data files as well as copy it in to the clipboard. Some of these functions have buttons on the <u>toolbar</u> to allow rapid access.

Edit or Add data   Add new data or alter existing data.

Change variable   Alter variable names (eg after pasting data into the spreadsheet).
name

Delete variables   Remove a variable from the data file.

Transform data   Convert data into log, square root or reciprocal, or return to original format.

Simple maths   Allows new data to be calculated from existing data.

Copy to clipboard Stores raw data or means and S.E. in to clipboard   (see Options menu for altering defaults).

Statistics Menu

The Statistics menu allows selection of the desired analytical test.   These tests are divided into parametric and non-parametric methods.   Where 't' and 'F' values are calculated, the exact probabilities will also be determined. Tests currently available include:-

      Mean, SE, SD and %CV
      Unpaired t-test
      Paired t-test
      1 x ANOVA
      2 x ANOVA
      Linear regression
      Comparison of regression lines (by ANOCoVA)
      Mann Whitney U-test
      Wilcoxon rank pairs test
      Chi Square test
      Kruskal Wallis ranked ANOVA
      Spearman rank correlation

Once the required test has been selected, you must click on the Calculate option of the Main menu to start the procedure.

Calculate

After selection of the desired statistic, the Calculate option initiates the procedure by opening the 'Chooser window' which allows selection of the required variables for the analysis.

The desired variables are placed into the 'Selected' list by either double clicking on the variable name or by clicking once on the name and then clicking on the 'Select' button.   Double clicking on the variable names listed in the 'Selected' list (or using the 'Remove' button) will return them to the unselected variable list.

When all the desired variables have been selected, clicking the 'OK' button will result in the display of the statistical result.   Closing the result window returns the user to the 'Chooser' window.   Other variables may then be selected for analysis or the process can be stopped by clicking the 'Exit' button.

After exiting from the 'Chooser' window the test results will be output to the printer.   <u>(See Printing data)</u>

Options Menu

This menu allows the user to alter certain defaults of the program.   These include:-

Decimal points   alter display output default from 3 down to no decimal points.

Missing values   Incomplete data points must be indicated by use of the missing values
                   indicator (and NOT a blank).   The default value is -1.

Clipboard data   This option allows either raw data, mean and S.E. or mean and S.D. values
                   to be copied into the clipboard with the copy command is used.

Path for temp files   Some functions need to temporarily write to disk.   This option
                   indicates to preferred drive (default is C:).

Toolbar visible   Clicking on this option shows or hides the toolbar.

Spreadsheet   To facilitate data entry, choosing the 'row' or 'column' option causes the
                   active cell to move across or down the spreadsheet after pressing the
                   Enter key.

Graph variables   You may select to graph variable means (as column, bar, line, etc graphs)
                   or draw a scatter plot of the variable values

Spreadsheet

Clicking this item of the Main Menu activates a spreadsheet for data entry or viewing.   Additional rows or columns can be created as an alternative to the data entry routine.   Double clicking on a cell may be necessary to allow the spreadsheet to commence data entry.

Data copied from Excel can be pasted directly into this spreadsheet.   Please note that the 'Paste' function removes all existing data before pasting in the clipboard data.   If you want to add new data to existing data use the File Merge option (see <u>File Menu</u>).   Care should also be taken with <u>missing values</u>.

The Copy button places all the raw data into the clipboard (regardless of the <u>Options Menu</u> setting).

New variables created within the spreadsheet are named by the program and not the user.   These names can be changed however by using the appropriate option in the <u>Edit Menu</u>.

The direction of active cell movement after pressing the Enter key, that is across or down, may be altered via the <u>Options Menu</u> setting).

The maximum number of variables and items per variable is stored in WIN.INI under [LabStats].   The default values are 50 by 150.   (See <u>Altering data file size)</u>

Graphs

This utility offers a quick way to visualise either your data points or graph the variable means.   As a consequence, it is necessary to click on the <u>Calculate</u> option before proceeding with the graphing option.   Use the <u>Options Menu</u> to select which type of graph format you want.   The default is to graph variable means.

The variables of interest are selected in a similar manner to that used to select variables for statistical analysis.   These selected variables become 'series 1' data.   Click on OK to proceed.

The data is initially presented in the default 'column' format.   Several graph features may be altered though use of the 'options' buttons on the right side on the screen.   To use these, click on the down arrow and selected to desired option.   New legend and X-axis labels may be entered (note: legend names must be less than 6 characters in length or it will not be visible).   If a 3-dimensional graph format is chosen, additional options will appear which allow rotation and altered viewing perspective of the graph.

A new series of data is overlayed on the present graph by clicking on the 'New Data' button and repeating the procedures described above. Sophisticated graph requirements may need use of specialist graphics software.

If graph type option is set to 'values', you will be promted to select one x-axis variable and up to four y-axis variables.   After clicking on OK, a scatter plot of each xy point is displayed with a different coloured symbol being used for each variable.   The 'join points' option connects variable points with a line. The scale of the X and Y axis can be altered by entering new maximum and minimum values.

Data entry

Selection of the 'File New' option will open a data entry window.   Available options are displayed on the various buttons with the inappropriate ones shown in grey rather than black print.

By clicking on the 'New Variable' button the cursor places itself in a text box to allow entry of the name.   After pressing 'enter' or clicking on the 'OK' button, the variable name is displayed in the variable list box and the cursor moves to the data value box.   The number to the left of the box is the item number of the data value about to be entered.   The item value is stored by simply pressing 'enter' or clicking on the 'Next' button.   The item counter increments by one and another value can be entered.   (see missing values)

A new variable can be entered at any time by clicking on the 'New Variable' button and repeating the process described above.

Should you want to check a previously entered value you can use the 'Previous' and/or 'Next' buttons to move back and forth through the currently selected variable's item values.   If a value has been incorrectly entered, highlight the current value and re-enter the new value.   Press 'enter' to ensure to value is stored in the data array.   A different variable's data can be checked by simply clicking on the required one in the variable list box.   The spreadsheet offers an alternate way to enter or check data values and may the best way to enter data for Chi-square analysis.

Missing values

If your data set is incomplete, it is important to distinguish the missing value from an 'empty' value.   The program uses 'empty' values to locate the end of a variables data points.   Hence if you know that a given variable should contain 9 data values but the program only reports a sample size of 3, then it is most likely that the forth item has been incorrectly entered as an 'empty' value.   If the value is genuinely missing, use the missing value indicator in its place (the default is -1).
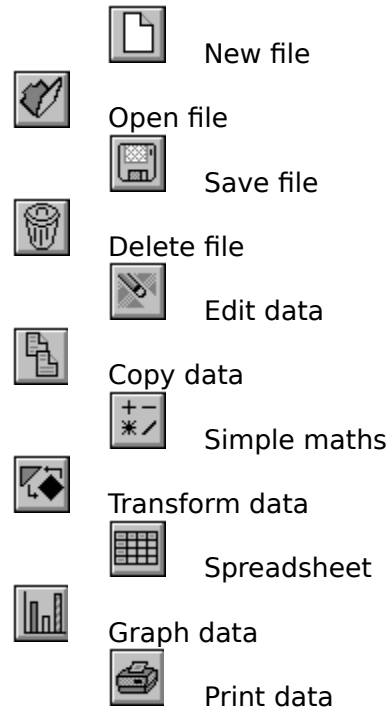
Any value (but an 'empty' value) can be used as the missing value indicator. The selection of the required value is outlined in the <u>Options Menu</u>.

Altering data file size

The default data file size is 50 variables, each of which can contain up to 150 items.   This can be readily altered however by editing the settings which are stored in WIN.INI under the heading [LabStats] using a text editor such as Window's Notebook.   The maximum settings will depend upon your own system's memory capability.

Toolbar

The toolbar can be turned on or off by selecting the option in the <u>Option Menu</u>.   The buttons offer a quick route to commonly used functions.   A brief outline of the the buttons and their use is shown below.

New file

Open file

Save file

Delete file

Edit data

Copy data

Simple maths

Transform data

Spreadsheet

Graph data

Print data

Printing Data

All information stored or calculated by 'Lab Stats' can be directed to the printer.

The variable values can be printed by selecting the <u>File Menu</u> Print option and then choosing 'Print data lists'.

Variable means, standard deviation, standard error, coefficient of variation and sample size can be printed by selecting the 'File Menu Print' option and then choosing 'Print variable statistics'.   Alternatively, click on the 'Printer button' on the <u>toolbar</u>.

Calculated results are produced by clicking on the 'Print' button in the result's window.   These results are stored until the 'Exit' button is clicked in the <u>Chooser window</u>, at which time all results are printed.   This allows results from one sort of analysis (eg t-tests) to be printed on the one page of paper.

Data file structure

Data entered into 'Lab Stats' is stored in two files.   Both files have the same file name given to the data when it is saved however one file receives a file extension of '.LSW'   (ie   *filename*.LSW and *filename*.).

The first file contains the numerical data in a single column with each variables data separated by an asterisk.   This marker informs the program to assign the next set of numbers to the next variable.   Alpha-numeric data may also be stored (eg names etc) if required.   The program detects variables whose items are not numeric and will automatically prevent selection of these in the analytical and graphical routines.

The names assigned to each variable are stored in the second file.

All data is stored as string characters.   Both files are deleted by the 'Delete file' option of the <u>File Menu</u>.

Mean, SE, n

Selection of this option (the default on start up) results in the calculation of each variable's mean, standard deviation, standard error, coefficient of variation and sample size.

These results are displayed in a 'results window' which is closed by double clicking in the small control box on the top left-hand side of the window.

This information can be printed by clicking on the 'printer' button of the toolbar or by selecting 'print variable summary' from the File Menu.

Unpaired T-Test

This test is used to determine whether two independent samples are drawn from the sample population.

After selecting the 'Unpaired t-test' option of the <u>Statistics Menu</u>, you then click the mouse pointer on <u>Calculate</u>.   The <u>Chooser window</u> then appears and you can click on the names of the required variables.   If anything other than 2 variables are selected, the user is warned of the error and can then alter the selection.

The calculated result is displayed in a <u>results window</u> which shows the variable names, means and standard errors.   The t-value, degrees of freedom and corresponding probability are shown.   The result can be printed by clicking the 'print' button.

<u>See Printing data</u>

Paired T-test

This test is used to compare groups of related or 'paired' individuals or things.   While this test is more sensitive to directional changes in the mean after treatment than the <u>unpaired t-test</u>, it does lose out by halving the number of degrees of freedom.   A common usage is in analysing self-pairing where two measurements are obtained on the same individual.

1 x ANOVA

This program is used for comparing 3 or more samples.   The advantage of this method of analysis over multiple t-testing is that the probability of committing a type 1 error, that is rejecting the null hypothesis (ie that the samples are the same) when it is true, is less.

The program can handle variables with different sample sizes.   The resultant 'F value', degrees of freedom and probability are displayed in a <u>results window</u>.

If a significant 'F value' is calculated, differences between the variables can be determined by Fischer's Least Significance Difference (LSD).   This procedure is performed by selecting the 'LSD' button.

The sums of squares (SS) and mean sum of squares (MS) values can be viewed by pressing the 'SS table' button.

2 x ANOVA

Two way analysis of variance allows the effects of two variables on a third to be analysed.   A significant 'F-value' for a column or row treatment indicates that there is a change in the third variable values due to that treatment.

The interaction between the row and column treatments is also calculated. It is advisable to graph the mean value of the third variable against the column treatments for each row to make interpretation of the interaction effect clearer.

The program is not designed to cater for unequal group sizes.   However, where there is one or two missing values, it will substitute the variable mean to allow an estimate of the 'F-value' to be made.

When selecting the variables of interest in the Chooser window, the program assumes that the order of selection is:-

R1C1	R1C2	R1C3	R1Cn
R2C1	R2C2	R2C3	R2Cn
RnC1	RnC2	RnC3	RnCn

with the column variables in row 1 being entered before column variables in row 2 and so on.   Once all the variables have been selected, the user is asked how many column treatments there are and thus it can calculate the number of rows.

Linear Regression

This program calculates the Pearson correlation coefficient, the line of best fit and the significance of the regression line.

The first variable entered in the <u>Chooser window</u> is assumed to be the X-variable.   This is clearly shown in the <u>result window</u>, together with the equation of the line, the correlation coefficient and level of significance.

Comparison of Regression Lines

The routine uses analysis of co-variance to determine whether two regression lines can be considered to be the same.   It assumes that the lines are linear (Pearsons correlation coefficients are calculated so that the need for data transformation can be determined).

The program will determine whether the lines are equivalent in terms of both slope and elevation.

It is required that the X and Y variables for line 1 be stored as two distinct variables.   Similarly for line 2.   At the <u>Chooser window</u>, select the line 1's X and Y variables first followed by line 2's X and Y variables.

The 'F-values' and associated significance levels are then displayed.

Mann-Whitney U Test

This test is an non-parametric method for determining whether two independent samples differ.   It does not require normality of distribution or homogeneity of variance for the two groups but does assume that at least ordinal measurements have been made (eg social economic status; high, middle and low).

The program can handle both large and small sample sizes although the smaller of the two groups must be selected first.

Appropriate tables are required to determine the level of significance associated with the calculated 'U-values'.   However, where larger samples sizes are used (ie > 20), 'Z-values' are determined and probabilities can be looked up in normal distribution tables.

Wilcoxon Ranked Pairs Test

This test is a non-parametric method for examining the differences between related or paired data.   Like the Sign Test, this procedure monitors changes in the direction of the differences within pairs, but it is more powerful as it also takes into account the magnitude of the change.

For small sample sizes, a set of appropriate tables are required to determine the level of significance.   For larger sample sizes (ie > 25), the Z value for a normal distribution table is calculated.

Chi Square Test

This test is a non-parametric method for comparing the frequency of occurence of discrete categories between two or more independent groups. For example, this test can be used to compare the voting preferences between the sexes.   The program can analyse 2 x 2 tables or larger 'contingency' tables.

For 2 x 2 tables, the program automatically determines whether the expected frequencies are sufficiently large to make the chi-square test valid. If not, the Fisher Exact Probability test is calculated instead.   Where the expected frequencies are big enough to allow chi-square estimates to be made, Yates correction is applied.

The purpose of the chi-square test as applied to a contingency table is to examine the hypothesis of independence between variables.   The χ_2 value does not serve as a measure of association.   However, an estimate of the level of association can be gained by calculating the contingency coefficient, or for 2 x 2 tables, Cramer's phi coefficient (Φ).   The level of significance (probability) for the χ_2 value is also calculated.

The best way to enter data for the Chi-square test is to use the <u>spreadsheet</u> method.   The program assumes the required data is   in the same column and row format as is displayed in the spreadsheet.   The columns are selected as 'variables' via the <u>Chooser window</u> with the rows being known by default.

Kruskal-Wallis ANOVA

The Kruskal-Wallis ANOVA by ranks is a very useful non-parametric test for deciding whether 3 or more independent samples are from different populations.   The test requires at least an ordinal measurement of the variables (eg social economic status; high, middle and low).

Once again, special tables are required to look up the associated levels of significance for the calculated value.   However, if more than 3 samples are analysed, Chi-squared tables may be used.

Spearman Rank Correlation Coefficient

This non-parametric test measures the association between data that can be ranked into two ordered series (ie at least ordinal measurements).

Special tables allow probabilities associated with the calculated 'r_s' value to be determined.   However, if there are 10 or more pairs of data, 't-values' can be calculated together with the probability values.