KRUSKAL-WALLIS TEST PROGRAM

Introduction

This program to perform a Kruskal-Wallis test is based on the description of this test given by Conover in his 1980 book “Practical Nonparametric Statistics”. The About box, as well as the References page in the online Help give the full reference to this book chapter.*

The first version of this program was implemented in 1986 on a Sinclair Spectrum computer and was partly based on the paper by E. Theodorsson-Norheim in Comp Methods Programs Biomed (1986). Since then, the code has been translated into Fortran and Pascal, and versions have been implemented for DOS and Windows operating systems. Somewhere along the line statistical functions from Numerical Recipes were added to replace the fixed Chi-square and t-values. The current version is developed in the Delphi developing environment for Windows and was extended with routines to read and write Excel sheets. It will run in all versions of Windows, provided that the decimal separator is configured as a point.

When you meet someone who is interested in this program, please do not give him or her a copy but ask them to request their own original. An e-mail to **biolab-services@amc.uva.nl** with the subject: **kwtest** will be answered automatically without delay. That way, we can keep track of the program and send all users information on bugs or updates. Please use my personal e-mail adress only for questions and suggestions.

February 2004, Jan Ruijter

---

* Disclaimer

This program to perform the Kruskal-Wallis test is based on the description given in Conover, Practical Nonparametric Statistics, 1980 (for complete reference: see the About box). By opening and using this software you acknowledge that you have read this book chapter, understand it, and agree with its conclusions. Therefore, you assume all responsibility and liability for the selection of this software program to achieve your intended statistical tests, and for the conclusions you draw from the results of these tests. The authors cannot be hold responsible for any consequences of the use of this program.

General Information

The Kruskal-Wallis Test is a non-parametric test to analyze the difference between more than 2 independent random samples and as such the Kruskal-Wallis test can be considered to be a nonparametric one-way ANOVA. Only when the Null Hypothesis, that the medians of all groups are equal, is rejected a MULTIPLE COMPARISON of groups is carried out to determine which groups differ from each other. Minimal requirement of the test is that the values of the test variable are at least measured on an ordinal scalebecause the test is based on the rank order of the observations. When only two groups are compared the result of the Kruskal-Wallis test is the same as the Mann-Whitney test. For further details, see the Reference page in the online Help

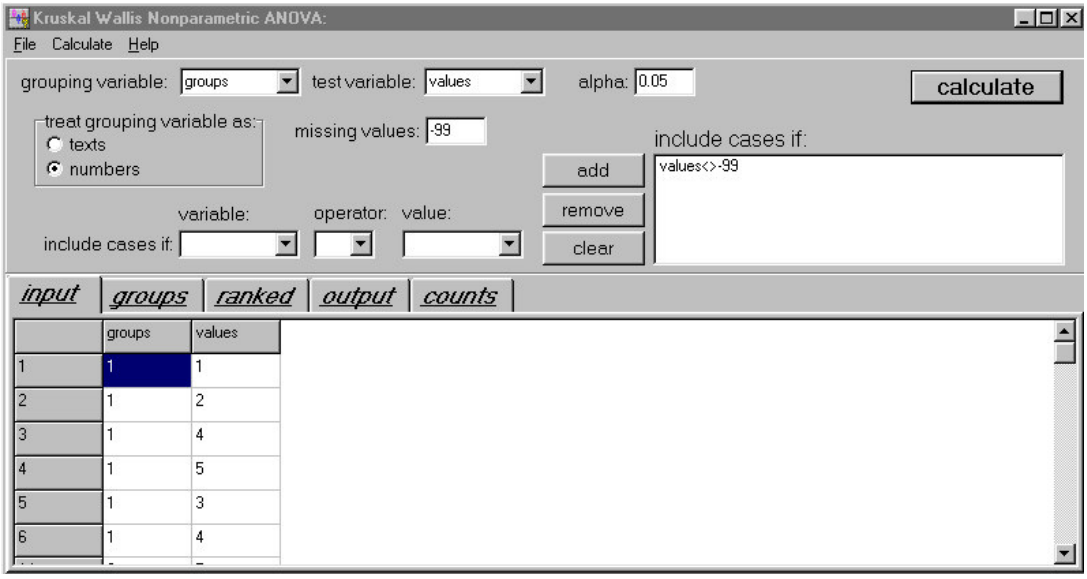


Figure 1: Screen shot of the Kruskal-Wallis test program. In the top part of the window you see the drop-down boxes to choose the grouping and test variables and the edit field to specify the significance level. The remainder of the header is dedicated to the selection of cases to include or exclude in the test. The bottom part of the window contains the Tab-pages displaying the input, calculations and output of the test

Help and Warnings

This Help text is also available from the main menu of the program. Additionally, when the program displays a warning, the help button on the warning window will most of the time display the appropriate section of the online Help.

The program in short

The basic operation of the Kruskal-Wallis Test program consists of:

1. Enter data manually or read data form a text file or an Excel worksheet
2. Select a grouping variable and a test variable
3. If you want to: define a selection to exclude cases from the calculation
4. Specify the significance level (default 0.05)
5. Press the calculate button

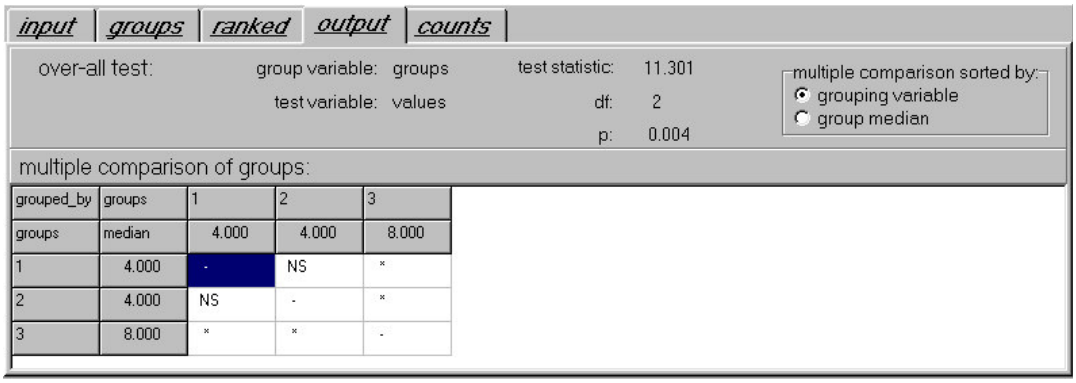
The five Tab-pages (Figure 1) in the lower part of the screen will then display the following output:

input page
After the calculation the data on the Input page are sorted in order of the Grouping variable.

groups page
In the upper panel of the Groups page you see the group identification values with the ranksum (not interesting but used in the test) and the Median per group.
The bottom half of the Groups page shows a table with the composition of each group. This table only shows the cases that meet the selection criteria (see: Define Selection).

ranked page
The Ranked page gives you the ranks assigned to each value (ties are assigned a mean rank for the group of ties. This ranking is sometimes informative because it shows the overlap between groups. The order of the data is the same as in the lower panel of the Groups page. Only values that meet the selection criteria are ranked.

output page
The top-part of the Output Page gives the result of the OVERALL test of the Null Hypothesis that all groups have the same median. The ouput gives the calculated test statistic, the degrees of freedom (number of groups minus 1) and the P-value of the test-statistic assumig that the latter is Chi2- distributed (see Reference).



grouped_by	groups	1	2	3
groups	median	4.000	4.000	8.000
1	4.000	-	NS	*
2	4.000	NS	-	*
3	8.000	*	*	-

Figure 2: Output page of the Kruskal-Wallis test program. In the top part of the page gives the result of the overall test of the null hypothesis that the medians of all groups are equal. When this null hypothesis is rejected, the bottom part of the page gives the result of the multiple comparison of groups. An asterisk indicates a pair of groups that is significantly different for the significance level alpha set at the top of the program window.

When the P-value is lower then the alpha value you gave on top of the main screen (default significance level of 0.05) the Overall test rejects the Null Hypothesis and only then a MULTIPLE COMPARISON of Groups is carried out to determine which groups differ from each other. The result of this comparison is then given in the bottom part of the output page. Interpretation of this output is often much easier when you sort the groups by their Median (radio buttons on the top right).
The Output page can be saved to a text file or to Excel; the multiple comparison table is saved as displayed.

counts page
The Counts page shows you a cross-tabulation of the frequency with which each value occurs in each group. This page was once added as a first step to construct a column scatter graph per group. This plan is still on the shelf.

Manual Input
NOTE: It is recommended to enter the data in an Excel worksheet and use "File - Read data from Excel file".
If you want to enter data directly into the K-W test:
1. Choose "File - Manual Input of data" from the main menu
2. in the 'Manual Input' dialog
    a. Give the number of variables (at least 2)
    b. Give the total number of cases
    c. Press OK
3. Enter the data in the input grid.

You may enter a missing value by using the value that you define in the missing value box on the main form.
Save Input
After Manual Input the menu item "File - Save manual Input to text" becomes enabled. The manual input will be saved to a space-delimited text file.

Read Text file
NOTE: It is recommended to enter the data in an Excel worksheet and use "File - Read data from Excel file".
If you want to read data from a text file, prepare a text file as follows (Figure 3, inset):
1. Start with a header row with variable names separated by spaces. The names themselves MAY NOT contain spaces.
2. Type the data as one row per case.
    - All values have to be separated by spaces
    - Coded values MAY NOT contain spaces.
    - Missing values have to be entered as defined missing value:
    - Do NOT leave an empty entry; each row should contain as many entries as there are variable names in the first row.
3. Finish with ONE empty row: press Enter ONCE at the end of the last data row
4. Save the file as a ASCII text file

When you use a spreadsheet program to enter the data, save the sheet as a space delimited text file (a *.prn file in Excel), after making sure that the column width is wide enough to accommodate all entries.
The K-W program reads the header row to determine the number of columns.
Thereafter it tries to read that number of values from each row. Spaces are considered delimiters between the column names and between the row entries.

Read Excel file
NOTE: Excel has to be running, and the book (spreadsheet) with your data should be opened. When data are read from Microsoft Excel spreadsheets, the decimal separator in Windows should be set to be a decimal point. Excel should NOT be in edit mode. To make sure of this: place the cursor outside the data range and press 'Enter'.

Prepare a worksheet with a row containing the variable names (at least a grouping variable and a test variable). Enter the data as a row of data per case (Figure 3). Values of

the grouping variable may be coded. Avoid completely empty rows. Cells with missing values can be left empty.

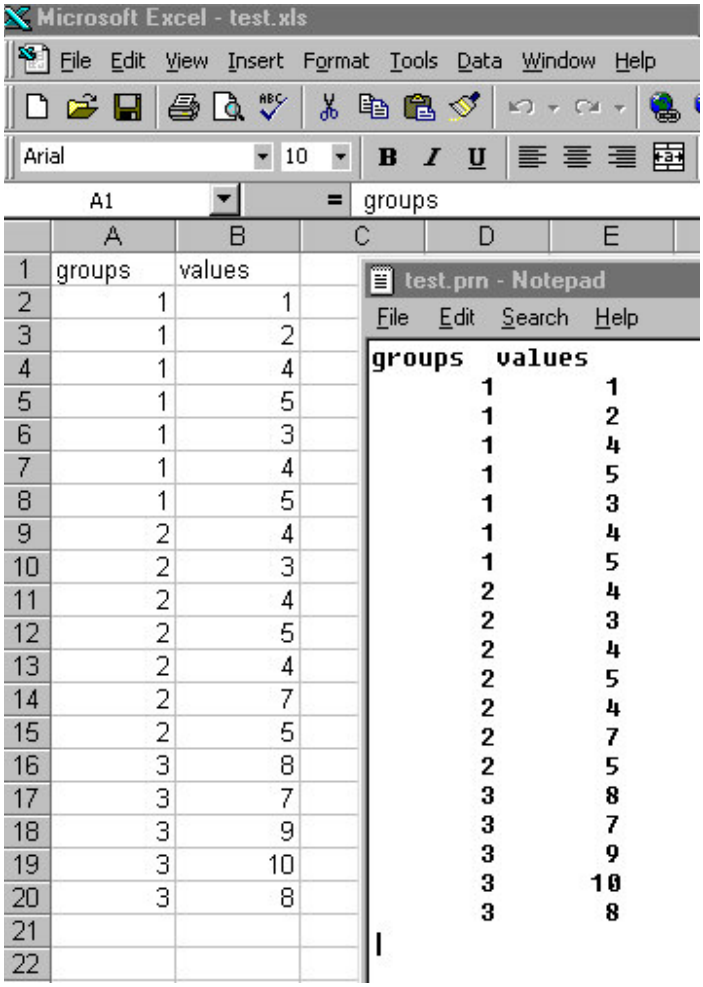


Figure 3: Examples of data input from an Excel sheet (background) and from a text file (inset).
Note that in both input modes the first line of data consists of variable names which may not contain spaces. The values of the grouping variable may be coded alphanumerically, but may not contain points. The values of the test variables should always be numerical. In the Excel sheet empty cells are allowed. In the text file missing values should be entered as a defined missing value, no empty entries are allowed in the text file. The cursor in the text file has to be positioned on an empty last line. This version of the Kruskal-Wallis test program expects decimal points.

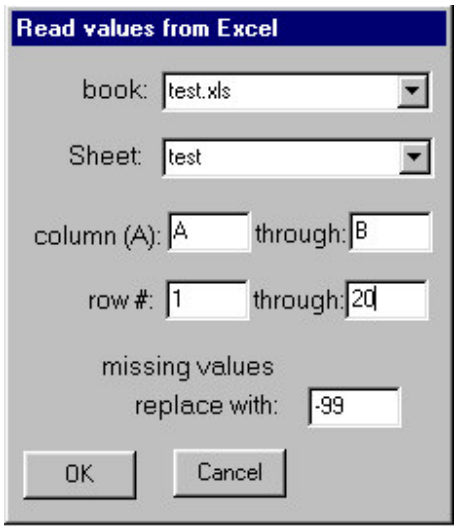


Figure 4: Read from Excel dialog. this dialog window pops-up when you choose the menu item "File - Read data from Excel". Excel has to be running and the book with your data has to be opened. You can choose the right book and sheet from the lists in the drop-down boxes. Give the columns and rows that define the data range that has to be read. When you have a defined code for missing values you have to give this value in the edit field.

To read data from an Excel sheet, choose “File - Read data from Excel file”. In the ‘Read values from Excel’ dialog (Figure 4):

1. Choose the Excel book that contains your data
2. Choose the sheet that contains your data
3. Give the column letters and row numbers of the range of cells that you want to read.
4. Give the value that should be substituted for empty cells (=defined missing value)

Grouping and Test variables

A grouping variable and a test variable have to be selected to run a Kruskal-Wallis test. Grouping variables can be text or numerical. Because the grouping variable has to be discrete the program tests for the presence of a decimal point: if a decimal point is found the variable is not accepted as a grouping variable. Therefore, a coded grouping variable MAY NOT contain points.

Test variables should be numerical and should code for events that are at least measurable at an ordinal scale. The program tests for the presence of letters: if found, the variable is not accepted as a test variable.

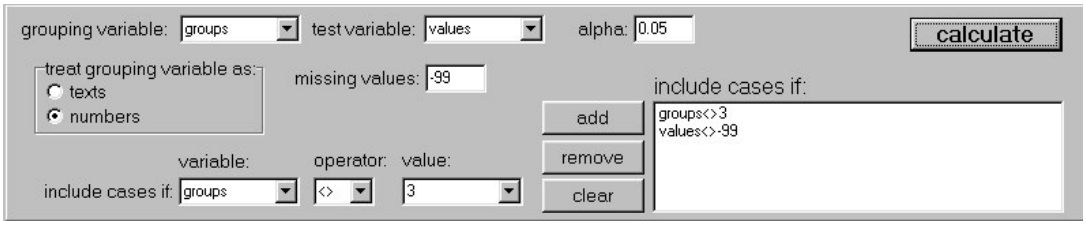


Figure 5: When not all cases are to be include in the test, you can define selection strings with the variable, operator and value drop-down boxes. Use this option when you have more than one defined missing value. The selection string for the default missing value is added by the program when you press the calculate button. Selection strings in the window will be combined with a logical AND. Do not forget that you need good reasons to exclude data from the analysis. This option is only added to this program to enable doing different tests on different variables in the same data file.

Define Selection

When you want to test a subset of your data you can define a selection.

1. Choose a selection variable. This may be any variable in your data set.
2. Choose an operator. Only the operators equal (=) and unequal (<>) are available.
3. Choose a value. The program expects a limited number of possible values for the selection variable. Choosing a continuous variable will result in a warning.
4. Press the ‘add’ button to add the selection to the selection strings window.

When you ‘add’ more than one selection string, the strings will be combined with a logical AND. The program itself adds one selection string when you press ‘calculate’ to exclude the defined missing values in the test variable (Figure 5).

You can check the effect of a selection on the Groups and Ranked Tab pages.

To remove a single selection string, highlight it in the selection window and press the ‘remove’ button. To clear all selections, press the ‘clear’ button.

Missing Values
When an observation failed, the missing value can be replaced by a defined value. During the test, cases with a missing value in the test variable will be ignored. The default missing value is -99 but you can define another value by entering it into the 'missing value' field on the main form or in the 'Read values from Excel' dialog.
NOTE: The program adds one selection string to the selection window when you press 'calculate' to exclude the defined missing values of the test variable from the calculations.
When more then one value is used to represent missing values or when you have missing values in your grouping variable, you can use a selection to exclude cases with these values from the test. (See: Define Selection)

Save results to Text file
Results of the Kruskal-Wallis test, as they are given on the 'output' page, can be saved to an ACSII text file:
1. Choose "File - Save results to Text" from the main menu
2. In the 'Save' dialog:
   a. Choose a directory
   b. Give a file name
   c. Press OK.

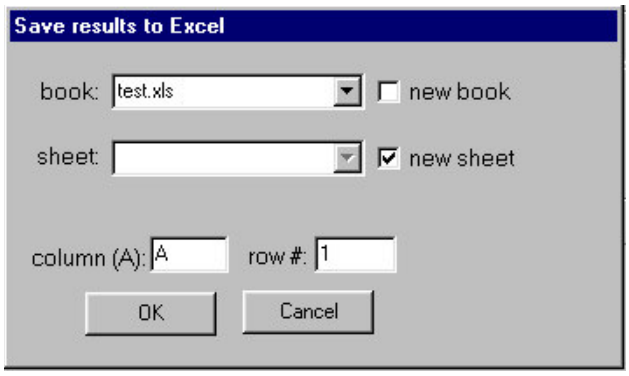


Figure 6: Save to Excel dialog. This dialog pops-up when you choose the menu item "File - Save to Excel". Excel has to be running and the sheet that you want to save to has to be opened. Because the program will overwrite the contents of the spreadsheet without warning, it is highly recommended to save to a new sheet. If you choose to save to an existing sheet, be sure to give a starting position in an empty part of the sheet.

Save to Excel file
NOTE: Excel must be running to save results to Excel
Results of the Kruskal-Wallis test, as they are displayed on the 'output' page can be saved to an Excel spreadsheet:
1. Choose "File - Save results to Excel" from the main menu
2. In the 'Save to Excel' dialog (Figure 6):
   a. Choose an Excel book and a sheet (a new sheet is recommended because data already present in the sheet will be overwritten)
   b. Give a column letter and a row number where you want to start writing
   c. Press OK

When you have more than 250 groups the results CANNOT be saved to Excel.
Figure 7 shows the results of the Kruskal-Wallis test as they appear in Excel.

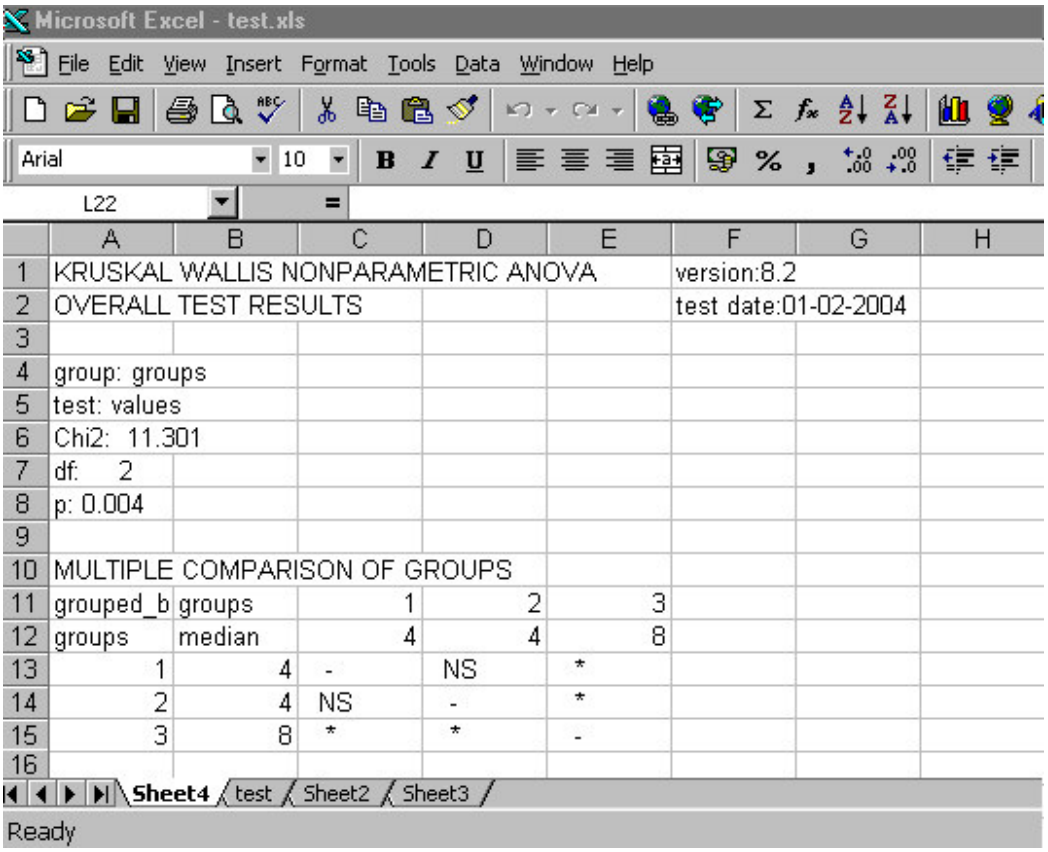


Figure 7: Results of the Kruskal-Wallis test program saved to Excel. The results are saved as they are displayed on the Output tab-page. Note that the program version and test-date are added to the output for future reference.

References

This implementation of the Kruskal-Wallis test is based on the description of this test in:

Conover W.J. Practical Nonparametric Statistics. 2nd edition, 1980 pages: 229 – 237. J. Wiley and Sons, New York.

A computer program in BASIC that performs the same statistical test has been published by:

Theodorsson-Norheim E. KRUSKAL-WALLIS test: BASIC computer program to perform non-parametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. Comp Methods Programs Biomed 23: 57-62, 1986

The P-value of the Chi2-statistic, as well as the t-value used in the multiple comparisons, are calculated with the statistical algorithms published in:

Press W.H., Flannery B.P., Teukolsky S.A. and Vettering W.T.
Numerical Recipes. The art of scientific programming. Cambridge Uninversity Press, Cambridge.