

# **Computer Program for the Kendall Family of Trend Tests**

Scientific Investigations Report 2005–5275

# **Computer Program for the Kendall Family of Trend Tests**

By Dennis R. Helsel, David K. Mueller, and James R. Slack

Scientific Investigations Report 2005–5275

**U.S. Department of the Interior**  
**U.S. Geological Survey**

**U.S. Department of the Interior**  
P. Lynn Scarlett, Acting Secretary

**U.S. Geological Survey**  
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, Virginia: 2006

For product and ordering information:  
World Wide Web: <http://www.usgs.gov/pubprod>  
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment:  
World Wide Web: <http://www.usgs.gov>  
Telephone: 1-888-ASK-USGS

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

*Suggested citation:*

Helsel, D.R., Mueller, D.K., and Slack, J.R., 2006. Computer program for the Kendall family of trend tests: U.S. Geological Survey Scientific Investigations Report 2005–5275, 4 p.

# Contents

Abstract.....1  
Introduction.....1  
The Tests .....1  
The Program .....2  
Tests and Input Formats.....2  
Program Output.....4  
Summary.....4  
References Cited.....4

## Table

1. Example data files for the Kendall family of trend tests.....3  
[data files and program available at <http://pubs.usgs.gov/sir/2005/5275/downloads/>]

# Computer Program for the Kendall Family of Trend Tests

By Dennis R. Helsel, David K. Mueller, and James R. Slack

## Abstract

The Seasonal Kendall (SK) test for trend was developed by the U.S. Geological Survey and has become the most frequently used test for trend in the environmental sciences. Recently the test was modified to form the Regional Kendall (RK) test for trend. In this form, trends at numerous locations within a region are tested to determine whether the direction of trend is consistent across the entire region. Computer code developed at the USGS in the 1980s to perform the SK test is no longer widely available. Other versions written by other scientists may or may not be easily available, and may require commercial software in order to be run. These other versions do not explicitly compute the RK test. Therefore, the original code for computing the SK test has been repackaged into a program that runs under the Windows operating system. This program may be used to verify that other implementations of the test give the same results as the original. The program also provides a means for computing the RK test and the simpler Mann-Kendall test for trend.

## Introduction

The Seasonal Kendall (SK) test for trend (Hirsch and others, 1982; Hirsch and Slack, 1984) was developed by the U.S. Geological Survey (USGS) in the 1980s to analyze trends in surface-water quality throughout the United States. This test has since become the most frequently used test for trend in the environmental sciences and has been applied to a variety of media at many different locations. Applications have included tests for trend in biologic community structure (Moore, 1987), estuarine salinity (Wiseman and others, 1990), lake water quality (Reckhow and others, 1993), and atmospheric chemistry (Cortes and Hites, 2000). Recently the test was modified by Helsel and Frans (in press) to form the Regional Kendall (RK) test for trend. In this form, trends at numerous locations within a region are tested to determine whether the direction of trend is consistent across the entire region.

Computer code developed at the USGS in the 1980s to perform the test is no longer widely available. The USGS currently uses a version that runs using S-Plus, a commercial

software package (Slack and others, 2003). Without access to S-Plus, no use can be made of this version. Other versions written by other scientists are available (Reckhow and others, 1993), but many also require commercial software in order to be run. None of these versions compute the Regional Kendall test as a specific adaptation. Therefore, the original code for computing the Seasonal Kendall test, written by James R. Slack of the USGS, has been repackaged into a DOS executable program that runs under current versions of the Windows operating system. This program may be used to verify that other implementations of the test give the same results as the original. The program also provides a means for computing the Regional Kendall test and the simpler Mann-Kendall (Mann, 1945) test for trend.

## The Tests

The Seasonal Kendall test performs the Mann-Kendall (MK) trend test for individual seasons of the year, where season is defined by the user. It then combines the individual results into one overall test for whether the dependent (Y) variable changes in a consistent direction (monotonic trend) over time. Developed by Hirsch and others (1982) and applied to the USGS collection of long-term water-quality records across the United States, the test was modified by Hirsch and Slack (1984) to account for serial correlation in the time series of the Y variable. Serial correlation violates the assumption of independence of data—that there is no short-term correlation between observations. When serial correlation occurs, p-values for significance tests will be too small, and trends may be incorrectly “found.” Hirsch and Slack (1984) suggest using the adjusted p-value whenever there are more than 10 years of data, as it commonly takes this much data to detect serial correlation, if present. The current code computes a p-value for the test both before and after the adjustment for serial correlation. An overall measure of the trend slope also is provided.

The Regional Kendall test extends this concept to spatial locations rather than seasons. A Mann-Kendall test is computed for individual locations and results are combined into one overall test for consistent regional trend (Helsel and Frans, in press). The test is applicable to data where observations

## 2 Computer Program for the Kendall Family of Trend Tests

have been made annually at numerous locations, such as at water wells, and one overall test is desired to determine whether the same trend is evident across those locations.

The Mann Kendall test computes Kendall's tau non-parametric correlation coefficient and its test of significance (Helsel and Hirsch, 2002) for any pair of X,Y data. If X is time, the test is a test for trend in the Y variable (Mann, 1945). However, the program may be used to compute the correlation between any set of paired observations, even when X is not time. In this case, the program no longer computes a trend test, but simply the Kendall's correlation coefficient measuring the association between X and Y.

### The Program

Kendall.exe (which can be downloaded from <http://pubs.usgs.gov/sir/2005/5275/downloads/>) performs the Seasonal Kendall, Regional Kendall, and Mann-Kendall tests for trend. The program runs using the Windows XP operating system within the DOS terminal window—its input and output uses DOS rather than Windows. After starting the program (see below) the user must type in the name of the input file. Either this file must be in the same directory as the program, or the entire file structure location of the data file must be input. No “browsing” option is given. The user also types in a name of the output file to be created by the program. Test results, including equations and p-values, are written to this output file. Diagnostic messages, including statistics about the input data and error messages, if any, are written to the DOS window. The program retains the somewhat archaic DOS input and output structure, without use of Windows dialog boxes, to minimize alterations of the original code.

To run the program:

1. Copy the “Kendall.exe” file into the directory that contains the input-data files.
2. Double click on the “Kendall.exe” file name or icon.
3. In the DOS window type the name of the input-data file.
4. In the DOS window type the name of the output file to which results are written.

Data can be adjusted for the effects of an ancillary variable such as streamflow before testing for trend (Hirsch and others, 1991). A LOWESS smooth is used to describe the relation between the Y variable and the ancillary variable. Residuals from the smooth therefore have the relation between Y and the ancillary variable subtracted, removing the effect of the ancillary variable from the Y data. When Y is concentration and streamflow is the ancillary variable, these residuals have been named “flow-adjusted concentrations” (Hirsch and others, 1991). However, the process can be extended to any set of three variables, in addition to its usual application.

The Kendall test is then computed on the residuals from the LOWESS smooth, to test for trend after removing the effect of the ancillary variable. This procedure is available for use with the Seasonal Kendall and Mann-Kendall tests.

### Tests and Input Formats

Input formats are a holdover from the original Seasonal Kendall code and seem archaic by today's standards. Four input structures are available. The first two produce the Seasonal Kendall (SK) test. With the first format (itype = 1), time is given in decimal year, and the number of seasons is subsequently specified. Each year will be divided into the specified number of equally spaced seasons by the program. With the second format (itype = 2), a season is designated for each observation in the input; uneven season lengths are therefore possible with this second format. The third format (itype = 3) produces the Regional Kendall (RK) test, and the fourth (itype = 4) results in the Mann-Kendall test. For the Seasonal Kendall test, a maximum of 120 seasons and 50 years of data are allowed. For the Regional Kendall test a maximum of 120 locations and 50 years of data are allowed. For the Mann-Kendall test a maximum of 500 pairs of X,Y data are allowed. Error messages are produced when these limits are exceeded.

The first line of each input-data file defines which data structure follows and which test is performed. This line is read in fixed format, with specific information located in specific columns. Formats beginning with the letter “i” are integers, followed by the number of columns used. Formats beginning with an “a” are alphanumeric text. The format of this first input line is:

Column	Format	Variable	Contents
1	i1	itype	Test and data structure: 1 = SK test: data as decimal year, data value(s) 2 = SK test: data as year, season, data value(s) 3 = RK test: data as year, site ID, data value 4 = Mann-Kendall test: data as x value, y value
3	i1	lws	Switch for computing LOWESS residuals. Not available for the RK (itype = 3) test. 0 = no. Test computed on Y data. 1 = yes. Test computed on LOWESS residuals.
5–7	i3	nseas	Number of seasons (maximum of 120). Used only for itype = 1; ignored for other itypes.
9–67	a60	title	Title which will be printed on output to identify this set of data.

Subsequent lines of the input-data file contain the data themselves, one line per observation, in the order required for the specified itype. Each line is read using free format, so data values can be separated by blanks or commas.

For itype = 1, the Seasonal Kendall test is computed on data supplied in two columns. The first column is the time, in decimal format (“dectime” -- 1997.50, and so forth). The fractional part of dectime designates the day of sampling by the fraction of the year it occurs in. Seasons are defined within the program as equally spaced segments of a year, based on the number of seasons specified by the user in columns 5–6. The second data column is the Y variable to be tested for trend. If more than one observation falls within a season, the program computes a median value of Y with this input type. This format is the original version of the Seasonal Kendall program as produced by Hirsch and Slack (1984). If the LOWESS option is selected, streamflow or some other ancillary data value is listed in a third column. Adjusted Y values are computed for all observations before the data are divided into seasons.

With itype=2, the Seasonal Kendall test is computed for data with specified seasons. Data are provided in three columns. The first column is the year of sampling as an integer. The second column represents the season as an integer. The program assumes that season 1 starts at the beginning of the

calendar year and that the largest integer number is the last season of the year. There is no requirement that the seasons represent an equal amount of time within a year. The third column is the Y data to be tested for trend. Multiple data values within a season can be entered, each on a separate line. The program will compute a median of these, which will be used in the trend test. If the LOWESS option is selected, streamflow or some other ancillary data value is listed in a fourth column. Adjusted Y values are computed for all observations before seasonal medians are computed.

With itype=3, the Regional Kendall test is computed. Data are given in three columns. The first column is the year of sampling as an integer. The second column is the site ID, unique for each location, represented as an integer number. A maximum of 120 locations may be designated, but no specific ordering is required for the location numbers. The third column is the Y data to be tested for trend. No option for adjustment by LOWESS is available with the RK test.

With itype=4, the Mann-Kendall test for trend is run. It may be used to compute Kendall’s tau and the associated trend line (Helsel and Hirsch, 2002) for any pair of X,Y values, even where X is not time. The first column is the X variable and the second column the Y variable. A maximum of 500 X,Y pairs can be entered. If a LOWESS adjustment is specified, the ancillary variable is listed in the third column.

**Table 1.** Example data files for the Kendall family of trend tests.

[mg/L, milligrams per liter; ft<sup>3</sup>/s, cubic feet per second; µeq/L, microequivalents per liter. Data files available at <http://pubs.usgs.gov/sir/2005/5275/downloads/>]

Test	itype	Input file name	Output file name	Description of data
Seasonal Kendall	1	SK1a.txt	SK1a_out.txt	Dissolved solids (mg/L) in the Cuyahoga River, Ohio.
		SK1b.txt	SK1b_out.txt	Natural logarithms of total phosphorus (tons/day) in the Maumee River, Ohio.
		SK1c.txt	SK1c_out.txt	Total phosphorus (mg/L) in the Whitewater River, Indiana (original Seasonal Kendall program data set).
Seasonal Kendall with LOWESS	1	SKL1a.txt	SKL1a_out.txt	Sk1a data plus streamflow (ft <sup>3</sup> /s).
		SKL1b.txt	SKL1b_out.txt	Sk1b data plus natural logarithms of streamflow (ft <sup>3</sup> /s).
Seasonal Kendall	2	SK2a.txt	SK2a_out.txt	Same data as Sk1a.
		SK2b.txt	SK2b_out.txt	Same data as Sk1b.
		SK2c.txt	SK2c_out.txt	Artificial data, 6 regularly spaced seasons (original Seasonal Kendall program data set).
Seasonal Kendall with LOWESS	2	SKL2a.txt	SKL2a_out.txt	Same data as SKL1a.
		SKL2b.txt	SKL2b_out.txt	Same data as SKL1b.
Regional Kendall	3	RK3a.txt	RK3a_out.txt	Nitrate (mg/L) in ground water, Adams County, Washington.
		RK3b.txt	RK3b_out.txt	Ammonium (µeq/L) in snowpack, Colorado and New Mexico.
Mann-Kendall	4	MK4a.txt	MK4a_out.txt	Same data as SK1a.
		MK4b.txt	MK4b_out.txt	Suspended sediment (tons/year) in Green River, Kentucky.
Mann-Kendall with LOWESS		MKL4a.txt	MK4a_out.txt	Same data as SKL1a.
		MKL4b.txt	MK4b_out.txt	MK4a plus annual sum of mean daily streamflow (ft <sup>3</sup> /s).

## 4 Computer Program for the Kendall Family of Trend Tests

Example input-data files for each type of test are available from the U.S. Geological Survey Web site where this document is located. These files include data sets that use the LOWESS option for the Seasonal Kendall and Mann-Kendall tests. File names for the input data and the associated output, along with a description of the input data, are listed in table 1. Two of these input files are examples that were included with the distribution of the original Seasonal Kendall program (Hirsch and Slack, 1984).

### Program Output

The output file for each test (available at <http://pubs.usgs.gov/sir/2005/5275/downloads/>) lists the correlation coefficient Kendall's tau, the Kendall test statistic S, the standard normal deviate Z determined from S and the variance of S, and the p-value for significance of the trend. Output for the Seasonal Kendall test provides a second p-value, which is adjusted to correct for covariance among seasons, such as could be caused if the time-series data are serially correlated. See Hirsch and others (1991) and Helsel and Hirsch (2002) for further information on these statistics. The Seasonal Kendall and Mann-Kendall output also include the slope and intercept of the Kendall's line describing the overall trend. For the Seasonal Kendall test, the intercept is set at the beginning of the first water year, defined as October through September. The Regional Kendall output includes only the slope of the overall trend because the intercept can vary by location. The estimated median trend is also reported in the Regional Kendall output as an increasing or decreasing amount per unit time.

### Summary

The original Seasonal Kendall computer code developed in the 1980s at the USGS has been repackaged to run using the Windows operating system. In addition, the Regional Kendall and Mann-Kendall variations of the test have been added so that each can be performed. This program may be used to compute the three tests and may be used as a benchmark for determining the accuracy of subsequent implementations of the Seasonal Kendall, Regional Kendall, and Mann-Kendall test procedures produced by others.

## References Cited

- Cortes, D.R., and Hites, R.A., 2000, Detection of statistically significant trends in atmospheric concentrations of semivolatile compounds: *Environmental Science and Technology* v. 34, p. 2826–2829.
- Helsel, D.R., and Frans, L.M., in press, The Regional Kendall test for trend: *Environmental Science and Technology*.
- Helsel, D. R., and Hirsch, R. M., 2002, Statistical methods in water resources: U.S. Geological Survey Techniques of Water-Resources Investigations, book 4, chap. A3, 524 p. Available online at <http://water.usgs.gov/pubs/twri/twri4a3/>
- Hirsch, R.M., Alexander, R.B., and Smith, R.A., 1991, Selection of methods for the detection and estimation of trends in water quality: *Water Resources Research* v. 27, p. 803–813
- Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of trend analysis for monthly water quality data: *Water Resources Research* v. 18, p.107–121.
- Hirsch, R.M., and Slack, J.R., 1984, A nonparametric trend test for seasonal data with serial dependence: *Water Resources Research* v. 20, p. 727–732.
- Mann, H.B., 1945, Non-parametric tests against trend: *Econometrica* v. 13, p. 245–259.
- Moore, C.R., 1987, Determination of benthic-invertebrate indices and water-quality trends of selected streams in Chester County, Pennsylvania, 1969–80: U.S. Geological Survey Water-Resources Investigations Report 85–4177, 62 p.
- Reckhow, K.H., Kepford, K., and Hicks, W. Warren, 1993, Methods for the analysis of lake water quality trends: U.S. Environmental Protection Agency report 841–R—93–003, 105 p.
- Slack, J.R., Lorenz, D.L., and others, 2003, USGS library for S-PLUS for Windows—Release 2.1: U.S. Geological Survey Open-File Report 2003–357. Available online at <http://water.usgs.gov/software/library.html>
- Wiseman, W.J., Swenson, E.M., and Power, J., 1990, Salinity trends in Louisiana estuaries: *Estuaries* v. 13, p. 265–271.