AXIS

A program for the statistical analysis of
circular data

Version 1.1

Pisces Conservation Ltd
IRC House
The Square
Pennington
Lymington
SO41 8GN
England

Pisces@irchouse.demon.co.uk
www.irchouse.demon.co.uk
www.pisces-conservation.com

Phone   44 (0) 1590 676622/674000
Fax   44 (0) 1590 675599

PISCES LICENCE AGREEMENT

This is a legal agreement between you the end user and PISCES Conservation Ltd. Lymington (PISCES). BY OPENING THIS PACKAGE YOU ARE AGREEING TO BE BOUND BY THE TERMS OF THIS AGREEMENT. IF YOU DO NOT AGREE TO THE TERMS OF THIS AGREEMENT PROMPTLY RETURN THE UNOPENED PACKAGE AND ALL ACCOMPANYING ITEMS (including written material) TO THE PLACE YOU OBTAINED THEM FOR A FULL REFUND.

1. GRANT OF LICENCE - This PISCES Licence Agreement ('Licence') permits you to use one copy of the PISCES software product acquired with this Licence (SOFTWARE) on any single computer, provided the SOFTWARE is in use on only one computer at any time. If you have multiple Licences for the SOFTWARE then at any time, you may have as many copies of the SOFTWARE in use as you have Licences. The SOFTWARE is 'in use' on a computer when it is loaded into the temporary memory (i.e. RAM) or installed into the permanent memory (e.g. hard disk, CD ROM. or other storage device) of that computer, except that a copy installed on a network server for the sole purpose of distribution to other computers is not 'in use'. If the anticipated number of users of the SOFTWARE will exceed the number of applicable Licences then you must have a reasonable mechanism or process in place to assure that the number of persons using the SOFTWARE concurrently does not exceed the number of Licences. If the SOFTWARE is permanently installed on the hard disk or other storage device of a computer (other than network server) and one person uses that computer more than 80% of the time it is in use then that person may also use the SOFTWARE on a portable or home computer.

2. COPYRIGHT - The SOFTWARE is owned by PISCES or its suppliers and is protected by all applicable national laws. Therefore, you must treat the SOFTWARE like any other copyrighted material (e.g. a book) except that if the software is not copy protected you may either (a) make one copy of the of the SOFTWARE solely for backup or archival purposes, or (b) transfer the SOFTWARE to a single hard disk provided you keep the original solely for backup or archival purpose. You may not copy the Product manual(s) or written materials accompanying the SOFTWARE.

3. DUAL MEDIA SOFTWARE - If the SOFTWARE package contains both 3-1/2" and 5-1/4" disks, then you may use only one set (either the 3-1/2" or 5-1/4') of the disks provided. You may not use the other disks on another computer or computer network, or lend, rent, lease, or transfer them to another user except as part of a transfer or other use expressly permitted by this PISCES Licence Agreement.

4. OTHER RESTRICTIONS - You may not rent or lease the SOFTWARE, but you may transfer your rights under this PISCES Licence Agreement on a permanent basis provided you transfer all copies of the SOFTWARE and all written materials, and the recipient agrees to the terms of this Agreement. You may not reverse engineer, decompile or disassemble the SOFTWARE. Any transfer must include the most recent update and all prior versions.

LIMITED WARRANTY

LIMITED WARRANTY - PISCES warrants that (a) the SOFTWARE will perform substantially in accordance with the accompanying Product Manual(s) for a period of 90 days from the date of receipt; and (b) any PISCES supplied hardware accompanying the SOFTWARE will be free from defects in materials and workmanship under nominal use and service for a period of one year from the date of receipt. Any implied warranties on the SOFTWARE and hardware are limited to 90 days and one (1) year respectively or the shortest period permitted by applicable law, whichever is greater.

CUSTOMER REMEDIES - PISCES'S entire liability and your exclusive remedy shall be, at PISCES option, either (a) return of the price paid or (b) repair or replacement of the SOFTWARE or hardware that does not meet PISCES'S Limited Warranty, and which is returned to PISCES with a copy of your receipt. This Limited Warrant is void if failure of the SOFTWARE or hardware has resulted from accident, abuse or misapplication. Any replacement SOFTWARE will be warranted for the remainder of the original warranty period or 30 days, whichever is longer.

NO OTHER WARRANTIES - TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, PISCES DISCLAIMS ALL OTHER WARRANTIES. EITHER EXPRESS OR IMPLIED, INCLUDING BUT LIMITED NOT TO IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WITH RESPECT TO THE SOFTWARE, THE ACCOMPANYING PRODUCT MANUAL (S) AND WRITTEN MATERIALS, AND ANY ACCOMPANYING HARDWARE. THE LIMITED WARRANTY CONTAINED HEREIN GIVES YOU SPECIFIC LEGAL RIGHTS.

NO LIABILITY FOR CONSEQUENTIAL DAMAGES - TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW PISCES AND ITS SUPPLIERS SHALL NOT BE LIABLE FOR ANY OTHER DAMAGES WHATSOEVER (INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, LOSS OF BUSINESS INFORMATION, OR OTHER PECUNIARY LOSS) ARISING OUT OF THE USE OF OR INABILITY TO USE THIS PISCES PRODUCT, EVEN IF PISCES HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN ANY CASE, PISCES'S ENTIRE LIABILITY UNDER ANY PROVISION OF THIS AGREEMENT SHALL BE LIMITED TO THE AMOUNT ACTUALLY PAID BY YOU FOR THE SOFTWARE.

This Agreement is governed by the laws of England.

Should you have any questions concerning this Agreement, or if you desire to contact PISCES for any reason, please use the address information enclosed in this product to contact PISCES or write to:

Axis
PISCES Conservation Ltd
IRC House, The Square
Pennington, Lymington
Hampshire, England
SO41 8GN Tel 01590 676622/674000

CONTENTS

I.   Introduction

Axis is a Windows® program that offers a range of graphical and analytical techniques commonly used by biologists, geologists and archaeologists for the analysis of circular data. Researchers in other fields also use these methods. *Axis* implements the principal graphical methods and statistical tests described by Fisher (1993) in *Statistical Analysis of Circular Data*. Periodic data of many kinds can be represented and analysed using the methods available within *Axis*. While compass bearings are the most obvious examples of circular data, other examples include the time of events over a 24-hour period, and the occurrence of activity over the lunar cycle.

At the back of this booklet, we include a worked example to help understand the operation of the program, and various methods used.

II.   Quick guide to running a data set

1. Prepare your data set using a spreadsheet such as Excel. The data are arranged with samples as columns and individual observations of direction as the rows. Give a title to each sample column. Values can be real or integer. (See the complete guide to making a data set in **section H** below)

2. Save these arrays as CSV (comma delimited text files) from the Save As menu. Take care if using Excel that the work sheet does not hold any other data or this will be included in the CSV files. It is good practice, before saving the data file, to select the first few rows and columns below and to the right of the data array and press 'Delete', to clear the cells of any spaces or data accidentally entered in them.

3. Close the spreadsheet data file before opening the data in *Axis*, otherwise you may get an '**I/O error 32 - access denied**' error message.

4. Start *Axis* and open your data files from the File menu.

5. The data will now open in the Raw Data window.

6. Undertake grouping by specifying a bin size and select axial or vectorial data. If this is not done *Axis* will assume your data are vectorial

7. Select a method from the drop-down menu at the top of the window and view the results using the tabs at the bottom.

8. Use File:Print or the Printer button on the toolbar to print the charts and Edit: Copy or the Copy button to copy it to the clipboard for pasting into another application.

9. Use the Save button on the toolbar to save the chart as an image file for use in another application, or as a TeeChart Pro (*.tee) file. See **File formats**.

III.   Installation and General instructions

A.   System requirements
1. A PC running Windows 9x, ME, NT, 2000 or XP. If you are using Windows NT over a shared network, you may need to ask your Network Administrator for permission to install software.
2. 4 MB of free hard disk space
3. SVGA screen

B.   Installation
1. Place the *Axis* CD in the CD drive.

2. The installation program should start automatically; follow the on-screen instructions.
3. If the CD does not auto-play, browse the CD in Windows Explorer or My Computer, and double-click Setup.exe in the root directory.

When installation is complete there will be an *Axis* group on your Start menu and a folder (directory) on your hard disk which holds the *Axis* program files. The default location for this folder is C:\Program Files\Axis – you will be offered the chance to install the files in a different location, if you wish. An uninstall option, available on the Start menu, will also be created to use should you wish to remove the program.

C.   General instructions

Start *Axis* in the normal Windows fashion either by clicking on the program icon or from the Start button. Along the top bar are a number of drop-down menus. These work in the same way as other standard Windows programs.
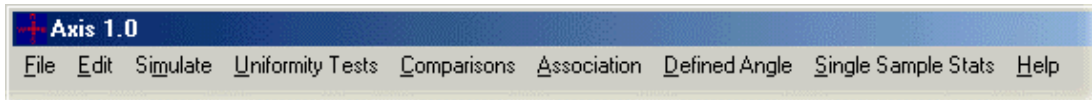
File: To open and save data sets.
Edit: To cut, copy and paste to and from the active window.
Simulate: To simulate sets of data
Uniformity Tests: To test how close a sample is to a uniform distribution
Comparisons: Tests whether 2 samples come from the same distribution
Association: Pairs of observations are positively associated if an increase in the angle of one observation is reflected in an increase in the angle of the second. A negative association occurs if an increase in angle of one observation is reflected by a decrease in the angle of the second.
Defined Angle: Testing a distribution for its fit to a pre-defined angle.
Single Sample Stats: The descriptive statistics for a single sample.
Help: to enter the Help system.

When the program is started, you will see a blank working area. From the File menu choose Open to load an existing data set. Once a data set is loaded or a new data set created the window will be filled with a form with tabs labelled Raw Data, Working Data and Summary. The Raw Data form shows the data loaded from the file, the Working Data shows the data set upon which the calculations will be performed and Summary gives statistics for the data set. You can switch between the different forms by clicking with the cursor on the tabs. Initially, the Raw Data sheet shows your data in hours, radians, months or whatever periodicity you are using; Working Data will display the same data in degrees, as used for plotting the charts.

D.   Obtaining help

For most active windows context sensitive help can be obtained by pressing F1, clicking on the help button or selecting the help drop down menu. If you have problems using the program which you cannot solve then contact Pisces Conservation at pisces@irchouse.demon.co.uk or by phone to +44 (0)1590 676622 during office hours (09.00 to 17.00). Alternatively, visit the Tech Support section of our website at **www.pisces-conservation.com/techframes.html**, or use the Help facility on the installation CD.

E.   Common errors and problems

Common error messages and their solutions are shown below:

1. " is not a valid floating point value, or **Invalid floating point operation** This will occur if the raw data holds columns or rows which sum to zero. (Normally, AXIS will have removed these rows from the data set while opening it). In some cases AXIS will identify the problem cell which should be edited.

Occasionally, this error can occur because a blank space has been accidentally entered into a cell outside the data matrix when it was being prepared in a spreadsheet program. To prevent this happening, it is good practice, before saving your data set as a .csv file, to highlight the first 10 or so blank rows and columns below and to the right of the data matrix, and press 'Delete'. This will clear the cells of any accidentally-entered contents.

2. I/O error 32 - access denied. This will occur if the data file you are trying to open is currently being used by another program - normally the spreadsheet which was used to organise the data. Close the file in other programs and try again.

3. When a data file is opened, all the data are dumped into the first cell, rather than opening in the grid properly.
This is usually because there are one or more blank cells in the second row of data. If the blanks are replaced by zeros, or another row with no blanks is put in the second position, the problem should disappear. Row 2 is the crucial one; blank cells are tolerated elsewhere.

4. I entered the data in a spreadsheet program, and saved them as a csv file. When I try to open it in Axis, the numbers are separated by ; not , so the analysis will not run.
This is because some non-British systems use "**;**" instead of "**,**" to separate values in a csv file. Open Windows Explorer, and change the file extension from csv to txt (i.e. Filename.csv becomes Filename.txt). Open the txt file in Word or another word-processing program. Use the Find/Replace function to replace every "**;**" with a "**,**" then save the file again. Change the txt file extension back to csv. It should now open and run perfectly in Axis.

5. My files do not appear to be saved.
This is often caused by the lack of the correct file extension. The program only shows files with the .csv extension. To see other files in the open file dialog type *.* into the file name box and press return. – **WARNING: this will show all the files in the directory - whether they are compatible with the program or not.** If you have saved a file without an extension either add the extension outside the program or open the file and save with the correct extension using Save As.

F.   Opening a data set

Use File|Open to start the file dialogue to select a data file for analysis. Data files are in the Comma-Delimited Text (*.csv) format: simple text files where the data in columns are separated by commas. This makes the data files easy to edit in a wide variety of spreadsheets and text editors, such as Excel, Lotus, Quattro, MS Word, Wordpad or Notepad. Once a file has been selected the screen will appear as follows:
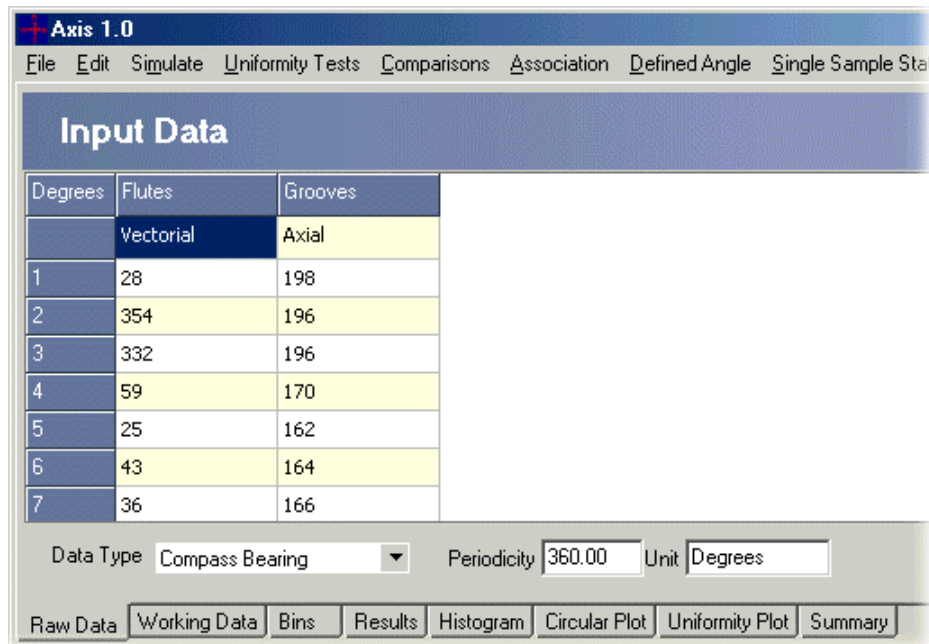
G. Demonstration data sets

4 demonstration data sets are provided with *Axis* (they can be found in C:\Program Files\Axis).
These data sets are taken from Fisher (1993).

ArrivalTimes.csv	Arrival times on a 24-hour clock of 254 patients at an intensive care unit, over a period of about 12 months.
BelfordB6.csv	Sets of cross-bed azimuths of palaeocurrents measured in the Belford anticline (New South Wales)
FeldsparLathsB2.csv	Measurement of long-axis orientation of 133 feldspar laths in basalt
GroovesAndFlutes.csv	Groove and tool marks, and flute marks, measured in the Murruin Creek area, southwest of Yerranderie, New South Wales

H. Creating and editing a data set

Data sets can be created both within and outside *Axis.* Refer to the example data sets installed
with Axis – and the sections below - to see the structure of the data.

While the simplest data set can consist of no more than a column of angles, *Axis* will need
more information if the data are to be handled correctly. First, it is necessary to define the data
as axial or vectorial. If you do not define the data type, it will be assumed that you have entered
vector data. The differences between these data types and how they are handled by *Axis* are
discussed below. While with simple angular measurements such as 70° or 340° it is obvious
that the periodicity is 360°, this is not the case with days over the lunar cycle or over a year.
Therefore you will need to inform *Axis* of your units of measurement and the periodicity of your
data in this chosen unit. Before giving simple instructions as to how to enter your data we will
discuss these issues in more detail below.

1. Axial and vectorial data

Vectorial data are represented as a directed line. For example the compass direction taken by a
tracked dolphin or the compass bearing of a sunflower at a certain time of day are vectorial data.
There is one clear direction in which the animal moves or the plant is pointing. In circular plots
within *Axis* we represent vectorial data as square points.

Axial data are used to describe the orientation of an undirected line for which there is no
distinction between the two ends. For example, a seam in a geological exposure can run at an
angle to the horizontal. If measured at the right-hand end this might give an angle of 10° above
the horizontal (80° from vertical clockwise). If measured at the left-hand end it would be 10°

below the horizontal (260° from vertical clockwise). Thus seams at angles of both 10° and 260° are at equivalent angles.

Axial data are analysed by first converting them to vectorial data. Each value is doubled and then reduced modulo 360°. For example the angles 80° and 260° become 80 x 2 = 160° and 260 x 2 = 520°; followed by division modulo 360° gives 160° = 160° and 520° = 160° (1 x 360° + 160°). Therefore the angles are now the same as required. However, after analysis any resulting angle must be back-transformed – 160°/2 = 80°.

2.   Inputting the periodicity

Periodic data of many kinds can be represented and analysed using the methods available within *Axis*. While compass bearings are the most obvious examples of circular data, other examples would be the time of events over a 24-hour period and the occurrence of activity over the lunar cycle. Most of the calculations undertaken by *Axis* use angles expressed in radians (2π radians = 360° or 1 period as defined by the user). It is therefore essential that you input information on the units of measurement and the periodicity.

Units and periods that are available with Axis include the following:
Hours - time over a 24-hour period.
Days - time over a 365-day year.
Days - time over a 29.5-day lunar cycle.
Degrees - angle round a 360° circle.
Radians - angle round a 2π period or circle.

3.   Preparing large data sets in a spreadsheet program

You do not need to create data sets within Axis; in fact the best way to enter large data sets is to use a spreadsheet such as Excel or Lotus 1-2-3, which will give access to a wide range of sorting and editing procedures to ease your task. A data set has the following format:

Degrees	Flutes	Grooves
	Vectorial	Axial
1	28	198
2	354	196
3	332	196
4	59	170
5	25	162

The normal arrangement of data within AXIS is to have the samples as columns and the angles as the rows. The top left-hand cell (A1) indicates the unit of measurement; the next two cells hold the column (sample) titles. The first cell of the second row is blank, then the next two cells indicate whether the data in each column is vectorial or axial – see section H1 above. The rows below hold the observations. Blank observations should **not** be entered as 0.

You should be aware that Excel has a maximum of 255 columns, which can prove difficult if you have a data set with a very large number of observations. Other spreadsheets (such as Corel Quattro) are able to handle larger numbers of columns. Numbers can be either integer or real; some methods may require integers, in most such cases the program will run with real data which will be automatically rounded.

When using Excel use the Save As function to save your data as a *.csv file. *Ensure that the work sheet you are saving only holds the tabulated data for analysis.*

Occasionally, errors occur because a blank space has been accidentally entered into a cell outside the data matrix. To prevent this happening, it is good practice, before saving your data set as a .csv file, to highlight the first 10 or so blank rows and columns below and to the right of the data matrix, and to press 'Delete'. This will clear the cells of any accidentally-entered contents.

The data set above looks like this if viewed in a word processor program, such as Word or Notepad. The columns are separated by commas:

Degrees,Flutes,Grooves
"",Vectorial,Axial
1,28,198
2,354,196
3,332,196
4,59,170
5,25,162

Note the "" and comma in Row 2, Cell 1, which will make the cell blank.

4. Data entry from within *Axis*

Data sets can also be created and edited within *Axis*. To create a new data set, select **File| New** from the drop-down menus. You will be presented with an empty 3-column by 3-row grid. The blue grid cells at the head of the columns starting from column 2 are used to hold (e.g.) sample names, and similarly the first column starting from row 2 is used to hold names for individual observations. To input text or data, click on the cell and type a number. Numbers may be either integer or real; some methods may require integers, in most such cases the program will run with real data, which will be automatically rounded.

The return key on your keyboard moves you sequentially through the grid and when pressed with the cursor in the lower right hand column will add a new column. To add a new row, press the Insert key on your keyboard. A row will be inserted below that currently selected. Please note, if the text within a box is selected (it will usually appear blue) insert will not work. Simply click on the cell again to select the cell and insert will be activated. Use the Delete key on your keyboard to remove a row.

5. Editing existing data

The raw data grid can be edited by using the mouse to click on a cell to select it and typing in a new value. Changes made to the raw data are automatically transferred to the working data when the Working Data tab is clicked. These changes will not alter a saved data file until **File|Save** is used. *Axis* will use the working grid thus created for all subsequent calculations. Note that output windows will not show calculations using the edited data until the methods have been re-run by selecting them in the normal fashion. The working data grid can also be edited. Changes made to the working grid will not be transferred to the raw data grid. Changed working data can be saved as a new data set using **File|Save as**.

## 6. Saving edited data or creating a new data file

When you choose to close the program, if a data set has been altered in any way, but not saved, you will be asked if you would like to save the data. A data set can be copied and saved under a different name by selecting **File|Save as**.

I. Maximum size of the data set

The maximum number of angles (rows) and samples (columns) is 1000.

J. Printing and exporting your results

The graphs and other results can be printed using File: Print, or the Print button on the toolbar. The Print Preview option will show the page layout and allow image size, margins, and paper orientation to be changed. If you have the full version of Adobe Acrobat (**not** Acrobat Reader) installed on your computer, you will be able to save the chart directly as a .pdf file by selecting Acrobat Distiller or Acrobat PDF Writer from the list of available printers in the Print dialog box. The graph option buttons on the Chart Toolbar are described in order from left to right below.
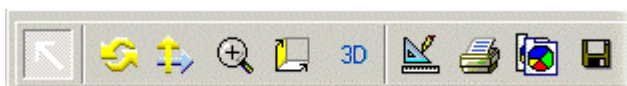
Normal Cursor - (Arrow) Use this button to zoom in onto an area of your graph. Move the pointer to a point above and to the left of the area you want to enlarge, hold the left-hand mouse button down and drag down to the right and release the button. To reverse the enlargement, click on the left hand mouse button and move up and to the left.

Rotate - (Two arrows in a circle) Use this button to rotate the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse. (Not available on circular plots).

Move - (Two arrows at right angles) Use this button to move the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse.

Zoom - (Magnifying glass) Use this button to zoom in onto the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse. See also the detailed explanation below

Depth - (3D box) Use this button to change the depth shown on the graph. Click the button then move onto the graph and hold down the left-hand mouse button and move the mouse. (Not available on circular plots).

3D - Use this to toggle the 3D view on and off. (Not available on circular plots).

Edit - (Set square and pencil) This button will offer a wide range of options to change the style of your graph. It is also used to export or copy your graph to file, and even to email it using the 'Send' button. For more information on chart editing use the TeeChart help system available from the Help button on the chart edit box.

Print - Use this button to print the graph

Copy - Use this option to copy the graph to the clipboard, to paste directly into other applications.

Save - Save the file in a variety of different formats.

File formats
The chart can be saved in a number of different file formats: Enhanced Metafile (*.emf), Bitmap (*.bmp), JPEG (*.jpg), or Native (*.tee). Each file format has advantages and disadvantages.

- The advantage of Enhanced Metafile is that, if pasted into, for instance, a Word document, it can be resized by dragging, without losing resolution.
- Bitmaps are a lossless method of saving; the stored file will not lose any of the original's detail. Because of this, bitmaps tend to be much larger than compressed files such as Enhanced Metafiles or JPEGs.
- JPEGs are file formats which can be compressed to take up less space - useful if you wish to send one by email, put it on a website, or paste it in to a document. If they are compressed too heavily, they can lose resolution and detail, and spoil colours.
- The Native (*.tee) format saves all the chart attributes, and the data series, rather than the image itself. This means that you can save the chart, and open it again at a later date to edit it, using the free TeeReader software supplied on the installation CD. Tee files tend to be very small indeed; often only 1-2 KB.

K.   Graph display options

To change the colours, axes, titles and many other features of any graph use Edit button above the graph.
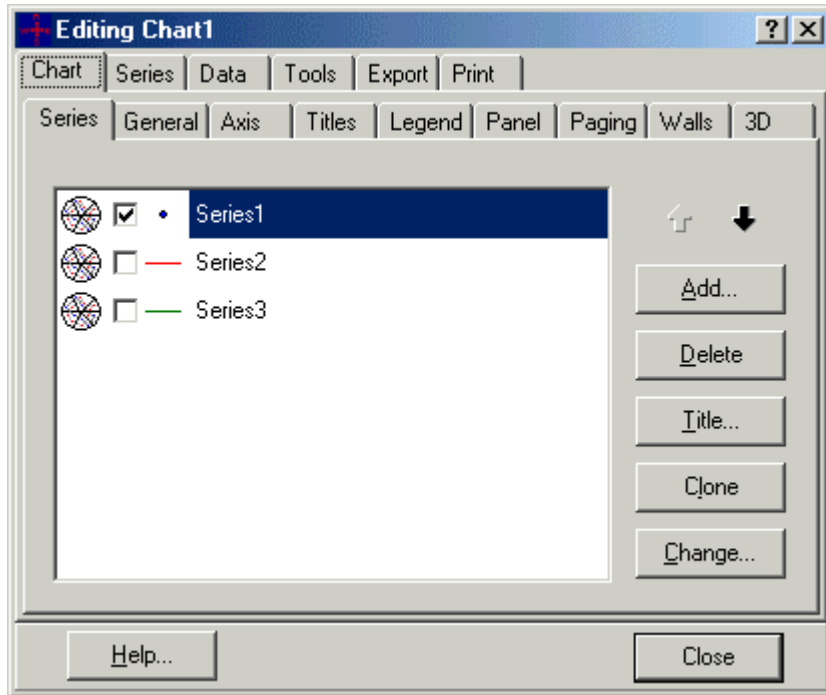
Fig 2. Graph display options from the Edit Graph button.

For help on customising graphs, adding annotations, etc., consult the TeeChart Help system, available by clicking the Help button on the Edit Graph dialog.

IV. Summarising statistics for each sample

Summarising statistics for each sample are obtained from the Single Sample Stats|Summary Stats drop-down menu and are displayed under the **Summary** tab. Any output which reads –99 indicates that the calculation could not be completed.
The following statistics are calculated

Observations: This is the number of measurements in each column of the data set.

Mean Direction: This is the mean for each sample. It is calculated as follows:

Given a series of n angular measurements θ_i :

$$C = \sum_{i=1}^{n} \cos(\boldsymbol{q}_i)$$

$$S = \sum_{i=1}^{n} \sin(\boldsymbol{q}_i)$$

The mean direction is given by:

$$\overline{\boldsymbol{q}} = \begin{cases} \tan^{-1}\left(S/C\right)............S > 0, C > 0. \\ \tan^{-1}\left(S/C\right)+\boldsymbol{p}..............,C < 0 \\ \tan^{-1}\left(S/C\right)+2\boldsymbol{p}....S < 0, C > 0 \end{cases}$$

Lower 95% mean: There is a 95% probability that the true mean direction is greater than this value. It is calculated by one of two methods depending upon sample size.
For n > 24

- Calculate the circular dispersion, δ (see below)
- Calculate the circular standard error

$$s = \sqrt{d/n}$$

- Then calculate the lower confidence interval as

$$lower 95\% = \overline{q} - \sin^{-1}(1.96w)$$

For n = 24

The lower confidence intervals are calculated by a bootstrap method by re-sampling the data 1000 times.

Upper 95% mean: There is a 95% probability that the true mean direction is less than this value. It is calculated by one of two methods depending upon sample size.
For n > 24

- Calculate the circular dispersion, δ (see below)
- Calculate the circular standard error

$$s = \sqrt{d/n}$$

- Then calculate the upper confidence interval as

$$lower 95\% = \overline{q} + \sin^{-1}(1.96w)$$

For n = 24

The upper confidence intervals are calculated by a bootstrap method by re-sampling the data 1000 times.

Mean Resultant Length: This is defined as:

$$\overline{R} = \frac{1}{n}\sqrt{C^2 + S^2}$$

See mean direction above for the definitions of C and S.

The mean resultant length ranges between 0 and 1. The value is 1 if all the directions in the data set are the same and is approximately zero for a uniform distribution.

Circular Variance: This is defined as:

$$V = 1 - \overline{R}$$

Circular Standard Deviation: This is defined as:

$$v = \left[-2\log(1-V)\right]^{1/2}$$

where V is the Circular Variance.

Median Direction: The median angle is a direction (or a point on a unit circle) that divides the data into two equal sized groups. Unfortunately, this is more difficult than it might seem. You cannot just find a midway between observations as on a linear scale, where observations can be always arranged in increasing order. For circular data we can rotate the data around the circle as we wish, since generally there is no preferable starting point. Only in the cases in which data are spread by an angle significantly less than $2p$ can one take the mid angle (or observation) if the number of observations (n) is odd, and average the two middle ones if it is even.

More generally the median is defined as the value of α that minimizes the function:

$$D(\boldsymbol{a}) = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{p} - \left| \boldsymbol{p} - \left| \boldsymbol{q}_i - \boldsymbol{a} \right| \right| \right)$$

The sample median always exists but it is not always unique.

Lower 95% Median: When n the number of observations is > 15 this is calculated by finding m defined as:

$$m = 1 + \text{int}(\frac{1}{2} n^{1/2} .1.96)$$

where int is the integer part the expression.

From the median value the lower interval is then found by counting m observations away in the lower direction.

For smaller values of n the tabulated values of Fisher & Powell(1989) are used.

Upper 95% Median: When n the number of observations is > 15 this is calculated by finding m defined as:

$$m = 1 + \text{int}(\frac{1}{2} n^{1/2} .1.96)$$

where int is the integer part the expression.

From the median value the upper interval is then found by counting m observations away in the higher direction.

For smaller values of n the tabulated values of Fisher & Powell(1989) are used.

Circular Dispersion: This is defined as:

$$\boldsymbol{d} = (1 - \boldsymbol{r}_2) / (2\overline{R}^2)$$

where

$$\overline{R} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \cos(\boldsymbol{q}_i - \boldsymbol{q})$$

$$\boldsymbol{r}_2 = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \cos 2(\boldsymbol{q}_i - \boldsymbol{q})$$

Skewness:

This gives a measured as the degree of skewness in the observed distribution and is defined as:

$$s = \left[\boldsymbol{r}_2 \sin\left(\hat{\boldsymbol{m}}_2 - 2\hat{\boldsymbol{m}}\right) \right] / \left(1 - \overline{R}\right)^{\frac{3}{2}}$$

where the terms are defined under circular dispersion above.

Kurtosis:

This gives a measured as the degree of kurtosis (peakness) in the observed distribution and is defined as:

$$k = \left[\boldsymbol{r}_2 \cos\left(\hat{\boldsymbol{m}}_2 - 2\hat{\boldsymbol{m}}\right) - \overline{R}^4 \right] / \left(1 - \overline{R}\right)^2$$

where the terms are defined under circular dispersion above.

# V.   Descriptive graphical methods

Circular data can be exceptionally confusing, far more so than linear data, so it is particularly useful to examine plots of the data. *Axis* offers a range of plots that are described below. More than one type of plot may be displayed on the circle simultaneously. For example you can show your data as simple plots of the observation round the circle, as a frequency histogram and as a smoothed curve simultaneously. The Multiple Plots chart allows to you plot more than one data set at once.

A.   Raw data plot - Circular Plots tab

The analysis of all data should start with an examination of the raw data plotted round the circle. This will indicate the general dispersal of your data and allow you to detect any unusual points. A simple circular plot is obtained by selecting the *Circular Plot* tab. The graph can be displayed in a number of forms described below. If you have entered data for a series of samples (columns) then you will need to select the sample for plotting using the drop-down list at the top of the left-hand panel. As we require a simple raw plot of the data, the Segments Type panel should be set to None. (This panel is used to plot circular histograms – see below)

The Point Type radio buttons will determine the type of raw plot produced as follows:

True Position: This shows each observation as a single point round the circle (below). Vector data are plotted as a blue diamond, axial data as a red circle. For axial data both ends of the line are shown.
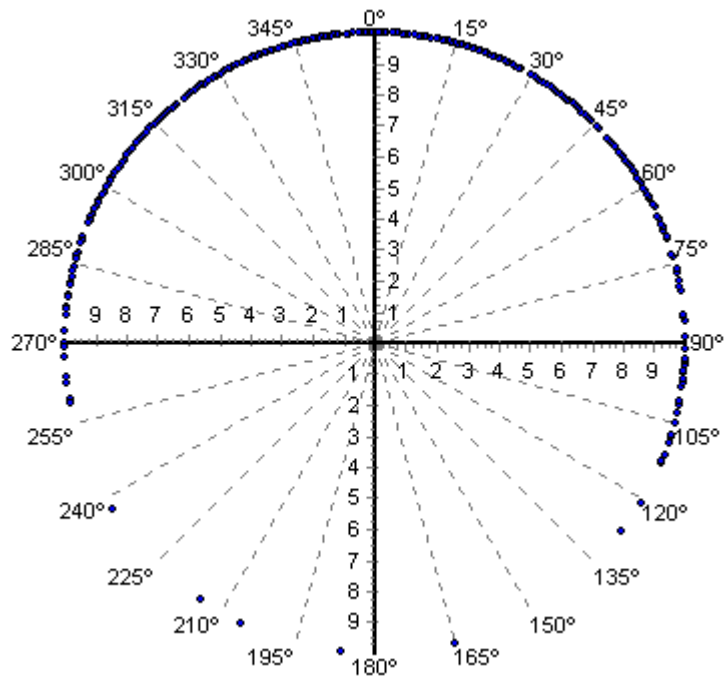
Bins inside: This shows the number of observations within each angular segment of the circle plotted inside the circle. If no bin size is selected on the *Working Data* tab the program will default to 1 degree for the bin size thus all observations between ≥1° and <2° will be plotted at 1.5°.

Bins outside: This shows the number of observations within each angular segment of the circle plotted outside the circle (below). If no bin size is selected on the *Working Data* tab the program will default to 1 degree for the bin size thus all observations between ≥1° and <2° will be plotted at 1.5°.
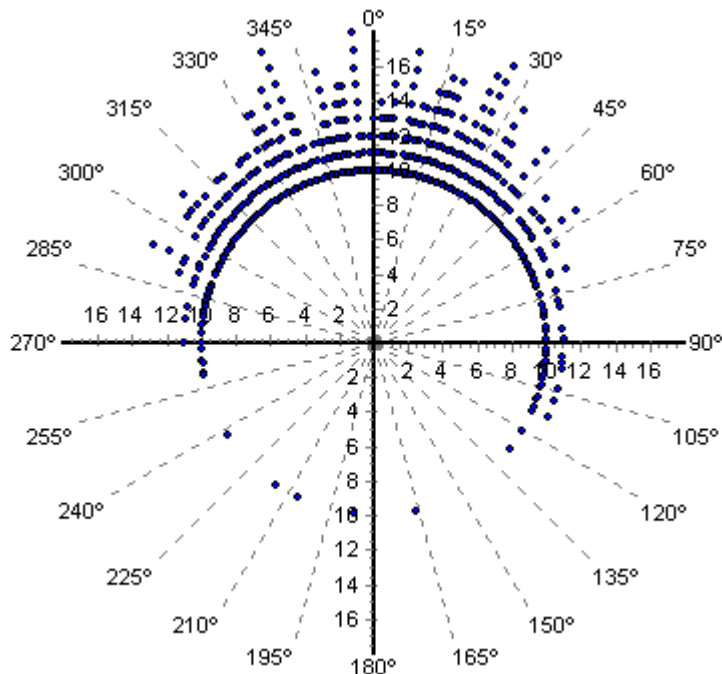
B.   Angular histograms – Rose diagrams

Circular data can be displayed on both circular and angular histograms. A simple angular histogram is obtained by wrapping a linear histogram round a circle, however it is more common to display angular data with each group as a sector, which is termed a rose diagram. Axis offers two display forms for a rose diagram as follows:

Frequency: A simple Rose diagram is obtained. Each segment on the histogram is centred on its midpoint in its grouping interval (below). The grouping interval is determined by the bin size on the *Working Data* tab. If no bin size is selected on the *Working Data* tab the program will default to 1 degree for the bin size, thus all observations between ≥1° and <2° will be plotted at 1.5°.
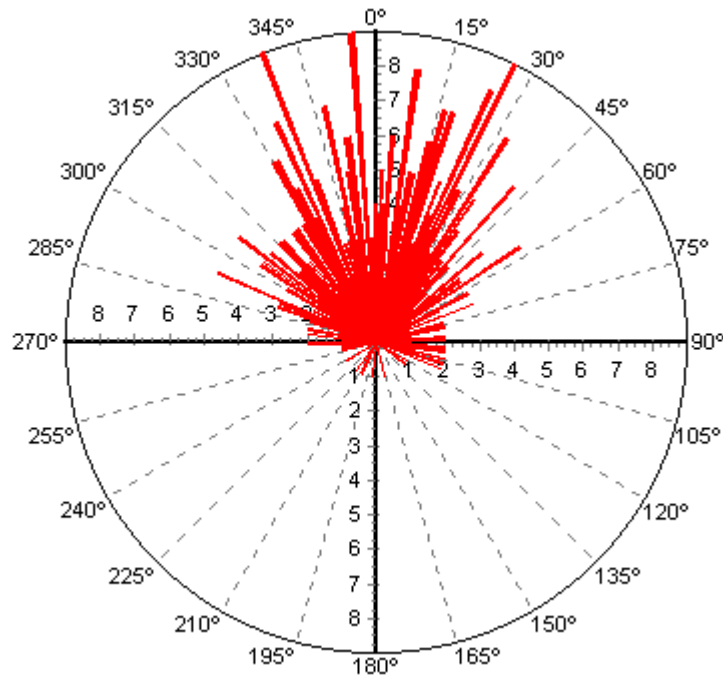

Square Root Frequency: The radius of each segment is plotted as relative to the square root of the frequency. This has the effect of making the area of the segment proportional to the frequency of the observations. The grouping interval is determined by the bin size on the *Working Data* tab. If no bin size is selected on the *Working Data* tab the program will default to 1 degree for the bin size thus all observations between ≥1° and <2° will be plotted at 1.5°.

C.   Smoothed plot

Activating the Show Smoothed Line tick box plots a smoothed line showing the distribution of observations round the circle (next page). The degree of smoothing is adjusted using the slider below the tick box, or by using the left and right arrow keys on your keyboard, for small adjustments, and the PgUp and PgDn keys for larger steps. It is important to choose the right degree of smoothing to show the structure of the plot fully; oversmoothing will lose details of smaller peaks, while undersmoothing will lead to the line having too many modes.

All aspects of the smoothed line can be altered: Click the Edit Chart button, double-click on the Series which makes the line (denoted by a green line). On the Format tab, click the Pen button (outlined in red below):
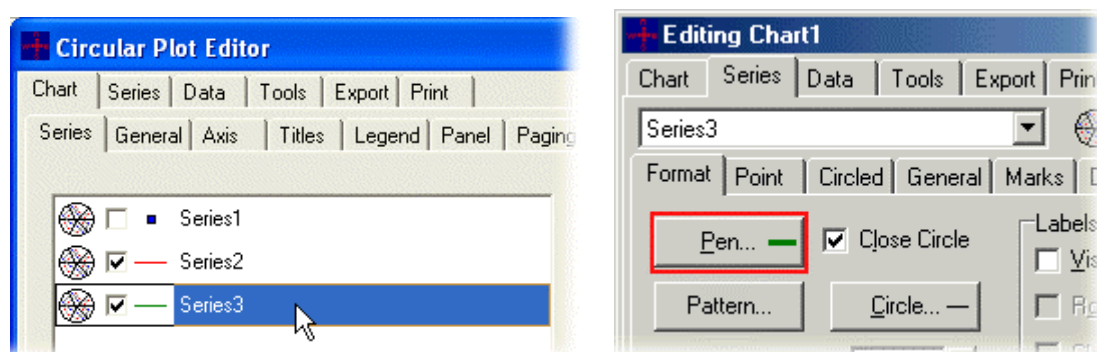
Here, you will be given options for altering the style, weight and colour of the line.
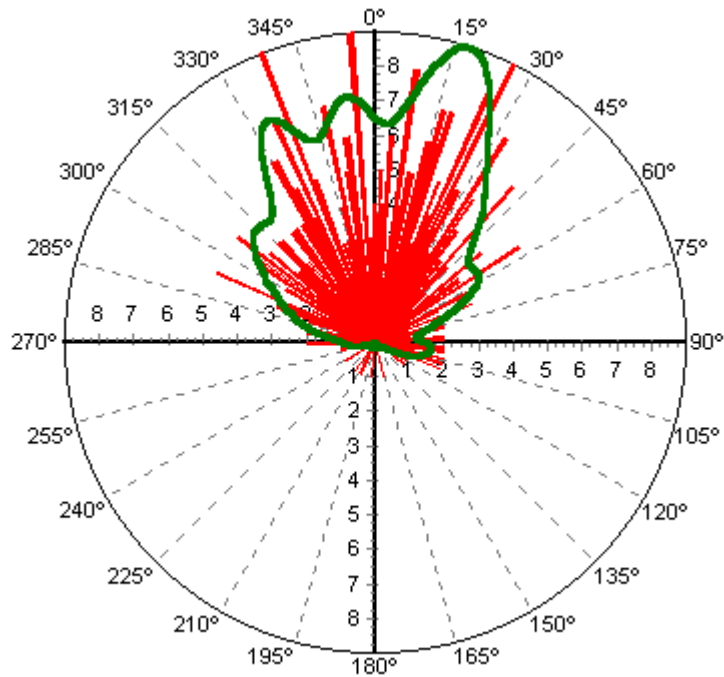
The smoothing is accomplished using a Fast Fourier Transform and the result should be viewed as a graphical technique to guide the eye to the main features of the distribution. While it is not possible to make any claim as to the statistical appropriateness of a particular smoothed curve they are useful to display the data. The data can be exported to other programs by using the Export option from the File menu. Choose the data tab and select Series 3.

D. Mean Direction and Confidence limits

Use the Mean Direction and 95% Confidence tick boxes to display the Mean direction of the

observations and the 95% confidence limits of the Mean on the chart. The colour and weight of the lines can be altered, using the Chart Edit button above the chart – in the Chart Edit dialog box double-click on the Series representing the line you wish to change; on the Format tab, select the Pen button (See C. Smoothed Plot, above) to change the line style, weight and colour.
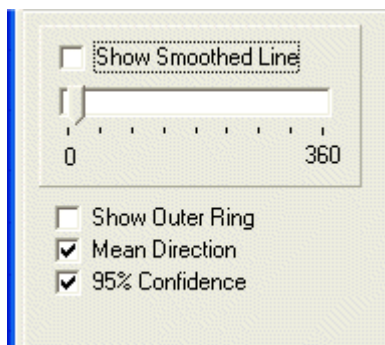
### E. Linear Histograms – Histogram tab

Linear histograms are plotted as if the data were linear instead of circular. Starting at an angle of 0° the data are placed into bins covering a set number of degrees. Each bar on the histogram is centred on its midpoint in these grouping intervals. The grouping interval is determined by the bin size on the *Working Data* tab. If no bin size is selected, the program will default to 1° for the bin size, thus all observations between ≥1° and <2° will be plotted at 1.5°. (below)
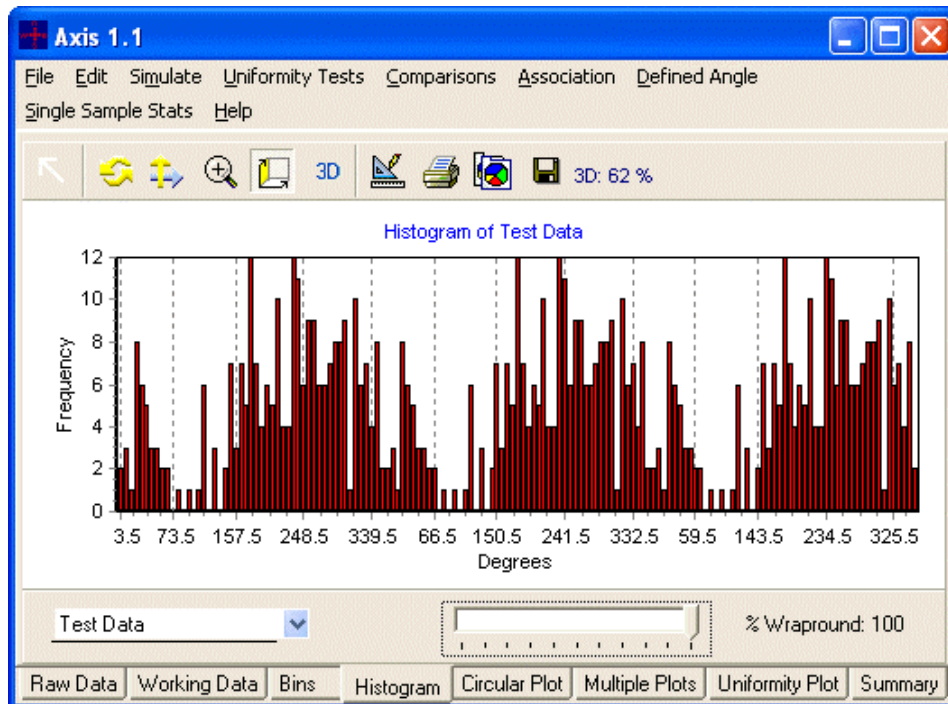
It is often easier to see the circular distribution better if the cycle is repeated. If the **% Wrapround** slider is moved to 100%, three complete cycles will be displayed. In other words the same data are plotted at either end of the distribution.If you have entered data for a series of samples (columns) then you will need to select the sample for plotting using the drop-down list on the left of the bottom panel.

F.   Multiple Plots tab

If you have a data set with multiple observations or samples, you can plot several samples on the same chart, using the Multiple Plots tab (below).
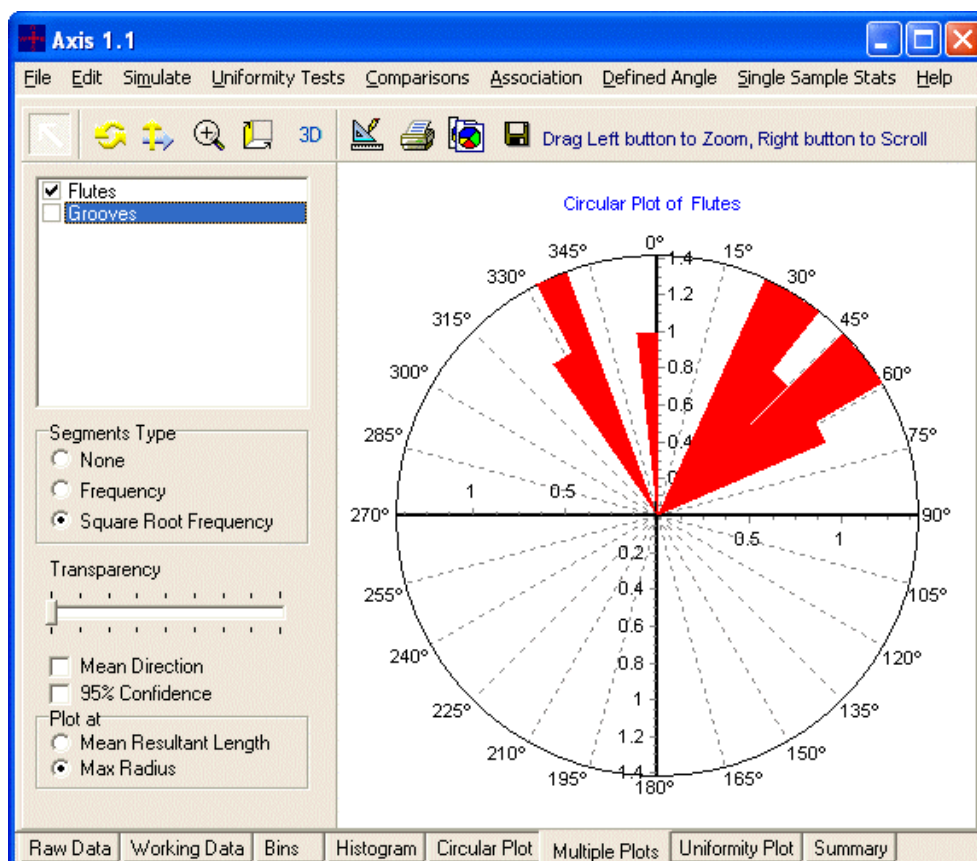
In the panel at the top left of the page, select the samples you wish to plot by ticking the relevant boxes – the sample will automatically be added to the chart. If bars on the rose diagram overlap (image left), use the Transparency adjustment so that both plots are visible. The Mean Direction and 95% Confidence limits can be added to the plot by ticking the relevant boxes – see Section D – **Mean Direction and Confidence limits**, above.
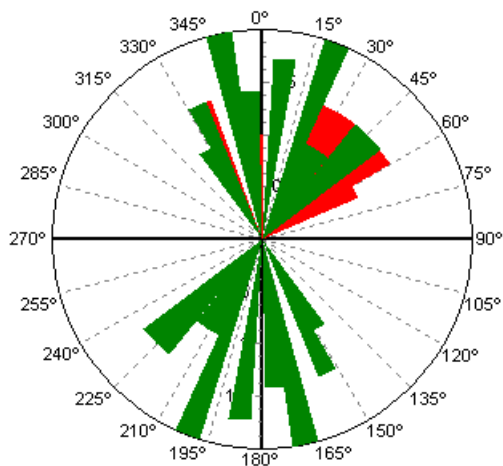
Using the radio box buttons below, the Mean Direction and Confidence limits can be plotted at the Mean Resultant Length (in effect, at the mean of the y-axis values) or the Max. radius – the maximum y-value of each sample.
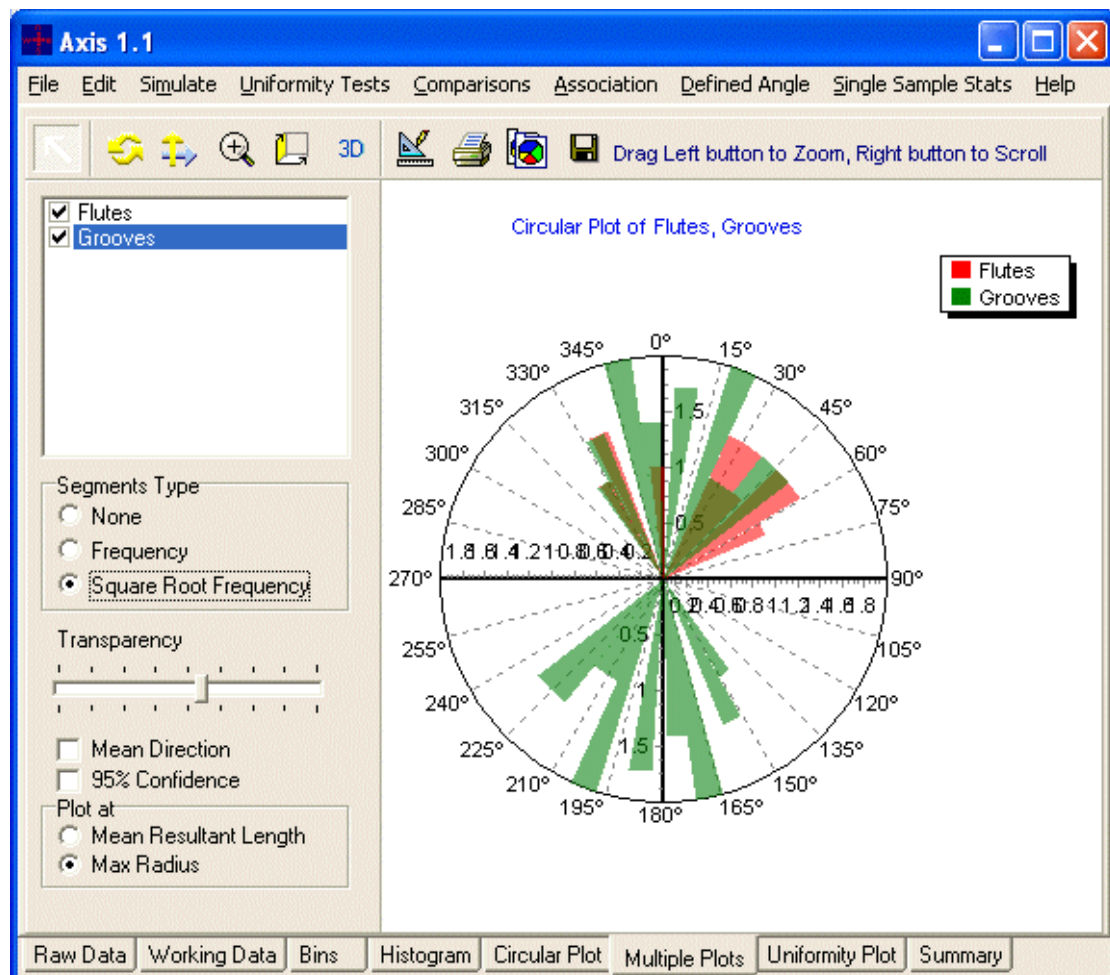
NOTE: The colours used by the program to show the various segments of the plot and their respective means and confidence limits are dynamically assigned. This means that, if you add a series of data or a Mean Direction to the plot, the colours of the segments and lines already plotted may be changed. You can over-ride this and choose the colour of each element of the plot manually; click the Chart Edit button, and then on the Series page, (below) double-click the colour of the series to change it.
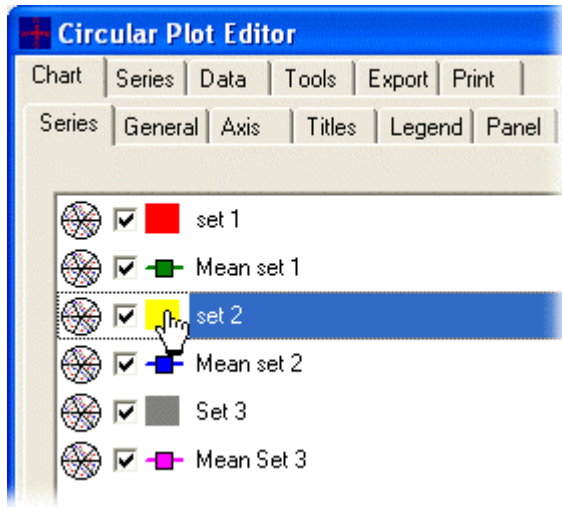
Please note that you should make all other alterations to the chart **before** you do this, since any further altering or resizing of the chart will cause the colours to return to their defaults.

VI.   Simulating distributions

Select the simulate drop-down menu to produce data sets conforming to either Uniform (random), Cauchy, Normal or von Mises distributions. Once you have selected the type of distribution required you will be presented with a pop-up window. In this window, you must specify the following:

1.   The number of points (observations) in each sample. The default is 500. Click on the box and type in a different number if required.

2.   The mean angle in radians (2p = 360°) of each sample. You can simulate a large number of samples each with different angles. Just type in a value in the box labelled Mean Angle (in radians) and press enter. The values you have entered will be displayed in the text box below.

3.   For some distributions you will need to input the Mean Resultant Length, or Kappa in the case of von Mises. The Mean Resultant Length is a real number greater than 0 and less than 1. Kappa can vary between 0 and infinity. As Kappa approaches 0 the distribution converges to the uniform. As Kappa approaches infinity the distribution tends to a point distribution.
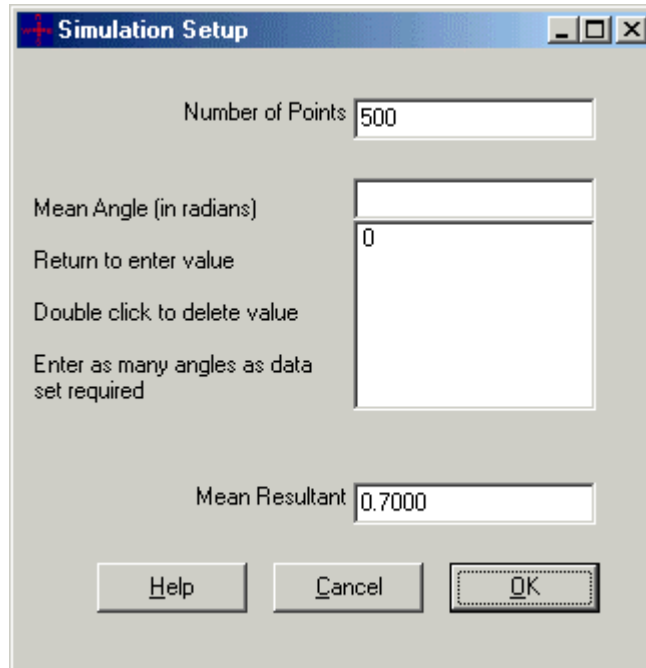
Simulation setup dialog box.

When you have specified the distributions for simulation press OK and the data will be entered into the data grid. You can then examine the linear and angular plots by clicking on their respective tabs.

VII.   Correlation between samples

These methods are concerned with statistically testing the association between two variables. They will only be available if the input data set comprises of at least 2 samples. Further, you should remember that the observations are considered paired so ensure that the observations in each row are correctly aligned. Two statistical procedures are offered.

T-linear: T-linear association is analogous to simple linear correlation. The degree of association is given by the statistic (Fisher & Lee, 1983,1986):

$$\hat{r}_T = \frac{\sum_{1 \le i \le j \le n} \sin(q_i - q_j)\sin(f_i - f_j)}{\left[\sum_{1 \le i \le j \le n} \sin^2(q_i - q_j)\sin^2(f_i - f_j)\right]^{1/2}}$$

The closer the statistic is to –1 or +1 the greater the degree of negative or positive association between the two variables. A significant degree of association is rejected if the value does not differ significantly from zero.

Randomisation Test: This is a general test of Rothman (1971) for the hypothesis that both circular distributions are independent. The critical values are obtained by a randomisation procedure in which the order of one of the samples is randomly changed, breaking any association between the observations.

After selecting the method you will be presented with a data selection window with two drop-down lists for the selection of the two samples. It is possible to select the same sample for each variable which should result in a correlation of +1.

VIII.   Testing for uniformity or randomness

The observed distribution can be tested for uniformity if you wish to test whether all directions are equally likely. A uniform distribution would also be expected if the directions were generated

by random chance. There are a number of methods to test for randomness. If you wish to test for any deviation from randomness then use an omnibus test. However, because the omnibus test will, given sufficient data, detect **any** deviation from randomness it will not be particularly sensitive to particular types of deviation from uniformity. For example, if you wish to test whether one particular direction occurs more frequently than by random chance alone so that the data have a distinct, single, mode, then the Rayleigh Test will be more appropriate.

A.   Graphical assessment

Before undertaking any statistical testing it is useful to examine the uniform probability plot of the data. This is selected by clicking on the **Uniformity Plot** tab. The sample that is plotted is selected using the left-hand drop down menu in the panel below the graph.
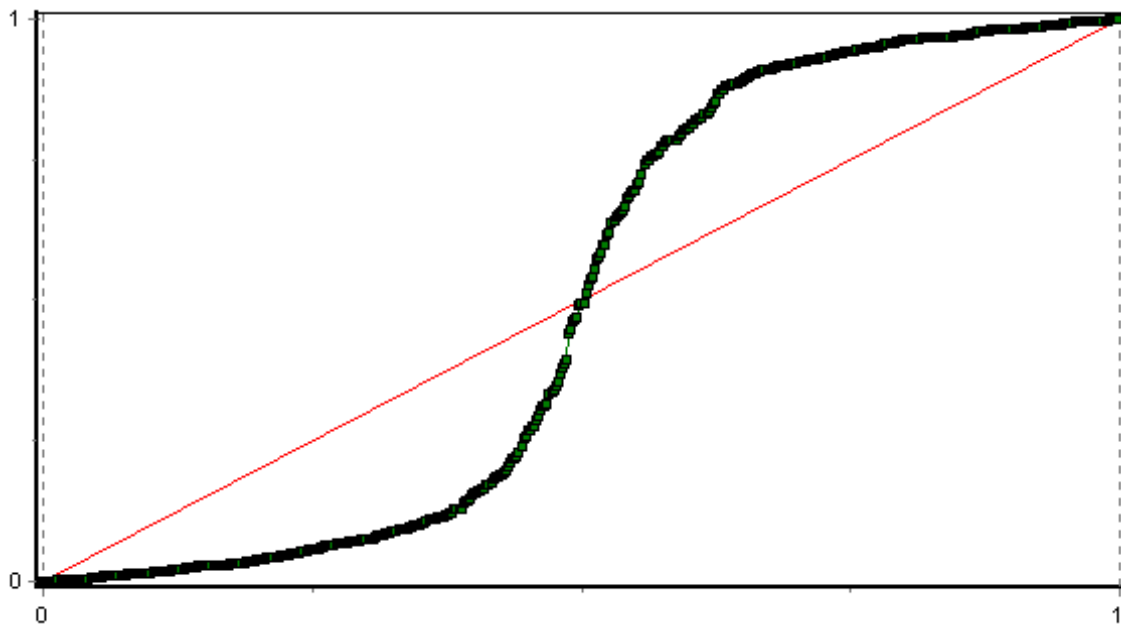
The angles for the n observations are sorted from smallest to largest and each divided by 2p

$$x_n = \mathbf{q}_{(n)} / 2\mathbf{p}$$

The plot comprises the paired values

$$\left[\frac{1}{(n+1)}, x_1 \right] \cdots\cdots\cdots\cdots \left[\frac{n}{(n+1)}, x_n \right]$$

If the data are from a uniform distribution then the points should lie along a 45° line passing through the origin. This line is drawn in red on the plot (below).

Because the data are circular the choice of point from which to start plotting is arbitrary. To avoid any artefact due to this choice a selected percentage of the points can be added onto each end of the distribution. This is done by taking a set percentage of the points from the lower end and adding 1 to each value in each pair and plotting them as addition points at the upper end. In similar fashion 1 is subtracted from the same percentage of points from the upper end. This percentage is chosen using the % Wraround slider situated on the panel below the graph.

B.   The omnibus test for randomness

Different procedures are used for ungrouped and grouped data.

1. Ungrouped omnibus test

This test is used to test to determine if the observations are randomly or uniformly distributed.

The angles for the n observations are sorted from smallest to largest and each divided by 2p

$$x_n = \theta_{(n)} / 2\pi$$

The following statistics are then calculated:

$$D_n^+ = \max\left[\frac{1}{n} - x_1, \frac{2}{n} - x_2, \ldots\ldots, 1 - x_n\right]$$

$$V_n = D_n^+ + D_n^-$$
$$D_n^- = \max\left[x_1, x_2 - \frac{1}{n}, \ldots\ldots, x_n - \frac{n-1}{n}\right]$$
$$V = V_n\left(n^{\frac{1}{2}} + 0.155 + 0.24\middle/ n^{\frac{1}{2}}\right)$$

The hypothesis that the sample was drawn from a uniform distribution is rejected if V is larger than a critical value tabulated by Arsham(1988).

2. Grouped omnibus test

This test is used to determine whether the observations are randomly or uniformly distributed. It is used when the observations have been recorded in say 5° or 10° intervals. The number of observations in each interval is tested using a Chi-squared test. If there are k intervals round the circle and n observations then on average there should be m = n/k observations per interval. Any large deviation from this average would indicate a lack of uniformity.

The test statistic Y is calculated using:

$$Y = \left(k/n\right)\sum_{i=1}^{k} n_i^2 - n$$

where n_i is the number of observation in the ith interval.

The hypothesis of uniformity is rejected if Y exceeds the critical value for a Chi-squared distribution with k-1 degrees of freedom.

C. The Rayleigh test; tests for randomness against a unimodal alternative

Frequently you will wish to determine if the observations tend to head in a single general direction against the alternative that they are randomly distributed. In other words a unimodal hypothesis is tested against a random hypothesis. The Rayleigh test is the appropriate method.

1. Rayleigh test – unspecified mean direction

Given a series of n angular measurements θ_i calculate:

$$C = \sum_{i=1}^{n} \cos(\theta_i)$$

$$S = \sum_{i=1}^{n} \sin(\theta_i)$$

and the mean resultant length as:

$$\overline{R} = \frac{1}{n}\sqrt{C^2 + S^2}$$

The null hypothesis of uniformity is rejected if the mean resultant is too large.

If the data are grouped the above test is applied with the each data point given the value of the midpoint of the grouping interval.

2.   Rayleigh test – specified mean direction

This is chosen using Selected Angle|Rayleigh Ungrouped or Selected Angle|Rayleigh Grouped. You will first be presented with a pop-up window to allow selection of the sample for analysis and the angle against which the sample will be compared.

The calculations are undertaken as follows: Given a series of n angular measurements θ_i calculate:

$$C = \sum_{i=1}^{n} \cos(\mathbf{q}_i)$$

$$S = \sum_{i=1}^{n} \sin(\mathbf{q}_i)$$

the mean resultant length is calculated.

$$\overline{R} = \frac{1}{n}\sqrt{C^2 + S^2}$$

Next using the mean direction, μ, and the user supplied angle θ the following test statistic is calculated:

$$\overline{R}_0 = \overline{R}\cos(\mathbf{m} - \mathbf{q})$$

The null hypothesis that there is no significant difference between μ, and θ is rejected if the test statistic is too large.

If the data are grouped the above test is applied with the each data point given the value of the midpoint of the grouping interval.

D.   The Watson one sample test for uniformity

This is a non-parametric test for randomness. First the n angular observations are divided by 360° to create the u_i transformed variables. Then the following test statistic is calculated:

$$U^2 = \sum u_i^2 - \frac{\left(\sum u_i\right)^2}{n} - \frac{2}{n}\sum i u_i + (n+1)\overline{u} + \frac{n}{12}$$

The test statistic is then looked up in a table of critical values to determine if there is a significant deviation from uniformity.

E.   Testing the median direction for a specified value

This is chosen using Selected Angle|Specified Median. You will first be presented with a pop-up window to allow selection of the sample for analysis and the median angle against which the sample will be compared.

Let m be the number of observations in the arc between the specified angle and the specified angle plus 180°, which are not equal to the specified angle. If the true median direction is equal to the specified angle then m should not differ much from (n-k)/2: Where n is the number of observations and k is the number of observations that equal the specified angle.

The test statistic is

$$Y = \frac{(2m - n + k)^2}{(n - k)}$$

The hypothesis that the specified angle equals the actual median is rejected if Y exceeds the critical value for a Chi-squared distribution with 1 degree of freedom.

IX.   Comparisons between samples

A.   Comparing the mean direction

From the Comparisons menu, choose Compare Mean Direction. A dialog box will appear, where you can choose which samples you wish to compare - see **Selecting samples to compare common mean/median** (below). The results of the comparison will be displayed on the Results tab, with the names of the samples compared listed in the grid underneath. A figure of -99 (or -99.00000) indicates that the result was not able to be calculated - most likely because the data set did not contain enough observations.

A minimum of 25 observations (rows) is required for this calculation. One of two methods is used, depending on the structure of the data. The P method is used when all the samples for comparison have a similar circular dispersion.

1.   The P method

This method is used when all samples have a similar circular dispersal.
Using the mean directions of each of r samples calculate

$$\hat{C}_p = \sum_{i=1}^{r} n_i \cos \hat{\boldsymbol{m}}_i$$

$$\hat{S}_p = \sum_{i=1}^{r} n_i \sin \hat{\boldsymbol{m}}_i$$

$$R_p = \sqrt{\hat{C}_p{}^2 + \hat{S}_p{}^2}$$

and the weighted average circular dispersal using

$$\hat{\boldsymbol{d}}_0 = \sum_{i=1}^{r} n_i \hat{\boldsymbol{d}}_i / N$$

where N is the total number of observations in all the samples.

The test statistic, Y, is then calculated

$$Y_r = \frac{2(N - R_p)}{\hat{d}_0}$$

and the null hypothesis that all the samples have a common direction is rejected if Y exceeds the critical value for a Chi-squared distribution with r-1 degrees of freedom.

2.   The M method

This method is used when the samples have dissimilar circular dispersals.
Using the mean directions of each of r samples calculate

$$\hat{C}_M = \sum_{i=1}^{r} n_i \cos \hat{m}_i / \hat{s}_i^2$$

$$\hat{S}_M = \sum_{i=1}^{r} n_i \sin \hat{m}_i / \hat{s}_i^2$$

The test statistic, Y, is then calculated

$$Y_r = 2\left(\sum_{i=1}^{r} \frac{1}{\hat{s}_i^2} - R_M \right)$$

and the null hypothesis that all the samples have a common direction is rejected if Y exceeds the critical value for a Chi-squared distribution with r-1 degrees of freedom.

B.   *Test for a common median*

From the Comparisons menu, choose Common Median Ungrouped. A dialog box will appear, where you can choose which samples you wish to compare - see Selecting samples to compare common mean/median. The results of the comparison will be displayed on the Results tab, with the names of the samples compared listed in the grid underneath.

To undertake this test you should have at least 10 observations (rows) in each sample. If there are insufficient observations, a figure of -99 (or -99.00000) will appear in the results.

The calculations are as follows.

1.   Calculate the median direction of the entire collection of N observations.
2.   For each of the r samples let m_i be the number of values less than the group median.
3.   Calculate M = m_1+........m_r
4.   Calculate the test statistic

$$P_r = \left\{ N^2 / [M(N - M)] \right\} \sum_{i=1}^{r} \frac{m_i^2}{n_i} - \frac{NM}{(N - M)}$$

The hypothesis that the median directions of the different samples are the same is rejected if P_r exceeds the critical value for a Chi-squared distribution with r-1 degrees of freedom.

C.   *Selecting samples to compare common mean/median*

When you select Common Median Ungrouped or Compare Mean Direction, you will be presented with the Choose Samples dialog box (below):
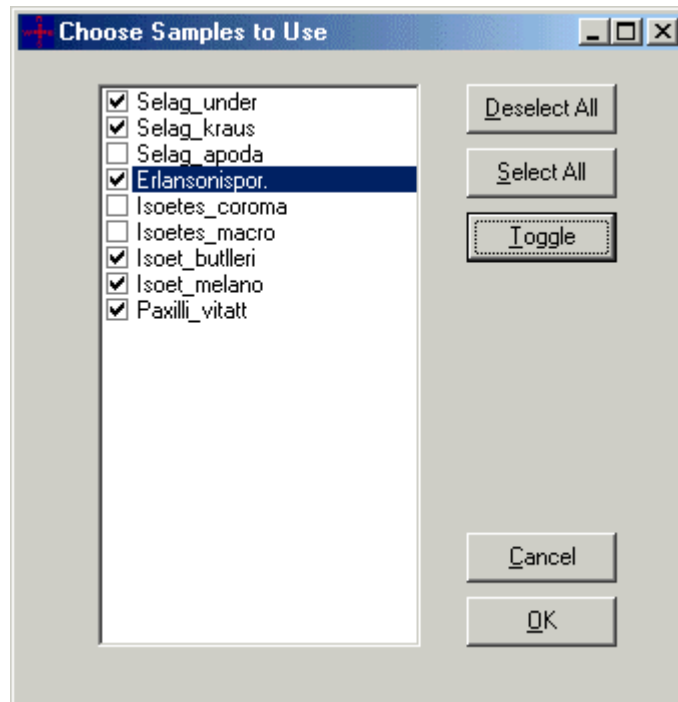
Tick the box next to a sample to select it, or use the Deselect All / Select All / Toggle buttons to make your choice, then press OK to perform the calculations.

X.   Worked example

To demonstrate *Axis*, we shall use the sample data set *ArrivalTimesB1.csv*, supplied with the program. Start *Axis*, at the file menu choose Open, and select file *ArrivalTimesB1.csv* from the folder C:\Program Files\Axis. These data record the arrival times of 254 patients at an intensive care unit, over a 12-month period.

Example 1
The arrival times may be regarded as circular measurements, with an arrival time of *m* minutes after midnight corresponding to a circular measurement of 360 x *m*/(24 x 60) degrees; thus 1 degree corresponds to 4 minutes. The plot (below) suggests a fairly steady stream of arrivals between 10am and 10pm, with smaller clusters of arrivals about 2am and 7am. For the plot below, 'Bins outside' is selected. You can experiment with different settings to see how the plot changes.
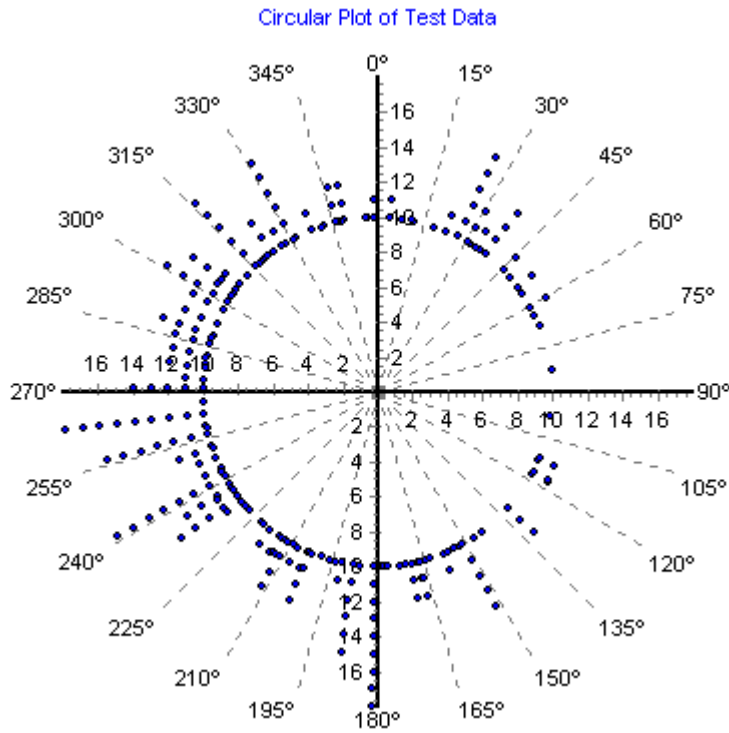
Example 1: Raw data plot

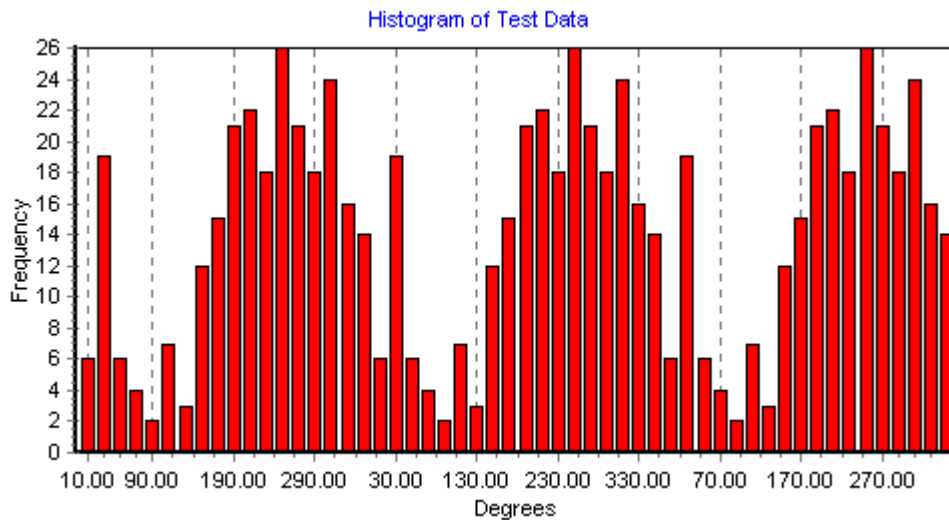## Example 2: Linear histogram

A linear histogram is constructed as if the data were linear, not angular. A starting point (e.g. 0° if the data are recorded in the range [0°, 360°], or –180° if they are in the range [-180°, 180°]), and a binwidth or grouping interval (say 5°, 10° or 20°) are selected.

For this example, we shall use a binwidth of 20°, or 80 minutes. On the Working Data tab, enter 20.00 in the box at the bottom left-hand corner, and press the Bin button to bin the data accordingly. The resulting histogram is shown below, and demonstrates the same temporal features as seen in Example 1 above.
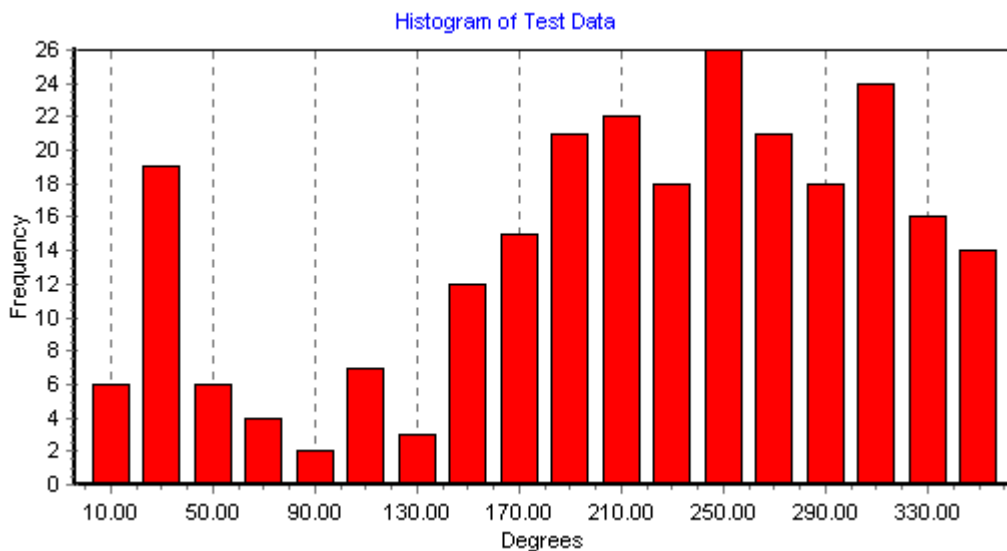
**Example 2a: Linear histogram**

To allow the eye to assess the circular distribution, the program allows you to repeat the 24-hour cycle on the linear histogram; simply use the **% Wrapround** slider to increase the amount of repetition of the cycle. When the slider is set to 100%, the histogram looks like the figure below:

Example 2a: Linear histogram - repeated

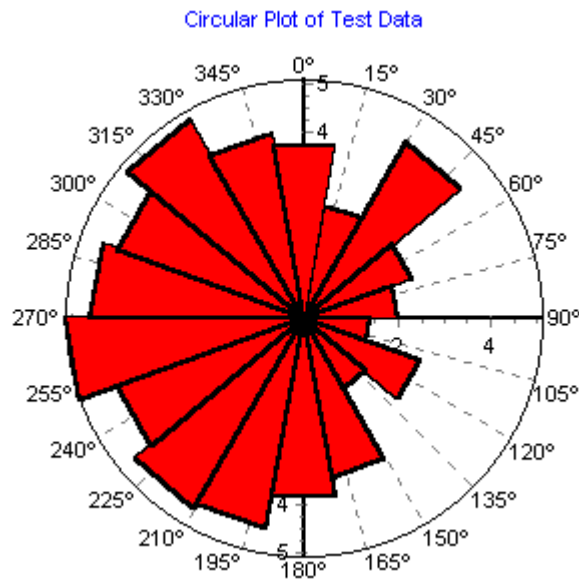**Example 3: Angular histogram – rose diagram**
In this instance, each group or bin is displayed as a sector; each sector in the histogram is centred at the midpoint of its grouping interval, with its radius proportional to the relative frequency (or the amount of data) in the group. The figure below uses the same data as Example 2, and can be viewed simply by clicking on the Circular Plot tab. Set the Segments Type option to Frequency, and Point Type to None. In this case, the bars of the histogram have been given borders for clarity: Click the Edit Chart button, double-click Series 2 to edit its properties. On the Format tab, click the Pen button. Select a colour for the border, and increase the border width to '2'.
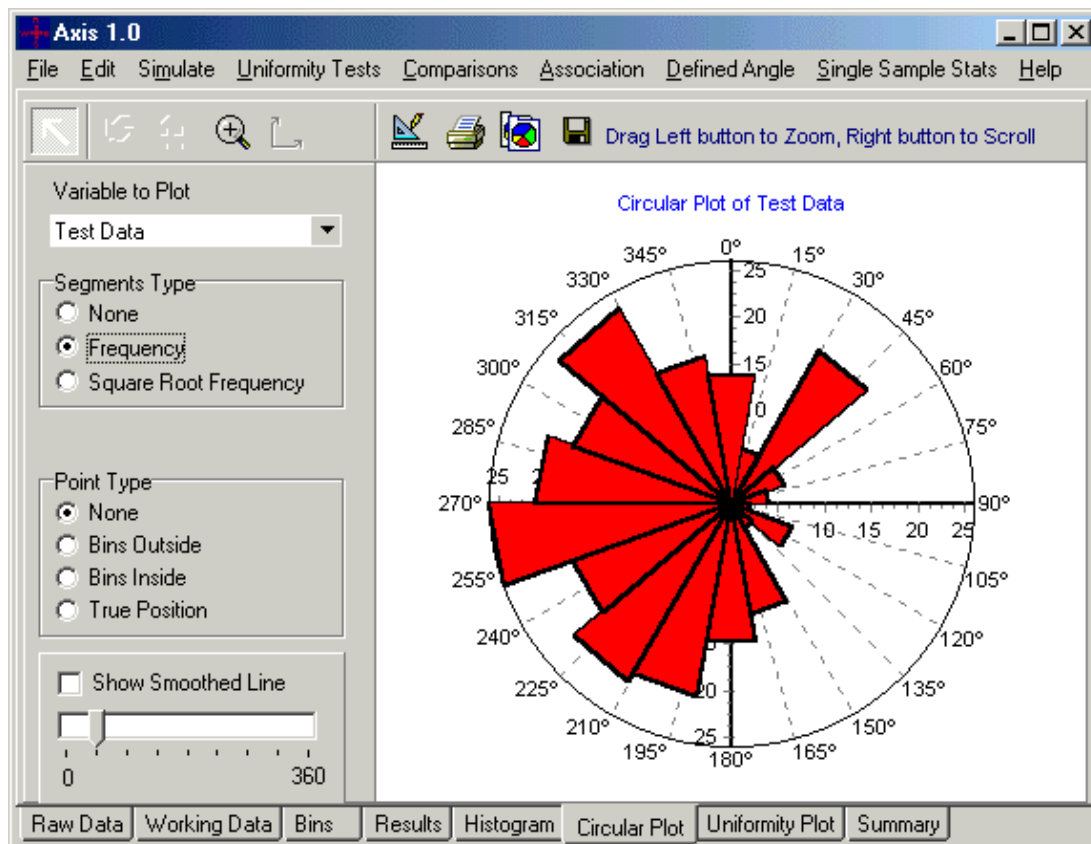
Example 3a: Angular histogram (rose diagram): relative frequency

In the figure above, because the relative frequency has been used, the magnitude of the group of arrivals at 2am is de-emphasised, and the group between 5am and 7am almost unnoticeable.

By setting the Segments Type to Square Root Frequency, the radius of each sector is plotted in proportion to the square root of its relative frequency, which reduces the range of the radii and allows the smaller peaks of the histogram to retain their prominence.

Example 3b: Angular histogram (rose diagram): square root of relative frequency

Example 4: The effect of varying bin size
One of the features of rose diagrams is their poor performance in revealing the shape of the underlying distribution. In view of this, it is recommended to use an arbitrary bin size of 10° or 20°, or possibly greater for very small data sets. Decreasing the binwidth does not necessarily produce an improvement in the quality of the plot. In the figure below, the data used above has been binned into 10°, rather than 20° groups. It can be seen that the effect is to increase the 'spikiness' of the plot, making it more difficult to grasp the broad features of the distribution:
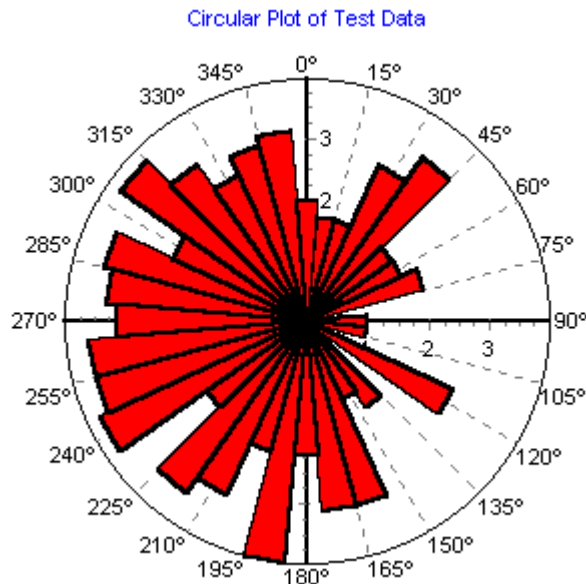
Example 4: decreased binwidth

XI.   Important references

Arsham, H. (1988) Kuiper's P-value as a measuring tool and decision procedure for the goodness-of-fit test. *J. Appl. Statist.* **15**, 131-135.

Fisher, N. I. (1993) *Statistical Analysis of Circular Data.* Cambridge University Press. 277 pp.

Fisher, N. I. & Lee, A, J (1983) A correlation coefficient for circular data. *Biometrika* **70**, 327-332.

Fisher, N. I. & Lee, A, J (1986) Correlation coefficients for random variables on a unit sphere or hypersphere. *Biometrika* **73**, 159-164.

Fisher, N. I. & Powell, C. McA. (1989) Statistical analysis of two-dimensional palaeocurrent data: Methods and examples. Aust. J. Earth Sci. **36**, 91-107.

Rothman E. D. (1971) Tests for coordinate independence for bivariate sample on a torus. Ann. Math. Statist. **42**, 1962-1969.