Journal of Applied Sciences Research, 2(11): 1019-1021, 20006

Analysis of Malaysian Wind Direction Data Using AXIS

[1]Abdul Ghapor Hussin, [1]Jedzry Fadzlin Jalaluddin, [2]Ibrahim Mohamed

[1]Centre for Foundation Studies in Science, University of Malaya, Malaysia.
[2]Institute of Mathematical Science, University of Malaya, Malaysia.

Abstract: Circular or directional data are rather special, but they do arise in surprisingly many different contexts such as within the geological, meteorological, biological, astronomical and medical sciences. The difficulty in the statistical analysis of circular data stems from the disparate topologies of the circle and the straight line: if angles are recorded in the range $[-\pi, \pi)$ radian or (0°, 360°], then direction close to the opposite end-points are near neighbours in a metric which respects the topology of the circle, but maximally distant in linear metric. For example, the "distance" between 340 and 20 angular degrees is more commonly thought of as 40 degrees, as opposed to the 320 degrees, a standard calculation would yield. AXIS is a statistical package for analysing such data which offer a range of graphical and analytical techniques. In this paper, we will discuss some of the descriptive graphical methods and summarising statistics of the Malaysian wind direction data using AXIS.

Key words: Directional data, AXIS software, wind directions.

INTRODUCTION

Data measured in the form of angels or two-dimensional orientations are to be found almost everywhere throughout applied science. They arise commonly in Biology, Geography, Geology, Geophysics, Medicine, Meteorology and Oceanography and in many other areas. Thus the need to analyse circular or directional data in this areas of research are indispensable. Further the methods advocated for usual linear data to perform exploratory data analysis or optimal statistical inference become misleading when applied to circular data due to the disparate topologies on the line and the circle. Hence new techniques for analysing circular data have emerged. However, these are often computationally intensive. Hence the need for a commercial software package for the statistical analysis of circular data is inevitable, as an analogue to MINITAB or S-Plus for statistical analysis of usual linear data.

The Axis Software: This package, called AXIS introduced in 2003, aims to take the first steps to bridge that gap. AXIS is so designed as to be of help to students, teachers, researchers, medical professionals and scientists in a wide variety of disciplines and offer a range of graphical and analytical techniques. AXIS is a stand-alone package implements the principle graphical methods and statistical tests described by Fisher[1] in Statistical Analysis of Circular Data. Periodic data of many kinds can be represented and analysed using the methods available within AXIS, which include the time of events over a 24-hours and the occurance of activity over the lunar cycle. Some of the features available are summarising statistics for each sample, descriptive graphical methods, simulating distributions, correlation between samples, testing for uniformity or randomness and comparisons between samples. As an illustration, the statistical analysis of Malaysian wind data has been performed using this package.

Malaysian Wind Direction Data: In this study, data of particular interest was obtained from the Malaysian Meteorological Services Department, consists the observations or records of maximum surface wind at Subang Airport station. The data was collected daily and measured by anemometer at 19.2m above ground level, at latitude 03°07 N and longitude 101°33E in the year 2004. Two sets of samples known as Southwest (for the period of June 2004 to Sept. 2004) and Northeast (for the period of Nov 2004 to March 2005) are considered.

Descriptive Statistics of Malaysian Wind Direction Data:

Circular plot: Circular or directional data can be displayed on both circular and angular histogram. 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. There are two display forms for a Rose diagram which are frequency and square root frequency. As an illustrations, Figures 1 and 2 shown a simple circular plot or also known as Rose diagram of Southwest and Northeast data sets respectively. A simple Rose diagram may be obtained from the frequency. Each segment on the histogram is centered on its midpoint in its grouping interval. The grouping interval is determined by the bin size on the working data tab. For 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. Also shown in the plot are the mean

Corresponding Author: Abdul Ghapor Hussin, Centre for Foundation Studies in Science, University of Malaya, Malaysia.

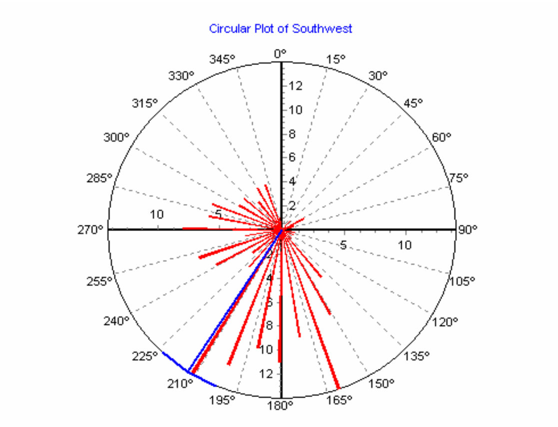

Fig. 1: Circular plot of Southwest data set.

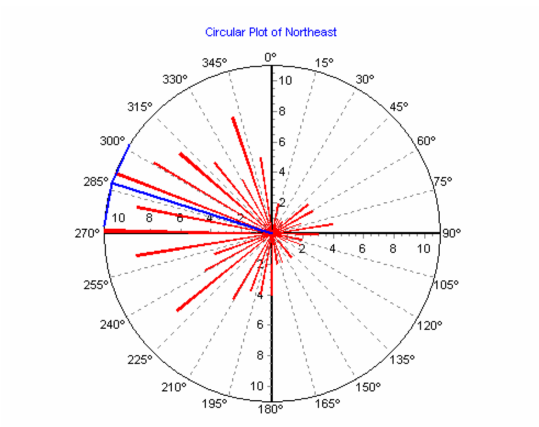

Fig. 2: Circular plot of Northeast data set.

direction of the observations and the 95% confidence limits of the mean. There is also an options for the plot of the linear histogram 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 centered on its midpoint in these grouping intervals. Again the grouping interval is determined by the bin size on the Working data tab.

Summarising statistics for each sample: Summarising statistics for each sample are obtained from the Single Sample Stats, i.e. Stats drop-down menu and are displayed under the summary tab. The following statistics are calculated as given in Table 1.

- There were 153 measurements for Southwest and 152 measurements for Northeast respectively.
- The mean direction is given by

$$\theta = \begin{cases} \tan^{-1}\left(\dfrac{S}{C}\right), & S>0, C>0 \\ \tan^{-1}\left(\dfrac{S}{C}\right)+\pi, & C<0 \\ \tan^{-1}\left(\dfrac{S}{C}\right)+2\pi, & S<0, C>0 \end{cases}$$

Table 1: Summarising statistics for each sample.

	Southwest	Northeast
Observations	153	152
Mean Direction	212.84	287.33
Lower 95% Mean	202.11	272.6
Upper 95% Mean	223.56	302.07
Mean Resultant Length	0.54	0.4
Circular Variance	0.46	0.6
Circular Std Dev	64.05	77.12
Median Direction	210	290
Circular Dispersion	79.03	146.67
Skewness	36.31	2.7
Kurtosis	-37.66	-30.47

where $S = \sum_{i=1}^{n} \sin(\theta_i)$ and $C = \sum_{i=1}^{n} \cos(\theta_i)$ The mean direction is also known the direction of the resultant vectors with given corresponding angles. For southwest data, the mean direction is 212.84° and for Northeast data, the mean direction is 287.33° respectively.

- The lower and (upper) 95% mean is given by $\{\theta \mp \sin^{-1}(1.96\sigma)\}$, where $\sigma = \sqrt{\dfrac{\delta}{n}}$ and δ is the circular dispersion. It is defined as a 95% probability that the true mean direction is greater and (less) than this value. The values is given by 202.11° and 272.6° for Southwest measurements whereas 223.56° and 302.07° for Northeast measurements.
- The mean resultant length is given by $\bar{R} = \dfrac{1}{n}\sqrt{C^2 + S^2}$ The values is 1 if all the directions in the data set are the same and approximately zero for the uniform distribution and defined as length of the directions of the resultant vectors at given angles. For the given data set the mean resultant is 0.54 and 0.4 respectively.
- The circular variance which is the measures of dispersion of circular data is defined as $V = 1 - \bar{R}$ and the values are 0.46 and 0.6 respectively.
- The circular standard deviation is analogue to standard deviation of data on the line and is given by $v = \sqrt{[-2\log(1-V)]}$. The values are 64.05 for Southwest and 77.12 for Northeast respectively.
- The median direction is a direction that divides the data into two equal sized groups. Unfortunately, this is more difficult that it might seen. We 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. For the given samples the median directions are 210° and 290° respectively.
- Circular dispersion is another measures of dispersion and is given by $\delta = \dfrac{1-\rho_2}{2\bar{R}^2}$, where $\bar{R} = \dfrac{1}{n}\sum \cos(\theta_i - \theta)$ and $\rho_2 = \dfrac{1}{n}\sum \cos 2(\theta_i - \theta)$

The values are 79.03 and 146.67 respectively.

- The measure of skewness for symmetric unimodal circular data set is nearly zero and further the measure of kurtosis for unimodal circular data set with a peak which is fitted well by a wrapped normal distribution is also nearly zero.

Testing for uniformity or randomness: The samples can be tested for uniformity, i.e. to test whether all directions are equal likely. A uniform distribution would also be expected if the directions were generated by random chances. By using the ungroup omnibus test, the test statistics are 5.10357 for Southwest dataset and 4.03981 for Northeast dataset respectively. Hence at 5% significant level, we reject the hypothesis that the samples were drawn from a uniform distribution.

Comparisons between samples: The test statistics for common median for Southwest and Northeast dataset equals 22.675 and test statistics for common mean equals 63.229. Hence the hypothesis that the median and mean are the same is rejected at 5% significant level. This is not surprise since the given datasets are the measurements of the wind directions of two different monsoons.

Conclusions: It is no doubt that this package offers a range of graphical and analytical techniques required for statistical analysis of circular data, but it has some limitation. To bridge this gap, the researcher required the package that offer others feature such as statistical inference, analysing grouped data sets, circular plots of the corresponding probability density functions as well as the regression of circular data. Hence the need of other statistical package to analyse circular data is indispensable.

REFERENCES

1. Fisher, N.I., 1993. Statistical Analysis of circular data. Cambridge University Press.
2. Fisher, N.I. and A.J. Lee, 1983. A correlation coefficient for circular data. Biometrika, 70: 327-332.
3. AXIS Manual Guide. 2003. Version 1.1.