

Two-sample and multisample testing of circular data

JERROLD H. ZAR

Department of Biological Sciences, Northern Illinois University,
DeKalb, Illinois 60115

Common data distributions on a circular scale are those of compass direction and clock time. While traditional descriptive statistics and hypothesis tests for interval scale data are not applicable to circular data, appropriate procedures are available for them (Batschelet, 1965, 1972; Mardia, 1972; Zar, 1974a, Chapter 22). This program computes the mean and angular deviation for each of several samples of circular data. The Rayleigh test is used to assess the significance of the mean. Then the null hypothesis that all population mean angles are the same is submitted to the Watson and Williams test.

Method. For each sample of data consisting of n angles, a_1 through a_n , we define

$$X = \frac{n}{\sum_{i=1}^n \cos a_i} / n$$

and

$$Y = \frac{n}{\sum_{i=1}^n \sin a_i} / n.$$

Then, the length of the mean vector is calculated as

$$r = (X^2 + Y^2)^{1/2}.$$

The mean angle, \bar{a} , is then found from the relationships: $\cos \bar{a} = X/r$ and $\sin \bar{a} = Y/r$.

Testing the significance of the mean angle amounts to testing the null hypothesis that the circular distribution is uniform against the alternate hypothesis that there is a "preferred" direction. Rayleigh's R is calculated as

$$R = nr,$$

and another test statistic is

$$z = R^2/n = nr^2.$$

Tables of critical values of r , R , and z are available (the most extensive being in Zar, 1974a, pp. 569-571); alternatively, one may calculate the associated probability directly (Zar, 1974b).

The angular deviation, a measure of angular dispersion, is best calculated as

$$s = (-2 \ln r)^{1/2} \text{ radians.}$$

The null hypothesis of all samples coming from populations with the same mean angle is then submitted to the Watson and Williams test (Watson & Williams, 1956), for which we compute

$$F = \frac{K(N-k) \left(\sum_{j=1}^k R_j - R \right)}{(k-1) \left(N - \sum_{j=1}^k R_j \right)},$$

This research is Contribution No. 545 from the Department of Biological Sciences, Northern Illinois University.

where

k = number of samples,

R_j = Rayleigh's R for Sample j ,

R = Rayleigh's R for all samples combined,

$$N = \sum_{j=1}^k n_j$$

(the sum of the k sample sizes), and K is a correction for bias (tabulated in Zar, 1974a, p. 574). The degrees of freedom associated with F are $k - 1$ and $N - k$. If $k = 2$, then an alternative test statistic is $t = \sqrt{F}$, with $N - 2$ df.

Input. Input to this program may be in any one of four forms: (1) degrees, (2) radians, (3) sectors, or (4) times of day. If data are in sectors, the circle may be designated as being divided into any number of equal sectors. For example, if it was stated that there were to be eight sectors, then an input datum of "1" would indicate 45 deg, "2" would be 90 deg, and so on; if there were 12 sectors, then input data of "1," "2," and "3" would designate 30 deg, 60 deg, and 90 deg, respectively. Input data representing times of day are expressed on the 24-h clock, so that 0600 = 90 deg, 1200 = 180 deg, and so on. The user may indicate the desired input format.

Output. For each of the k samples of data, the following are given as printed output: sample number, sample size, r , R , z , probability of z (which is the same as the probability of r or of R), the mean angle in both angles and radians, the sine and cosine of the mean angle, and the angular deviation in both angles and radians. Then, the following printed output is obtained for the combination of all k samples (if $k > 1$): N , r , R , z , the probability of z , mean angle in both angles and radians, the sine and cosine of the mean angle, and the angular deviation in both angles and radians. In addition, if the input data are sectors or times of day, then all means and angular deviations are given in sectors or times of day, respectively. Then F (also t , if $k = 2$) is given, with its degrees of freedom, and the probability of F (or t) is computed by the routine of Dunlap and Duffy (1975). The input data may be printed at the user's option.

Computer and language. This program is written in FORTRAN IV and was prepared and tested on the Northern Illinois University IBM 360/67 computer.

Restrictions. The present program is limited to $k = 10$ and $N = 2,000$, but these limitations may be changed readily by modifying two FORTRAN statements.

Availability. A source program listing and documentation for program use may be obtained without charge from Jerrold H. Zar, Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115.

REFERENCES

- BATSCHLET, E. *Statistical methods for the analysis of problems in animal orientation and certain biological rhythms*. Washington, D.C.: American Institute of Biological Sciences, 1965.
- BATSCHLET, E. Recent statistical methods for orientation data. In S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, & R. E. Belleville (Eds.), *Animal orientation and navigation*. Washington, D.C.: National Aeronautics and Space Administration, 1972. Pp. 61-91.
- DUNLAP, W. P., & DUFFY, J. A. FORTRAN IV functions for calculating exact probabilities associated with z , χ^2 , t , and F values. *Behavior Research Methods & Instrumentation*, 1975, 7, 59-60.

- MARDIA, K. V. *Statistics of orientation data*. New York: Academic Press, 1972.
- WATSON, G. S., & WILLIAMS, E. J. On the construction of significance tests on the circle and the sphere. *Biometrika*, 1956, 43, 344-352.
- ZAR, J. H. *Biostatistical analysis*. Englewood Cliffs, N.J: Prentice-Hall, 1974. (a)
- ZAR, J. H. Probabilities of Rayleigh's test statistics for circular data. *Behavior Research Methods & Instrumentation*, 1974, 5, 450. (b)