Probabilities for the Kolmogorov-Smirnov one-sample test statistic*

JERROLD H. ZAR
Department of Biological Sciences
Northern Illinois University, DeKalb, Illinois 60115

The Kolmogorov-Smirnov one-sample test is appropriate for testing goodness of fit of a hypothesized population distribution to a sampled population distribution (Bradley, 1968; Siegel, 1956; Zar, 1974).

If F(X) is the hypothesized cumulative frequency of the variable X [i.e., F(X) is the number of cases expected to be less than or equal to X], and S(X) is the observed cumulative frequency of X [i.e., S(X) is the number of cases in the sample which are less than or equal to X], then

$$D = [\max | S(X) - F(X) |]/n,$$

where n is the sample size.

A computer program is available which computes, for a given value of D, the probability that the observed sample came from a population having the hypothesized distribution. This program is especially useful as a subroutine in conjunction with programs which perform the Kolmogorov-Smirnov goodness of fit test. It also has been used to prepare the extensive table of critical values of D appearing in Zar (1974).

The probabilities of D were computed using Eq. 3.0 of Birnbaum and Tingey (1951) and a subroutine for logarithms of factorials.

Input. Any number of values of the Kolmogorov-Smirnov D, each with its associated sample size.

Output. The probability, under the null hypothesis, of a D at least as large as each D submitted.

Computer and Language. The program was prepared in FORTRAN IV on the Northern Illinois University IBM 360/67 computer.

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

REFERENCES

Birnbaum, Z. W., & Tingey, F. H. One-sided confidence contours for probability distribution functions. Annals of Mathematical Statistics, 1951, 22, 592-596.

Bradley, J. V. *Distribution-free statistical tests.* Englewood Cliffs, N.J.: Prentice-Hall, 1968. Pp. 296-303.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956. Pp. 47-52.

Zar, J. H. *Biostatistical analysis.* Englewood Cliffs, N.J.: Prentice-Hall, 1974. Pp. 471-474.

*Contribution No. 513 from the Department of Biological Sciences, Northern Illinois University.

Probabilities of Rayleigh's test statistics for circular data*

JERROLD H. ZAR
Department of Biological Sciences
Northern Illinois University DeKalb, Illinois 60115

Common circular data distributions include those involving compass directions or clock time. Means of variables such as angular directions or times of day may be computed by appropriate methods (Batschelet, 1965, pp. 12-13; Mardia, 1972, pp. 20-21; Zar, 1974, pp. 313-314). Subsequently, one may ask whether the observed mean angle or time of day reflects a significant concentration. This consists of testing the null hypothesis of a uniform population distribution against the alternate hypothesis of a "preferred" direction or time in the population sampled. This may be accomplished by employing Rayleigh's test (Batschelet, 1965, pp. 28-29; Mardia, 1972, pp. 133-136; Zar, 1974, pp. 316-318).

For a sample of size n, Rayleigh's R is a convenient test statistic for testing the null hypothesis of uniformity:

$$R = nr,$$

where r is the length of the mean vector. The statistic

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

is also employed for this purpose. Indeed, r itself is suitable for the hypothesis testing.

A computer program has been developed which computes the probability of a given value of z (or R, or r) for a specified sample size, n. This program is especially useful as a subroutine in conjunction with a program (e.g., Zar, 1969) which computes descriptive statistics, such as means, for circular data. It has also been used to prepare extensive tables of critical values of r, R, and z (Zar, 1974, pp. 569-571).

Computer Method. The probabilities of z are computed using Eq. 6 of Durand and Greenwood (1958), as this procedure has been found to give slightly more accurate results than their Eq. 4, distinctly better results than the Pearson curve approximation (Stephens, 1969), and very much better results than the chi-square approximation (Stephens, 1969).

Input. Any number of values of Rayleigh's z, each with its associated sample size. Comments in the program source listing explain simple modifications to allow r or R to be used as input instead of z.

Output. The probability that the sample data came from a uniformly distributed population of circular data (i.e., H_0: the population mean vector = 0).

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

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

REFERENCES

Batschelet, E. *Statistical methods for the analysis of problems in animal orientation and certain biological rhythms.* Washington, D.C: American Institute of Biological Sciences, 1965.

Durand, D., & Greenwood, J. A. Modifications of the Rayleigh test for uniformity in analysis of two-dimensional orientation data. Journal of Geology, 1968, 66, 229-238.

Mardia, K. V. *Statistics of orientation data.* New York: Academic Press, 1972.

Stephens, M. A. Tests for randomness of direction against two circular alternatives. Journal of the American Statistical Association, 1969, 64, 280-289.

Zar, J. H. A FORTRAN IV program for statistics of circular distributions. Behavioral Science, 1969, 14, 80-81.

Zar, J. H. *Biostatistical analysis.* Englewood Cliffs, N.J.: Prentice-Hall, 1974.

Contribution No. 514 from the Department of Biological Sciences, Northern Illinois University.