

PROGRAM ABSTRACTS/ALGORITHMS

Watson's Nonparametric Two-Sample Test

JERROLD H. ZAR

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

Watson (1962) introduced a distribution-free test of the null hypothesis that two samples come from the same population. The computed statistic, U^2 , depends on the relative ranks of the observations in the two samples, and not on the starting point of the measurement scale. Thus, while it may also be used for data measured on a linear scale, the test is especially useful for data on a circle, for which conventional hypothesis testing is inappropriate.

Common data from a circular scale of measurement are compass directions or clock times; parametric testing for such data is possible (see Batschelet, 1965, 1972; Mardia, 1972; Zar, 1974, 1976). Watson's nonparametric test is described in Batschelet (1965, pp. 35-36), Mardia (1972, pp. 201-203), and Zar (1974, pp. 324-326). A computer program has been developed that performs Watson's (1962) calculation of U^2 , appropriate for either nontied data or data tied within and/or between samples.

The most extensive tables of critical values of U^2 are in Zar (1974, pp. 575-576). Stephens (1965) gives Pearson curve approximations for some large sample sizes. For large samples, a normal deviate may be calculated as:

$$Z = \frac{[(N+1)/12N] - U^2}{[(n_1 - 1)(n_2 - 1)(N+1)/360Nn_1n_2]^{1/2}}$$

where n_1 and n_2 are the two sample sizes, and $N = n_1 + n_2$ (see Mardia, 1972, p. 202; Stephens, 1965).

Input. The input data are the observed measurements

from two samples. The input format is at the user's option.

Output. For each of the two samples, the frequency and cumulative relative frequency of each datum is given. The Watson U^2 and the two sample sizes are given; with these, one can consult a table of critical values.

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

Restrictions. The present program is limited to sample sizes of 99, but this limitation may be changed readily by modifying one FORTRAN statement.

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 Galler, S. R., 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.
- MARDIA, K. V. *Statistics of directional data*. New York: Academic Press, 1972.
- STEPHENS, M. A. Significance points for the two-sample statistic $U^2_{M,N}$. *Biometrika*, 1965, **52**, 661-663.
- WATSON, G. G. Goodness of fit tests on a circle. II. *Biometrika*, 1962, **49**, 57-63.
- ZAR, J. H. *Biostatistical analysis*. Englewood Cliffs, N.J: Prentice-Hall, 1974.
- ZAR, J. H. Two-sample and multisample testing of circular data. *Behavior Research Methods & Instrumentation*, 1976, **8**, 329-330.