

a similar qualitative conclusion. If one suspects a violation of the assumptions of parametric methods, then the nonparametric procedures may be preferred in order to have greater trust in the resulting  $P$  value, but one should not expect a nonparametric procedure to fundamentally alter the interpretation of an experiment.

## LITERATURE CITED

- Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* **47**:2–15.
- Potvin, C., and D. A. Roff. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology* **74**:1617–1628.

*Ecology*, 76(6), 1995, pp. 1998–2000  
© 1995 by the Ecological Society of America

STATISTICAL SIRENS: THE ALLURE OF NONPARAMETRICS<sup>1</sup>

DOUGLAS H. JOHNSON

Northern Prairie Science Center, National Biological Service, Jamestown, North Dakota 58401 USA

**Abstract.** Although nonparametric statistical methods have a role to play in the analysis of data, often their virtues are overstated and their deficiencies overlooked. A recent Special Feature in *Ecology* advocated nonparametric methods because of an erroneously stated advantage that they require no assumptions regarding the distribution underlying the observations. The present paper points out some often-ignored features of nonparametric tests comparing two means, and advocates parameter estimation as a preferred alternative to hypothesis testing in many situations.

**Key words:** hypothesis testing; nonparametric methods; normal distribution; parameter estimation; Wilcoxon-Mann-Whitney test vs.  $t$  test.

How often have you read something like, “Our data were not normally distributed, so we used nonparametric methods.”? The reasoning is that nonparametric methods require few, if any, assumptions. In a recent Special Feature article in *Ecology*, Potvin and Roff (1993:1619) made the point explicit: “The main advantage of nonparametric methods over their parametric counterparts is the *absence of assumptions regarding the distribution underlying the observations*” (emphasis added). Numerous authors have made similar statements, but I focus on the Potvin and Roff article because it was intended as an update for ecologists. My purpose here is to indicate that their characterization is incorrect and the implied advice is misleading.

The situations for which nonparametric methods are commended vary, and include correlation, regression, and more, but I concentrate on the comparison of two means. The usual (parametric) method is the  $t$  test, which can be employed either when variances within the two groups are the same (ordinary Student's  $t$ ) or when they differ (Welch-Satterthwaite modification).

The nonparametric counterpart is the Wilcoxon rank sum or the equivalent Mann-Whitney test (WMW test).

Several points bear emphasis. First, data do not need to be normally distributed in order to apply the  $t$  test. Only the means need to be, and that property is assured by the Central Limit Theorem, even for relatively small samples, for all but the most perverse data. This is exemplified in Fig. 1, which shows at the upper left a very nonnormal (in fact, a uniform) distribution of original data. Random samples of size  $N = 2, 4,$  and  $8$  demonstrate that the distribution of averages based on even those small sample sizes rapidly approaches normality.

Second, statements are often made about means of distributions differing, based on nonparametric tests such as WMW, although Potvin and Roff did not make this mistake. The WMW test actually tests the hypothesis that the two distributions are *identical*, not that they have the same mean (e.g., Gibbons 1985). In particular, variances must be the same if the test is to compare means; as Hollander and Wolfe (1973:71) stated, “we assume that the two populations do not differ in dispersion.” To compare means, the WMW test requires the assumption that the two distributions are

<sup>1</sup> Manuscript received 23 May 1994; revised 16 December 1994; accepted 27 January 1995; final version received 21 February 1995.

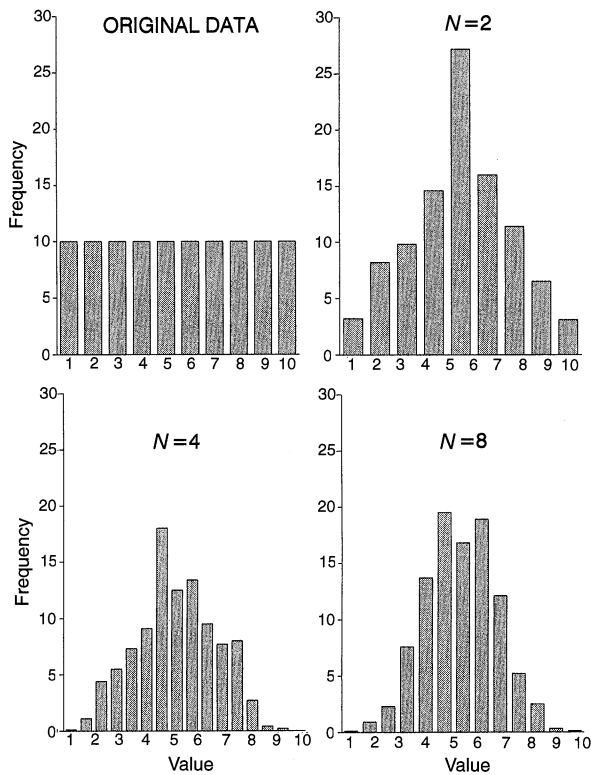


FIG. 1. Uniform distribution of values, and distributions of means based on random samples of size  $N = 2, 4,$  and  $8$  from that population, indicating how means from a nonnormal distribution can rapidly approach normality.

identical in shape and scale, differing only in their means. This assumption can be harder to justify than the asymptotic normality demanded by the  $t$  test, and is rarely evaluated (Petranka 1990). A significant test statistic from, say, the WMW procedure indicates that the two distributions differ in some way; it does not suggest how they differ—mean, variance, shape, etc. Petranka (1990) provided an example of two distributions that had identical means and medians: the  $t$  test indicated no difference between means, whereas the WMW test was significant. If the distributions have different variances, the Welch-Satterthwaite version of the  $t$  test performs well (Wang 1971) and is more valid than the WMW test (Fligner and Policello 1981, Stewart-Oaten 1995).

Third, although nonparametric methods can be used for estimating parameters, they are better adapted to testing hypotheses and used mostly for that purpose. By their very nature, nonparametric methods do not specify an easily interpreted parameter (Simberloff 1990). And parameter estimates are generally more useful than hypothesis tests. Almost all null hypotheses tested truly are false; the only real question is whether

the sample gathered is large enough to make the test statistic significant. For example, does the density of a plant species in one study area differ from that in another? Of course. Densities might be 5000 plants/ha in one area and 4999.9 plants/ha in the other, but that is a real difference, which will be detected (the difference between sample means will become significant) once the samples are large enough. As Yoccoz (1991: 106) noted, "most biologists and other users of statistical methods seem still to be unaware that significance testing by itself sheds little light on the questions they are posing." Overemphasis on statistical hypothesis testing may be due to confusing that activity as an "inductive or even descriptive procedure" with the deductive logic involved in hypothesis testing in "strong inference" (Quinn and Dunham 1983).

The emphasis on hypothesis tests raises the issue of biological significance, as contrasted with statistical significance. Biological significance implies importance in some sense. Statistical significance means that the result was unlikely due to chance; if the null hypothesis is true, an improbable event has occurred. Differences of certain magnitudes are said to be not biologically significant, although they were shown to be statistically significant (does that mean that the samples were too large?). And some authors talk of differences that are biologically significant, even though they did not meet usual  $\alpha$  criteria (does that mean that the differences are important but perhaps not real?). It is rarely sufficient to know that two parameters differ; estimates of their values are needed for useful application.

More meaningful than a test comparing two means are estimates of the difference between means, along with an assessment of one's confidence in that difference. If ecologists are to be taken seriously by decision makers, they must provide information useful for deciding on a course of action, as opposed to addressing purely academic questions. What Roberts (1990:382) said about business applies equally well to ecology: "[S]ignificance tests are irrelevant to the manager who must make the business decision."

Returning to the plant density example, estimates of the difference between the two areas would approach 0.1 plants/ha, the true value, as sample sizes grow large. The estimated difference, along with a confidence interval for it, can be brought to bear on a decision. Neither the  $t$  statistic—the ratio of the estimate to its estimated standard error—nor the significance level of the  $t$  value is useful or even interesting.

Nonparametric methods have an important role to play, especially in the analysis of ordinal data. My concern is only that they are too freely adopted for inappropriate purposes. Glass et al. (1972:237) referred to "a largely unnecessary hegira to non-parametric statistics" that took place in education and the social sci-

ences during the 1950s and that ecology now seems in danger of repeating. Parameters are generally of most interest, so we should provide estimates of those parameters that are meaningful and applicable to making real decisions. If the data we have do not meet assumptions underlying the standard techniques, and those assumptions are in fact necessary, then alternatives should be considered, such as transforming the data to better meet the assumptions (Green 1979:43–54, Atkinson and Cox 1988) or using robust parametric methods (Huber 1981, Bickel 1988), which are less sensitive to violations of the assumptions.

If ecologists are careful about randomly sampling the populations about which they want to draw inferences, standard parametric methods will ordinarily be adequate; if they are not, nonparametric methods will not protect them from sailing off course (Box et al. 1978, Stewart-Oaten 1995).

#### ACKNOWLEDGMENTS

This note benefitted from comments by C. E. McCulloch, W. E. Newton, T. L. Shaffer, N. G. Yoccoz, and two anonymous referees. B. R. Euliss prepared the figure.

#### LITERATURE CITED

- Atkinson, A. C., and D. R. Cox. 1988. Transformations. Pages 312–318 in S. Kotz and N. L. Johnson, editors. *Encyclopedia of statistical sciences*. Volume 9. John Wiley & Sons, New York, New York, USA.
- Bickel, P. 1988. Robust estimation. Pages 157–163 in S. Kotz and N. L. Johnson, editors. *Encyclopedia of statistical sciences*. Volume 8. John Wiley & Sons, New York, New York, USA.
- Box, G. E. P., W. G. Hunter, and J. S. Hunter. 1978. *Statistics for experimenters*. John Wiley & Sons, New York, New York, USA.
- Fligner, M. A., and G. E. Policello II. 1981. Robust rank procedures for the Behrens-Fisher problem. *Journal of the American Statistical Association* **76**:162–168.
- Gibbons, J. D. 1985. *Nonparametric statistical inference*. Second edition. Marcel Dekker, New York, New York, USA.
- Glass, G. V., P. D. Peckham, and J. R. Sanders. 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research* **42**:237–288.
- Green, R. H. 1979. *Sampling design and statistical methods for environmental biologists*. John Wiley & Sons, New York, New York, USA.
- Hollander, M., and D. A. Wolfe. 1973. *Nonparametric statistical methods*. John Wiley & Sons, New York, New York, USA.
- Huber, P. J. 1981. *Robust statistics*. John Wiley & Sons, New York, New York, USA.
- Petranka, J. W. 1990. Caught between a rock and a hard place. *Herpetologica* **46**:346–350.
- Potvin, C., and D. A. Roff. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics. *Ecology* **74**:1617–1628.
- Quinn, J. F., and A. E. Dunham. 1983. On hypothesis testing in ecology and evolution. *American Naturalist* **122**:602–617.
- Roberts, H. V. 1990. Applications in business and economic statistics: some personal views. *Statistical Science* **5**:372–402.
- Simberloff, D. 1990. Hypotheses, errors, and statistical assumptions. *Herpetologica* **46**:351–357.
- Stewart-Oaten, A. 1995. Rules and judgments in statistics: three examples. *Ecology* **76**:2001–2009.
- Wang, Y. Y. 1971. Probabilities of the Type I errors of the Welch tests for the Behrens-Fisher problem. *Journal of the American Statistical Association* **66**:605–608.
- Yoccoz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin of the Ecological Society of America* **72**:106–111.

## ERRATA

In an article by C. Potvin and D. A. Roff (“Distribution-free and robust statistical methods: viable alternatives to parametric statistics?”, *Ecology* **74**(6): 1617–1628) it should be understood that a false positive means accepting  $H_1$  if  $H_1$  is wrong. We recognize that the use of “positives” could be ambiguous. There is an error in Table 2 (p. 1619), due to a misinterpretation by SYSTAT of a dot indicating a missing value. Correct values for diameters are  $U = 563$ ,  $P = 0.046$ ; for log diameter  $t = -2.00$ ,  $P = 0.05$ ; and for square root of diameter  $t = -1.991$ ,  $P = 0.05$ . This unfortunate mistake, however, does not invalidate the claim that upon violation of the normality assumption WMW may prove more powerful than a  $t$  test. Another subset of the maple–ash data set, available upon request to the

authors, shows a similar behavior with sum of ranks for maple of 686 and ash of 854. The corresponding  $U$  statistic is 251 with  $P = 0.03$ . When analyzed by  $t$  test, values are  $-1.983$ ,  $P = 0.05$ . The authors stand by their opinion that, given the existence of distribution-free statistical methods, these should be given preference if violation of the assumptions of parametric tests are suspected.

In an article by R. G. Shaw and T. Mitchell-Olds (“ANOVA for unbalanced data: an overview,” *Ecology* **74**(6):1638–1645) there is a mistake in Table 1, p. 1642. The correct marginal mean for size class 2 is 105.87 rather than 108.87. We would like to acknowledge Drs. D. J. Fairbairn and S. M. Smith for pointing out these mistakes.