Ecology, 74(6), 1993, pp. 1617–1628

DISTRIBUTION-FREE AND ROBUST STATISTICAL METHODS: VIABLE ALTERNATIVES TO PARAMETRIC STATISTICS?[1]

CATHERINE POTVIN AND DEREK A. ROFF
Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montréal, Québec, Canada H3A 1B1

Abstract. After making a case for the prevalence of nonnormality, this paper attempts to introduce some distribution-free and robust techniques to ecologists and to offer a critical appraisal of the potential advantages and drawbacks of these methods. The techniques presented fall into two distinct categories, methods based on ranks and "computer-intensive" techniques. Distribution-free rank tests have features that can be recommended. They free the practitioner from concern about the underlying distribution and are very robust to outliers. If the distribution underlying the observations is other than normal, rank tests tend to be more efficient than their parametric counterparts. The absence, in computing packages, of rank procedures for complex designs may, however, severely limit their use for ecological data.

An entire body of novel distribution-free methods has been developed in parallel with the increasing capacities of today's computers to process large quantities of data. These techniques either reshuffle or resample a data set (i.e., sample with replacement) in order to perform their analyses. The former we shall refer to as "permutation" or "randomization" methods and the latter as "bootstrap" techniques. These computer-intensive methods provide new alternatives for the problem of a small and/or unbalanced data set, and they may be the solution for parameter estimation when the sampling distribution cannot be derived analytically. Caution must be exercised in the interpretation of these estimates because confidence limits may be too small.

INTRODUCTION

The majority of statistical methods used in biology assume that the error variances are homogeneous and normally distributed, but as early as 1947 this assumption was seriously doubted. Geary (1947) suggested that "amends might be made in the interest of the new generation of students by printing in leaded type in future editions of existing textbooks and in all new text-books, 'Normality is a myth; there never was, and never will be, a normal distribution.'" Biologists (Hampel et al. 1986, Austin 1987, Biondini et al. 1988) have recently provided evidence indicating that nonnormality is prevalent in ecological data. Moreover, only rarely does one have sufficient data to test the assumption adequately or to estimate a reasonable alternative (Forsythe 1972, Mardia 1980, Seaman and Jaeger 1990). It is therefore legitimate to ask whether assuming normality is meaningful and whether testing normality and, if necessary, correcting for nonnormality are worth the trouble.

Violation of the normality assumption has three main causes. First, the distribution underlying the observations may be other than normal, e.g., symmetric with long tails or kurtose. Pearson (1931) and Geary (1947) showed that the actual probability of differences between mean and variance may differ considerably from the probability derived from standard tables if the universal distribution is not normal. Second, outliers may be present. We distinguish between outliers and "blunders," which should be eliminated de facto. Analysis of variance is robust against "spotty data" sprinkled at random in a data set in the sense that outliers will not lead to too many false positives, i.e., type II error, but gross errors, either concentrated or spread, will increase the mean squares substantially, thus decreasing the chances of detecting real effects and increasing the number of false negatives, i.e., type I error (Tukey 1962, Hettmansperger and McKean 1978). Finally, error variances may be heterogeneous. This type of heterogeneity is frequent in ecological studies, especially field studies, where the true error variance is likely to vary from one observation to the next. Because the error variance will affect certain treatments in an unpredictable way, it will result in a loss of power (Cochran 1947, Wu 1985). Heteroscedasticity may lead to large errors in significance testing, by artificially dis-

[1] For reprints of this Special Feature, see footnote 1, p. 1615.

torting the significance levels (Cochran 1947). It is apparent that, if the assumption of normality does not hold, significance levels may be distorted, the methods may suffer loss of power, and the estimates obtained based on this assumption may be wildly inaccurate (Huber 1964, 1973). Users frequently disregard assumptions of parametric tests (Conover and Iman 1976), and their results are correspondingly incorrect or unreliable.

Within the framework of parametric analyses, some solutions for the problem of nonnormality have been proposed. The best known include trimming, winzorization, and data transformation when outliers are present (Tukey 1962, Winer 1971, Berry 1987, Burbidge et al. 1988) and the change of scale to reestablish normality by means of data transformation (Bartlett 1947, Morris 1985). Weighted analysis may be an appropriate solution for error heterogeneity (Winer 1971). A considerable amount of research is aimed at developing robust methods. These methods are not entirely distribution free but are less sensitive to the assumption of normality than are the usual parametric methods. Most robust techniques are specifically designed to be relatively insensitive to outliers (Huber 1977, Efron 1979). An alternative avenue is to consider statistical methods that are distribution free, i.e., that make no assumption regarding the distribution underlying the observations.

After making a case for the prevalence of nonnormality, this paper attempts to introduce some distribution-free and robust techniques to ecologists and to offer a critical appraisal of the potential advantages and drawbacks of these methods. The techniques presented fall into two distinct categories, methods based on ranks and "computer-intensive" techniques.

Methods Based on Ranks

A large portion of the literature on nonparametric statistics deals with methods based on ranks, which analyze the ranks of the observations, rather than their actual values (Shirley 1987). We will first consider distribution-free methods used as substitutes for analysis of variance (ANOVA). In contrast with ANOVA, different rank tests should be used depending on whether the design is one- or two-way. One of the most commonly used nonparametric methods for the one-way layout is the Kruskal–Wallis (KW) test. The basic model under consideration is

$$X_{ij} = \mu + T_j + e_{ij}, \qquad P[e_{ij} \leq x] = F(x), \qquad (1)$$

where μ is the population mean, T_j the effect of the jth level of treatment T, and e_{ij} the error variables. The KW test assumes that the errors are mutually independent and come from the same continuous population. The difference in the assumptions of the KW compared with ANOVA is that KW does not assume that the data follow any particular distribution. It is a distribution-free method, while ANOVA assumes that the errors are random NORMAL variables. The KW test shares with ANOVA the assumption that the error variances are homogeneous, a property known as homoscedasticity. Finally, although formally the errors should come from a continuous population, modifications accounting for ties are available. In rank tests, the null hypothesis of no treatment effect will be accepted if the samples come from populations having the same location, i.e., populations for which the medians and quartiles coincide (Hollander and Wolfe 1973). In the KW test, the observations are ranked across the whole data set, from smallest to largest or vice versa (Zar 1984), and the corresponding statistic is based on the sum of the ranks for each sample. Exact probabilities have been computed for the Kruskal–Wallis statistic and are used for significance testing. When the treatment effect is represented by only two levels ($k = 2$), the Kruskal–Wallis test reduces to a Mann–Whitney test, also known as a Wilcoxon rank sum test (WMW).

Methods to accommodate the two-way layout have been developed for the analysis of randomized block designs (Hollander and Wolfe 1973, Conover 1980). Consider the model,

$$X_{ij} = \mu + T_j + e_{ij}, \qquad P[e_{ij} \leq x] = F_i(x), \qquad (2)$$

where the index i refers to the ith block. Since the law of errors varies across blocks, the observations should be ranked within each block, which contrasts with the KW where the observations are ranked across the whole data set. In that case, the most frequently used test is the Friedman; the assumptions of the test are that, within each block, the errors are mutually independent and come from the same population. On the other hand, if the law of errors within blocks varies solely by a parameter of location, that is, $F_i(x) = F(x - B_i)$, then by putting $e_{ij}{}^* = e_{ij} - B_i$, model 2 can be rewritten as:

$$X_{ij} = \mu + B_i + T_j + e_{ij}{}^*, \qquad P[e_{ij}{}^* \leq x] = F(x). \qquad (3)$$

In the presence of such an additive model, the Quade test is often used. This test ranks the observations within the blocks while including information on the ranking of the blocks themselves (Thompson and Ammann 1989). Because of the difference in assumptions, the Friedman test considers the block only as a nuisance, while the Quade test enables determination of the effect of the blocks themselves. In that respect the latter test is more useful for the analysis of two factorial experiments. Tardif (1987) demonstrated that, assuming normally distributed data, the Friedman test is to be preferred when more than eight observations per block are compared; otherwise the Quade test is more efficient.

The main advantage of nonparametric methods over their parametric counterparts is the absence of assumptions regarding the distribution underlying the observations. However, to be acceptable alternatives, distribution-free techniques should prove to be powerful. Pitman (1949) developed the concept of the asymptotic relative efficiency (ARE) to compare the large-sample power of two sequences of tests. ARE's have been used to compare several nonparametric tests with their parametric equivalents (Hodges and Lehmann 1955). When the assumptions of the parametric tests are met, nonparametric tests are less efficient than their parametric counterpart (Tukey 1962, Conover 1980). However, when the underlying distribution is nonnormal, nonparametric tests prove generally more efficient than those assuming normality of the observations (Table 1). A word of caution is needed in the interpretation of the ARE's: they pertain to the efficiency of the tests for large samples and not to the efficiency of the tests when sample sizes are small. However, Monte-Carlo simulations for small samples are generally in agreement with the results of the ARE's.

To illustrate some of these points, we will compare the sizes of some 30 saplings of ash and maple measured in a forest of southern Québec. A look at size distribution (Fig. 1) shows that sapling diameter follows a skewed distribution. Skewness is, however, more pronounced for maple than for ash, and transformation fails to normalize completely the size distribution. When the sapling sizes of the two species are compared, the large-sample approximation of the Mann–Whitney test indicates a highly significant difference, the sum of ranks for ash (1058) being bigger than that for maple (771). The basic hypothesis in the WMW test is that the two samples (ash and maple) come from the same population and we can clearly reject it (Table 2). For raw data, the t test uncovered a difference that was only marginally significant with $P = .051$; therefore the analysis of untransformed data with a parametric method leads to ambiguous results. On the other hand, transformation confirms the result of the WMW tests (Table 2). In our example, the WMW test proves more powerful than its normal counterpart. The reported ARE's for this test (Table 1) indicate that for uniform and double exponential distributions, the WMW was equally or more efficient than the t test, and the same conclusion was drawn for heavy–tailed populations (Hettmansperger and McKean 1978). Our results suggest a similar behavior of the WMW for a skewed distribution.

Table 1. Asymptotic relative efficiency (ARE) to compare the large-sample power of central nonparametric methods and their parametric equivalent.

Underlying distribution	Kruskal-Wallis	Friedman $k = 2$	Friedman $k = 4$	Friedman $k = 10$	Kendall
Normal	0.955	0.637	0.764	0.868	0.912
Uniform	1.000	0.667	0.800	0.909	1.000
Double exponential	1.500	1.000	1.200	1.364	1.266

Table 2. Analyses for the difference in sapling diameter of ash and maple from a forest of southern Québec. Raw data were analyzed by means of the large-sample approximation of the Wilcoxon-Mann-Whitney U test (WMW) corrected for ties and by means of Student's t test. Log- and square-root-transformed data were also analyzed by t test.

	WMW		t test	
	U	P	t	P
Diameter (cm)	23.82	.001	−1.993	0.051
Log diameter	···	···	−2.460	0.017
Square root of diameter	···	···	−2.480	0.016

The above example uses the simplest possible situation, that in which only two samples are compared. Most often, biologists are studying far more complex situations; therefore, to be useful, rank tests need to be able to accommodate complex designs. The KW is readily extended to more complex experimental designs, including trends (Shorack 1967, Shirley 1977, House 1986, Williams 1986), covariance analysis (Shirley 1981), and cross-over designs (Jones and Kenward 1988). Adaptations of the WMW test have been used for the analysis of ordered categorical data (Emerson and Moses 1985) and could be applied to the ecological analysis of survival data (Lan and Wittes 1985). Simple multiple-comparison procedures are available for rank tests (see Hollander and Wolfe 1973, Shirley 1987). As in the case of the KW test, extensions of the Friedman test are available. The treatment of repeated-measures designs has received wide attention from statisticians (Koch 1970, Gill 1978, Koch et al. 1980, Wei and Lachin 1984, Davis and Wei 1988, Carr et al. 1989); the Friedman statistic is used to analyze the within-subject effects, whereas the test of the between-subject effect relies on the KW (Potvin et al. 1990). Extension of the Friedman tests are available for analysis of ordinal data (Carr et al. 1989) and incomplete univariate (Davis and Wei 1988) and multivariate (Wei and Lachin 1984) longitudinal data. Friedman-type statistics have also been developed to accommodate unbalanced designs with missing data (Wittkowski 1988), whereas the Quade test can be used in analysis of covariance (Quade 1982). In the ecological literature, the Friedman and Quade tests have been compared in their ability to detect habitat selection (Alldredge and Ratti 1986).

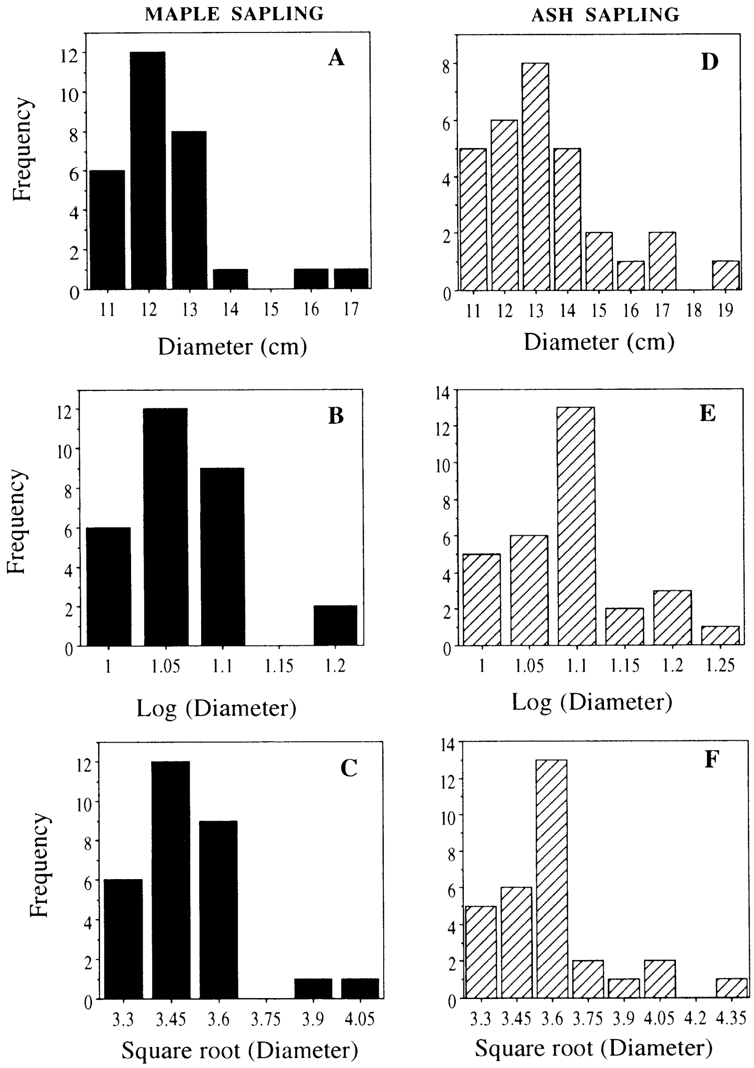

FIG. 1. Effect of log (B, E) and square root (C, F) transformation of the frequency distribution of maple (n = 29) and ash (n = 30) sapling diameters (A, D).

Rank-ordered methods have also been developed to assess the relationships between two variables. Sokal and Rohlf (1981) make a clear distinction between regression methods, which are used to describe the dependence of a variable Y on an independent variable X, and correlation, which tries to explain the interrelation between two variables. In linear regression, the model to describe the observations is:

$$Y_i = \alpha + \beta x_i + e_i,$$

where α and β are unknown parameters, x_i are known constants, and e_i is random error. The same equation pertains to parametric and nonparametric regression with differences in the assumptions of the model. In parametric regression, the errors are normally distributed random variables, while distribution-free regression requires only that the e_i's be independent and identically distributed. In both cases, α estimates the intercept of the regression line and β the slope. The least-square estimator of the slope β is the weighted average of the individual slope estimator:

$$S_{ij} = (Y_j - Y_i)/(x_j - x_i).$$

Several nonparametric estimators are available and were reviewed in Collomb (1981) and Dietz (1989). One of the best known nonparametric slope estimators, the Theil estimator, is the median of the S_{ij}. The preferred use of this estimator was recommended by Dietz (1989). A highly desirable property of the Theil estimator is that it is robust to gross errors (Hollander and

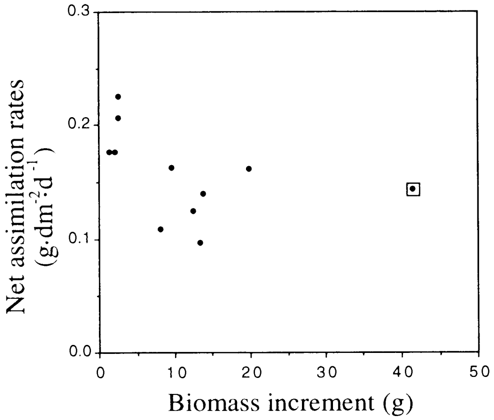

FIG. 2. Relationship between net assimilation rates and biomass increment in two species of grasses: *Echinochloa crus-galli* and *Eleusine indica*. Data are from Potvin and Strain (1985). The re-analysis of the data removed ⊡ as an outlier.

TABLE 3. Parametric and nonparametric, correlation coefficients between net assimilation rate and biomass increment. Data are from Potvin and Strain (1985). The coefficients were computed (A) with the entire data set illustrated in Fig. 2 and (B) once the extreme biomass increment value of 41.42 was removed.

	Product–moment correlation coefficient		Spearman correlation	
	r	P	r_s	P
A)	−0.3983	.2236	−0.5968	.0507
B)	−0.5965	.0667	−0.6444	.0428

Wolfe 1973). This is a property common to all statistics based on the median (Seaman and Jaeger 1990). The median is the value of the observation that splits a distribution in half. In other words, 50% of the observations are smaller than the median and 50% are larger. Therefore the presence of an outlier with an extreme value would not change the median although it would alter the mean. Conversely, parametric regression estimators are very sensitive to the presence of outliers. Donoho and Huber (1983) have shown that a single bad observation can cause least square regression estimators to behave badly. In fact, improved insensitivity to outliers was a motivation for the development of nonparametric regression methods (Huber 1973). The ARE of the Theil estimator of slopes is identical to that of the KW test in the case of equally spaced, no-replication design (see Table 1). Sen (1968) provides a formula to compute ARE in other cases. Graphical procedures are also available for computation of estimated nonparametric regression (Edgeworth 1923, Brown 1980, Fisher 1983, Buckley and Eagleson 1989).

Two well-known nonparametric measures of correlation are available: Spearman's r_s and Kendall's τ. An important difference between these methods and their parametric equivalent, the product–moment correlation coefficient, is that the latter describes only the linear part of the relation between the two variables. This may explain why nonparametric correlation coefficients are frequently used in ecology. Assumptions of nonparametric tests of independence are that the bivariate observations be mutually independent, each coming from the same continuous population. Robustness to outliers, a widespread advantage of nonparametric techniques, is illustrated, using Spearman's r_s, by the study of the interdependence between net assimilation rate (NAR) and biomass increment (ΔW) in two species of grasses (Fig. 2). Data are from Potvin and Strain (1985). The correlation between the two variables was analyzed with parametric and nonparametric coefficients. The nonparametric method suggests that there might be a significant negative correlation between the variables but no significant correlation is uncovered by the parametric coefficient (Table 3). A look at the raw data (Fig. 2) indicates that the last value of biomass increment could be an outlier, as it is four times bigger than any other observations. When the analyses were preformed on the data set once the outlier was removed (Table 3), both coefficients found a significant negative correlation, thus confirming the results obtained in the first place by Spearman's r_s.

Overall distribution-free rank tests have several features that can be recommended. First, these methods free the practitioner from concerns about the underlying distribution. There is evidence that, when the distribution underlying the observations is other than normal, rank tests are equally or more efficient than parametric tests. The use of distribution-free methods protects the practitioner against making the false decisions than can result from a distorted significance level due to nonnormality. Rank-ordered techniques have yet another advantage. As stated earlier, variance estimates based on ranks are less sensitive to the values of outliers than are those based on the original data. This was elegantly demonstrated by Hettmansperger and McKean (1978). Outliers might lead to the rejection of a true null hypothesis in ANOVA's and, to biologists, this is the most costly type of error. The robustness of distribution-free methods should thus be viewed as a major advantage.

Computations beyond basic rank tests are simple to understand, which is seen as an advantage of nonparametric methods (Hollander and Wolfe 1973). In a computer age, it is important to note that the Mann–Whitney, Kruskal-Wallis, and Friedman tests are readily available in most computer packages and can therefore be conveniently used. Likewise, distribution-free correlation coefficients are available in standard computer packages. Conversely, the use of a distribution-free regression estimator is impaired by the absence of available computer programs. Extensions of the KW and Friedman tests to accommodate complex designs have not been implemented in computer packages as "canned" models. It requires a thorough understanding of the procedure to accommodate it to a specific design, thereby reducing the availability of these methods in ecological experiments. It has been argued that using only the ordinal structure of the data resulted in a net loss of information compared with the original data in which the information on the interval is retained (Wainer and Thissen 1976). When ranked data are used to test hypotheses about populations, there is no loss of information, as the relative location of the distributions is not changed. If the aim is to estimate parameters of a population, then using ranks instead of raw data may lead to an information loss. The solution is then to back-transform the differences in the original scale of measurement. The advantages and disadvantages of nonparametric methods have been recently discussed by several authors (Petranka 1990, Seaman and Jaeger 1990, Simberloff 1990, Toft 1990). To biologists, the absence, in computing packages, of distribution-free procedures for complex designs may be one more serious limitation to the use of rank tests.

We conclude this section by briefly introducing an alternative method to rank-ordered statistics, rank transformation (RT). Rank transformation was proposed by Conover and Iman (1981) as a way to bridge the gap between parametric and nonparametric statistics. Rank transformations are applicable to a wide variety of techniques: balanced (Hora and Conover 1984) and unbalanced (Akritas 1990) two-way layout, randomized complete block (Iman et al. 1984) and simple repeated measures designs (Kepner and Robinson 1988, Thompson and Ammann 1989), correlation, regression, and discriminant analysis (Conover and Iman 1981). Critical discussion of RT by Noether and Fligner (in Conover and Iman 1981) is available and we would encourage interested ecologists to read it. The arguments will not be summarized here. It proceeds very simply in two steps: First, data are replaced by their ranks across the whole data set, and second, the usual, most appropriate parametric test (t test, F test, etc.) is applied to the RT data. Rank transformation is not itself a nonparametric method. The rationale for using RT is quite the opposite: data transformed into ranks are thought more likely to satisfy the assumptions of the parametric model than would the original data, especially the assumption of homoscedasticity. As mentioned in the introduction, heterogeneity of error variances is frequent in ecological studies, especially field studies, and may lead to costly errors in decision making. Unfortunately, both rank and parametric tests assume that the error variance should be homogeneous. RT potentially corrects for heteroscedasticity because the variance of rank data is automatically stable (Bartlett 1947). It should be viewed as the main advantage of this technique. Therefore, RT takes advantage of the positive properties of rank data and makes them available for complex designs.

Randomization (Permutation) Methods

An entire body of novel distribution-free methods has been developed in parallel with the increasing capacities of today's computers to process large quantities of data. These techniques either reshuffle or resample a data set (i.e., sample with replacement) in order to perform their analyses. The former we shall refer to as "permutation" or "randomization" methods and the latter as "bootstrap" techniques. We begin this overview of some computer-intensive techniques by looking at "randomization" methods.

Consider the following problem. We have two sets of observations, set A comprising seven observations and set B comprising nine observations, and we wish to test the hypothesis that the two means are not significantly different. This data set has several "problems." The sample size is small, which reduces the power of traditional statistical methods and makes it impossible to test the normality of the observation. Furthermore, the data are unbalanced, which in the present case presents no difficulties, but could do so in more complex designs. Randomization allows us to analyze this data set regardless of these problems. If normality were assumed, the hypothesis could be tested with the t statistic, but we can avoid any assumption about the underlying distribution by combining all 16 observations into a single data set and considering all 11 440 ways of partitioning this set into two sets of seven and nine observations. For each partition, we compute the absolute difference between the two means, then simply count how many cases exceed the observed difference. If fewer than 5% do so, we conclude that the two means are significantly different. As noted by Efron (1979, from whom this example is drawn), 40 yr ago this solution would have been "unthinkable," but computers have made it entirely feasible to contemplate performing millions of numerical operations to analyze a data set.

We need not consider all possible permutations in a randomization test (in fact, it would not be feasible to do so in the χ^2 problem described below). In the example given in the preceding paragraph, after constructing the combined data set, we could draw seven observations at random, and without replacement, from this set to form one randomized set, letting the remaining nine observations form the second set. If we repeat this procedure a large number of times (say 5000) and score the number of cases in which the difference between the means is greater than the observed value, the estimate of the probability that the observed value is greater than expected by chance, P, is given by n/N, where n is the number of randomizations in which the difference is greater than the observed and N is the total number of randomizations. The standard error of P is $\sqrt{[P(1 - P)/N]}$. Generally, 5000 randomizations are sufficient to make the standard error of P negligible and therefore to make our conclusion very likely the same that would be obtained if we considered all possible permutations. (More randomizations may be required if the probability is close to the level of significance.) The important difference between this method and that of considering all permutations is that, although the number of possible permutations increases very rapidly with sample size, the number of randomizations required stays the same. Furthermore, it is generally far easier to construct computer algorithms that will produce randomizations of a data set than to construct algorithms that will iterate through all possible permutations.

As shown in the preceding example, randomization methods represent a solution to the problem of small sample size or unbalanced data. The benefits of using "randomization" methods in the analysis of unbalanced data within the framework of analysis of variance are clearly explained by Manly (1991), and can be viewed as an alternative to traditional approaches such as those discussed by Shaw and Mitchell-Olds (1993). The analysis of small samples is problematic with standard parametric methods because of the difficulties of testing for normality, and with nonparametric methods, because of loss of power. The use of randomization analysis to overcome the problem of small sample size can be illustrated with reference to the χ^2 test. The use of χ^2 is not recommended when too many cells have small expectations (Cochran 1954). This problem may arise even when total sample sizes are large, for example in the analysis of geographic heterogeneity of mitochondrial DNA variants. Bentzen et al. (1988) sampled 244 American shad (*Alosa sapidissima*) from 14 separate rivers and identified 10 different mitochondrial DNA (mtDNA) genotypes. They wished to address the question, "Is there statistically significant variation in genotype frequencies among rivers?" Unfortunately, 66% of the 140 cells had expected frequencies <1, and only 9.3% had expected values >5, so the χ^2 test was inappropriate. The standard solution to this problem is to combine cells, a procedure that necessarily loses information and reduces power accordingly. Applying the standard solution in this case indicated marginal heterogeneity among the samples ($\chi^2 = 22.96$, where the critical value at the 5% level is 22.36), but application of a randomization test to the original data set revealed highly significant heterogeneity ($P < .001$, Roff and Bentzen 1989).

The disadvantages of randomization tests are that they are relatively time consuming. Recently a new "canned" program (STAT XACT by CYTL) became available that is a breakthrough for practitioners. Randomization tests are reviewed in detail by Edgington (1987) and Manly (1991), who provide computer listings for the method applied to various types of ANOVA's and correlation analysis. Various algorithms for the analysis of genetic data, including the χ^2 analysis discussed above, are now available (McElroy et al. 1992, Zaykin and Pudovkin 1992).

Randomization analyses have been applied to several ecological problems, which illustrate the range of situations in which it may be a useful alternative. Examples are Mantel's test and its extensions for the analysis of association between genetic, behavioral, morphological, ecological, and geographic distances (Mantel 1967, Manly 1986, Smouse et al. 1986, Jackson and Somers 1989, Legendre and Fortin 1989); tests of similarities among the songs of neighboring birds (Shackell et al. 1988); the analysis of animal movements (Ko and Zeh 1988, Solow 1989); the detection of associations within communities (Strauss 1982); and the detection of density dependence (Pollard et al. 1987, den Boer and Reddingius 1989, Reddingius and den Boer 1989). Randomization tests are not, however, a panacea and should not be adopted without proper consideration. Consider for example, the problem of detecting density dependence in time-series data. Conceptually, we might tackle this problem by constructing random permutations of the time series and comparing the statistic measuring density dependence with the set of values obtained from the randomizations (Pollard et al. 1987, Reddingius and den Boer 1989), but if the population has several cohorts, any autocorrelation between years that may be present will be destroyed by the randomization procedure (Arditi 1989), so the test, in its present form, may only be appropriate for univoltine species. A similar argument has been advanced against the use of randomization tests for the detection of cycles of extinctions in the fossil record (Quinn 1987). For a randomization test that preserves the autocorrelation structure of the data see Legendre et al. (1990).

The Jackknife and Bootstrap

The estimation of many statistics in ecology is difficult because their sampling distributions cannot be derived exactly; it may be difficult not only to estimate the bias in the parameter to be estimated but also to estimate confidence limits. The jackknife was originally introduced by Quenouille (1949) as a method of bias reduction, but has since become a general tool for estimating both the value of a statistic and its confidence region. The bootstrap was introduced in 1979, by Efron, as an alternative to the jackknife. The term "jackknife" derives from the jackknife's role as an all-purpose tool; "bootstrap" arises from the saying "to pull oneself up by one's own bootstraps" and reflects the use of the one available sample to give rise to many others (Diaconis and Efron 1983).

A good example of an ecological statistic that cannot be derived exactly is the per-capita rate of increase, r. This statistic is of considerable significance in both theoretical and applied biology, but although r has been estimated for many years, no exact method is available for estimating the bias in the estimate or its confidence region (Meyer et al. 1986). The rate of increase is a function of several very different distributions, which makes the derivation of its sampling distribution very difficult. Furthermore, it may not be possible to characterize one or more of the component functions adequately with sample sizes available from natural populations. Empirical data suggest that the sampling distribution is likely to be highly skewed, so parametric methods will be inappropriate (Antonovics and Ellstrand 1984). The jackknife and bootstrap procedures may overcome these difficulties. Apart from their use in estimating rates of increase (Lenski and Service 1982, Meyer et al. 1986), the jackknife and bootstrap have proved valuable in the estimation of genetic distance measures (Mueller 1979, Trexler 1988), heritability (Henrich and Travis 1988), niche overlap indices (Mueller and Altenberg 1985), the Gini coefficient of inequality (Dixon et al. 1987), population estimates from mark–recapture data (Manly 1977, Burnham and Overton 1979), diversity indices (Zahl 1977, Heltshe and Forrester 1985), the standard error on a functional evenness index (Troussellier and Legendre 1981), and species richness measures (Smith and van Belle 1984).

The jackknife estimate is obtained by consideration of all possible data sets in which one value has been eliminated from the original set (there will obviously be n such estimates), whereas the bootstrap procedure involves computing a large number of estimates by random sampling with replacement from the original data set. The general construction of the jackknife and bootstrap estimates is as described below. To obtain a jackknife estimate of some required parameter, θ, we first define the following quantities.

- n: sample size,
- $\hat{\theta}$: estimate of θ using all n values,
- $\tilde{\theta}_i$: estimate of θ using all n values except the ith datum.

In the case of estimating r, for example, where there are n animals, the n values of $\tilde{\theta}_i$ are obtained by successive deletion of animals 1 through n from the samples. The quantity $S_{n,i} = n\hat{\theta} - (n - 1)\tilde{\theta}_i$ is called a pseudovalue; there are n pseudovalues. The jackknife estimator $\tilde{\theta}$ is simply the mean of the n pseudovalues,

$$\tilde{\theta} = (1/n)\,\Sigma\, S_{n,i}.$$

An estimate of the variance is given by

$$\mathrm{Var}\,\tilde{\theta} = [\Sigma\,(S_{n,i} - \tilde{\theta})^2]/[n(n - 1)].$$

To obtain a bootstrap estimate, we first, from the original sample of n values, draw n values with replacement and from this sample compute the estimate of θ. For example, in the case of estimating r, n animals are drawn with replacement from the original sample, and the estimate of r computed. We repeat this procedure k times to obtain k bootstrap replicates; the ith replicate is designated θ_i^*. The bootstrap estimate of θ is given by

$$\theta^* = (1/k)\,\Sigma\,\theta_i^*.$$

This estimate will, in general, be biased. A bias-adjusted estimate of θ is given by

$$\hat{\theta}_A = 2\hat{\theta} - \theta^*,$$

where $\hat{\theta}$ is the estimate obtained from the original sample (Meyer et al. 1986). The variance is calculated in the usual manner as

$$\mathrm{Var}\,\theta^* = [\Sigma\,(\theta_i^* - \theta^*)^2]/(k - 1).$$

For a review of the statistical motivation and theory underlying the jackknife and bootstrap, see Miller (1974) and Efron (1981, 1982).

The two procedures can be illustrated with reference to the estimation of population size using the Petersen Index for mark–recapture data,

$$\hat{N} = \frac{Mn}{m},$$

where M is the number of marked animals released from the first sample, n is the number of animals in the second sample, m is the number of marked animals in the second sample, and $\hat{N}$ is the estimated population size. Both this estimator and its estimated variance are known to be biased (Bailey 1951), so estimation by the jackknife or bootstrap procedure may provide more reliable estimates.

The useful feature of this example is that for the jackknife estimator there are only two possible values of θ_i, a sample minus an unmarked animal, and a sample minus a marked animal. Because there are $n - m$ unmarked animals in the sample and m marked animals, successive deletions of single data from the sample will give $n - m$ instances of a sample minus an unmarked animal and m instances of a sample minus a marked animal. If we rank the data set according to mark status, unmarked animals followed by marked, we can denote the two possible values of $\hat{\theta}_i$ as $\hat{N}_{1,i}$ ($1 \le i \le n - m$) and $\hat{N}_{2,i}$($n - m < i \le n$), respectively. Thus we have

$$\hat{N}_{1,i} = \frac{M(n-1)}{m}$$

$$\hat{N}_{2,i} = \frac{M(n-1)}{(m-1)}.$$

We can similarly denote the two pseudovalues as $S_{1,i}$ and $S_{2,i}$, giving

$$S_{1,i} = n\hat{N} - (n-1)\hat{N}_{1,i}$$
$$S_{2,i} = n\hat{N} - (n-1)\hat{N}_{2,i}.$$

The jackknife estimator, $\tilde{N}$, is then

$$\tilde{N} = \frac{1}{n}\sum S_{n,i} = \frac{1}{n}\left(\sum_{i=1}^{n-m} S_{1,i} + \sum_{1=n-m+1}^{n} S_{2,i}\right)$$

$$= \frac{1}{n}[(n-m)S_{1,i} + mS_{2,i}],$$

and the variance is

$$\text{Var } \tilde{N} = \frac{\sum_{i=1}^{n-m}(S_{1,i} - \tilde{N})^2 + \sum_{i=n-m+1}^{n}(S_{2,i} - \tilde{N})^2}{n(n-1)}.$$

As described above, to obtain a bootstrap estimate, we randomly select, with replacement, n animals from the sample comprising m marked and $n - m$ unmarked animals, and repeat this procedure k times. In this case, the number of marked animals in a single bootstrap sample could range from 0 to n. (Samples with no marked animals pose a problem, as the population estimate is undefined for such samples; for the present illustrative example we shall ignore this problem.) If the number of marked animals in the ith replicate is m_i*, then the ith bootstrap replicate is

$$N_i^* = \frac{Mn}{m_i^*},$$

and the bootstrap estimate is

$$N^* = (1/k)\sum_{i=1}^{k} N_i^*.$$

The bias-adjusted bootstrap is

$$N_A = 2\hat{N} - N^*.$$

The variance is

$$\text{Var } N^* = \frac{\Sigma (N_i^* - N^*)^2}{(k-1)}.$$

The above example is presented only to illustrate the jackknife and bootstrap procedures; we do not advocate their use for the Petersen Index without further analysis, because it cannot be assumed a priori that either or both the jackknife and bootstrap will provide good estimates (Miller 1974). Before they are used to estimate values from real data sets, their performance must be examined with simulated data sets where the actual values are known and the underlying assumptions can be varied. Although it is a general finding that these methods reduce bias in the estimator, confidence limits may be too small (see, e.g., Mueller and Altenberg 1985, Dixon et al. 1987), so caution must be exercised in the interpretation of the estimates. Provided this caveat is kept in mind, the jackknife and bootstrap can be highly useful tools in ecological, behavioral, and evolutionary studies.

Applications and Further Readings

Despite prejudices, traditional nonparametric methods are only slightly less powerful than their parametric counterparts and are buffered against distortion in significance testing due to nonnormality (Hollander and Wolfe 1973, Noether 1987). We hope we have made a case for wider use of distribution-free methods by ecologists. The question to ask is, "In which cases should distribution-free methods be favored over parametric techniques?" Mild violations of normality may lead to a loss of power of parametric methods (Conover and Iman 1976). For certain cases of nonnormality, namely highly skewed, U-shaped, and bimodal distributions, distribution-free methods are more powerful (Shirley 1981). Distribution-free methods are also suggested for nonnormal, heavy-tailed distributions, and when the data set contains outliers (Hettmansperger and McKean 1978). When the distribution underlying the data cannot be tested, as in the case of small sample sizes, nonparametric statistics may be preferred, especially if nonnormality is suspected. Maybe the most important positive attribute of rank tests is their robustness regarding outliers.

The efficiency reported for rank analyses is asymptotic, i.e., based on sample sizes that tend to infinity. Therefore, rank analyses are not necessarily the answer for small sample size; in such cases randomization methods can be employed. Distribution-free methods are a possible solution for the analysis of ordinal or categorical data and should be compared to other methods such as linear logistic models or logit and probit analyses, discussed by Trexler and Travis (1993). An ecological study is frequently concerned not only with testing hypotheses but also with parameter estimation. Many parameters of concern to ecologists cannot be estimated exactly because the sampling distribution cannot be derived analytically. In such cases the jackknife or bootstrap may be the solution. Several texts (Hollander and Wolfe 1973, Lehmann 1975, Conover 1980, Zar 1984) provide clear and detailed discussions of rank ordered techniques, while computer-intensive techniques are described in Edgington (1987), Manly (1991), and Efron and Tibshirani (1991).

Another frequent problem in ecological data sets is heteroscedasticity. Neither rank tests nor randomization methods can protect against distortion of significance levels due to unequal variances; both types of method assume that the data come from the same population. A solution to that problem is the use of rank transformation, which often stabilizes the variance. However, RT cannot be considered as a distribution-free method.

ACKNOWLEDGMENTS

We are thankful to Drs. M. Bradford, D. J. Fairbairn, M. J. Lechowicz, P. Legendre, and especially to S. Tardif for many helpful comments on a first draft of this manuscript.

LITERATURE CITED

Akritas, M. G. 1990. The rank transform method in some two-factor designs. Journal of the American Statistical Association **85**:73–78.

Alldredge, J. R., and J. T. Ratti. 1986. Comparison of some statistical techniques for analysis of resource selection. Journal of Wildlife Management **50**:157–165.

Antonovics, J., and N. C. Ellstrand. 1984. Experimental studies of the evolutionary significance of sexual reproduction. I. A test of the frequency-dependent selection hypothesis. Evolution **38**:103–115.

Arditi, R. 1989. Avoiding fallacious significance tests in stepwise regression: a Monte Carlo method applied to meteorological theory for the Canadian lynx cycle. International Journal of Biometeorology **33**:24–26.

Austin, M. P. 1987. Models for the analysis of species response to environmental gradient. Vegetatio **69**:35–45.

Bailey, N. J. 1951. On estimating the size of mobile populations from recapture data. Biometrika **38**:293–306.

Bartlett, M. S. 1947. The use of transformations. Biometrics **3**:39–52.

Bentzen, P., W. C. Leggett, and C. G. Brown. 1988. Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). Genetics **118**:509–518.

Berry, D. A. 1987. Logarithmic transformations in ANOVA. Biometrics **43**:439–456.

Biondini, M. E., P. W. Mielke, Jr., and K. J. Berry. 1988. Data-dependent permutation techniques for the analysis of ecological data. Vegetatio **75**:161–168.

Brown, B. M. 1980. Median estimates in simple linear regression. Australian Journal of Statistics **22**:154–165.

Buckley, M. J., and G. K. Eagleson. 1989. A graphical method for estimating the residual variance in nonparametric regression. Biometrika **76**:203–210.

Burbidge, J. B., L. Magee, and A. L. Robb. 1988. Alternative transformations to handle extreme values of the dependent variable. Journal of the American Statistical Association **83**:123–127.

Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. Ecology **60**:927–936.

Carr, G. J., K. B. Hafner, and G. G. Koch. 1989. Analysis of rank measures of association for ordinal data from longitudinal studies. Journal of the American Statistical Association **84**:797–804.

Cochran, W. G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics **3**:22–38.

———. 1954. Some methods for strengthening the common 2 tests. Biometrics **10**:417–451.

Collomb, G. 1981. Estimation nonparamétrique de la régression: revue bibliographique. International Statistical Review **49**:75–93.

Conover, W. J. 1980. Practical nonparametric statistics. Wiley, New York, New York, USA.

Conover, W. J., and R. L. Iman. 1976. On some alternative procedures using ranks for the analysis of experimental designs. Communications in Statistics—Theory and Methods **A5**(14):1349–1368.

Conover, W. J., and R. L. Iman. 1981. Rank transformation as a bridge between parametric and nonparametric statistics. American Statistician **35**:124–133.

Davis, C. S., and L. J. Wei. 1988. Nonparametric methods for analysing incomplete nondecreasing repeated measurements. Biometrics **44**:1005–1018.

den Boer, P. J., and J. Reddingius. 1989. On the stabilization of animal numbers. Problems of testing 2. Confrontation with data from the field. Oecologia (Berlin) **79**:143–149.

Diaconis, P., and B. Efron. 1983. Computer-intensive methods in statistics. Scientific American **248**:116–128.

Dietz, E. J. 1989. Teaching regression in a nonparametric statistics course. American Statistician **43**:35–40.

Dixon, P. M., J. Weiner, T. Mitchell-Olds, and R. Woodley. 1987. Bootstrapping the Gini coefficient of inequality. Ecology **68**:1548–1551.

Donoho, D. L., and P. J. Huber. 1983. The notion of breakdown point. Pages 157–184 *in* P. J. Bickel, K. A. Doksum, J. L. Hodges, and C. A. Belmont, editors. A festschrift for Erich Lehmann. Wadsworth, Belmont, California, USA.

Edgeworth, F. Y. 1923. On the use of medians for reducing observations relating to several quantities. Philadelphia Magazine (sixth series) **56**:1074–1088.

Edgington, E. S. 1987. Randomization tests. Marcel Dekker, New York, New York, USA.

Efron, B. 1979. Computers and the theory of statistics: thinking the unthinkable. SIAM Review **2**:460–480.

———. 1981. Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics **9**:139–172.

———. 1982. The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, 38. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Efron, B., and R. Tibshirani. 1991. Statistical analysis in the computer age. Science **253**:390–395.

Emerson, J. D., and L. E. Moses. 1985. A note on the Wilcoxon–Mann–Whitney test for 2 × k ordered tables. Biometrics **41**:303–309.

Fisher, N. I. 1983. Graphical methods in nonparametric statistics: a review and annotated bibliography. International Statistical Review **51**:25–58.

Forsythe, A. B. 1972. Robust estimation of straight line regression coefficients by minimizing pth power deviations. Technometrics **14**:159–166.

Geary, R. C. 1947. Testing for normality. Biometrika **34**:209–242.

Gill, J. L. 1978. Design and analysis of experiments in the animal and medical sciences. Volume 2. Iowa State University Press, Ames, Iowa, USA.

Hampel, F. R., E. M. Ronchetti, P. J. Bousseeuw, and W. A. Stahel. 1986. Robust statistics: the approach based on influence functions. Wiley, New York, New York, USA.

Heltshe, J. F. and N. E. Forrester. 1985. Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. Ecology **66**:107–111.

Henrich, S., and J. Travis. 1988. Genetic variation in reproductive traits in a population of *Heterandria formosa* (Pisces: Poeciliidae). Journal of Evolutionary Biology **1**: 275–280.

Hettmansperger, T. P., and J. W. McKean. 1978. Statistical inference based on ranks. Psychometrika **43**:69–79.

Hodges, J. L., and E. L. Lehmann. 1955. The efficiency of some nonparametric competitors of the t-test. Annals of Mathematical Statistics **27**:324–335.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric statistical methods. Wiley, New York, New York, USA.

Hora, S. C., and W. J. Conover. 1984. The F-statistics in the two-way layout with rank-score transformed data. Journal of the American Statistical Association **79**:668–673.

House, D. E. 1986. A nonparametric version of William's test for a randomized block design. Biometrics **42**:187–190.

Huber, P. J. 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics **35**:73–107.

———. 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Annals of Statistics **1**:799–821.

———. 1977. Robust covariances. Swiss Federal Institute of Technology, BK No. **00344**:165–191.

Iman, R. L., S. C. Hora, and W. J. Conover. 1984. Comparison of asymptotically distribution-free procedures for the analysis of complete blocks. Journal of the American Statistical Association **79**:674–685.

Jackson, D. A., and K. M. Somers. 1989. Are probability estimates from the permutation model of Mantel's test stable? Canadian Journal of Zoology **67**:766–769.

Jones, B., and M. G. Kenward. 1988. Design and analysis of cross-over trials. Routledge, Chapman and Hall, New York, New York, USA.

Kepner, J. L., and D. H. Robinson. 1988. Nonparametric methods for detecting treatment effect in repeated-measures designs. Journal of the American Statistical Association **83**: 456–461.

Ko, D., and J. E. Zeh. 1988. Detection of migration using sound location. Biometrics **44**:751–764.

Koch, G. G. 1970. The use of non-parametric methods in the statistical analysis of complex split plot experiment. Biometrics **26**:105–128.

Koch, G. G., I. A. Amara, M. E. Stokes, and D. Gillings. 1980. Some views on the parametric and nonparametric analysis for repeated measurements and selected bibliography. International Statistical Review **48**:249–265.

Lan, K. K. G., and J. T. Wittes. 1985. Rank tests for survival analysis: a comparison by analogy with games. Biometrics **41**:1063–1069.

Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio **80**:107–138.

Legendre, P., N. L. Oden, R. R. Sokal, A. Vandor, and J. Kim. 1990. Approximate analysis of variance of spatially autocorrelated regional data. Journal of Classification **7**:53–75.

Lehmann, E. L. 1975. Nonparametrics: statistical methods based on ranks. Holden-Day, San Francisco, California, USA.

Lenski, R. E., and P. M. Service. 1982. The statistical analysis of population growth rates calculated from schedules of survivorship and fecundity. Ecology **63**:655–662.

Manly, B. F. J. 1977. A simulation experiment on the application of the jackknife with Jolly's method for the analysis of capture–recapture data. Acta Theriologica **22**:215–223.

———. 1986. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. Researches in Population Ecology **28**:201–218.

———. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, London, England.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Research **27**:209–220.

Mardia, K. V. 1980. Tests of univariate and multivariate normality. Pages 279–320 *in* P. R. Krishnaiah, editor. Handbook of statistics. Volume 1. North-Holland, Amsterdam, The Netherlands.

McElroy, D., P. Moran, E. Bermingham, and I. Kornfield. 1992. REAP: an integrated environment for the manipulation and phylogenetic analysis of restriction data. Journal of Heredity **83**:157–158.

Meyer, J. S., C. G. Ingersoll, L. L. McDonald, and M. S. Boyce. 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. Ecology **67**:1156–1166.

Miller, R. G. 1974. The jackknife—a review. Biometrika **61**:1–15.

Morris, G. E. L. 1985. The presentation of treatment responses from block experiments after analysis of variance of transformed data. Annals of Applied Biology **107**:571–580.

Mueller, L. D. 1979. A comparison of two methods for making statistical inferences on Nei's measure of genetic distance. Biometrics **35**:757–763.

Mueller, L. D., and L. Altenberg. 1985. Statistical inference on measures of niche overlap. Ecology **66**:1204–1210.

Noether, G. E. 1987. Sample size determination for some common nonparametric tests. Journal of the American Statistical Association **82**:645–647.

Pearson, E. S. 1931. The analysis of variance in cases of non-normality. Biometrika **23**:114–133.

Petranka, J. W. 1990. Caught between a rock and a hard place. Herpetologica **46**:346–350.

Pitman, E. J. G. 1949. Application of the method of mixtures to quadratic forms in normal variates. Annals of the Mathematical Society **20**:552–560.

Pollard, E. K., H. Lakhani, and P. Rothery. 1987. The detection of density-dependence from a series of annual censuses. Ecology **68**:2046–2055.

Potvin, C., M. J. Lechowicz, and S. Tardif. 1990. The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures. Ecology **71**:1389–1400.

Potvin, C., and B. R. Strain. 1985. Effects of CO_2 enrichment and temperature on growth in two C_4 weeds, *Echinochloa crus-galli* and *Eleusine indica*. Canadian Journal of Botany **63**:1495–1499.

Quade, D. 1982. Nonparametric analysis of covariance by matching. Biometrics **38**:597–611.

Quenouille, M. H. 1949. Approximate tests of correlation in time-series. Journal of the Royal Statistical Society, B **11**:68–84.

Quinn, J. F. 1987. On the statistical detection of cycles in extinctions in the marine fossil record. Paleobiology **13**: 465–478.

Reddingius, J., and P. J. den Boer. 1989. On the stabilization of animal numbers. Problems of testing 1. Power estimates and estimation errors. Oecologia (Berlin) **78**:1–8.

Roff, D. A., and P. Bentzen. 1989. The statistical analysis of mitochondrial DNA polymorphisms: 2 and the problem of small samples. Molecular Biology and Evolution **6**:539–545.

Seaman, J. W., Jr., and R. G. Jaeger. 1990. Statistical dogmaticae: a critical essay on statistical practice in ecology. Herpetologica **46**:337–346.

Sen, P. K. 1968. Estimates of the regression coefficient based on Kendall's Tau. Journal of the American Statistical Association **63**:1379–1389.

Shackell, N. L., R. E. Lemon, and D. A. Roff. 1988. Song similarity between neighboring American Redstarts (*Setophaga ruticilla*): a statistical analysis. Auk **105**:609–615.

Shaw, R. G., and T. Mitchell-Olds. 1993. ANOVA for unbalanced data: an overview. Ecology **74**:1638–1645.

Shirley, E. 1977. A non-parametric equivalent of Williams' test for contrasting increasing dose levels of a treatment. Biometrics **33**:386–389.

———. 1981. A distribution-free method for analysis of covariance based on ranked data. Applied Statistics **30**: 158–162.

———. 1987. Applications of ranking methods to multiple comparison procedures and factorial experiments. Applied Statistics **36**:205–213.

Shorak, G. R. 1967. Testing against ordered alternatives in model I analysis of variance: normal theory and non-parametric. Annals of Mathematical Statistics **38**:1740–1752.

Simberloff, D. 1990. Hypotheses, errors, and statistical assumptions. Herpetologica **46**:351–357.

Smith, E. P., and G. van Belle. 1984. Nonparametric estimation of species richness. Biometrics **40**:119–130.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoologist **35**:627–632.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. Freeman, San Francisco, California, USA.

Solow, A. R. 1989. A randomization test for independence of animal locations. Ecology **70**:1546–1549.

Strauss, R. E. 1982. Statistical significance of species clusters in association analysis. Ecology **63**:634–639.

Tardif, S. 1987. Efficiency and optimality results for tests based on weighted rankings. Journal of the American Statistical Association **82**:637–644.

Thompson, G. L., and L. P. Ammann. 1989. Efficacies of rank-transform statistics in two-way models with no interaction. Journal of the American Statistical Association **84**: 325–330.

Toft, C. A. 1990. Reply to Seaman and Jaeger: an appeal to common sense. Herpetologica **46**:357–361.

Trexler, J. C. 1988. Hierarchical organization of genetic variation in the sailfin molly, *Poecilia latipinna* (Pisces: Poeciliidae). Evolution **42**:1006–1017.

Trexler, J. C., and J. Travis. 1993. Nontraditional regression analyses. Ecology **74**:1629–1637.

Troussellier, M., and P. Legendre. 1981. A functional evenness index for microbial ecology. Microbial Ecology **7**:283–296.

Tukey, J. W. 1962. The future of data analysis. Annals of Mathematical Statistics **33**:1–67.

Wainer, H., and D. Thissen. 1976. Three steps towards robust regression. Psychometrika **41**:9–34.

Wei, L. J., and J. M. Lachin. 1984. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. Journal of the American Statistical Association **79**:653–661.

Williams, D. A. 1986. A note on Shirley's nonparametric test for comparing several dose levels with a zero-dose control. Biometrics **42**:183–186.

Winer, B. J. 1971. Statistical principles in experimental design. McGraw-Hill, New York, New York, USA.

Wittkowski, K. N. 1988. Friedmann-type statistics and consistent multiple comparisons for unbalanced designs with missing data. Journal of the American Statistical Association **83**:1163–1170.

Wu, L. 1985. Robust M-estimation of location and regression. Pages 316–388 *in* N. B. Tuma, editor. Sociological methodology. Jossey-Bass, San Francisco, California, USA.

Zahl, S. 1977. Jackknifing and index of diversity. Ecology **58**:907–913.

Zar, J. H. 1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Zaykin, D. V., and A. I. Pudovkin. 1992. Two programs to estimate significance of χ^2 values using pseudo-probability test. Journal of Heredity, *in press*.