Distance-Based Tests for Homogeneity of Multivariate Dispersions

Marti J. Anderson

Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand
email: mja@stat.auckland.ac.nz

SUMMARY. The traditional likelihood-based test for differences in multivariate dispersions is known to be sensitive to nonnormality. It is also impossible to use when the number of variables exceeds the number of observations. Many biological and ecological data sets have many variables, are highly skewed, and are zero-inflated. The traditional test and even some more robust alternatives are also unreasonable in many contexts where measures of dispersion based on a non-Euclidean dissimilarity would be more appropriate. Distance-based tests of homogeneity of multivariate dispersions, which can be based on any dissimilarity measure of choice, are proposed here. They rely on the rotational invariance of either the multivariate centroid or the spatial median to obtain measures of spread using principal coordinate axes. The tests are straightforward multivariate extensions of Levene's test, with *P*-values obtained either using the traditional *F*-distribution or using permutation of either least-squares or LAD residuals. Examples illustrate the utility of the approach, including the analysis of stabilizing selection in sparrows, biodiversity of New Zealand fish assemblages, and the response of Indonesian reef corals to an El Niño. Monte Carlo simulations from the real data sets show that the distance-based tests are robust and powerful for relevant alternative hypotheses of real differences in spread.

KEY WORDS: Dissimilarity; Heterogeneity; Levene's test; Multivariate analysis; Permutation tests; Principal coordinates; Robust methods.

1. Introduction

A fundamental concept in the analysis of multivariate data is the extent to which groups of observations may differ in their relative dispersions. For example, stabilizing selection acting on organisms should reduce the multivariate heterogeneity of their morphometric characteristics (Bumpus, 1898) and increases or decreases in the multivariate dispersion of ecological species data can be an important symptom of environmental stress (Warwick and Clarke, 1993; Chapman, Underwood, and Skilleter, 1995). Traditional likelihood-based tests of homogeneity of variance–covariance matrices (e.g., Box, 1949) are, however, extremely sensitive to departures from the assumption of multivariate normality (Layard, 1974; Olson, 1974; O'Brien, 1992) and also require that the number of observations per group be relatively large compared to the number of variables. Despite the development of certain more robust procedures (Tiku and Balakrishnan, 1985; O'Brien, 1992), such methods nevertheless still require, like the traditional tests, that dispersion be measured essentially using either Euclidean or Mahalanobis distance, which may not always be appropriate.

In many biological applications, the assumption of normality is unreasonable. This is particularly so in the analysis of ecological community data, where variables consist of counts of abundances of individual species (Taylor, 1961; McArdle, Gaston, and Lawton, 1990). In such data sets, the number of variables often exceeds the number of observations, zeros are common due to rare or patchy species, and abundance distributions tend to be strongly right skewed. Furthermore, community dissimilarity is generally better measured by non-Euclidean measures, such as the Jaccard or Bray–Curtis coefficients (Faith, Minchin, and Belbin, 1987; Clarke, 1993). Thus, a robust test for homogeneity of multivariate dispersions is needed, which allows dispersions to be measured on the basis of any distance or dissimilarity measure of choice and which is robust to skewed or zero-inflated data.

Several robust dissimilarity-based tests for equality of multivariate locations have been described (Mielke, Berry, and Johnson, 1976; Smith, Pontasch, and Cairns, 1990; Clarke, 1993; Pillar and Orlóci, 1996; Gower and Krzanowski, 1999; Legendre and Anderson, 1999; Anderson, 2001). These tests derive their validity and robustness through the use of permutation or randomization procedures, but are nevertheless sensitive to differences in dispersion among groups. Thus, an analogous dissimilarity-based test of dispersions is therefore needed, both in its own right and as a complement to such tests of location effects.

In a univariate context, Levene's test for homogeneity—essentially an analysis of variance (ANOVA) done on deviations from group means—was found to be quite robust to departures from normality (Levene, 1960). Brown and Forsythe (1974) suggested that medians be used instead of means for Levene's test, and this approach was found to be the most robust and useful of 56 possible tests for homogeneity in extensive simulations done by Conover, Johnson, and Johnson (1981). In a multivariate context, Van Valen (1978) proposed

a multivariate analogue to Levene's test as an ANOVA on the Euclidean distances of individual observations to their group centroid. O'Brien (1992) and Manly (1994) suggested that this approach could be made more robust by replacing centroids with multivariate medians, where they defined the multivariate median as the median for each variable within each group.

Haldane (1948), Gower (1974), and Brown (1983) noted that there is more than one way to define a multivariate median. More particularly, Gower (1974) provided an algorithm for obtaining the spatial median of a group of points, which he called the "mediancentre," defined as that point in the multivariate space such that the sum of the distances from each observation in the group to that point is minimized. This spatial (or geometric) median is invariant to rotational changes in the axes, while the vector of medians of individual variables is not. In contrast, the vector of medians of individual variables is invariant to monotonic transformations of the variables, but the spatial or geometric median is not (Haldane, 1948). The property of rotational invariance possessed by the spatial median, a property also shared by the centroid, is exploited here in order to develop a robust test of dispersion on the basis of any dissimilarity measure.

A dissimilarity-based multivariate generalization of Levene's test is proposed, following the ideas of Van Valen (1978), O'Brien (1992), and Manly (1994) who used Euclidean distances. Two possible test statistics are suggested: the ANOVA F-statistic comparing distances to centroids or that which compares distances to spatial medians. A unique feature is that the analysis may be based on any dissimilarity measure of choice, through the use of principal coordinates (Gower, 1966). P-values are then obtained either using the traditional F-distribution, or by permutation of appropriate residuals.

The utility of the technique is demonstrated first by reference to a set of well-behaved multivariate morphometric data on sparrows due to Bumpus (1898), for which it was found to be equally as powerful as the traditional likelihood test. Its use, including empirical type I error and power, is then demonstrated for two sets of ecological community data: the percentage cover of Indonesian coral species responding to an El Niño event (Warwick, Clarke, and Suharsono, 1990) and spatial variation in abundances of New Zealand temperate reef fishes (Anderson and Millar, 2004). For the coral and fish data sets, the traditional test would be impossible, due to the number of variables, and direct Euclidean-based approaches would be unreasonable, due to the excessive number of zeros and the skewness of individual variables.

Retrospective power analyses for individual studies have been justifiably criticized in the literature (e.g., Hoenig and Heise, 2001; Lenth, 2001). However, the use of simulations to compare power among several statistical methods for particular alternative hypotheses is valid and useful, especially for multivariate tests such as those examined here, where potential alternative hypotheses have so many facets (Olson, 1976; Rencher, 1998). Although to do extensive power comparisons was not the aim here, measures of empirical power for several sets of data having different degrees of heterogeneity and different error distributions yielded some useful initial insights into the relative power of these tests.

2. Description of Tests

2.1 *Levene's Test and Some Multivariate Analogues*

In the univariate case, let x_{ij} be a set of $j = 1, \ldots, n_i$ observations in each of $i = 1, \ldots, g$ groups. Levene's test statistic is then the ANOVA F-ratio comparing the g groups, calculated on the absolute deviations $z_{ij} = |x_{ij} - \bar{x}_{i.}|$ from the group means $\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$. The z's are not independent: for example, in the most extreme case, $n = 2$, the two values of z are equal. However, nonindependence of the z's is not a serious problem in most cases with reasonable sample sizes, as the correlation between them for two observations in the same group is sufficiently small, of order $O(N^{-2})$ where $N = \sum_{i=1}^{g} n_i$ (Levene, 1960; Miller, 1968).

For the multivariate case, let x_{ij} be the vector which denotes the point for the jth observation in the ith group in the multivariate space of p variables. Furthermore, let $\Delta(\cdot, \cdot)$ denote the Euclidean distance between two points. The centroid vector c_i for group i is defined as the point that minimizes the sum of squared distances to points within that group, i.e., $\sum_{j=1}^{n_i} \Delta^2(x_{ij}, c_i)$.

One multivariate analogue to Levene's test is to perform ANOVA on the Euclidean distances from individual points within a group to their group centroid,

$$z_{ij}^c = \Delta(x_{ij}, c_i). \tag{1}$$

A P-value for the F-statistic calculated on distances to centroids (F_c) may be obtained either by using the traditional F-distribution ($F_c(t)$, which assumes the errors in the z^c's are approximately normal), or by using a permutation procedure ($F_c(p)$). For the latter, permutation of the least-squares (LS) residuals, $r_{ij}^c = (x_{ij} - c_i)$ is appropriate in this context (Anderson and Robinson, 2001). For data with normally distributed errors, the permutation test and normal-theory test will give similar results (Manly, 1997). It has been suggested that the normal-theory test gains its validity by virtue of its agreement with the permutation test (Fisher, 1936; Kempthorne, 1955; Edgington, 1995).

A more robust version of Levene's test, suggested by Brown and Forsythe (1974), is to analyze deviations from medians instead. One multivariate analogue of this would be to calculate ANOVA on distances from the spatial median,

$$z_{ij}^m = \Delta(x_{ij}, m_i). \tag{2}$$

The spatial median m_i is defined as the point that minimizes the sum of distances to points within that group, i.e., $\sum_{j=1}^{n_i} \Delta(x_{ij}, m_i)$ (Gower, 1974; Brown, 1983). Once again, a P-value associated with the test statistic (F_m) may be obtained either by using the traditional F-distribution ($F_m(t)$) or by using a permutation procedure ($F_m(p)$). In this case, however, it is permutation of the least-absolute-deviation (LAD) residuals, $r_{ij}^m = (x_{ij} - m_i)$, that is appropriate (Cade and Richards, 1996).

Note that the use of an appropriate permutation procedure for the test of either F_c (using LS residuals) or F_m (using LAD residuals) avoids making any particular assumption regarding the distribution of the distances, z^c or z^m, respectively. Only the exchangeability of points in the multivariate space under the null hypothesis of equal dispersions after centering (on either the group centroid or on its spatial median)

is assumed. The relative robustness of using permutation of residuals for the univariate case has been demonstrated for the classical variance ratio test (Boos and Brownie, 1989; Boos, Janssen, and Veraverbeke, 1989) and for Levene's test (Francis and Manly, 2001).

2.2 Extension to Any Dissimilarity Using Principal Coordinates

The approach may be extended to base the analysis on any distance or dissimilarity measure of choice through the use of principal coordinates (Gower, 1966). Let $D = [d_{\ell\ell'}]$ be a square symmetric matrix of distances calculated between every pair of observations, $\ell = 1, \ldots, N$ and $\ell' = 1, \ldots, N$. In the case of the Euclidean distance measure,

$$d_{\ell\ell'} = \Delta(\boldsymbol{x}_\ell, \boldsymbol{x}_{\ell'}) = \sqrt{\sum_{k=1}^{p}(x_{\ell k} - x_{\ell'k})^2}.$$

Alternatively, some other more appropriate measure may be calculated as the basis for the analysis, such as the measure proposed by Bray and Curtis (1957),

$$d_{\ell\ell'} = \frac{\displaystyle\sum_{k=1}^{p}|x_{\ell k} - x_{\ell'k}|}{\displaystyle\sum_{k=1}^{p}(x_{\ell k} + x_{\ell'k})},$$

or a scale-invariant binomial deviance (Anderson and Millar, 2004),

$$d_{\ell\ell'} = \sum_{k=1}^{p}\frac{1}{s_k}\left\{x_{\ell k}\log_e\left(\frac{x_{\ell k}}{s_k}\right) + x_{\ell'k}\log_e\left(\frac{x_{\ell'k}}{s_k}\right) - s_k\log_e\left(\frac{1}{2}\right)\right\},$$

where $s_k = (x_{\ell k} + x_{\ell'k})$. Either of these latter two functions may be meaningful for measuring ecological dissimilarity between sites when data consist of species abundances (nonnegative integers). Note also that the Bray–Curtis and binomial deviance measures are semimetric and so, like many other functions used in ecological applications, do not fulfill the triangle inequality (Gower and Legendre, 1986). Legendre and Gallagher (2001) demonstrated how several other ecologically meaningful dissimilarity measures (such as the chi-squared, chord, or Hellinger distance) can be obtained through the calculation of Euclidean distance on suitably transformed variables. For measures such as Euclidean distance or Bray–Curtis, some form of standardization or transformation of variables may also be done before calculating distances, so that variables have equal weight or are placed on similar scales (Faith et al., 1987; Clarke, 1993).

To obtain principal coordinates, first let matrix $A = [a_{\ell\ell'}]$, where $a_{\ell\ell'} = -\frac{1}{2}d_{\ell\ell'}^2$. Centering this matrix in the manner of Gower (1966) gives

$$G = [g_{\ell\ell'}] = [a_{\ell\ell'} - \bar{a}_{\ell\cdot} - \bar{a}_{\cdot\ell'} + \bar{a}_{\cdot\cdot}],$$

where $\bar{a}_{\ell\cdot}$ is the mean for row ℓ, $\bar{a}_{\cdot\ell'}$ is the mean for column ℓ', and $\bar{a}_{\cdot\cdot}$ is the overall mean of the values in matrix A. Next, spectral decomposition of the G matrix yields

$$G = \sum_{\ell=1}^{N}\lambda_\ell \boldsymbol{q}_\ell \boldsymbol{q}_{\ell'}^T,$$

where $\lambda_1 \geq \cdots \geq \lambda_N$ are the ordered eigenvalues of G and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N$ are the corresponding orthonormal eigenvectors. Principal coordinate axes (column vectors) are then obtained by scaling each axis $\boldsymbol{q}_\ell$ by the square root of its corresponding eigenvalue, $\boldsymbol{u}_\ell = (\lambda_\ell)^{1/2}\boldsymbol{q}_\ell$. Gower (1966) has shown how Euclidean distances among points using principal coordinate axes scaled in this way replicates the original inter-point dissimilarities in matrix D.

Now, unless the dissimilarities are indeed distances (Euclidean embeddable), matrix G may not be nonnegative definite and so some eigenvalues may be negative. This generally occurs when the distance function used is not a full metric, but is only semimetric. Negative eigenvalues may also be produced by metric coefficients that are not Euclidean. Note that we still have trace$(G) = \sum_{\ell=1}^{N}\lambda_\ell$, even if some eigenvalues are negative. In this case, the axes of matrix Q can be split into two sets, $Q = [\boldsymbol{q}_1 \cdots \boldsymbol{q}_r \mid \boldsymbol{q}_{r+1} \cdots \boldsymbol{q}_N]$, such that $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$ and $0 > \lambda_{r+1} \geq \cdots \geq \lambda_N$, corresponding to the nonnegative and the negative eigenvalues, respectively. For eigenvectors corresponding to nonnegative eigenvalues, $\ell = 1, \ldots, r$, we denote scaled axes as $\boldsymbol{u}_\ell^+ = (\lambda_\ell)^{1/2}\boldsymbol{q}_\ell$. For eigenvectors $\ell = r + 1, \ldots, N$ corresponding to negative eigenvalues, we may scale by the square root of the absolute value of λ_ℓ and subsequently multiply by $(-1)^{1/2}$, recognizing that these correspond to axes in imaginary space, i.e., $(-1)^{1/2}\boldsymbol{u}_\ell^- = (|\lambda_\ell|)^{1/2}\boldsymbol{q}_\ell$.

The two sets of principal coordinate axes can be considered separately and then recombined after squaring to calculate distances between any two points in the principal coordinate space, with squared distances calculated from imaginary axes contributing negatively. Thus, let $U = [U^+ \mid U^-]$ be the $N \times N$ matrix of principal coordinate axes, where row $\boldsymbol{u}_{ij}^+$ is a row giving coordinates along the $1, \ldots, r$ real axes for the jth observation point in the ith group and row $\boldsymbol{u}_{ij}^-$ gives the coordinates for that point along the $r + 1, \ldots, N$ imaginary axes (if any). The original dissimilarity between two points x_{ij} and $x_{i'j'}$ can be recovered in the principal coordinate space using Euclidean distances, as

$$d_{ij,i'j'} = \sqrt{\Delta^2(\boldsymbol{u}_{ij}^+, \boldsymbol{u}_{i'j'}^+) - \Delta^2(\boldsymbol{u}_{ij}^-, \boldsymbol{u}_{i'j'}^-)}.$$

Furthermore, we can calculate a centroid for each of the $i = 1, \ldots, g$ groups in each of the real and imaginary spaces as $\boldsymbol{c}_i^+$ and $\boldsymbol{c}_i^-$, respectively, in the usual way. Then, the distance (or dissimilarity) from the ijth point to its centroid in the full principal coordinate space is

$$z_{ij}^c = \sqrt{\Delta^2(\boldsymbol{u}_{ij}^+, \boldsymbol{c}_i^+) - \Delta^2(\boldsymbol{u}_{ij}^-, \boldsymbol{c}_i^-)}, \tag{3}$$

where we will consider only the positive square root. Similarly, using spatial medians instead of centroids, we have

$$z_{ij}^m = \sqrt{\Delta^2(\boldsymbol{u}_{ij}^+, \boldsymbol{m}_i^+) - \Delta^2(\boldsymbol{u}_{ij}^-, \boldsymbol{m}_i^-)}, \tag{4}$$

with $\boldsymbol{m}_i^+$ being the spatial median for the ith group using real axes and $\boldsymbol{m}_i^-$ being the spatial median for the ith group using imaginary axes.

The test for homogeneity of dispersions then simply consists of doing univariate one-way ANOVA on the z's (defined as either (3) or (4)), with or without the use of permutations, as described in Section 2.1 above. Although neither (3) nor (4) precludes the possibility of either z^c or z^m being complex values, in practice this would be highly unlikely to occur, provided a reasonable dissimilarity measure was chosen. If D contains Euclidean distances on the original variables to begin with, then (3) will be equal to (1) and (4) will be equal to (2).

3. Applications

3.1 *Stabilizing Selection in Sparrows*

Bumpus (1898) recorded the morphological characteristics of sparrows in Rhode Island after a severe storm. The subset of data given in Manly (1994) and analyzed here consists of five of these variables (total length, alar extent, and the lengths of the beak and head, the humerus, and the keel of the sternum, all in mm) for female sparrows only, 21 of which survived the storm and 28 of which did not survive. Under the general theory of stabilizing selection, birds having morphologies distant from the average sparrow should be more susceptible to being killed in the storm. The prediction is, therefore, that the group of birds that died should have a greater multivariate dispersion in their morphometric characteristics than the birds that survived.

A principal component analysis of the normalized data indeed shows a slight pattern of greater dispersion for the sparrows that had died (Figure 1, top). A test of the null hypothesis of no difference in the multivariate dispersions of the two groups, using deviations from spatial medians and 9999 permutations of LAD residuals, suggested there was some evidence for a difference ($F_m = 3.818$, $P = 0.0470$), with the average Euclidean distance to the spatial median being 1.72 for the survivors and 2.23 for those that died (Figure 1, bottom). The P-value obtained using the traditional F-distribution was similar, $P = 0.0567$, and nearly equivalent results were also obtained using distances to centroids, $F_c = 3.869$, with either the tables ($P = 0.0551$) or permutation of LS residuals ($P = 0.0541$). On the other hand, the traditional likelihood test (Box, 1949; see Rencher, 1998, p. 138–140 for details) provided no evidence against the null hypothesis of equal variance–covariance matrices ($\log M = -5.893$, $F = 0.692$, $P = 0.795$). This result was probably due to the strong similarity between groups in the correlation structure among the variables.

Monte Carlo simulation was used to empirically estimate the power and type I error for each of the tests. Data were simulated from a multivariate normal distribution, which was reasonable for these morphometric variables. Means and variances for each variable for each group and the pooled sample correlation matrix were estimated from the raw data. For type I error, 5000 data sets were simulated (each having sample sizes like the original data of $n_1 = 21$ and $n_2 = 28$) where the means and variances estimated for group 1 were used for all observations (i.e., the null hypothesis was true). For each simulation, the data were first normalized for each variable and then all four distance-based tests were calculated on the basis of the Euclidean distance measure, along with Box's M-statistic, with 999 permutations being used to obtain P-values
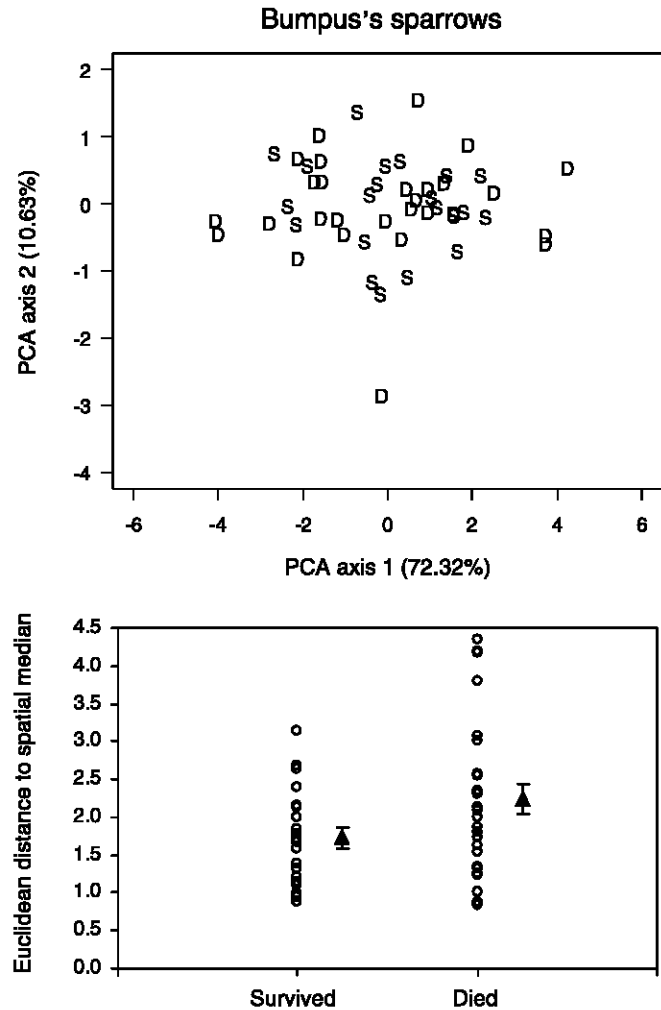

Bumpus's sparrows

Figure 1. Principal component ordination of the sparrow data (top), showing sparrows that either survived (S) or died (D), and plot of the Euclidean distances from individual points in the full dimensional space to their group spatial median (bottom), along with the group average $\pm$ 1 SE.

in the case of the permutation procedures. The number of rejections of the null hypothesis (using a significance level of $\alpha = 0.05$) was counted, yielding empirical rates of type I error. Note that the rejection rate follows a binomial distribution. Thus, for 5000 trials, the 99% confidence interval for the type I error is (0.042–0.058). The same basic approach was used to examine power, but for this, the 5000 data sets were drawn using different mean and variance parameters for each group, as estimated from the data, although the same pooled correlation matrix was used for all simulations. The mean vectors for groups 1 and 2 were

$$\mu_1 = [157.4, 241.0, 31.43, 18.50, 20.81]$$
$$\mu_2 = [158.4, 241.6, 31.48, 18.45, 20.84]$$

and the vectors of variances for groups 1 and 2 were

$$\sigma_1^2 = [11.05, 17.50, 0.531, 0.176, 0.575]$$
$$\sigma_2^2 = [15.07, 32.55, 0.728, 0.434, 1.321].$$

Table 1

Empirical type I error and power for the tests in each of three example data sets. The tests done were Box's M test (where possible) and the F-statistic using either distances to centroids (Fc) or medians (Fm), with P-values obtained using either the tables (t) or an appropriate permutation procedure (p).

		$Fc(t)$	$Fm(t)$	$Fc(p)$	$Fm(p)$	Box's M
Sparrows	Type I error	0.060	0.047	0.055	0.055	0.054
	power	0.518	0.485	0.504	0.505	0.492
Fish	Type I error	0.060	0.046	0.049	0.047	
	power	1.000	1.000	1.000	1.000	
Corals	Type I error	0.063	0.026	0.053	0.047	
	power	0.977	0.947	0.957	0.946	

for these reasons, and also because it is not generally considered useful for measuring the ecological dissimilarity among species assemblages (Faith et al., 1987; Clarke, 1993). Also, the traditional test statistic simply cannot be calculated for these data because of the large number of variables. These data were analyzed on the basis of the scale-invariant binomial deviance dissimilarity (Anderson and Millar, 2004).

A plot of the first two principal coordinate axes from the dissimilarity matrix (also called a metric multidimensional scaling or MDS plot) showed an apparent pattern of greater dispersion of the fish assemblages at Home Point compared to those at Leigh, although not many other patterns could easily be seen on this plot (Figure 2, top). The test provided very strong support for a difference in dispersion among the
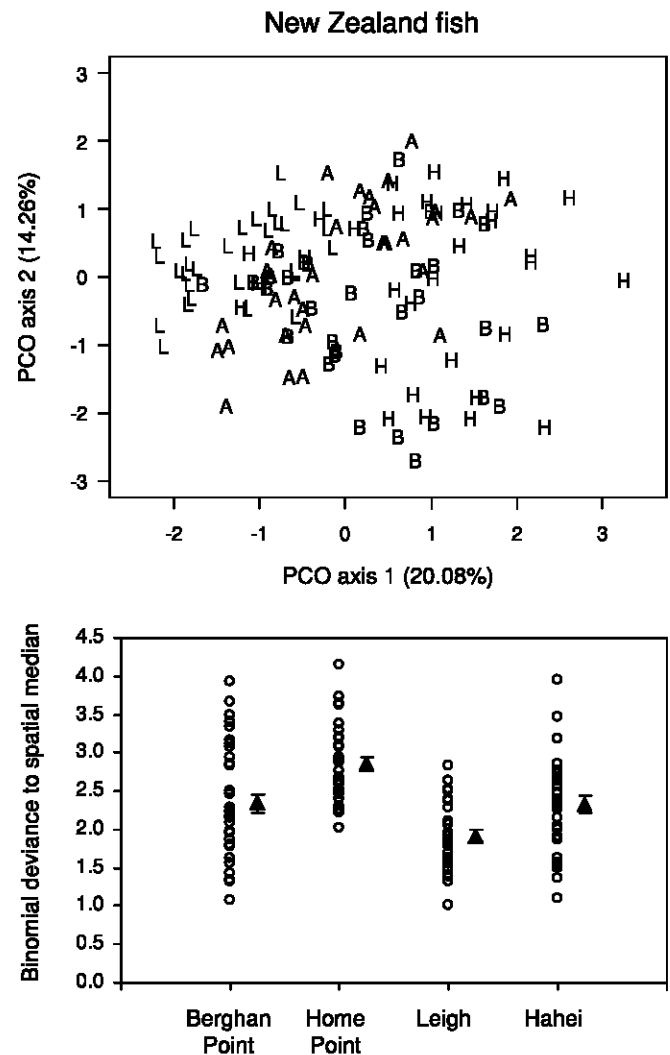
Thus, the null hypothesis of equal dispersions was false in this case, and the number of rejections (again at $\alpha = 0.05$) was counted for each method, yielding empirical measures of power.

A common correlation matrix was used even under a false null hypothesis because these tests, by their very nature, are not expected to have any power to detect differences in correlation structure among groups. They are designed only to detect differences in the overall spread of points in the multivariate space, not differences in either shape or direction. Permutation tests for correlation structure have been discussed by Krzanowski (1993), but are not considered in detail here.

The results indicated that the type I error for all of the test procedures is intact for this situation (Table 1). The empirical power of all of these tests was comparable, between 0.485 and 0.518.

3.2 Spatial Variation in New Zealand Fish

Spatial variation in temperate reef fish assemblages was examined in a study along the northeastern coast of New Zealand (Anderson and Millar, 2004). There were eight sites (separated by hundreds of meters) within each of four locations along the coast (separated by hundreds of kilometers): Berghan Point, Home Point, Leigh, and Hahei. At each site, SCUBA divers swam along each of 10 transects (25 m × 5 m), recording abundances of individual fish species. Abundances were pooled at the site level for analysis. Sampling has been repeated in each of 4 years (2000–2004) and yearly variation is not significant (see Anderson and Millar, 2004, for details), resulting in a total of $n_i = 32$ multivariate observations per location. At Berghan Point, one more site was added in two of the years, thus yielding a sample size of $n_1 = 34$ for this location alone. A total of 57 fish species were recorded. Fundamental interest lies in comparing the variability in the fish assemblages among the four locations. More particularly, the historical and present biogeography of the coastline leads to the prediction that there should be greater biodiversity (and thus greater multivariate dispersion) of assemblages at the northern locations (i.e., Berghan Point and Home Point), particularly at Home Point, which is most exposed to the influences of the East Auckland current.

Fish abundance data such as these are highly skewed and contain many zeros. Euclidean distance is inappropriate here

Figure 2. Principal coordinate ordination of the fish data (top), showing assemblages from Berghan Point (B), Home Point (H), Leigh (L), and Hahei (A), and plot of the dissimilarities, as measured by the scale-invariant binomial deviance, from individual points in the full dimensional space to their group spatial median (bottom), along with the group average ± 1 SE.

Table 2
Pairwise comparisons (with no adjustments for multiple tests) for the New Zealand fish data set using
binomial deviance dissimilarities from spatial medians and 9999 permutations of LAD residuals. Also given is
the empirical power for each of the four distance-based test procedures, as in Table 1.

	t	P	$Fc(t)$	$Fm(t)$	$Fc(p)$	$Fm(p)$
Home Pt. vs. Leigh	8.008	0.000	1.000	1.000	1.000	1.000
Home Pt. vs. Hahei	3.843	0.000	0.984	0.981	0.983	0.982
Home Pt. vs. Berghan Pt.	3.312	0.002	0.931	0.918	0.920	0.919
Hahei vs. Leigh	3.077	0.004	0.713	0.689	0.693	0.688
Berghan Pt. vs. Leigh	2.843	0.007	0.883	0.871	0.870	0.869
Berghan Pt. vs. Hahei	0.111	0.913	0.117	0.105	0.110	0.107

four groups using either distances to centroids ($F_c = 15.812$) or distances to spatial medians ($F_m = 13.967$), regardless of whether the traditional F-distribution or permutation procedures were used ($P < 0.0001$ in all cases). Thus, there were important differences in the biodiversity of fish assemblages as measured by multivariate spread. The average distance to the spatial median was significantly larger for assemblages at Home Point, followed by those at Berghan Point and Hahei, which did not differ appreciably, with the smallest average dispersion being found at Leigh (Table 2; Figure 2, bottom).

Type I error and power for the four methods were examined using the same general approach as that described for the sparrow data, except that here data were simulated from a multivariate Poisson-lognormal distribution (Aitchison and Ho, 1989). This was appropriate for these data, being highly overdispersed counts. Means and variances estimated from the original data were set to be either the same as those estimated for Berghan Point for all four groups (type I error), or the same as those estimated separately for each group (power). Power was also investigated for each of the pairwise tests. A single pooled correlation matrix obtained from the full data set was used in all cases. For each simulation, the same sample sizes as occurred in the original data were generated and analyses were done on the basis of the scale-invariant binomial deviance dissimilarity measure.

In the comparison of all four groups, power was equally high (100%) for all methods (Table 1). The type I error of the test was slightly, but not badly, inflated for the test based on distances to centroids, using either the traditional or permutation P-values. The empirical power for each pairwise comparison generally indicated the following rank order of the tests: $F_c(t) > F_c(p) > F_m(p) > F_m(t)$ (Table 2). These differences were very slight, however, and within the range of error expected under simulation.

3.3 Response of Indonesian Corals to El Niño

Warwick et al. (1990) described a study of coral assemblages at South Tikus Island, Indonesia. The percentage cover of 75 coral species along each of 10 replicate transects was obtained in each of several years from 1981 to 1988. An oceanographic El Niño event occurred in 1982–1983. Primary interest lies in determining and characterizing the effect of this event on the coral assemblages. These data included a great many zeros, due to the occasional appearance of rare species. For this reason, and to reduce the importance of some of the

more common and prolific species, Warwick et al. (1990) performed multivariate analyses on the matrix of Bray–Curtis dissimilarities, calculated from square root-transformed data.

There appeared to be much greater variability in the coral assemblages for the 1983 survey, compared to the relative dispersion observed either before (1981) or after (1985) the El Niño event (Figure 3, top). The overall tests confirmed this general observed pattern ($F_c = 9.097$, $F_m = 6.292$, $P < 0.001$ in all cases). Furthermore, pairwise comparisons demonstrated the effect to be due substantially to the multivariate dispersion in 1983 being higher than that observed in other years (Table 3; Figure 3, bottom). More detailed analyses revealed that this was largely a consequence of coral bleaching and loss of species in 1983. This led Warwick and Clarke (1993) to propose that increased multivariate dispersion, as measured by the Bray–Curtis dissimilarity, may in general be a sign of increased environmental stress.

Once again, type I error and power associated with the tests were measured using simulations. In this case, one of the most salient characteristics of the original data was the predominance of zeros. Therefore, data were simulated for each variable within each group in a two-step process. The basic idea was first to model the presence/absence of a species, and then to model its quantitative percentage cover, conditional on its presence. First, a Bernoulli random value was drawn, according to the probability of a nonzero value occurring as estimated from the data. Second, if a 1 was drawn in this first step, then a normal random variable with mean and variance estimated from the nonzero values only was then drawn and truncated to the nearest integer, otherwise the value was set at zero. This truncated normal mixture produced zero-inflated data quite similar to the observed data.

The type I error of the test (using parameters equal to those observed in 1981 for all four groups) was somewhat inflated (although not badly) for $F_c(t)$ (Table 1). In contrast, the test using the $F_m(t)$ was quite conservative. Both test statistics were robust, however, when P-values were instead obtained using a permutation procedure.

The pairwise comparisons in this example (Table 3) gave quite a lot more range in empirical power values than in the previous example. Although $F_c(t)$ demonstrated the greatest rejection rates (Table 3), this was probably due, to some extent, to its lack of robustness in maintaining type I error (Table 1). Either of the test statistics, when used with
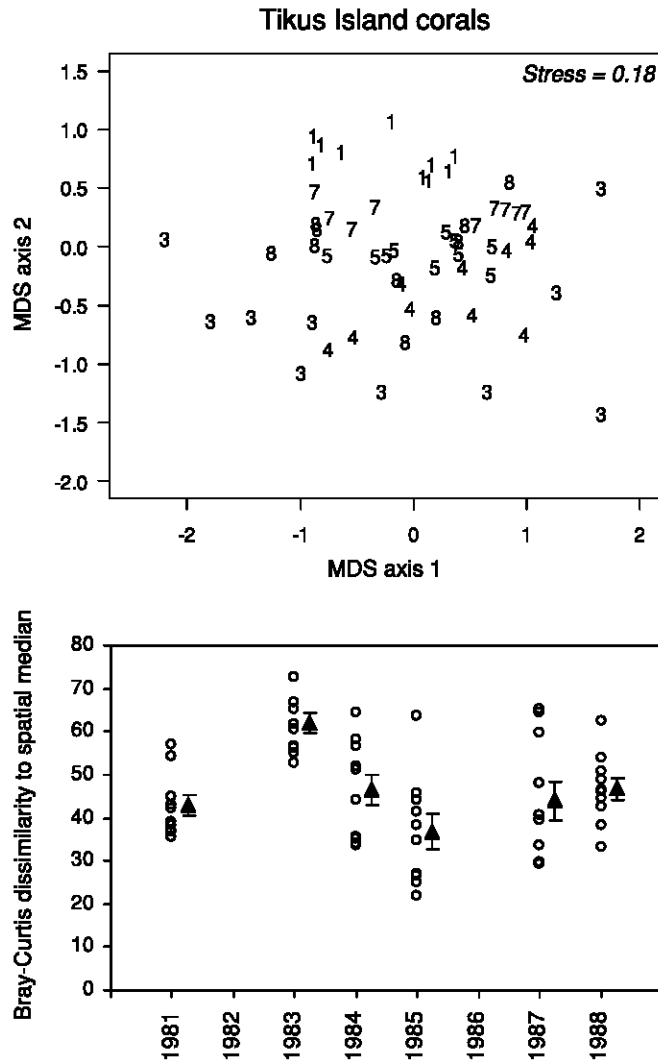
Figure 3. Nonmetric multidimensional scaling ordination of the coral data (top), where assemblages from a given year in the interval 1981–1988 are indicated using the corresponding integer 1–8, and plot of the dissimilarities, as measured by Bray–Curtis on square root-transformed data, from individual points in the full dimensional space to their group spatial median (bottom), along with the group average ± 1 SE.

permutation techniques, however, provided correct type I error and adequate power, while the approach using F_m combined with the traditional F-distribution to obtain P-values was generally a slightly more conservative test (Tables 1–3).

4. Discussion

The distance-based tests described here provide rigorous tests for homogeneity of multivariate dispersions when the traditional test is either inappropriate or impossible. This includes situations when data are nonnormal, when the measure of dispersion of interest is not the Euclidean distance, and/or when the number of variables exceeds the number of observations. The tests were found to be equally as or more powerful than

Table 3

Pairwise comparisons (with no adjustments for multiple tests) for the Tikus Island coral data set using Bray–Curtis dissimilarities from spatial medians and 9999 permutations of LAD residuals. Also given is the empirical power for each of the four distance-based test procedures, as in Table 1.

	t	P	$Fc(t)$	$Fm(t)$	$Fc(p)$	$Fm(p)$
'83 vs. '81	5.883	0.000	0.980	0.950	0.968	0.949
'83 vs. '85	5.464	0.000	0.988	0.971	0.978	0.968
'83 vs. '88	4.487	0.001	0.869	0.792	0.827	0.794
'83 vs. '84	3.656	0.002	0.672	0.564	0.618	0.571
'83 vs. '87	3.517	0.000	0.807	0.712	0.750	0.711
'88 vs. '85	2.070	0.047	0.329	0.255	0.294	0.276
'84 vs. '85	1.782	0.099	0.542	0.446	0.483	0.461
'81 vs. '85	1.295	0.213	0.205	0.164	0.183	0.181
'87 vs. '85	1.173	0.264	0.260	0.180	0.216	0.190
'88 vs. '81	1.105	0.222	0.113	0.069	0.108	0.094
'84 vs. '81	0.835	0.387	0.338	0.233	0.295	0.259
'88 vs. '87	0.513	0.566	0.076	0.046	0.068	0.057
'84 vs. '87	0.420	0.670	0.124	0.083	0.095	0.087
'87 vs. '81	0.220	0.824	0.115	0.071	0.098	0.081
'88 vs. '84	0.055	0.953	0.127	0.080	0.109	0.097

the traditional likelihood-based statistic when data were multivariate normal and alternative hypotheses were restricted to differences in overall dispersion or spread (variances) only. It is readily acknowledged that the distance-based tests considered here would not be sensitive to alternative hypotheses regarding differences in correlation or covariance structure among groups. For some robust methods directed at such alternatives, see O'Brien (1992) and Krzanowski (1993).

One distance-based approach described here utilizes the multivariate median defined in a particular way: a spatial median that has invariance under rotation of the axes (Gower, 1974; Brown, 1983). This allows the analysis to be based on non-Euclidean dissimilarities, as distances to spatial medians can, therefore, be calculated in a principal coordinate space. Although it would be possible in Euclidean space to calculate distances to an elementwise multivariate median, as suggested by O'Brien (1992) and Manly (1994), the lack of rotational invariance for the median defined in this way makes it impossible to consider generalizing its use for any dissimilarity function through the framework of principal coordinates. The deepest location (Struyf and Rousseeuw, 2000), defined as the point with maximal halfspace depth (Tukey, 1975), also lacks rotational invariance. The flip side of this issue is that the spatial median, in its turn, like the centroid, is not invariant to transformations of the variables. Therefore, the choice of transformation will have very important repercussions on the proposed distance-based tests of dispersion for a given set of data. This is not terribly surprising, however, and may even be considered to be a reasonable quality for the proposed tests, as it is well known that univariate tests of homogeneity are also affected strongly by transformations. Although perhaps less well recognized, the choice of dissimilarity measure will also have extremely important effects on the observed patterns and tests of multivariate dispersion. Thus, careful thought regarding such choices is warranted.

Simulations based on real data sets have shown that all of the proposed tests perform reasonably well, even in the presence of large numbers of variables, serious overdispersion, truncation, or zero inflation. The tests that provided the best balance between power and robustness, at least for these data sets, were those which relied on permutation procedures to obtain P-values. Thus, the tests using an F-ratio to compare distances to centroids (F_c) or distances to spatial medians (F_m), with P-values obtained using permutation of LS or LAD residuals, respectively, are both recommended. While the latter may provide a slight advantage for its potentially greater robustness in some circumstances, the former may provide slightly greater power and may provide a way to examine dispersions in response to a more complex model partitioning (e.g., for multiway ANOVA designs), but this is clearly a topic for future research.

REFERENCES

Aitchison, J. and Ho, C. H. (1989). The multivariate Poisson-lognormal distribution. *Biometrika* **76**, 643–653.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32–46.

Anderson, M. J. and Millar, R. B. (2004). Spatial variation and effects of habitat on temperate reef fish assemblages in northeastern New Zealand. *Journal of Experimental Marine Biology and Ecology* **305**, 191–221.

Anderson, M. J. and Robinson, J. (2001). Permutation tests for linear models. *Australian and New Zealand Journal of Statistics* **43**, 75–88.

Boos, D. D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances. *Technometrics* **31**, 69–82.

Boos, D. D., Janssen, P., and Veraverbeke, N. (1989). Resampling from centered data in the two-sample problem. *Journal of Statistical Planning and Inference* **21**, 327–345.

Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36**, 317–346.

Bray, J. R. and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* **27**, 325–349.

Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society, Series B* **45**, 25–30.

Brown, M. B. and Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association* **69**, 364–376.

Bumpus, H. C. (1898). The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. Biological Lectures, Marine Biology Laboratory, Woods Hole, Massachusetts, 11th Lecture, pp. 209–226.

Cade, B. S. and Richards, J. D. (1996). Permutation tests for least absolute deviation regression. *Biometrics* **52**, 886–902.

Chapman, M. G., Underwood, A. J., and Skilleter, G. A. (1995). Variability at different spatial scales between a subtidal assemblage exposed to the discharge of sewage at two control locations. *Journal of Experimental Marine Biology and Ecology* **189**, 103–122.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18**, 117–143.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981). A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351–361.

Edgington, E. S. (1995). *Randomization Tests*, 3rd edition. New York: Marcel-Dekker.

Faith, D. P., Minchin, P. R., and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* **69**, 57–68.

Fisher, R. A. (1936). The coefficient of racial likeness and the future of craniometry. *Journal of the Royal Anthropological Institute* **66**, 57–63.

Francis, R. I. C. C. and Manly, B. F. J. (2001). Bootstrap calibration to improve the reliability of tests to compare sample means and variances. *Environmetrics* **12**, 713–729.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Gower, J. C. (1974). Algorithm AS 78: The mediancentre. *Applied Statistics* **23**, 466–470.

Gower, J. C. and Krzanowski, W. J. (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *Applied Statistics* **48**, 505–519.

Gower, J. C. and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* **3**, 5–48.

Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35**, 414–415.

Hoenig, J. M. and Heise, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations in data analysis. *American Statistician* **55**, 19–24.

Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association* **50**, 946–967.

Krzanowski, W. J. (1993). Permutational tests for correlation matrices. *Statistics and Computing* **3**, 37–44.

Layard, M. W. J. (1974). A Monte Carlo comparison of tests for equality of covariance matrices. *Biometrika* **61**, 461–465.

Legendre, P. and Anderson, M. J. (1999). Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* **69**, 1–24.

Legendre, P. and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**, 271–280.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *American Statistician* **55**, 187–193.

Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics*, I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (eds), 278–292. Stanford, California: Stanford University Press.

Manly, B. F. J. (1994). *Multivariate Statistical Methods: A Primer*, 2nd edition. Boca Raton, Florida: Chapman and Hall.

Manly, B. F. J. (1997). *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edition. London: Chapman and Hall.

McArdle, B. H., Gaston, K. J., and Lawton, J. H. (1990). Variation in the size of animal populations: Patterns, problems and artefacts. *Journal of Animal Ecology* **59**, 439–454.

Mielke, P. W., Berry, K. J., and Johnson, E. S. (1976). Multi-response permutation procedures for a priori classifications. *Communications in Statistics—Theory and Methods* **5**, 1409–1424.

Miller, R. G. (1968). Jackknifing variances. *Annals of Mathematical Statistics* **39**, 567–582.

O'Brien, P. C. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics* **48**, 819–827.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association* **69**, 894–908.

Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin* **83**, 579–586.

Pillar, V. D. P. and Orlóci, L. (1996). On randomization testing in vegetation science: Multifactor comparisons of relevé groups. *Journal of Vegetation Science* **7**, 585–592.

Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. New York: John Wiley and Sons.

Smith, E. P., Pontasch, K. W., and Cairns, J. (1990). Community similarity and the analysis of multispecies environmental data: A unified statistical approach. *Water Research* **24**, 507–514.

Struyf, A. and Rousseeuw, P. J. (2000). High-dimensional computation of the deepest location. *Computational Statistics & Data Analysis* **34**, 415–426.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature* **189**, 732–735.

Tiku, M. L. and Balakrishnan, N. (1985). Testing the equality of variance-covariance matrices the robust way. *Communications in Statistics—Theory and Methods* **14**, 3033–3051.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians* **2**, 523–531.

Van Valen, L. (1978). The statistics of variation. *Evolutionary Theory* **4**, 33–43 (Erratum *Evolutionary Theory* **4**, 202).

Warwick, R. M. and Clarke, K. R. (1993). Increased variability as a symptom of stress in marine communities. *Journal of Experimental Marine Biology and Ecology* **172**, 215–226.

Warwick, R. M., Clarke, K. R., and Suharsono. (1990). A statistical analysis of coral community responses to the 1982–83 El Niño in the Thousand Islands, Indonesia. *Coral Reefs* **8**, 171–179.