Why You Cannot Transform Your Way Out of Trouble for Small Counts

David I. Warton

School of Mathematics and Statistics and Evolution & Ecology Research Centre, UNSW Sydney, New South Wales 2052, Australia
*email: David.Warton@unsw.edu.au

SUMMARY. While data transformation is a common strategy to satisfy linear modeling assumptions, a theoretical result is used to show that transformation cannot reasonably be expected to stabilize variances for small counts. Under broad assumptions, as counts get smaller, it is shown that the variance becomes proportional to the mean under monotonic transformations $g(\cdot)$ that satisfy $g(0) = 0$, excepting a few pathological cases. A suggested rule-of-thumb is that if many predicted counts are less than one then data transformation cannot reasonably be expected to stabilize variances, even for a well-chosen transformation. This result has clear implications for the analysis of counts as often implemented in the applied sciences, but particularly for multivariate analysis in ecology. Multivariate discrete data are often collected in ecology, typically with a large proportion of zeros, and it is currently widespread to use methods of analysis that do not account for differences in variance across observations nor across responses. Simulations demonstrate that failure to account for the mean–variance relationship can have particularly severe consequences in this context, and also in the univariate context if the sampling design is unbalanced.

KEY WORDS: Community composition; Ecology; Generalized linear models; Multivariate analysis; Variance stabilizing transformation.

1. Introduction

In modern statistics, count data are conventionally analyzed by specifying a distributional assumption that implies a mean–variance relationship, typically done via generalized linear models (McCullagh and Nelder, 1989) and its various extensions. However, many applied researchers with little statistical training persist with the older approach of transforming data and fitting a linear model assuming equal variance. Indeed, transformation is the only method of dealing with heteroscedasticity that is taught in most introductory statistics courses, and is still occasionally advocated in the applied literature (Ives, 2015).

A particular motivation for revisiting this issue is the analysis of multivariate abundance data in ecology (Warton et al., 2015) as in Table 1. Counts of the abundance of different taxonomic groups are recorded at different sites, and we are primarily interested in making simultaneous inferences across many taxonomic groups in a community, concerning the effect of some site treatment. A key property for the purpose of this article is that the response variable is discrete with many zeros, since not every taxon is observed at every site. While extensions of the generalized linear model have recently been proposed by several authors (Warton, 2011; Ovaskainen et al., 2017; Hui et al., 2017), the convention in ecology has been to use algorithmic approaches to analysis, constructing pairwise dissimilarities and using those as a basis for analysis (e.g., Anderson, 2001), without reference to an underlying statistical model. Several articles on this algorithmic approach have been published in this journal (Anderson, 2006; Reiss et al., 2010; Li et al., 2011; Gijbels and Omelka, 2013), and in some cases seen considerable uptake in ecology. Such methods can be understood as being related to the classical transform-linear model approach, in the sense that the distance measures typically used for analysis (as in Anderson, 2006; Li et al., 2011; Gijbels and Omelka, 2013) do not account for differences in variance across observations nor response variables, and data transformation is relied on to account for such changes in variability. For example, the commonly used Bray–Curtis distance (Bray and Curtis, 1957) between multivariate observations $(y_{i1}, \ldots, y_{ip})$ and $(y_{i'1}, \ldots, y_{ip'})$ can be written as follows:

$$d_{ij} = \frac{\sum_{j=1}^{p} |y_{ij} - y_{ij'}|}{\sum_{j=1}^{p} (y_{ij} + y_{ij'})}$$

and while there is an overall standardization on the denominator, absolute differences are summed across responses and across pairs of observations without any normalization for changes in variance. The issue of the behavior of small counts under transformation is especially relevant here, given that this type of data includes a high proportion of zeros and many rare species with very small means (e.g. 36% of the data in Table 1 are zero).

Conventional wisdom (Bartlett, 1947; Cameron and Trivedi, 2013, Section 3.7.3) suggests that the transformation approach can work reasonably well if counts are large, but that it will encounter difficulties if counts take small values. For the example of Table 1, this is evident in a “fanning out” of residuals as fitted values increase from the origin.
Transforming Small Counts

Counts of invertebrates, classified to 24 Orders, at ten sites in Sydney, Australia, from a project monitoring the effects of revegetation on ecological communities. The first two sites (first two rows) are controls, the remaining eight sites have been subject to revegetation efforts, and it is of interest to study whether there is an effect on the community and what the nature of the effect is. Data from Anthony Pik, Macquarie University.

Table 1

| Acarina | Amphipoda | Araneae | Blattodea | ... | Soleolifera | Thysanoptera | Tricladida |
|---------|-----------|---------|-----------|-----|-------------|--------------|-----------|
| 21      | 0         | 12      | 3         | ... | 2           | 0            | 0         |
| 8       | 0         | 5       | 4         | ... | 2           | 0            | 4         |
| 70      | 11        | 1       | 0         | ... | 17          | 0            | 1         |
| 306     | 21        | 3       | 0         | ... | 40          | 0            | 0         |
| 98      | 159       | 7       | 0         | ... | 13          | 0            | 0         |
| 112     | 52        | 13      | 1         | ... | 5           | 0            | 0         |
| 87      | 31        | 3       | 0         | ... | 4           | 1            | 0         |
| 33      | 1         | 7       | 0         | ... | 0           | 0            | 0         |
| 405     | 156       | 1       | 0         | ... | 26          | 0            | 0         |
| 320     | 376       | 10      | 0         | ... | 25          | 3            | 1         |

(Figure 1a), under any transformation. The central contribution of this article is to formalize this idea, showing that under broad conditions on the distribution of the counts and on the transformation applied, if counts are sufficiently small it becomes impossible to transform to stabilize variances—in the limiting case, the variance becomes proportional to the mean, irrespective of data transformation. The obvious consequence is that if counts are small we need not consider data transformation as a solution. How small is subjective, but we suggest that if the predicted mean is less than one for some observations then it can reasonably be expected that their variance will be underestimated compared to those of larger values. In such situations, assuming variances are equal when they have not in fact been stabilized can affect inference and lead to poor decision making, as illustrated by simulation in Section 4.

2. Main Result

Define a count variable $Y$ to be a discrete-valued random variable taking non-negative values only over some set $X$, including the value zero, with probability function $p(y; \mu, \theta)$ for mean $\mu$ and possible nuisance parameters $\theta$. Consider also a monotonic transformation $g(y)$ such that $g(0) = 0$ and $g(y) > 0$ if $y > 0$. Our main result is below.

**Theorem 1.** If $\lim_{\mu \to 0} \Pr(Y > 0) \xrightarrow{D} X$ for some (possibly degenerate) distribution $X$ for which $E[g(X)]$ and $E[g(X)^2]$ are defined, then

$$\lim_{\mu \to 0} \frac{\text{Var}[g(Y)]}{E[g(Y)]} = \frac{E[g(X)^2]}{E[g(X)]} = C$$

for some positive, finite constant $C$.

Theorem 1 implies that under a broad class of discrete distributions and transformations, as $\mu$ approaches zero, the slope of the mean–variance relationship (on log–log axes) approaches one, irrespective of transformation $g(y)$. Thus, Theorem 1 formalizes the notion that it is not possible to transform rare counts to stabilize the variance.

![Figure 1](image.png)

**Figure 1.** $\log(y+1)$-transforming the counts of Table 1 does not stabilize the variance, as shown in: (a) a residuals versus fits plot; (b) a sample mean–variance plot. Notice in (a) the fanning out near the origin, and in (b) the line of points in the mean–variance plot with a mean of 0.5 or below with a slope close to one. These patterns are also present under other transformations of the data.
Proof. First note that if \( y \geq 0 \) then \( \lim_{\mu \to 0} p(0) = 1 \), and after further noting that \( g(0) = 0 \), \( \lim_{\mu \to 0} E[g(Y)]=0 \). (Similarly, the limiting variance of transformed counts is zero.) Now consider the ratio of transformed variance and mean:

\[
\lim_{\mu \to 0} \frac{\text{Var}[g(Y)]}{E[g(Y)]} = \lim_{\mu \to 0} \frac{\sum_{y \in X} g(y)^2 p(y)}{\sum_{y \in X} g(y)p(y)}
\]

Excluding zero from the summation and dividing numerator and denominator by \( 1 - p(0) \):

\[
\lim_{\mu \to 0} \frac{\sum_{y \in X \setminus 0} g(y)^2 p(y)/(1 - p(0))}{\sum_{y \in X \setminus 0} g(y)p(y)/(1 - p(0))} = \frac{E[g(X)^2]}{E[g(X)]}
\]

3. Conditions under which Theorem Applies

The conditions in Theorem 1 are quite general and cover most cases of practical interest. In fact, they cover all analyses the author has previously undertaken with count data to date.

Firstly, consider the conditions on the form of distribution of \( Y \) in Theorem 1 that \( \lim_{\mu \to 0} P(Y=0)=0 \). Most common count distributions, including the Poisson and negative binomial (under the NB-2 parameterisation, as in Cameron and Trivedi, 2013, \( p(y; \mu, \theta) = \binom{y + \theta - 1}{\theta} (1 + \theta)^{-y} (1 + \frac{1}{\theta})^{-\theta} \)), have \( \lim_{\mu \to 0} P(Y=1|Y>0) = 1 \), that is, \( X \) is degenerate at the value one. Other cases of note, but which do not lead to \( X \) degenerate at one, include the negative binomial under the NB-1 parameterisation (Cameron and Trivedi, 2013, \( p(y; \mu, \theta) = \binom{y + \theta - 1}{\theta} (1 + \theta)^{-y} (1 + \frac{1}{\theta})^{-\theta} \)) and zero-inflated distributions, \( p(y; \pi, \theta) = (1 - \pi) + \pi f(y; \theta) \) for some probability function \( f(y; \theta) \) defined over \( X \), if we let \( \pi \to 0 \) while keeping \( \theta \) fixed. Counterexamples that do not have a limit for \( Y \) as \( \mu \to 0 \) can be constructed, for example, in the zero-inflated case we could let \( \theta \) change as a function of \( \pi \) in such a way that \( f(y; \theta) \) had no limiting distribution. But it is difficult to imagine a practical use for such an example.

Secondly, consider the conditions on the form of the transformation in Theorem 1. The conditions that \( g(y) > 0 \) if \( y > 0 \) and that \( g(0) = 0 \) are satisfied by most common transformations applied to count data, including log \((y+1)\), \( y^p \) for any positive \( p \) \((p = 0.5 \) or \( 0.25 \) are particularly common in ecology), and variance stabilizing transformations for the Poisson and negative binomial. The variance stabilizing transform has classical origins (Bartlett, 1947), it satisfies \( g(y) = V(y)^{-1/2} \) for mean–variance function \( V(\cdot) \), and is so-called because its delta method approximation to the transformed variance is a constant. Transformations that do not satisfy \( g(0) = 0 \) will still have an asymptotic variance that is a linear function of its mean, and could be trivially modified to satisfy the conditions of Theorem 1 by subtracting the constant \( g(0) \). The condition that \( E[g(X)] \) and \( E[g(X)^2] \) are both defined is readily satisfied if \( X \) is degenerate at one, since this requires only that \( g(1) \) be finite. In other cases, there are certainly transformations one could construct that do not have finite expectations under transformation. But again, it is difficult to imagine a practical use for such examples, as a transformation that does not result in finite first and second moments could hardly be expected to be useful for variance stabilization!

The example data of Table 1 illustrate the point, with a sample mean–variance relationship showing a linear trend with a slope close to one at small values (mean less than one), even after transformation (Figure 1b). This mean–variance plot also has a peculiar (and unrelated) pattern at larger means, with sample variances for the control often being quite small, likely due to there only being two replicates in the control, hence considerable variability in sample estimators of mean and variance.

4. Implications

The obvious consequence of this result is that when counts are small, it is simply not possible to transform to stabilize the variance. Assuming the variance is constant when it has not been stabilized can affect inference and lead to poor decision making, in univariate (Miller, 1986) and multivariate (Warton et al., 2012) settings.

But exactly how small is too small? Simulation was used to explore the mean–variance relationship under transformation, for Poisson and negative binomial counts (with shape parameter one) under a few different transformations (Figure 2). While Theorem 1 applies to the relationship between the transformed mean and variance, Figure 2 plots the transformed variance against the untransformed mean (to refer results back to the untransformed scale), on the understanding that the proportionality result of Theorem 1 implies proportionality in the limit on Figure 2 also. We also include broken lines for the relationship between transformed variance and untransformed mean count, as predicted via Theorem 1. Transformed variances were estimated from a million randomly generated counts at each of 50 different means regularly spaced (on the log scale) between 0.05 and 50.

A few results are apparent. Firstly, Theorem 1 seems to be a good approximation for the mean–variance relationship only for very small counts, with mean less than 0.2, say. Secondly, comparing Figure 2a and b, a given transformation can have quite different effects on data from different distributions, so clearly any given transformation must be used with care. Finally, an increasing mean–variance relationship is seen under all transformations, and both distributions considered, until a mean of at least one. These results suggest a rough rule-of-thumb that if many predicted counts are less than one, variance stabilization cannot be expected to work well. Even under the “variance-stabilizing” transformation, the variance did not actually stabilize until the mean was about five. The variance stabilizing transformation was never intended for counts with a mean of one or less, indeed Bartlett (1947) described mean counts in the 2–10 range as “very small.”

4.1. Implications for Linear Models

The implications of heteroscedasticity in linear models are well-known (e.g., Miller, 1986). Estimates of standard errors are typically biased, an important exception being in perfectly balanced designs. The direction and extent of bias...
Figure 2. Variance of transformed counts, under different data transformations, as a function of the untransformed mean, for (a) Poisson counts, (b) negative binomial counts (NB2 parameterization), as estimated from a million random counts at each value for the mean. The asymptotic mean–variance relationship, as predicted by Theorem 1, is also included as a thin broken line, and coincides well with the results for small means. In the Poisson case, the asymptotic relationship is the same for the variance-stabilizing and $y^{1/4}$ transformations.

depends on the degree of heteroscedasticity and on the sampling design. Specifically, it depends not just on the degree to which variances differ across observations, it also depends on the degree to which observations with larger variances are under- or over-sampled compared to observations with smaller variances (leading respectively to under- or over-estimation of the standard error). When standard errors are biased this will affect the coverage probability of confidence intervals, rejection rate of tests, and even if using resampling (e.g. a permutation test) to control error rates, it will have predictable and largely undesirable effects (Warton et al., 2016). A simple coverage probability simulation demonstrates this idea (Figure 3a).

Univariate Poisson counts were simulated under the design of Table 1, that is, two independent samples of sizes two ("control") and eight ("treatment"), with interest in a 95% confidence interval for the treatment effect. The mean of the treatment counts took 50 different values regularly spaced (on the log scale) between 0.05 and 50, and the mean of the control counts was five times larger. A linear model was fitted using the variance stabilizing transform ($\sqrt{y}$), and Wald confidence intervals constructed. Under data transformation, the treatment effect is estimated with bias. This transformation bias was adjusted for by estimating coverage probability for the "true" treatment effect on the transformed scale, the difference in means on transformed counts as estimated from a million observations. Linear model results were compared to those of a Poisson regression model either using a Wald interval or a confidence interval constructed by inverting the likelihood ratio statistic (Hilbe, 2007, Section 5.2.1.4). Coverage probability at the 95% level was estimated from 5000 simulated datasets. The code used for these simulations is available as Supplementary Material.

Linear models performed very poorly for small counts, with coverage as low as 60% (Figure 3a). This happened because for small counts the variance could not be stabilized, so the pooled variance had a standard error that was too small, resulting in undercoverage. The generalized linear model performed surprisingly well given the small sample size ($n = 10$), but for small means, the Wald confidence interval was often

Figure 3. Simulations illustrating some implications of failure to model the mean–variance relationship, when investigating a treatment effect in an unbalanced design as in Table 1. (a) Coverage probabilities of 95% confidence intervals for a treatment effect for a univariate Poisson response. Exact confidence intervals should lie in the gray band 95% of the time. (b) Power of two multivariate tests from a simulation with Poisson counts, where one "effect variable" was set to a mean of ten in the control group, and the remaining 23 variables had no treatment effect. Note from (a) that linear models had coverage as low as 60% for small counts, and from (b) that the distance-based approach (PERMANOVA) had lower power, unless the mean of the effect variable was large. Both properties arise from failure to stabilize the variance at low means, and can be resolved by including an appropriate assumption on the mean–variance relationship in analyses (as done here using Poisson regressions).
very wide due to boundary problems (Væth, 1985). This issue was largely resolved when estimating the confidence interval by inverting a likelihood ratio statistic (Figure 3a).

4.2. Implications for Multivariate Analysis
A second and important implication of the result in this article is in multivariate analysis of count data in ecology, as in the motivating example (Table 1). It is especially relevant here to discuss behavior as \( \mu \to 0 \), since a large proportion of counts are zero, with not all taxonomic groups being observed at all sites, and many being rarely observed. There has been a history of transforming data and then applying a pairwise dissimilarity measure to characterize differences between multivariate observations (Anderson, 2006; Reiss et al., 2010; Li et al., 2011; Gijbels and Omelka, 2013), using dissimilarities that imply the quite strong assumption of equal variances both across observations and across response variables (Warton et al., 2012). This can be quite problematic, as illustrated in a simple power simulation (Figure 3b).

Power was computed for two different multivariate tests for a treatment effect, mimicking the data of Table 1, but with a treatment effect in only one out of the 24 response variables. This “effect variable” was set to have a mean abundance of ten in the control group, and its mean in the treatment group was set the sample mean from the original data (Table 1). The remaining 23 responses were simulated to have constant means, equal to sample estimates from the original data, which ranged from 0.2 to almost 5000, with all but one taxon having means below 150. In total 24 different simulations were run, treating each response as the effect variable once. Data were simulated to be independent and Poisson, further work suggests results are largely unaffected by correlation across variables, and that including overdispersion only exaggerates the effects seen here. Power was computed from 5000 simulated datasets and plotted against the mean of the effect variable for the following tests. A common test from ecology, PERMANOVA using the Bray–Curtis distance (Anderson, 2001), was computed on square-root transformed counts (this being the variance stabilizing transform), via the adonis function in the vegan package (Oksanen et al., 2015). This was compared to a likelihood ratio test from a generalized linear model fitted across all 24 response variables, assuming the data were Poisson and independent, computed using anova.manyglm from the mvabund package (Wang et al., 2012). A permutation test was used in each case for exact inference, based on the 45 possible permutations of the design.

The distance-based test had relatively poor power for all but large treatment means (Figure 3b). The distance-based test statistic did not account for differences in variance between response variables, meaning it paid little attention to taxa with lower mean hence lower variance, and transformation was unable to address this issue when the mean was small. Similar behavior has been shown in a different simulation setting (Warton et al., 2012).

4.3. Violations of Mean–Variance Assumptions
Using a generalized linear model has clear advantages when the mean–variance relationship can be modeled correctly. Even a variance-stabilizing transformation, while clearly inferior for small means, can improve matters considerably if the right transformation is chosen. But what happens if you misspecify the mean–variance relationship, or the data transformation intended to stabilize variances?

The simulations of Figure 3 were repeated but for negative binomial data and with different overdispersion parameters across samples, \( \theta = 1 \) and 0.25 for control and treatment groups, respectively. Negative binomial regression models were fitted via the manyglm function in the mvabund package (Wang et al., 2012), incorrectly assuming constant overdispersion. When fitting a linear model (Figure 4a) or using PERMANOVA (Figure 4b), the commonly used \( \log(y + 1) \) transformation was employed rather than a variance-stabilizing transformation—thus all models considered were misspecified. As previously, 5000 simulated datasets were used to estimate rejection rates, and multivariate tests used all 45 permutations of treatment labels.

Figure 4. Simulations illustrating some implications of misspecification of the mean–variance relationship, or variance-stabilizing transformation, when investigating a treatment effect in an unbalanced design as in Table 1. (a) Coverage probabilities of 95% confidence intervals for a treatment effect for a univariate negative binomial response with different overdispersion across samples. Exact confidence intervals should lie in the gray band 95% of the time. (b) Power of two multivariate tests from a simulation with negative binomial counts with differing overdispersion across samples, and one “effect variable” set to a mean of ten in the control group, and the remaining 23 variables had no treatment effect. All models have been misspecified, either assuming common overdispersion or using a data transformation that does not stabilize the variance. Note from (a) that no method consistently maintained nominal coverage probability, and from (b) that power curves did not converge to the significance level (0.05) as the mean count approached ten, suggesting both tests were invalid in this setting.
Figure 5. Diagnostic tools detecting misspecification of the model. Multivariate data were simulated from negative binomial distributions with differing overdispersion across samples, as for Figure 4, which is evident in: (a) a residuals versus fits plot, with more variation in residuals in the control group; (b) a sample mean–variance plot, with different mean–variance trends across samples.

All tests performed poorly in both the univariate (Figure 4a) and multivariate (Figure 4b) setting. Neither the linear model nor negative binomial regression maintained nominal coverage probability when estimating the treatment effect in univariate simulations (Figure 4a)—for small means, the linear model had coverage probability below 60%, and for large means, all intervals were conservative, exceeding 95% coverage. The linear model did not maintain nominal coverage probabilities for large values of the mean because model misspecification led to estimates of the standard error that were biased upward. Negative binomial regression also had this issue, but in addition, the sample size was too small for accurate inference when using large sample critical values from a normal distribution (Szöcs and Schäfer, 2015, resampling can help with this issue). In multivariate simulations, while similar general patterns were observed as for the Poisson simulations (Figure 3), the concerning feature in this plot was that as the treatment mean of the effect variable approached ten (the value it was set to in the control group), none of the test statistics approached the nominal rejection rate of 5% (Figure 4b). The difference in variances across samples seemed to considerably inflate error rates.

The effects of misspecification in Figure 4 can be rather dramatic and concerning, emphasizing the importance of carefully checking that the model fitted (or transformation used) is appropriate for the data at hand. Useful diagnostic tools are residuals proposed by Dunn and Smyth (1996), using the probability integral transform (Figure 5a), sample mean–variance relationships (as in Figure 5b), and standard model selection approaches. Such methods are imperfect, especially in cases like our univariate simulations (Figure 4a), where there were only ten observations in a given dataset. In the multivariate setting however the advantage of having many response variables is that there are many observations that can be used to check distributional assumptions, for example, for the simulated data of Figure 4b we can clearly see a change in the amount of overdispersion across the two groups (Figure 5).

5. Summary
In summary, while it remains relatively common in the applied sciences to transform counts and assume equal variance, it has been shown that it is simply not feasible to transform rare counts to satisfy equal variance assumptions (Theorem 1)—when there are many zeros, hence fitted values can be less than one, variances will differ appreciably across observations irrespective of transformation. The implications of ignoring this heteroscedasticity are especially severe when sampling is unbalanced (Figure 3a), or in multivariate settings when failing to account for differences in variance across response variables (Figure 3b), a practice that is unfortunately quite common in ecology at the moment. An effective solution to the problem in most cases is to assume an appropriate mean–variance relationship in the analysis model, typically using the generalized linear model and its extensions, using diagnostic tools to guide model specification.

6. Supplementary Materials
The code used to undertake simulations is available as an R script file at the Biometrics website on Wiley Online Library.

ACKNOWLEDGEMENTS
Thanks to the UNSW Eco-Stats group for feedback on an initial draft. The author is supported by the Australian Research Council (FT120100501 and DP150100823).

REFERENCES
Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26, 32–46.
Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253.
Bartlett, M. S. (1947). The use of transformations. *Biometrics* 3, 39–52.
Bray, J. R. and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27, 325–349.
Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Dunn, P. and Smyth, G. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics* **5**, 236–244.

Gijbels, I. and Omelka, M. (2013). Testing for homogeneity of multivariate dispersions using dissimilarity measures. *Biometrics* **69**, 137–145.

Hilbe, J. M. (2007). *Negative Binomial Regression*. Cambridge: Cambridge University Press.

Hui, F. K., Warton, D. I., Ormerod, J. T., Haapaniemi, V. and Taskinen, S. (2017). Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics* **26**, 35–43.

Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution* **6**, 828–835.

Li, J., Ban, J. and Santiago, L. S. (2011). Nonparametric tests for homogeneity of species assemblages: A data depth approach. *Biometrics* **67**, 1481–1488.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman & Hall.

Miller, R. G., Jr (1986). *Beyond ANOVA, Basics of Applied Statistics*. New York: John Wiley & Sons.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O’Hara, R. B., et al. (2015). *vegan: Community Ecology Package*. R package version 2.3-2.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., et al. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters* **20**, 561–575.

Reiss, P. T., Stevens, M. H. H., Shehzad, Z., Petkova, E. and Milham, M. P. (2010). On distance-based permutation tests for between-group comparisons. *Biometrics* **66**, 636–643.

Szöcs, E. and Schäfer, R. B. (2015). Ecotoxicology is not normal. *Environmental Science and Pollution Research* **22**, 13990–13999.

Væth, M. (1985). On the use of Wald’s test in exponential families. *International Statistical Review* **53**, 199–214.

Wang, Y., Naumann, U., Wright, S. T. and Warton, D. I. (2012). *mvabund*—An R package for model-based analysis of multivariate abundance data. *Methods in Ecology and Evolution* **3**, 471–474.

Warton, D. I. (2011). Regularized sandwich estimators for analysis of high dimensional data using generalized estimating equations. *Biometrics* **67**, 116–123.

Warton, D. I., Blanchet, F. G., O’Hara, R., Ovaskainen, O., Taskinen, S., Walker, S. C., et al. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology and Evolution* **30**, 766–779.

Warton, D. I., Lyons, M., Stoklosa, J. and Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution* **7**, 882–890.

Warton, D. I., Wright, S. T. and Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution* **3**, 89–101.

*Received August 2016. Revised April 2017. Accepted April 2017.*