MODELS FOR POWER OF DETECTING TRENDS—A REPLY TO LINK AND HATFIELD

Tim Gerrodette[1]

Link and Hatfield (1990) have criticized the analysis of Gerrodette (1987), which investigated the statistical power of detecting linear or exponential trends in ecological data. The analysis was not limited to detecting temporal changes in the size of animal populations, although to my knowledge it has only been applied in that way; it could also be used, for example, in designing sampling along spatial gradients. Link and Hatfield concluded that the power of detecting a trend had been overestimated by Gerrodette's methods. However, Link and Hatfield's comments do not invalidate the original analysis; rather, they suggest an extension of it. Here I reply to the main points raised by Link and Hatfield, and extend the original results slightly by considering other approaches that relax some of the restrictive assumptions of the 1987 paper.

(1) *Selection of proper statistical distribution.* Gerrodette (1987) employed the standard normal distribution (referred to here as the z distribution) to estimate the power of linear regression analysis to detect trends. Link and Hatfield (1990) argued that the t distribution would be more appropriate to compute power for small sample sizes. The difference arises from considering different statistical models. Link and Hatfield considered the classical regression problem, where the investigator has simply n data points. With no information about the variance of each point, the error variance must be estimated from the deviations of the observed points from the regression line (i.e., from the residual sum of squares). Because a slope and intercept have been computed from the data already, there remain $n - 2$ degrees of freedom (df) for estimating the variance. The measurement of temporal or spatial gradients in ecology may often involve small n (5–10, say), and this means that the t distribution should be used.

In many ecological applications, however, the dependent or response variable (abundance A_i in this case) is not a simple, easily measured quantity. Gerrodette (1987) considered, but did not clearly distinguish, a

[1] National Marine Fisheries Service, Southwest Fisheries Science Center, P.O. Box 271, La Jolla, California 92038 USA.

more complex situation in which estimates of A_i resulted from an extended sampling effort, so that estimates of A_i and also var(A_i) would be made at each of n points. For example, a program to monitor population size, where estimates of A_i and var(A_i) come from annual surveys in year i, would typically produce data of this kind. The df for significance testing may not be clearly defined in this situation, but with reasonably large sampling effort, we may consider that the var(A_i) have been estimated with "large" df. This leads to using the z distribution.

Intermediate cases occur if there are "simple" estimates (i.e., with no associated variance estimates) but with replication or near replication at each "point." With k true replicates at each of n "points," there are $n - 2$ df to test lack-of-fit and $n(k - 1)$ df to estimate true error. If these can be pooled, a total of $kn - 2$ df can be used to estimate variance. Similar considerations apply even if there are not true replicates, but the data can be arranged in n groups of near replicates (not necessarily equal in size) based on values of the independent variable; in such a case there are effectively only n "points" available for fitting the regression line, but a possibly much larger number of values for estimating the error variance.

The general issue is that the degrees of freedom to be used in testing the significance of a regression line are not always simply $n - 2$. Determination of the appropriate df, and hence whether the t or z distribution should be used to estimate power, depends on the nature of the data. One may argue that since the t approaches the z distribution for large df, the t distribution could be used in all cases. In principle this is correct, but there are important practical reasons for using the z distribution when it is suitable. Computational routines are widely available. Indeed, because the z distribution is readily tabled, it is possible to solve many problems, at least approximately, with just a hand calculator. In the original paper (Gerrodette 1987), I attempted to emphasize this practical application. The probability that a trend *will* be detected, given that it is present, is a calculation so infrequently considered that even "quick and dirty" estimates of power will show that many proposed experimental designs have unacceptably low power.

Calculations of power using the t distribution are slightly more involved. For a t test at the α significance level with df degrees of freedom, the power of the test is

$$\Pr\{nct(\mathrm{df},\delta) < t_{\alpha,\mathrm{df}}\}$$

for $\delta < 0$,

$$\Pr\{nct(\mathrm{df},\delta) > t_{1-\alpha,\mathrm{df}}\}$$

for $\delta > 0$, and

$$\Pr\{nct(\mathrm{df},\delta) < t_{\alpha/2,\mathrm{df}}\} + \Pr\{nct(\mathrm{df},\delta) > t_{1-\alpha/2,\mathrm{df}}\}$$

in the two-sided case, where δ is the noncentrality parameter of the noncentral t distribution nct. Unfortunately, Link and Hatfield (1990) tried to use the central t distribution to compute power, and thus incorrectly concluded that power could only be estimated by Monte Carlo simulation. Tables of the noncentral t distribution are available (Resnikoff and Lieberman 1957, Owen 1965, Park 1972), but the function is more easily implemented by computer routines in IMSL (function TNDF), SAS (function PROBT), and some specialized programs for computing power (Goldstein 1989).

In the present case, the noncentrality parameter δ is equal to

$$\delta = \frac{b}{\sigma_b},$$

where b is the true slope of the regression line. Note the similarity of this formula with the usual test statistic for the regression line

$$T = \frac{\hat{b}}{s_{\hat{b}}}.$$

The test statistic T, however, uses estimates for the slope of the regression line and its standard deviation derived from the data. A calculation of power, on the other hand, is conditional on specified population values, not sample estimates, a point sometimes confused.

Note also that δ as defined above is a dimensionless number that measures the difference of the slope of the regression line from the null hypothesis H_0: $b = 0$ in standardized units. For this reason, δ can be thought of as a measure of the strength of some effect one is trying to detect through the regression analysis, i.e., an effect size index (Cohen 1988).

To sum up, using the z distribution to calculate power is a valid approach if the number of df is effectively "large." If the number of df is "small," particularly in the classical regression case with $n - 2$ df, the t distribution will provide more accurate estimates of power. Because t values are larger than corresponding z values for any finite df, calculations of power based on the t distribution will always be lower than those based on the z distribution. Similarly, calculations based on the t distribution will indicate higher rates of change needed to be detectable, as well as larger sample sizes and smaller coefficients of variation needed to detect a given rate of change. The numerical difference between calculations of power using the z and t distributions may or may not be important, depending on several factors in addition to df (factors discussed in Gerrodette 1987). A specific example is shown in Fig. 1 with $n = 10$, $\alpha = .05$, for the linear model with the coefficient of variation (CV) proportional to the inverse of the square root of abundance (see Gerrodette 1987), with power plotted as a function of the rate of change.
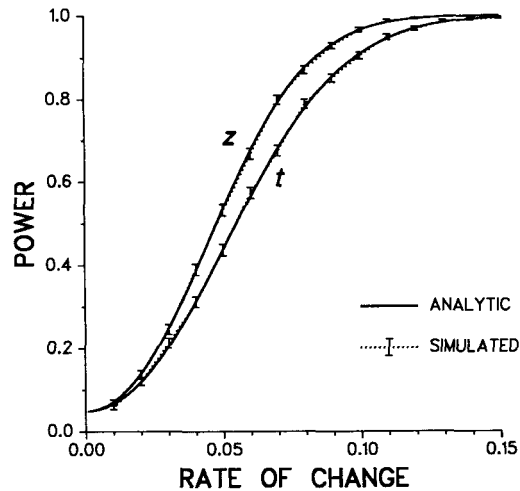

Fig. 1. Power of linear regression as a function of r, the rate of change, estimated by analytic formula and Monte Carlo simulation, and based on either the z or t distribution. Calculations are based on the linear model with CV $\propto A^{-\frac{1}{2}}$, $\mathrm{CV}_1 = 0.2$, $n = 10$, and $\alpha = .05$, two-tailed (cf. Gerrodette 1987). The simulation results are so close to the analytic results that they are difficult to distinguish. Error bars show 95% binomial confidence intervals for 5000 simulations.

(2) *Accuracy of analytic results.* The models considered in Gerrodette (1987) involved violations, to various degrees, of the normal and homoscedastic assumptions of linear regression. These violations were explicitly recognized in the original paper. Because of such violations, the formulae presented did not compute power exactly, but only estimated, or approximated, power. Link and Hatfield (1990) have confirmed this, showing that the formulae are not exact. The question remains, however, whether the approximations are sufficiently accurate to be useful in practice in an experimental design context. Both simulations (discussed in the next paragraph) and analytic comparison (Appendix) show that they are.

Link and Hatfield (1990) claimed that analytic estimates of power for the models discussed in Gerrodette (1987), even if based on the t distribution, were invalid, and they recommended Monte Carlo simulations. However, Link and Hatfield's calculations of power were invalid because they were based on the central t distribution, rather than the noncentral t, as noted in the previous section. I used simulations to verify the original analytic results, but did not report the simulations at that time. Simulation results using either the z or t distribution generally agree with analytic formulae within the uncertainty of the Monte Carlo estimator (Fig. 1). Data for each run of the Monte Carlo simulations were randomly drawn from normal

distributions with means A_i and variances var(A_i) as given in Gerrodette (1987). Analytic values of power in Fig. 1 were computed with either the z or t distribution functions using $\delta = b/\sigma_b$, where b and σ_b were computed from equations in Gerrodette (1987). The close agreement of the analytic formulae with the simulations, in spite of known violations of assumptions of the regression model, is probably due to two factors: linear regression is a robust procedure, and the change in abundance is small (on a relative basis) in typical trend problems.

Therefore, the supposed overestimation of power that Link and Hatfield (1990) discussed is not due to any problem with the analysis presented in Gerrodette (1987), but relates entirely to a different model based on the t distribution. Link and Hatfield's simulation results ("actual" in their figures) are indistinguishable from the results, simulated or analytic, based on the t distribution properly computed (compare Fig. 1 with their Fig. 1). Furthermore, their simulated estimate of power for the sea otter (*Enhydra lutris*) example (Gerrodette 1987: 1368–1370) is virtually identical to the analytic calculation of power using the t distribution (0.424 vs. 0.420). Monte Carlo simulation is certainly a useful tool, and could profitably be used in situations less tractable than the simple models of my original paper. For linear (additive) and exponential (multiplicative) models of change with equally spaced sampling, however, the analytic approximations are quite adequate, and simulations are unnecessary.

(3) *More general approaches.* The analysis of Gerrodette (1987) was based on unweighted linear regression with equal sampling effort at equally spaced points. These are fairly restrictive assumptions. More generally, the slope and variance of a weighted regression line, using the inverse of the point variances $V_i = \text{var}(A_i)$ as weights, are

$$b = \frac{\sum \frac{1}{V_i} \sum \frac{x_i A_i}{V_i} - \sum \frac{x_i}{V_i} \sum \frac{A_i}{V_i}}{\sum \frac{1}{V_i} \sum \frac{x_i^2}{V_i} - \left(\sum \frac{x_i}{V_i}\right)^2}$$

and

$$\sigma_b^2 = \frac{\sigma^2 \sum \frac{1}{V_i}}{\sum \frac{1}{V_i} \sum \frac{x_i^2}{V_i} - \left(\sum \frac{x_i}{V_i}\right)^2},$$

where all sums are taken from $i = 1, \ldots, n$. These expressions follow directly from standard equations for weighted regressions (Draper and Smith 1981). Then we may estimate power for arbitrary x_i and V_i as

$$1 - \beta = \Phi\left(z_\alpha - \frac{b}{\sigma_b}\right)$$

for a decreasing trend ($b < 0$), as

$$1 - \beta = \Phi\left(\frac{b}{\sigma_b} - z_{1-\alpha}\right)$$

for an increasing trend ($b > 0$), and as

$$1 - \beta = \Phi\left(z_{\alpha/2} - \frac{b}{\sigma_b}\right) + \Phi\left(\frac{b}{\sigma_b} - z_{1-\alpha/2}\right)$$

for a trend in either direction. If, as before, we assume that the variance at each point is V_i, we may take $\sigma^2 = 1$ (i.e., it is not estimated), and let Φ be the distribution function for the standard normal distribution. If the V_i are treated as weighting factors, we estimate the common error variance σ^2 from the residuals about the regression, replace z_α by $t_{\alpha, df}$, and let Φ be the distribution function for the noncentral t distribution.

Finally, to test for simple patterns in the A_i, we may dispense with regression entirely and use the method of linear contrasts, as in ANOVA. Given independent estimates of abundance $\hat{A}_i$, $i = 1, \ldots, n$, with associated variances $\hat{V}_i$, we can test for a linear trend by choosing contrast coefficients c_i proportional to $x_i - \bar{x}$, computing

$$Z = \frac{\sum c_i \hat{A}_i}{(\sum c_i^2 \hat{V}_i)^{1/2}},$$

and comparing Z to a standard normal distribution. Again, this procedure implicitly assumes that the $\hat{V}_i$ have been estimated with "large" df: with different assumptions the test could be based on the t distribution. Power for this test is computed in a manner similar to the above, e.g., as

$$1 - \beta = \Phi\left[\frac{\sum c_i A_i}{(\sum c_i^2 V_i)^{1/2}} - z_{1-\alpha}\right]$$

for a one-sided test for an increasing trend, where A_i and V_i are generated by the alternative hypothesis for which power is being computed. Calculations of power by this method are identical to the results in Gerrodette (1987), except for slight differences in some cases resulting from using the variances V_i directly rather than using the mean variance as an approximation (see Appendix). However, the contrast method is considerably more flexible: it allows the x_i to be unequally spaced (for example, if surveys were not conducted at equal intervals), it allows the V_i to follow any pattern (for example, if sampling effort or survey methods changed), and it allows other, nonlinear patterns in the A_i to be tested by choosing other coefficients c_i, subject to the condition that $\sum c_i = 0$.

Reporting of statistical power is widely neglected in ecological work (Peterman 1990). Although not a difficult concept, power is unfamiliar to many ecologists, and its calculation and use are easily misunderstood. For example, Link and Hatfield (1990) correctly point out that the formula for the confidence interval and standard error given on p. 1368 of Gerrodette (1987) is in error. The formula should apply to b, the estimated slope of the regression line, not r, the rate of change,

and I would add that the formula is out of place in any case because it does not directly refer to power. However, Link and Hatfield go on to provide methods for estimating r. Their procedure gives an estimate of r for a particular set of data, but the answer is irrelevant for a power analysis. This is most easily seen by noting that the equations they give for r do not involve either α or β, the probabilities of statistical error of main interest in a power analysis. Link and Hatfield have answered the question: "What is the best estimate of r, given the data at hand?" However, the question of interest for a power analysis is: "What is the minimum r that will give a significant result, with controlled Type 1 and 2 errors, if the experiment were to be carried out under the stated conditions?"

Acknowledgments: I thank K. P. Burnham for critical comments and for suggesting the method of linear contrasts applied to this problem. C. Lennert provided useful statistical discussions.

Literature Cited

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum, Hillsdale, New Jersey, USA.

Draper, N. R., and H. Smith. 1981. Applied regression analysis. John Wiley & Sons, New York, New York, USA.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology 68:1364–1372.

Goldstein, R. 1989. Power and sample size via MS/PC-DOS computers. American Statistician 43:253–260.

Link, W. A., and J. S. Hatfield. 1990. Power calculations and model selection for trend analysis: a comment. Ecology 71:1217–1220.

Owen, D. B. 1965. The power of Student's t-test. Journal of the American Statistical Association 60:320–333.

Park, H. 1972. Tables of percentage points of the non-central t-distribution. Technical Report Number 152, Department of Operations Research and Department of Statistics, Stanford University, Stanford, California, USA.

Peterman, R. M. 1990. The importance of reporting statistical power: the forest decline and acidic deposition example. Ecology 71:2024–2027.

Resnikoff, G. J., and G. J. Lieberman. 1957. Tables of the non-central t-distribution. Stanford University Press, Stanford, California, USA.

APPENDIX

Under the models proposed by Gerrodette (1987), Link and Hatfield (1990, Appendix) gave an exact expression for the error variance σ^2 (which they labelled $\sigma_b{}^2$) as

$$\sigma^2 = \frac{3(n + 1)^2 \Sigma V_i - 12(n + 1) \Sigma i V_i + 12 \Sigma i^2 V_i}{n(n + 1)(n - 1)},$$

where all sums are taken from 1 to n. Gerrodette (1987) approximated the same quantity (which he labelled $\sigma_{res}{}^2$) as the simple arithmetic average of the variances $\Sigma V_i/n$. Link and Hatfield's expression can be written as

$$\sigma^2 = \frac{\Sigma i^2 V_i - (n + 1) \Sigma i V_i + \left(\frac{n + 1}{2}\right)^2 \Sigma V_i}{n\frac{(n + 1)(n - 1)}{12}}$$

$$= \frac{\Sigma V_i(i - \bar{\imath})^2}{\Sigma (i - \bar{\imath})^2}.$$

This is an average of the V_i weighted by the squared deviations of the i's from their mean. Because of the symmetrical nature of the series $(i - \bar{\imath})^2$, $i = 1, \ldots, n$, and the regular pattern of V_i ($V_i \propto A_i{}^q$, $q = 1, 2, 3$), the arithmetic mean will equal the weighted mean for $q = 1$ with the linear model and for $q = 2$ with the exponential model, and will differ only slightly for other combinations of q and model with n in the small-to-moderate range under consideration.