Statistical Power in Analyses of Population Trend Data

David T. Cobb, *Florida Game and Fresh Water Fish Commission, Rt. 7, Box 3055, Quincy, FL 32351*

Gary L. Sprandel, *Florida Game and Fresh Water Fish Commission, Rt. 7, Box 3055, Quincy, FL 32351*

Douglas E. Runde, *WTC 1A5, Weyerhauser Corporation, 32901 Weyerhauser Way S., Federal Way, WA 98002*

Abstract: We developed a Monte Carlo simulation approach to examine statistical power in analysis of population trend data. Our stepwise approach was to perform a regression analysis to test the null hypothesis that the slope of the time series regression line was equal to 0 (i.e., H_0:b = 0 for population count data collected over *i* years), to use Monte Carlo simulations to calculate the statistical power of the test of H_0:b = 0 when H_0 was not rejected, and to estimate sample size requirements within and across years to detect a population trend at a specified power, Type I error, and coefficient of variation. To demonstrate this approach and illustrate important considerations when conducting power analysis, we analyzed 5 sets of shorebird count data collected by a single observer in the International Shorebird Survey, Marco River, Florida, in 1975 and 1980 to 1987. Our approach to determining statistical power in analysis of trends in population count data offers improvements over previously described methods because it is a straightforward approach to simultaneously evaluating the relationship between variance, sample size, effect size, alpha, and power, and it allows assessment over a range of sample sizes, providing a means for planning and evaluating sampling designs for trend tests at multiple levels of statistical precision.

Proc. Annu. Conf. Southeast. Assoc. Fish and Wildl. Agencies 50:294-302

The power (1-β) of a statistical test of the null hypothesis (H_0) is the probability that a false H_0 will be correctly rejected (Rotenberry and Wiens 1985, Cohen 1988). The underlying issues and principles of statistical power analysis have been reviewed by Cohen (1988), Peterman (1990*a*), and Cobb et al. (1994). Power analysis has received considerable attention in other fields, but wildlife scientists have usually overlooked its use in planning and evaluating research (Cobb et al. 1994). Cobb et al.'s (1994) review of the use of power analysis by wildlife scientists confirmed our suspicion that wildlife researchers/scientists do not regard power analysis as a critical component of experimental design or analysis. Researchers may be uninitiated in the use of power analysis (Peterman 1990*a*) or may not report power results because they are unacceptable (Cobb et al. 1994). Power analysis should be an integral component of experimental design and sound statistical analyses (Cohen 1988, Mangel 1993). Kraemer and Thiemann (1987), Cohen (1988), and Peterman (1990*a*) provided excellent discussions of power and sample size calculations for many statistical methods used by wildlife scientists. We describe a simulation-based approach to power analysis in tests of trends in population count data.

Population trend data typically are analyzed using normal-theory regression (Rawlings 1988) to test $H_0{:}b = 0$ or using a nonparametric approach (Titus et al. 1990). The analysis should include an estimate of the power of the test slope coefficient, particularly if H_0 is not rejected. Gerodette (1987) presented an approach for estimating the power of linear regression to detect a trend at a specified power. This approach was re-examined by Link and Hatfield (1990) and appears inappropriate for practical monitoring of populations when sample sizes are low (Gerrodette 1991) or data sets include multiple annual surveys. Kendall et al. (1992) described a method for estimating the power of tests to detect changes in low-density, dispersed animals using sign data recorded along survey transects.

The Florida Game and Fresh Water Fish Commission (FGFWFC) expends considerable financial and staff resources to conduct population surveys of numerous species of game and nongame wildlife. Only recently have FGFWFC staff begun questioning the statistical power of these monitoring programs and trend data. Our objective was to develop a practical tool for planning programs for monitoring populations and analyzing abundance data under conditions typically faced by management agencies. To address this need, we used PC-SAS^R to develop a stepwise approach for performing regression analyses to test $H_0{:}b = 0$ for population count data collected over *i* years, using Monte Carlo simulations to calculate the statistical power of the test of $H_0{:}b = 0$, and estimating samples size requirements within and across years to detect a population trend at a specified power and coefficient of variation (CV). We illustrate our approach with an example using International Shorebird Survey (ISS) data collected at Marco Island, Florida.

This work was supported through the Nongame Wildlife Trust Fund of the Florida Game and Fresh Water Fish Commission's Bureau of Nongame Wildlife. We appreciate reviews and comments on the manuscript from D. R. Eggeman, K. M. Enge, B. A. Harrington, P. Jodice, B. A. Millsap, M. R. Riggs, M. D. Samuel, and T. E. O'Meara. International Shorebird Survey data from Marco Island, Florida, were collected by T. Below and provided for use by B. A. Harrington, Manomet Observatory, Massachusetts.

Methods

Our approach to power analysis uses Monte Carlo simulation in PC-SAS® to model 2 parts of a 3-part regression and power analysis package (Fig. 1). Input are annually replicated count data from dBase IV or SAS data files.

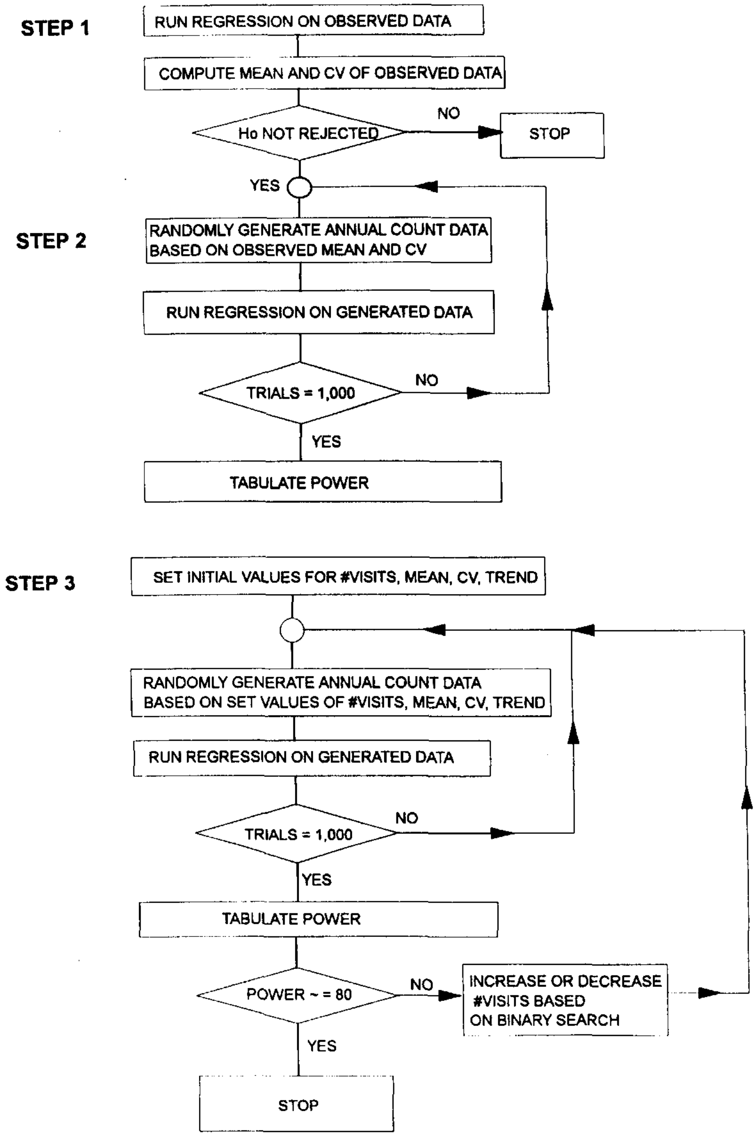

Figure 1. Flow diagram for Monte Carlo simulation of statistical power analysis.

Step 1. Conduct *N* annual population surveys for j years, resulting in an *N* x j matrix of count data. Assuming adherence to assumptions of least squares regression and adequate sample sizes (Madansky 1988, Rawlings 1988), data are modeled as $Y_i = \hat{a}+\hat{b}\, j_i+\varepsilon_i$, where Y_i is the mean of *N* survey counts in year *I*. A trend in observed field data is tested as H_o:b = 0 using standard least squares regression of Y on j. Annual mean, CV, standard deviation (SD), and sample size values are written to an output data set and an average CV and SD are calculated over all years.

Step 2. To calculate the power of non-significant results in Step 1, we use Monte Carlo simulations to generate sample data within limits specified by annual mean and SD values. For each year, values are used to generate annual survey count data sets with the same number of annual replicates as in the observed data set. In each of 1,000 iterations, a random deviate representing a mean generated count is derived for each year from a gamma distribution (Evans et al. 1993); hereby, the variation inherent in each annual survey is maintained. A standard regression analysis is run on each of the 1,000 iterations of generated data sets. H_0:b = 0 for each regression line is tested using a *t*-test statistic, based on the noncentral *t* distribution, calculated from the parameter (PARMEST) and standard error (STDERR) values in the variance-covariance matrix. The number of model iterations in which H_0:b = 0 is rejected (i.e., 1-β) is then counted at α = 0.1, 0.05, and 0.01 and output in tabular form.

Step 3. As a critical component of survey planning our final procedure is used to estimate the number of survey counts/year required to detect a specified trend from an initial population level over j years at a specified power (1-β). The procedure uses a binary search (Wirth 1976) to determine the number of surveys needed to achieve an approximate match (i.e., ± 2%) on the specified power. Because the solution space is from 1 to *q*, where *q* is the maximum number of surveys in a year, the possible number of macro iterations is $\log_2(q)$. Macro iterations continue until the number of surveys resulting in the specific power is determined. A table is printed with the number of surveys in each iteration and the resulting power. Use of the binary search procedure assumes that the potential solution space (i.e., min and max number of surveys) can be defined and that variation in power derived from 1,000 simulations is within 1 percentage point.

Generation of sample-population count data within the macro follows procedures in Step 2, but with the same CV and number of surveys for all years. Required variables for Step 3 include an initial estimate of the number of required surveys (VISITS), α (ALPHA) and 1-β (POWER), number of model iterations for generating sample data set (TRIALS), and definition of the minimum and maximum number of surveys. Input parameters include the population coefficient of variation (CVWITHIN), initial population mean (STARTMU), desired rate of annual change (+ or −) in the population (CHANGE), and years over which the population sampling is modeled (YEARS). The population CV may be approximated as the average CV for all years (see Step 1). An annual mean population value is calculated following an exponential model, $N_j = N_1(1 + r)^{j-1}$, where *r* is the rate of change. The annual mean population value and population CV are then used to calculate the gamma distribution variables for generation of a sample data set of count data for each year. Least-squares regression is conducted on the derived mean annual count, parameter estimates and standard error values are extracted from the variance-covariance matrix, the *t*-test statistic and associated probability level are calculated, and a power table is printed. Power tables in Step 3 include three categories: trials in which a change at the correct slope (+ or −) was detected at ≤ the specified α, trials in which a change at the incorrect slope was detected at ≤ the specified α, and trials in which no significant change was detected.

Example. To demonstrate the utility of our power analysis approach and important considerations when conducting power analysis we analyzed black-bellied plover (*Pluvialis squatarola*), dunlin (*Calidris alpina*), red knot (*C. canutus*), willet (*Catoptrophorus semipalmatus*), and total shorebird count data from the ISS, Marco River, Florida, in winters 1975 and 1980 to 1987. Analyses included evaluation of CVs, effect sizes (Cohen 1988), exponential rates of change and cumulative total change in numbers, standard least-squares regression analysis to test H_o:b = 0, analysis to determine the statistical power of nonsignificant trends, and the number of annual samples required to detect a + or − 5% annual trend, dependent upon results from Step 1, over 10 years with $1\text{-}\beta = 0.80$ at $\alpha = 0.05$. Effect size was expressed as f^2 (Cohen 1988) or SS_{model}/SS_{error}, the degree to which a trend was present in the population count data.

Results

The annual ranges in CVs were higher in red knot and dunlin count data (Table 1). Counts of black-bellied plovers, dunlins, red knots, willets, and total shorebirds exhibited incremental exponential rates of change (r) of 0.29, 0.28, 0.95, −0.08, and 0.22, respectively (Table 2). Increases in counts of black-bellied plovers, red knots, and total shorebirds were significant ($P < 0.015$); increases in counts of dunlins and willets were not significant ($P > 0.2$). Population CVs for dunlins and willets were high, with low indices of effect size.

At $\alpha = 0.1$, 0.05, and 0.01, the power to find a significant trend in regression on the dunlin ($r = 0.28$) and willet ($r = -0.08$) counts (i.e., results from Step 2) was 15%, 4%, and 0%, and 7%, 3%, and 0%, respectively. Based on the observed trends and variance in dunlin and willet counts, we determined the number of annual surveys

Table 1. Mean number of black-bellied plovers (BBPL), dunlins (DUNL), red knots (REKN), willets (WILL), total shorebirds (total), and associated coefficients of variation (CV) recorded in the International Shorebird Survey, Marco River, Florida, winters 1975 and 1980 to 1987.

	BBPL		DUNL		REKN		WILL		Total		
Year	$\bar{x}$	CV	$\bar{x}$	CV	$\bar{x}$	CV	$\bar{x}$	CV	$\bar{x}$	CV	N
1975	10		30		0		60		283		1
1980	29	60.0	133	66.2	17	86.1	38	61.1	1129	60.0	9
1981	38	31.2	125	76.0	85	50.5	56	38.9	939	26.9	6
1982	25	42.9	45	134.7	9	204.7	36	90.0	809	71.2	5
1983	34	14.7	110	44.3	32	120.3	35	57.7	818	24.2	6
1984	57	40.5	180	98.2	7	142.2	68	106.2	1291	59.9	6
1985	58	40.4	113	55.6	46	164.2	46	58.9	1716	61.6	7
1986	63	32.7	125	79.8	276	64.3	28	31.0	1520	22.4	6
1987	78	66.3	214	147.2	208	140.2	31	84.9	1411	66.6	6

required to detect a +5% and −5% annual change over 10 years at $\alpha = 0.05$ and $1 - \beta = 0.80$, respectively. An annual winter sampling intensity of 40 and 21 surveys would be required to detect the specified trends at $\alpha = 0.05$ and $1 - \beta = 0.80$, respectively (i.e., results from Step 3), for dunlin and willets. At $\alpha = 0.2$, an increase from 5% to 20% in the probability of our analyses indicating a statistically significant trend when no trend actually exists (i.e., Type I error), an annual sampling intensity of 20 and 10 surveys would be required to detect the specified trends at $1 - \beta = 0.80$.

Discussion

Power is dependent on the interaction between variance, as reflected in the CV, alpha, effect size, and the sample size (Cohen 1973, 1988; Bernstein and Zalinski 1983; Toft and Shea 1983; O'Brien and Lohr 1984; Rotenberry and Wiens 1985; Kraemer and Thiemann 1987; Peterman 1990*a*; Cobb et al. 1994). In general, statistical power is positively related to alpha, effect size, and sample sizes, and inversely related to the CV. Our analysis of shorebird data from Marco Island illustrates several relationships among these variables when testing for trends in population count data.

For conservation purposes, we considered the observed trends in counts of black-bellied plovers, dunlins, red knots, and total shorebirds to be of a magnitude important to managers (Table 2). We considered all of these annual changes to be large (i.e., > 20%), but only the trend in counts of dunlins was non-significant. The black-bellied plover data had a small population CV, especially for shorebird count data, and a large effect size index. Red knot data had a large population CV and a large effect size index. Total shorebird count data had a medium size population CV (i.e., ≈ 50%) and a relatively large effect size. The indication that observed trends in these counts were significant, therefore, was not surprising.

At first inspection of the data, we believed that the observed 28% annual increase

Table 2. Estimates of parameters of change in the mean number of black-bellied plovers (BBPL), dunlins (DUNL), red knots (REKN), willets (WILL), and total shorebirds (total) counted during the International Shorebird Survey, at Marco River, Florida, in winters 1975 and 1980 to 1987.

Species	CVWITHIN[1]	P	ES[2]	% Change[3]	r[4]	$1-\beta$[5]
BBPL	41	0.0001	0.50	+710	+0.29	
DUNL	88	0.2240	0.03	+188	+0.28	0.04
REKN	122	0.0018	0.22	+16,900	+0.95	
WILL	66	0.4493	0.01	−30	−0.08	0.03
Total	49	0.0126	0.13	+252	+0.22	

[1]CVWITHIN = mean of coefficients of variation in annual samples expressed as a percentage.

[2]ES = f^2 = effect size calculated as (SS_{model}/SS_{error}) with a range from 0–1.

[3]Overall percentage change in abundance from 1975 to 1987 calculated from regression equation.

[4]r = observed annual exponential rate of change.

[5]Power calculated for only non-significant results at $\alpha = 0.05$.

in counts of dunlin was of a magnitude that should have been statistically significant in an ISS survey (Howe et al. 1989); however, our analysis showed that detection of a trend with a population CV = 88% and a small effect size index (0.03) was difficult. Thus, even with a mean annual sample size of ≈6 surveys, a trend in dunlin counts of 28% could not be detected over 13 years at $\alpha = 0.05$. Average annual sampling effort needed to be 5.3 times and 3.3 times greater to detect the observed trend at $\alpha =$ 0.05 and 0.2, respectively. The results exemplify the serious implication that sample variance has for population monitoring efforts. In this case, the required sampling intensity is likely prohibitive for long-term monitoring by conservation agencies responsible for a multitude of species and species groups.

Results from analyses of the willet count data also are not quantitatively surprising, but demonstrate serious implications for conservation monitoring. While the effect size index was relatively small (0.02, Table 2), the population CV was large (66%). With a mean annual sampling effort of ≈ 6 surveys and a population CV of 66%, we were not surprised at the lack of a significant trend. We believe that a 30% decline in willet use of Marco Island over 13 years should be detectable if it actually occurred. Based on these data, managers could not conclude that the trend indicated by the data actually occurred, and any management actions based on these data would be unfounded. To collect data that allow testing with sufficient power upon which to base management decisions, minimum average annual sampling effort should have been 21 and 10 surveys to detect the observed trend at $\alpha = 0.05$ and 0.2, respectively. If managers are willing to increase their Type I error rate to 20%, sufficient sampling seems feasible.

As suggested by Peterman (1990*a*, 1990*b*) and Cobb et al. (1994), we believe scientists monitoring trends in wildlife populations must rigorously design monitoring efforts to have sufficient power to distinguish among alternative explanations of their data, conduct pilot studies to estimate CVs before long-term projects are initiated, estimate the power of sampling approaches used and examine through simulation the effects of imprecision and inaccuracies, abandon the arbitrary tradition of setting $\alpha =$ 0.05 and consider desired power and acceptable β when setting α, calculate power for several effect sizes (including that obtained in the original analysis when H_o is not rejected), use 80% as an initial objective for power analysis, report the biological ramification of Type II error, and not assert that a trend does not exist when an analysis fails to reject H_o and $1\text{-}\beta$ is low. Conservation agencies should balance the objective of not missing a trend against the object of having a high level of certainty in trends that are detected. In managing declining species, we suggest that it is biologically safer to be conservative (i.e., increased power and a higher alpha).

We believe that our approach to determining statistical power in analysis of trends in population count data using Monte Carlo simulations addresses many of these issues and offers several important improvements over previously described methods. First, the analysis can provide biologically significant results by ensuring managers that population monitoring surveys are designed to detect biologically significant changes in abundance. Secondly, use of the model requires the researcher to have an in-depth knowledge of the variance components in each data set. We believe the importance of inter-observer, inter-year, inter-species, and other sources of variance in count data is underestimated by wildlife researchers. For locally transient or flocking species this underestimation may lead to overly optimistic planning when using population counts to detect trends. Thirdly, the assumptions and mathematical algorithms in our model hold along the broad range from low (i.e., ≥3) to high number of years and from low to high number of surveys per year (i.e., ≥2). Finally, the model provides in one package all the analytical tools needed to plan and evaluate population sampling designed to test for trends at multiple levels of statistical precision, and it is a straightforward approach to simultaneously evaluating the relationships between variance, sample size, effect size, alpha, and power. This package is available from the senior author upon request or can be accessed on the World Wide Web at **http://www.freenet.fsu.edu/~sprandel.**

Literature Cited

Bernstein, B. B. and J. Zalinski. 1983. An optimum sampling design and power tests for environmental biologists. J. Environ. Manage. 16:35–43.

Cobb, D. T., G. L. Sprandel, and M. Williams. 1994. A review of power analysis in assessing impacts to wildlife populations. Proc. Symp., issues and tech. in the manage. impacted wildl. 6:153–158.

Cohen, J. 1973. Statistical power analysis and research results. Am. Educ. Res. J. 10:225–230.

———. 1988. Statistical power analysis for the behavioral sciences, 2nd ed. Lawrence Erlbaum Assoc. Hillsdale, N.J. 567pp.

Evans, M., N. Hastings, and B. Peacock. 1993. Statistical distributions, 2nd ed. John Wiley and Sons, Inc. N.Y. 170pp.

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

———. 1991. Models for power of detecting trends: a reply to Link and Hatfield. Ecology 72:1889–1892.

Howe, M. A., P. H. Geissler, and B. A. Harrington. 1989. Population trends of North American shorebirds based on the International Shorebird Survey. Biol. Conserv. 49:185–199.

Kendall, K. C., L. H. Metzgar, D. A. Patterson, and B. M. Steele. 1992. Power of sign surveys to monitor population trends. Ecol. Appl. 2:422–430.

Kraemer, H. C. and S. Thiemann. 1987. How many subjects? Statistical power analysis in research. Sage Publ., Newbury Park, Calif. 120pp.

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

Madansky, A. 1988. Prescriptions for working statisticians. Springer-Verlag, New York, N.Y. 295pp.

Mangel, M. 1993. Effects of high-seas driftnet fisheries on the northern right whale dolphin *Lissodelphis borealis*. Ecol. Appl. 3:221–229.

O'Brien, R. G. and V. I. Lohr. 1984. Power analysis for linear models: the time has come. SAS Users Group Internatl. Conf. Proc. 9:840–846.

Peterman, R. M. 1990*a*. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47:2–15.

———. 1990*b*. The importance of reporting statistical power: the forest decline and acidic deposition example. Ecology 71:2024–2027.

Rawlings, J. O. 1988. Applied regression analysis: a research tool. Wadsworth and Brooks/Cole, Pacific Grove, Calif. 553pp.

Rotenberry, J. T. and J. A. Wiens. 1985. Statistical power analysis and community-wide patterns. Am. Nat. 125:164–168.

Titus, K., M. R. Fuller, and D. Jacobs. 1990. Detecting trends in hawk migration data. Pages 105–113 *in* J. R. Sauer and S. Droge, eds. Survey design and statistical methods for the estimation of avian population trends. U.S. Fish and Wildl. Serv., Biol. Rep. 90(1).

Toft, C. A. and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical power. Am. Nat. 122:618–625.

Wirth, N. 1976. Algorithms + data structures = programs. Prentice-Hall Inc., Englewood Cliffs, N.J. 366pp.