

A POWER ANALYSIS FOR DETECTING TRENDS¹

TIM GERRODETTE

Southwest Fisheries Center Honolulu Laboratory, National Marine Fisheries Service, NOAA,
 Honolulu, Hawaii 96822-2396 USA

Abstract. A power analysis allows estimation of the probability of detecting upward or downward trends in abundance using linear regression, given number of samples and estimates of sample variability and rate of change. Alternatively, the minimum number or precision of samples required to detect trends with a given degree of confidence can be computed. The results are applicable to an experimental situation in which samples are taken at regular intervals in time or space. The effects of linear and exponential change and of having sample variability be a function of abundance are investigated. Results are summarized graphically and, as an example, applied to the monitoring of the California sea otter population with aerial surveys.

Key words: experimental design; linear regression; power analysis; trends.

INTRODUCTION

A common proximate goal in ecological research is to determine whether the magnitude of some quantity is tending to increase or decrease. The quantity of interest might be population size, productivity, diversity, or mortality rate, for example. A linear regression of estimated abundance of the quantity against time or distance is commonly used to evaluate such a trend. When designing a program to detect trends, several related questions often arise. What is a sufficient number of samples? How precise must the samples be? What is the probability of detecting a trend if it is present?

This paper provides answers to these questions in terms which, it is hoped, are applicable to a wide range of ecological studies. The effects of several different factors are investigated, including the precision of the estimates of abundance, the dependence of precision on changes in abundance, the nature and magnitude of the actual rate of change in abundance, the asymmetry between upward and downward trends, and the levels of Type 1 and Type 2 statistical errors. Results are presented in both numerical and graphical form.

The ability of a statistical procedure to distinguish a situation different from the null hypothesis is called the power of that procedure. The estimation of statistical power is important for several reasons. Epistemologically, power analysis is important for the interpretation of results when the null hypothesis is not rejected (Toft and Shea 1983). Operationally, power analysis is important during the planning of experiments to avoid wasted time and effort on a program that is unlikely to yield useful information. Peterman and Routledge (1983), for example, showed that a proposed 48 million smolt/yr release would be insufficient to gain new information about the smolt-adult salmon

relationship, while an alternative 88 million smolt experiment had a good chance of succeeding. Holt et al. (1987) used a power analysis to optimize the design of a large-scale program to detect changes in dolphin stocks over vast areas of the eastern tropical Pacific Ocean.

METHODS

By definition, a trend is detected when the regression has a slope significantly different from zero. The conclusion that a trend in abundance is occurring, when in fact it is not, is termed a Type 1 error, while the conclusion that no trend is occurring, when in fact it is, is a Type 2 error. The probabilities of making Type 1 and 2 errors are labelled α and β , respectively. Power is defined as $1 - \beta$.

I consider two simple models of change, linear and exponential. Let A_i represent abundance as a function of i , an index of time or distance. The linear model is

$$A_i = A_1[1 + r(i - 1)]. \quad (1)$$

For each increment in i , abundance increases by a constant absolute amount rA_1 . The parameter r thus expresses this constant increment of change as a fraction of the initial abundance A_1 . The exponential model is

$$A_i = A_1(1 + r)^{i-1}, \quad (2)$$

where r is the finite fractional rate of change per time or distance unit. A plot of A_i against i is an exponentially increasing (or decreasing) set of points; a plot of $\ln A_i$ against i is a straight line with slope $\ln(1 + r)$, the instantaneous rate of change.

Each abundance A_i is estimated by the sample abundance estimate \hat{A}_i . For the linear model (Eq. 1),

$$\hat{A}_i = A_i + \epsilon_i,$$

while for the exponential model (Eq. 2),

$$\ln \hat{A}_i = \ln A_i + \epsilon_i,$$

where the ϵ_i are normal, independent random variables

¹ Manuscript received 16 May 1986; accepted 22 August 1986; final version received 5 January 1987.

TABLE 1. Theoretical dependence of the precision of an abundance estimate, as measured by the coefficient of variation (CV), on abundance (A), for some common methods of estimating abundance.

Method	Assumptions	Dependence	Reference
Quadrats/strip transects	random	$cv \propto 1/\sqrt{A}$	Seber (1982:22)
	clumped	$cv \propto 1/\sqrt{A}$	Seber (1982:25)
Line transects	random or clumped	$cv \propto 1/\sqrt{A}$	Burnham et al. (1980)
Distance sampling	random	cv constant*	Moore (1954)
	clumped	cv constant	Eberhardt (1967)
Catch per unit effort (CPUE)	random	$cv \propto 1/\sqrt{A}$	de la Mare (1984)
Single mark-recapture	Petersen method assumptions	$cv \propto \sqrt{A}$	Seber (1982:60)

* "cv constant" means the coefficient of variation does not depend on abundance, although it does depend on other parameters.

with mean zero. For the exponential model, note that this implies that the \hat{A}_i are lognormally distributed. I consider three different assumptions about how the variance of the abundance estimate could be related to abundance: $\text{Var}(\hat{A}_i)$ is proportional to A_i , to A_i^2 , or to A_i^3 . These relations can also be expressed in terms of the coefficient of variation, the ratio of the standard deviation to the mean: $cv(\hat{A}_i)$ is proportional to $1/\sqrt{A_i}$, is constant with respect to A_i , or is proportional to $\sqrt{A_i}$, respectively. The nature of the dependence of $cv(\hat{A}_i)$ on A_i will depend on the species, on the quantity whose abundance is of interest, and on the method used to estimate that quantity. The three assumptions made above span a range of functional relationships, and correspond to the theoretical dependences of $cv(\hat{A}_i)$ on A_i for common methods of estimating abundance (Table 1).

Let n be the number of points used in the regression (the number of samples or estimates of abundance). In the case of linear change, the regression line is fitted to the points $\{y_i\} = \{A_1, A_2, \dots, A_n\}$, whereas for exponential change it is fitted to the points $\{y_i\} = \{\ln A_1, \ln A_2, \dots, \ln A_n\}$. Let

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

be the usual estimator of the slope of the regression line. Then \hat{b} is normally distributed with mean $\mu_b = b$ and variance $\sigma_b^2 = \sigma_{res}^2/n\sigma_x^2$, where b is the true slope of the regression, σ_{res}^2 is the variance of the residuals and σ_x^2 is the variance in the independent variable x (Freund 1962:316). To reduce statistical errors to the specified levels α and β , we must have

$$|\mu_b| - z_\beta\sigma_b \geq z_{\alpha/2}\sigma_b$$

or

$$b^2 n \sigma_x^2 \geq (z_{\alpha/2} + z_\beta)^2 \sigma_{res}^2, \tag{3}$$

where z_α is the value of a standardized random normal variable such that the area under one tail of the prob-

ability density function beyond z_α is α . If the null hypothesis is one-tailed, inequality 3 is modified to

$$b^2 n \sigma_x^2 \geq (z_\alpha + z_\beta)^2 \sigma_{res}^2. \tag{4}$$

To relate these equations to the problem of detecting trends, we need to express b , σ_x^2 , and σ_{res}^2 in terms of r , cv , and n .

For the linear model (Eq. 1), the slope of the regression line is

$$b = \frac{dA_i}{dt} = rA_i. \tag{5}$$

If the estimates of abundance are taken at regular intervals in time or space, the independent variable x can, without loss of generality, be renumbered 1, 2, 3, ..., n . Then

$$\sigma_x^2 = \frac{(n+1)(n-1)}{12}. \tag{6}$$

The variance of the residuals about the regression line depends on how $cv(\hat{A}_i)$ changes with A_i . Let cv_1 be the coefficient of variation at the initial abundance A_1 . If $cv(\hat{A}_i)$ is proportional to $1/\sqrt{A_i}$, an estimate of the residual variance is

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\hat{A}_i) \\ &= \frac{1}{n} \sum cv_1^2 A_1 A_i \\ &= \frac{cv_1^2 A_1^2}{n} \sum [1 + r(i-1)] \\ &= (cv_1 A_1)^2 \left[1 + \frac{r}{2}(n-1) \right]. \end{aligned} \tag{7}$$

Substituting Eqs. 5, 6, and 7 into Eq. 3,

$$\begin{aligned} r^2 n(n-1)(n+1) &\geq 12 cv_1^2 (z_{\alpha/2} + z_\beta)^2 \\ &\quad \cdot \left[1 + \frac{r}{2}(n-1) \right]. \end{aligned} \tag{8}$$

If $cv(\hat{A}_i)$ is constant over the sampled range of A_i , write cv without the subscript. Then

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\hat{A}_i) \\ &= \frac{1}{n} \sum cv^2 A_i^2 \\ &= \frac{cv^2 A_1^2}{n} \sum [1 + r(i-1)]^2 \\ &= (cv A_1)^2 \left\{ 1 + r(n-1) \right. \\ &\quad \left. \cdot \left[1 + \frac{r}{6}(2n-1) \right] \right\}. \end{aligned} \tag{9}$$

Substituting Eqs. 5, 6, and 9 into Eq. 3,

$$\begin{aligned} r^2 n(n-1)(n+1) &\geq 12cv^2(z_{\alpha/2} + z_\beta)^2 \\ &\cdot \left\{ 1 + r(n-1) \left[1 + \frac{r}{6}(2n-1) \right] \right\}. \end{aligned} \tag{10}$$

If $cv(\hat{A}_i)$ is proportional to $\sqrt{A_i}$,

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\hat{A}_i) \\ &= \frac{1}{n} \sum \frac{cv_1^2}{A_i} A_i^2 \\ &= \frac{cv_1^2 A_1^2}{n} \sum [1 + r(i-1)]^2 \\ &= (cv_1 A_1)^2 \left\{ 1 + \frac{3r}{2}(n-1) \right. \\ &\quad \left. \cdot \left[1 + \frac{r}{3}(2n-1) + \frac{r^2}{6}n(n-1) \right] \right\}. \end{aligned} \tag{11}$$

Substituting Eqs. 5, 6, and 11 into Eq. 3,

$$\begin{aligned} r^2 n(n-1)(n+1) &\geq 12cv_1^2(z_{\alpha/2} + z_\beta)^2 \\ &\cdot \left\{ 1 + \frac{3r}{2}(n-1) \left[1 + \frac{r}{3}(2n-1) \right. \right. \\ &\quad \left. \left. + \frac{r^2}{6}n(n-1) \right] \right\}. \end{aligned} \tag{12}$$

Next consider the exponential model (Eq. 2). The slope of the regression line is

$$b = \frac{d \ln A_i}{di} = \ln(1+r), \tag{13}$$

while the variance of x remains as before. Since the abundance estimates are log-transformed, we make use of the identity

$$\text{Var}(\ln y) = \ln \left[\frac{\text{Var}(y)}{[E(y)]^2} + 1 \right],$$

where y is lognormally distributed, which identity can be derived from the moment generating function of

the normal distribution. If $cv(\hat{A}_i)$ is proportional to $1/\sqrt{A_i}$,

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\ln \hat{A}_i) \\ &= \frac{1}{n} \sum \ln \left\{ \frac{\text{Var}(\hat{A}_i)}{[E(\hat{A}_i)]^2} + 1 \right\} \\ &= \frac{1}{n} \sum \ln \left[\frac{cv_1^2}{(1+r)^{i-1}} + 1 \right]. \end{aligned} \tag{14}$$

Substituting Eqs. 6, 13, and 14 into inequality 3 gives

$$\begin{aligned} [\ln(1+r)]^2 n(n-1)(n+1) \\ \geq 12(z_{\alpha/2} + z_\beta)^2 \left\{ \frac{1}{n} \sum \ln \left[\frac{cv_1^2}{(1+r)^{i-1}} + 1 \right] \right\}. \end{aligned} \tag{15}$$

If $cv(\hat{A}_i)$ is constant with respect to A_i , we write cv without the subscript as before, and

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\ln \hat{A}_i) \\ &= \ln(cv^2 + 1) \end{aligned} \tag{16}$$

Note that this is the only case in which abundance estimates are truly homoscedastic. Substituting Eqs. 6, 13, and 16 into inequality 3,

$$\begin{aligned} [\ln(1+r)]^2 n(n-1)(n+1) \\ \geq 12(z_{\alpha/2} + z_\beta)^2 [\ln(cv^2 + 1)]. \end{aligned} \tag{17}$$

Finally, if $cv(\hat{A}_i)$ is proportional to $\sqrt{A_i}$,

$$\begin{aligned} \hat{\sigma}_{res}^2 &= \frac{1}{n} \sum \text{Var}(\ln \hat{A}_i) \\ &= \frac{1}{n} \sum \ln [cv_1^2(1+r)^{i-1} + 1]. \end{aligned} \tag{18}$$

Substituting Eqs. 6, 13, and 18 into inequality 3 gives

$$\begin{aligned} [\ln(1+r)]^2 n(n-1)(n+1) \\ \geq 12(z_{\alpha/2} + z_\beta)^2 \\ \cdot \left\{ \frac{1}{n} \sum \ln [cv_1^2(1+r)^{i-1} + 1] \right\}. \end{aligned} \tag{19}$$

RESULTS

In this analysis, the detection of a trend has five parameters: n , the number of samples; r , the rate of change of the quantity being measured; cv , the coefficient of variation, a measure of precision; and α and β , the probabilities of Type 1 and 2 errors. Eqs. 8, 10, 12, 15, 17, and 19 relate these five parameters under different assumed models of change (linear or exponential) and different models of dependence of sample precision on abundance (cv constant, proportional to $\sqrt{A_i}$, or proportional to $1/\sqrt{A_i}$). Given any four of the parameters, the fifth can be found. If we are interested

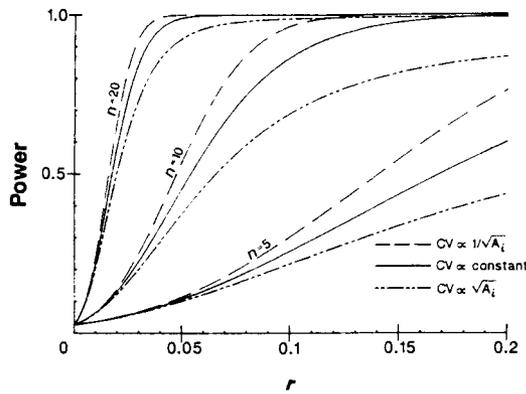


FIG. 1. Power of linear regression as a function of r , n , and the dependence of $cv(A_i)$ on A_i . Curves shown use the linear model with $cv_1 = 0.2$ and $\alpha = 0.05$.

in the power of a certain procedure, for example, we solve for z_β , find β , and then compute power as $1 - \beta$.

In most cases, explicit solutions for r , n , or cv are not possible; iterative solutions on a computer are required. However, approximate solutions can be carried out on a hand calculator. For example, using the approximations

$$n(n - 1)(n + 1) \approx n^3$$

for moderately large integer n , and

$$\ln(1 + a) \approx a$$

for small a , Eq. 17 becomes

$$r^2 n^3 \geq 12cv^2(z_{\alpha/2} + z_\beta)^2, \quad (20)$$

which can easily be explicitly solved for any of the five parameters. In fact, for small to moderate values of r , n , and cv , Eq. 20 serves as a useful approximation for any of the equations. For the common case where $\alpha = \beta = .05$, $(z_{\alpha/2} + z_\beta)^2 = (1.960 + 1.645)^2 = 13.0$, and an even simpler form useful for quickly relating r , n , and cv is

$$r^2 n^3 \geq 156cv^2.$$

Figs. 1-3 present some of the detailed results graphically. In general, the figures show that power increases as r increases in absolute value (a stronger trend), as n increases (more samples), and as cv decreases (greater precision).

Fig. 1 shows how power is affected by the dependence of estimate precision on abundance. There are three sets of curves, for $n = 5, 10$, and 20 . Power increases as n , the number of points used in the regression, increases. Within each set are three curves corresponding to the three assumptions of dependence of variability on abundance. There can be substantial differences in power among the different models, and these differ-

ences increase with cv (here set at 0.2). Within each set, power is highest—a trend is easiest to detect—when $cv(A_i)$ is proportional to $1/\sqrt{A_i}$, but this order is reversed when $r < 0$, that is, when detecting a declining trend. Fig. 1 suggests that, at least for some combinations of parameters, it is important to know how precision will change as a function of abundance if power is to be correctly calculated. The three models explored here certainly do not exhaust possible functions, but they do span a range of possibilities. The choice of the most appropriate model of dependence may be made on theoretical (e.g., Table 1) or on empirical (e.g., Holt et al. 1987) grounds.

Fig. 2 contrasts linear and exponential models of change and the asymmetry between detecting increasing and decreasing trends. As in Fig. 1, there are three sets of curves corresponding to $n = 5, 10$, and 20 . Within each set are four curves representing linear and exponential models, for increasing and decreasing trends. The power curves for exponential trends lie between the power curves for linear trends. In other words, between two trends that begin with the same amount of change, it is easier to detect an increasing trend in a quantity changing by constant proportional amounts than in one changing by constant absolute amounts, but the opposite holds for decreasing trends. Regardless of whether change is linear or exponential, decreasing trends are easier to detect than increasing ones, and the difference is particularly marked for linear change. For a linear decrease, note that abundance A_i becomes 0 for $i \geq 1 + 1/(-r)$. A trend will not necessarily be detected before A_i becomes 0 if the coefficient of variation is high.

Fig. 3 shows the effects of estimate precision on power, and contrasts one- and two-tailed tests of significance. The five pairs of curves, one for each value of cv , show that power decreases rapidly as the coefficient

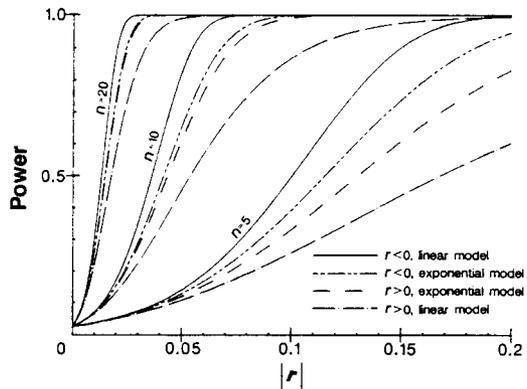


FIG. 2. Power of linear regression as a function of n , r (either >0 or <0), and linear vs. exponential models of change. Curves shown use the cv constant model with $cv = 0.2$ and $\alpha = .05$.

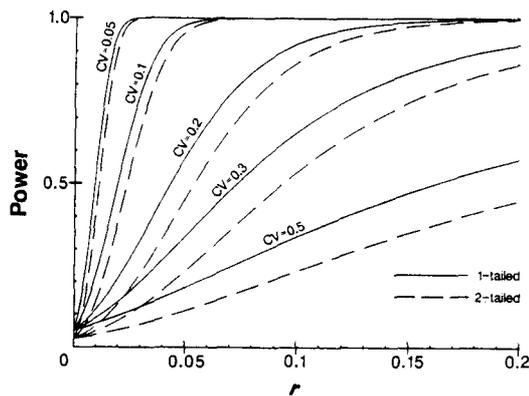


FIG. 3. Power of linear regression as a function of r , CV , and 1- vs. 2-tailed tests. Curves shown use the linear model with CV constant, $n = 10$, and $\alpha = .05$.

of variation increases; thus, the ability to detect a trend depends greatly on how precisely the quantity of interest can be measured. The solid lines show the increase in power if the null hypothesis is one-sided—in this case, that the trend is not increasing. Of course, this calculation of power applies only to trends that do in fact differ in the direction of the one-sided alternative hypothesis. A one-tailed test has no power to detect differences in the opposite direction. More concretely, if the one-sided null hypothesis H_0 : “the trend is not increasing” is tested against the alternative H_1 : “the trend is increasing,” it is not possible to conclude that the trend is decreasing.

With minor reparameterization the results can be applied to some related problems of detecting trends. For example, the trend may be stated as an overall change in abundance of a given magnitude rather than as a rate of change per interval. That is, our question may be something like “What are the chances of detecting a trend if there is an 80% change in abundance between the first and last samples?” The results of this paper can be used if the 80% change is occurring uniformly between the first and last samples. Let R be the overall fractional change in abundance, and n the number of samples, as before. Then calculate r as

$$r = \frac{R}{n-1} \quad (21)$$

for linear change, and as

$$r = (R + 1)^{1/(n-1)} - 1 \quad (22)$$

for exponential change. This is the rate of change r , per unit time or distance, that will lead to the specified overall change. Another related problem is a null hypothesis that the rate of change is some nonzero r^* . The equations and figures above still apply if r is replaced by $(r - r^*)$. The $100(1 - \alpha)$ percent confidence limits for an observed trend may be calculated as $r \pm z_{\alpha/2} SE_r$, where the standard error is estimated by

$$SE_r = \sqrt{\frac{12\hat{\sigma}_{res}^2}{n(n+1)(n-1)}}$$

and $\hat{\sigma}_{res}^2$ is calculated from Eq. 7, 9, 11, 14, 16, or 18, depending on choice of model.

If there is more than one sample taken per time or distance interval, the variance of the mean will be reduced. The coefficient of variation of the mean of m independent replicate estimates is CV/\sqrt{m} , where CV is, as above, the coefficient of variation of a single estimate. If estimates are taken close together in time or space, however, they may not be independent, and CV/\sqrt{m} may be an underestimate. Correction for autocorrelation can in many cases be made by assuming the first-order autoregressive model

$$\epsilon_i = \rho\epsilon_{i-1} + \delta_i,$$

where $|\rho| < 1$ and the δ_i are independent random normal variables with mean zero and constant variance σ^2 . Under this model, the observed variance of the error terms is

$$\sigma^2(\epsilon_i) = \frac{\sigma^2}{1 - \rho^2},$$

and estimation of CV may require a corresponding adjustment. Analysis of autocorrelated data requires modified regression techniques not discussed in this paper (see, for example, Neter et al. 1983).

NUMERICAL EXAMPLE

After being reduced to near extinction at the beginning of this century, the sea otter (*Enhydra lutris*) population in central California has gradually recovered. As the otter population has grown, so has pressure to manage the species. Otters strongly influence the composition of the nearshore community (Estes et al. 1978, VanBlaricom and Estes 1987) and, more particularly, greatly reduce the abundance of species highly valued by humans (lobster, crab, and abalone). There is considerable controversy about how to manage these marine mammals, but an important factor to consider in any management scheme is whether the population is currently increasing, decreasing, or remaining stable.

In order to judge the feasibility of monitoring sea otter population trends using aircraft, the United States Fish and Wildlife Service conducted a series of seven replicated strip transects during the winter of 1981–1982 to determine the precision of aerial counts. On the basis of these replicates, the coefficient of variation of a single count is estimated to be 0.13 (J. A. Estes, *personal communication*). The coefficient of variation of strip transects is proportional to the inverse of the square root of abundance (Table 1), and population growth is likely to be exponential. For these reasons, Eq. 15 is selected as the most appropriate equation to analyze the power of detecting future changes in population size. It is assumed that the data will be analyzed at the $\alpha = .05$ significance level (two-tailed).

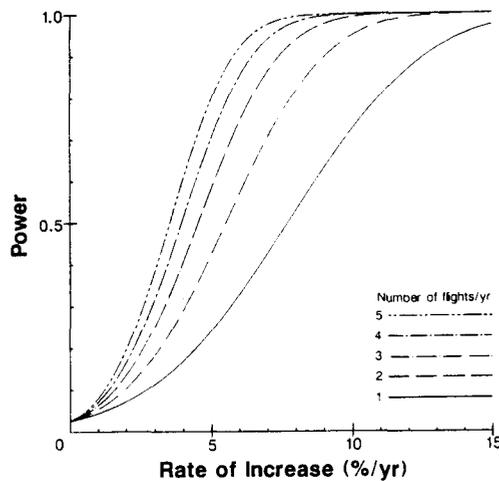


FIG. 4. Power curves for detecting various rates of annual increase in population size of sea otters in central California using five annual aerial surveys. More flights per year permit more precise estimates of population size, hence greater power to detect a given rate of increase.

Suppose a 5-yr monitoring program is contemplated. Fig. 4 shows power curves for detecting various rates of annual increase with different numbers of flights per year, under the assumption that the variance of the mean count is reduced in proportion to the number of flights. If a single flight is made each year, for example, the probability of detecting a 10%/yr increase in population size is 0.72. With 2 flights/yr, the coefficient of variation of the mean count is $0.13/\sqrt{2} = 0.092$, and the power increases to 0.95. Suppose we wish to know the probability of detecting a 30% increase in population size at the end of 5 yr. From Eq. 22, with $R = 0.3$ and $n = 5$, we calculate $r = 0.068$. With 1 flight/yr, the probability is only 0.4 of detecting the 30% increase, assuming the increase occurred smoothly over 5 yr. The probability of detecting a decrease of the same size is only slightly higher (cf. Fig. 2).

Let us ask how precise our samples must be to detect this 6.8%/yr increase in population size. To reduce Type 2 error to 0.05 or less (that is, to detect the trend in abundance with 95% confidence if it is occurring), we calculate that cv must be 0.061 or less. To achieve this precision by replication alone would require $(0.13/0.061)^2 = 4.5$, or 5 flights/yr. Precision might also be increased in other ways, such as careful training of observers, flying only under optimal sighting conditions, etc.

Next suppose that the time period is not fixed, but that we wish to know how many years will be required to detect a trend. Fig. 5 summarizes the situation if we again assume that Type 2 error must be reduced to 0.05 or less. For example, with 1 flight/yr and an annual rate of increase of 6.8%, the graph shows that $n = 7.9$. Thus, eight annual surveys (over a period of 7 yr) are

required to detect the trend under the conditions given. Note that, because the number of annual surveys can only be a whole number, the ordinate in Fig. 5 should be read as the next largest integer. The number of annual surveys required to detect a trend rises sharply as the annual rate of change becomes small. Annual rates of increase in the otter population of $< 2\%/yr$ would be very difficult to detect using aerial surveys.

We might also consider whether annual estimates are the optimal frequency of sampling. If population size is changing slowly, it might be better to conduct surveys only every 2nd or 3rd yr. As the interval between surveys increases, the effective rate of change per interval increases, and the number of surveys therefore decreases. Table 2 shows the number of surveys required for different intervals between surveys, assuming a 5%/yr increase in population size and 2 flights/yr. The number of surveys required, and therefore the number and cost of flights, could be reduced to half by conducting surveys once every 3 yr instead of annually in this example. But Table 2 also shows that the number of years which will have elapsed by the time the trend is detected increases from 7 to 9 years if surveys are conducted only once every 3 yr. Moreover, the total increase in population size which will have occurred by the time the trend is detected changes from 41% to 55%. How these less tangible costs, including the cost of ignorance of population size two years out of three, are weighed against the savings in actual flight costs depends on the research goals. For example, Allen (1980) pointed out that, when dealing with possible declines in populations of species which are already rare or endangered, the total percentage change which

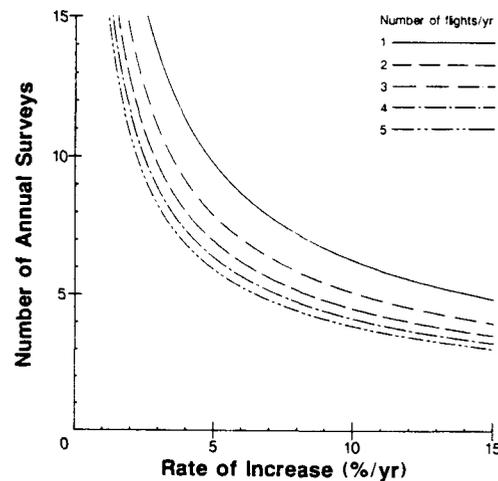


FIG. 5. Minimum number of annual surveys required to detect various rates of annual increase in population size of California sea otters. More flights permit more precise estimates of population size during each survey, hence fewer annual surveys required.

TABLE 2. Effects of different survey intervals on number of surveys required to detect a 5%/yr increase in sea otter population size. Also shown are several other quantities relevant to deciding on an optimal frequency of sampling this population.

No. yr between surveys (<i>t</i>)	No. surveys required* (<i>n</i>)	Effective % change/ interval ($1.05^t - 1$)	No. yr to detection† [$t(n - 1)$]	Total % change‡ [$1.05^{t(n-1)} - 1$]
1	8	5.0	7	41
2	5	10.3	8	48
3	4	15.8	9	55
4	4	21.6	12	80
5	3	27.6	10	63

* The required number of surveys is calculated from Eq. 15, assuming 2 flights/yr.

† The number of years that will have elapsed by the time the trend is detected.

‡ The total percentage change that will have occurred by the time the trend is detected.

will have occurred by the time a significant trend is detected may be the most critical factor to consider.

DISCUSSION

Statistical power is the probability that an analysis will reject a null hypothesis which is, indeed, false. Thus power is calculated as $1 - \beta$, where β is the probability of Type 2 error. In general, for any statistical test, power is a function of sample size (n), the probability of Type 1 error (α), and the magnitude of the difference between the null hypothesis and reality (the "effect size," Cohen 1977). Here effect size is quantified by the rate of change parameter r . In addition, we must consider that, in many ecological applications, measurement error is not trivial. This means that a practical power analysis must consider the uncertainty or variability associated with each estimate of abundance. That variability is parameterized by cv , the coefficient of variation of the estimate of abundance. These five parameters (cv , n , r , α , and β) are related by the equations derived in the Methods section. Which of the five parameters is of primary interest depends on the application.

There are two general situations in which a power analysis is useful: in experimental design, and in interpretation of results. When an experiment or sampling program is being designed, the questions most likely to be asked are: How many samples will be needed? How precise must the samples be? What is the probability of detecting a trend? These questions are answered by solving for n , cv , and β , respectively. Allen and Kirkwood (1976) analyzed the feasibility of detecting the results of experimental manipulations of whale stocks; they derive an equation similar to Eq. 17 and conclude that such a program is not practicable. When power analysis is used to aid in the interpretation of results, particularly when the null hypothesis is not rejected, the questions most likely to be of interest are: How large a trend could have been detected? What was

the probability of detecting it? In this case, we solve for r and β .

The relations among these five parameters depend on assumptions made about the ecological process producing the trend and the techniques used to detect it. Models investigated explicitly in this paper include the assumptions that changes in abundance take place in constant increments (the linear model) or at a constant rate (the exponential model), and that the coefficient of variation is either constant, proportional to the square root of abundance, or proportional to the inverse of the square root of abundance. To apply the results of this paper one must decide which model most closely describes the situation at hand.

In general, the results show that the detection of trends depends strongly on the number and precision of the samples. Given a reasonable number of samples (say 10–15), cv must be sufficiently low, or r must be sufficiently high, to detect trends with controlled Type 2 error. We cannot control r , except indirectly by sampling at greater intervals, but we can control cv , at least the portion of it due to measurement error. cv can be reduced by expending greater sampling effort—more traps or greater areal coverage, for example. It can also be reduced by taking replicate measurements, since the mean of several independent measurements will have a lower variance than the measurements themselves, as illustrated by the sea otter example. It is difficult to generalize about what to expect the coefficient of variation to be, because that depends on the particular species involved, on what the quantity of interest is, and on what technique is used to estimate its abundance. Eberhardt (1978) discusses variability in a variety of techniques used in population studies.

A more subtle issue concerns balancing Type 1 and Type 2 errors. Other things being equal, there is a trade-off between the levels of these two kinds of error. What error levels are acceptable depends on the research goals (Rotenberry and Wiens 1985). It is common practice to adopt .05 as the largest significance (α) level at which the null hypothesis is rejected, but there is no such common practice with regard to β , which is, by analogy, the level at which the null hypothesis is accepted. All too often, the null hypothesis is implicitly accepted without any consideration of power. If one is attempting to decide between a null hypothesis and a definite alternative, symmetry would dictate that α and β should be equal. The problem is that β , and its complement, power, is often not easily estimated. This paper allows β to be estimated for one commonly used statistical procedure, linear regression.

Some assumptions of this analysis may limit application of the results. The most restrictive assumption is that points are equally spaced along the x axis. This means that samples are assumed to be taken at regular intervals of time (daily, monthly, annually) or distance (every metre, kilometre, etc.). The results are robust to mild violations of this assumption, but if samples

are clustered near either end of the x axis, quite different calculations of power could result. It is also assumed that the estimates of abundance are taken similarly each time, so that a single coefficient of variation suffices to describe all samples, and that, as a corollary, all samples are given equal weights in the regression. In some applications weighted regressions will be more appropriate, or will give lower estimates of variance (de la Mare 1984).

Another important assumption is that estimates of abundance are independent. This is a particular concern for time series, where estimates temporally close to each other may tend to be more similar than estimates far apart in time. Such positive autocorrelation means that residual variance may be seriously underestimated. This has important consequences. If studies are designed based on an underestimated variance, the power of detecting a trend will be less than planned (inflated Type 2 error). And if data with positive autocorrelation are analyzed by ordinary least squares, a trend may appear to exist when in fact it does not (inflated Type 1 error). Harris (1986) reported simulations that violated the assumption of independence, as well as other assumptions of detecting trends. He emphasized that for data collected under typical field conditions, analytical calculations will tend to underestimate the standard error of the regression line. The prudent investigator, therefore, should consider that the equations of this paper give minimum estimates of the number of samples n required, minimum estimates of the magnitude of the trend r which could be detected, and maximum estimates of the coefficient of variation cv which would permit it to be detected.

The a priori specification of variability of the samples is one of the most difficult problems when power analysis is used in experimental design. There are two general sources of variability: measurement error, and real variation in the parameter. In an ecological context, real variation means that, even if we could measure without error, points would never lie precisely along a straight line, either because of natural environmental variability in time or space, or because the processes that produce the trend do not act in an exactly linear or exponential manner. Although these two sources of variability are distinct conceptually, they are often difficult to separate in practice. Fortunately, it is not necessary to separate them for this analysis; both types of variability can be combined in the parameter cv . The problem lies in specifying cv correctly to reflect both types. If replicate measurements are taken, the resulting variance is an estimate of measurement error, and is a minimum estimate of the variability to be expected about the regression line because it does not include real variation. If cv is estimated from residual variance about the regression line, this will include real variability as well as measurement error, but of course such information will be available only after data have been collected. This is important

for interpreting results, but for planning purposes, we would like to have an estimate of expected precision before the sampling program is begun. To make planning more realistic, some additional variability due to the environment should be included. If separate estimates of variance due to measurement error and to real variation are available, they can be combined. Let cv_m be the coefficient of variation of measurement error and cv_r the coefficient of variation due to real variability in the rate of change parameter r . Then the combined cv that would be appropriate to use in the equations above is

$$cv = \sqrt{cv_m^2 + cv_r^2 + cv_m^2 cv_r^2}.$$

To make the discussion less abstract, consider the detection of a trend in population size over time. The population dynamics of many insects, birds, small mammals, and marine animals with pelagic larvae are very sensitive to environmental conditions. The recruitment or production of new individuals can fluctuate substantially from one year to the next, depending on how favorable the environment is; in other words, there is real variation in r . If we wish to use the power analysis of this paper to detect a mean trend in population size, it will be important that the estimate of cv include this real variation, which may well be more important than measurement error. If we are unable to estimate the real variation in r , application of the results will be limited for such species. On the other hand, the population dynamics of large-bodied mammals and birds, certain fish, and some other organisms are characterized by damped variation in the population growth rate because reproduction is spread over many age classes. Such species approximate more closely the assumption of a constant growth rate. cv will mainly reflect measurement error, and since measurement error will usually be easier to estimate than future environmental variability, the results of this paper will be more readily applied to such species.

The use of linear regression analysis to detect a trend assumes that the change in abundance is linear, or has been made linear by suitable transformation of the axes. Frequently, however, the processes underlying the trend are but dimly perceived. The two simple models of change considered here, linear and exponential, are two ways in which abundance of a quantity could change regularly over time and space. In the absence of more detailed information to the contrary, the linear model is usually assumed in practice. The exponential model is frequently used for growth or decay processes. If another model is known to be more appropriate for a particular situation, a power analysis similar to the one presented here could be carried out for such another model. The usual situation, however, is that little is known about how the quantity of interest is changing over time or space—indeed, this is the motivation of the sampling.

ACKNOWLEDGMENTS

This work was initiated under NOAA Contract Number 80-ABC-00207 to Daniel Goodman and continued under a National Research Council Research Associateship at the Southwest Fisheries Center La Jolla Laboratory. I thank Douglas P. DeMaster for originally suggesting the problem. Earlier versions of the manuscript benefited from the criticism of K. R. Allen, D. P. DeMaster, J. A. Estes, D. Goodman, D. A. Meeter, W. F. Perrin, D. A. Roff, G. T. Sakagawa, and D. R. Strong.

LITERATURE CITED

- Allen, K. R. 1980. Conservation and management of whales. Washington Sea Grant, Seattle, Washington, USA.
- Allen, K. R., and G. P. Kirkwood. 1976. Is experimental management of whale stocks practicable? Paper ACMRR/MM/SC/Cmt. 3, FAO Scientific Consultations on Marine Mammals, Bergen, Norway.
- Burnham, K. P., D. R. Anderson, and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. *Wildlife Monographs* 72:1-202.
- Cohen, J. 1977. *Statistical power analysis for the behavioral sciences*. Academic Press, New York, New York, USA.
- de la Mare, W. K. 1984. On the power of catch per unit effort series to detect declines in whale stocks. Report of the International Whaling Commission 34:655-661.
- Eberhardt, L. L. 1967. Some developments in 'distance sampling.' *Biometrics* 23:207-216.
- . 1978. Appraising variability in population studies. *Journal of Wildlife Management* 42:207-238.
- Estes, J. A., N. S. Smith, and J. F. Palmisano. 1978. Sea otter predation and community organization in the western Aleutian Islands, Alaska. *Ecology* 59:822-833.
- Freund, J. E. 1962. *Mathematical statistics*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- Harris, R. B. 1986. Reliability of trend lines obtained from variable counts. *Journal of Wildlife Management* 50:165-171.
- Holt, R. S., T. Gerrodette, and J. B. Cologne. 1987. Research vessel survey design for monitoring dolphin abundance in the eastern tropical Pacific. United States National Marine Fisheries Service Fishery Bulletin 85:435-446.
- Moore, P. G. 1954. Spacing in plant populations. *Ecology* 35:222-227.
- Neter, J., W. Wasserman, and M. H. Kutner. 1983. *Applied linear regression models*. Richard D. Irwin, Homewood, Illinois, USA.
- Peterman, R. M., and R. D. Routledge. 1983. Experimental management of Oregon coho salmon (*Oncorhynchus kisutch*): designing for yield of information. *Canadian Journal of Fisheries and Aquatic Sciences* 40:1212-1223.
- Rotenberry, J. T., and J. A. Wiens. 1985. Statistical power analysis and community-wide patterns. *American Naturalist* 125:164-168.
- Seber, G. A. F. 1982. *The estimation of animal abundance and related parameters*. Second edition. MacMillan, New York, New York, USA.
- Toft, C. A., and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. *American Naturalist* 122:618-625.
- VanBlaricom, G. R., and J. A. Estes, editors. 1987. *The community ecology of sea otters*. *Ecological Studies, Analysis and Synthesis* 65, *in press*.