*North American Journal of Fisheries Management* 8:181–190, 1988


# Sample Size Requirements for Detecting Changes in Some Fisheries Statistics from Small Trout Lakes

ERIC A. PARKINSON

*British Columbia Fisheries Branch, 2204 Main Mall*
*University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada*

JONATHAN BERKOWITZ

*Pacific Datametrics Consultants, 6530 Dunbar Street*
*Vancouver, British Columbia V6N 1X8, Canada*

CHRISTOPHER J. BULL

*British Columbia Fisheries Branch, 3547 Skaha Lake Road*
*Penticton, British Columbia V2A 7K2, Canada*

*Abstract.*—The purpose of this paper is to determine sample sizes needed to use *t*-tests to detect changes in indices of angling effort, catch per unit effort (CPUE), fish length, fish growth, and fish age. The results are presented in graphs of sample size versus detectable change for data from small trout lakes in southern British Columbia. Sample size requirements are most difficult to meet for indices of CPUE, effort, and age, but the cost of data acquisition appears reasonable even for small lakes with less than 1,000 angler-days/year. Alternative choices for significance levels (0.01–0.20) and power (0.80–0.95) result in almost a six-fold variation in sample size requirements.

There are fundamental differences in the problems facing a biologist who manages a single large lake and one who manages many small lakes. In the first case, management often involves extensive data collection on the population dynamics of individual stocks (e.g., Benson and Bulkley 1963; Rieman and Bowler 1980) combined with numerous constraints on management action as a result of political considerations and the complexity of the system (Hilborn 1987). Managers of small lakes cannot enjoy the luxury of detailed data collection on each stock, but are freer to engage in active manipulation of individual lakes (i.e., active adaptive management; Walters 1986) because the consequences of individual mistakes are less serious for small lakes than for large ones.

Trout fisheries in small lakes are a good example of a finely subdivided management responsibility. Managers often have a considerable degree of control over important driving variables such as recruitment and harvest rate on each lake. In many cases, however, this potential control of individual lakes is not exercised because of the difficulty in collecting sufficient information to determine the effectiveness of changes in management strategy. Often the cost of implementing a management change on small lakes is minimal (e.g., regulation change) or small (increased fry stocking) relative to the cost of evaluating the effect of the management change. In contrast to larger lakes, where a conventional creel census can be used to determine effort, catch, and catch per unit effort (CPUE) at a cost that is well below the value of the fishery, the cost of a conventional creel census on a small lake can easily exceed the value of the fishery.

An alternative to a conventional creel census is to estimate values for indices of selected fisheries statistics from a minimal amount of data. The choice of indices will reflect the goals of management. In British Columbia lakes, a major goal is to increase fishing effort while maintaining angling quality. An index of effort is, therefore, essential. Indices of quality that are directly perceived by the angler include fish size, CPUE, and angler crowding (effort). In addition, estimates of growth rate and age at capture help to interpret changes in fish size. The purpose of this paper is to determine the sample sizes needed to detect a given change in these fishery statistics; data from small lakes (<300 hectares) in southern British Columbia are used as an example. This information is then used to evaluate the feasibility of collecting adequate management data on each of a large number of small lakes.

Statistical convention is to set the significance level of a test ($\alpha$, the probability of a type-I error) at either 0.05 or 0.01 and to set the power of the test ($1 - \beta$, $\beta$ being the probability of a type-II error) at the desired alternative hypothesis at 0.80. In this paper, we explore the consequences of some alternative choices of both $\alpha$ and $\beta$ and discuss situations in which these alternative values may be more appropriate.

## Methods

Data were collected during creel surveys and aerial counts of boats on 25 lakes (Table 1). These lakes lie on the southern interior plateau and on the southern interior highland regions of British Columbia at a latitude range of about 49–50°N and a longitude range of 119–120°W (regions D and E: Northcote and Larkin 1956). There are no population centers on any of these lakes and all lie within a 50-km radius of each other.

An index of fishing effort was calculated from midday (1100–1500 hours) aerial counts of boats. Counts were made between 1979 and 1985 on two groups of lakes (Table 1): 23 and 37 counts on groups 1 and 2, respectively. These groupings reflect geographic proximity, which allowed boats on the lakes within each group to be counted on the same days at approximately at the same times. Fishing was the major boating activity on these lakes, and boats with anglers were easily identified.

Boat counts were also made from the ground five times a day on four lakes during May–August 1985 to determine within-day variation in boat counts. The total number of days per lake on which boats were counted was 47 on Headwater Lake, 74 on Lady King Lake, 75 on Pinaus Lake, and 71 on Postill Lake. Catch-per-unit-effort data were collected on these same four lakes on the same days that the boat counts were made. Anglers were interviewed by fishing party at the end of their trips. The number of anglers, hours fished, and number of fish caught were recorded for each party.

Fish length and age data were collected from angler catches on three lakes (Alleyne, Kentucky, Bluey) during the May and July holiday weekends of 1985. Ages were determined from scales by the methods of Tesch (1968).

TABLE 1.—Some characteristics of the study lakes in southern British Columbia.

| Type of data collected | Lake | Elevation (m) | Area (hectares) | Mean depth (m) | Total dissolved solids (mg/L) | Fish species present[a] |
|---|---|---|---|---|---|---|
| Fish lengths, ages | Alleyne[b] | 990 | 49 | 16.8 | 308 | RBT |
| | Bluey[b] | 980 | 26 | 11.9 | 310 | RBT |
| | Kentucky[b] | 990 | 36 | 16.5 | 232 | RBT |
| Catch per unit effort, boat counts | Headwater | 1,260 | 57 | 3.4 | 66 | RBT, BT, LNS |
| | Lady King | 1,010 | 6 | 3.7 | 190 | RBT |
| | Pinaus[c] | 1,010 | 161 | 23.1 | 140 | RBT |
| | Postill | 1,400 | 91 | 5.5 | 60 | RBT |
| Aerial counts | | | | | | |
| Group 1 | Allison | 820 | 71 | 14.9 | 241 | RBT, BT |
| | Borgeson | 840 | 10 | 9.8 | 340 | RBT, BT |
| | Boss | 1,020 | 24 | | | RBT, RSS |
| | Dry | 840 | 16 | 7.3 | 243 | RBT, BT |
| | Kump | 1,130 | 14 | | | RBT |
| | Laird | 790 | 28 | 10.1 | 276 | RBT, BT, RMW, LNS, LSS, PS, NS, RSS |
| | Link | 1,080 | 19 | 3.7 | 62 | RBT |
| | Osprey | 1,100 | 37 | 4.6 | 78 | RBT |
| | Yellow | 760 | 31 | 17.4 | 352 | RBT, BT |
| Group 2 | Aberdeen | 1,280 | 254 | 4.8 | 43 | RBT |
| | Crooked | 1,350 | 40 | | | RBT |
| | Dee | 1,360 | 37 | | 78 | RBT |
| | Haddo | 1,250 | 80 | | 44 | RBT |
| | Hidden | 670 | 132 | 17.7 | 149 | RBT |
| | Island | 1,360 | 42 | 4.0 | 168 | RBT |
| | Nicklen | 1,310 | 65 | | | RBT |
| | Oyama | 1,371 | 260 | 6.7 | 54 | RBT |
| | Swalwel | 1,340 | 242 | 5.9 | 63 | RBT, PS |

[a] Fish species: RBT—rainbow trout *Salmo gairdneri*, BT—brook trout *Salvelinus fontinalis*, RMW—rocky mountain whitefish *Prosopium williamsoni*, LNS—longnose sucker *Catostomus catostomus*, LSS—largescale sucker *Catostomus macrocheilus*, PS—prickly sculpin *Cottus asper*, NS—northern squawfish *Ptychocheilus oregonensis*, RSS—redside shiner *Richardsonius balteatus*.

[b] Also included in group 1.

[c] Also included in group 2.

We assume that a $t$-test will be used to compare the value of a fishery statistic before and after a change in management strategy. The required sample size ($N$) for each sample period (i.e., total sample size is $2N$) then depends on the standard deviation (SD) and the desired detectable change ($c$) of the index, as well as on a constant ($k$) that varies with the significance level ($\alpha$) and the power ($1 - \beta$) of the test:

$$N = k(\text{SD}/c)^2 \quad (1)$$

(Snedecor and Cochran 1980). In a specific situation, it is easier to visualize the effect of $N$ on the detectable change if the detectable change is expressed as a percentage ($p$) of the mean ($\bar{X}$):

$$c = \bar{X}p/100, \quad (2)$$

and substituted into equation (1):

$$N = 100^2k(\text{SD}/\bar{X})^2/p^2. \quad (3)$$

A value of $\text{SD}/\bar{X}$ is generated for each set of samples (e.g., fish lengths from one lake) within the preliminary data set. The average $\text{SD}/\bar{X}$ of all these preliminary data sets is then used to generate $N$ as a function of $p$. Graphs of $N$ versus $p$ were generated for each data set in this study with equation (3). Exact values of $N$ for other data sets can be calculated for some combinations of $\alpha$ and $\beta$ by using the values of $k$ given in Table 2.

A one-tailed test was assumed in all examples. Because $N$ is proportional to $k$, $N$ for a two-tailed test can be calculated by multiplying $N$ for a one-tailed test by the ratio of the $k$ values for the two- and one-tailed tests (e.g., for $\alpha = 0.05$ and $\beta = 0.2$, multiply by 15.70/12.37 = 1.24; Table 2).

## Results

### *Length*

Fish length at age was the least variable of the four statistics evaluated. Values of $\text{SD}/\bar{X}$ ranged from 0.083 to 0.128 and averaged 0.107 (Table 3). A change of less than 10% can be detected with a sample of 20 fish (Figure 1) when $\alpha = 0.05$ and $\beta = 0.20$. If $\alpha$ is raised to 0.2, the required sample size for a given detectable change drops by 54%, whereas lowering $\alpha$ to 0.01 increases the sample size by about 61%. If $\beta$ is changed from 0.2 to 0.05, the required sample size at $\alpha = 0.05$ rises by 77%. Because $N$ is directly proportional to $k$, these proportional changes in $N$ in response to changes in $\alpha$ and $\beta$ are identical for all values of $\text{SD}/\bar{X}$.

Lengths of all fish in angler catches, irrespective of age, were somewhat more variable than length at age and tended toward a multimodal frequency distribution because of the presence of more than one age-class. Values of $\text{SD}/\bar{X}$ ranged from 0.127 to 0.318 (Table 3) and were highest for those lakes where the catch was distributed among two or three age-classes and lowest where the catch consisted mostly of a single age-class. The required sample sizes for a given $p$, $\alpha$, and $\beta$ are about four times those of length at age because $\text{SD}/\bar{X}$ has approximately doubled (Figure 2).

TABLE 3.—Mean lengths of age-2, age-3, and all rainbow trout harvested from three lakes in southern British Columbia.

| Lake | Date | Age | $N$ | Mean length ($\bar{x}$, mm) | SD | SD/$\bar{x}$ |
|---|---|---|---|---|---|---|
| Alleyne | May 20, 1985 | 2 | 28 | 275 | 29 | 0.107 |
| | | 3 | 23 | 356 | 42 | 0.118 |
| | July 1, 1985 | 2 | 10 | 275 | 34 | 0.124 |
| | | 3 | 8 | 395 | 51 | 0.128 |
| Bluey | July 1, 1985 | 2 | 29 | 251 | 30 | 0.120 |
| Kentucky | May 20, 1985 | 2 | 63 | 249 | 22 | 0.087 |
| | | 3 | 14 | 361 | 33 | 0.093 |
| | July 1, 1985 | 2 | 12 | 241 | 20 | 0.083 |
| Grand mean | | | | | | 0.107 |
| Alleyne | May 20, 1985 | all | 53 | 317 | 59 | 0.187 |
| | July 1, 1985 | all | 19 | 333 | 75 | 0.224 |
| Bluey | July 1, 1985 | all | 31 | 250 | 32 | 0.127 |
| Kentucky | May 20, 1985 | all | 82 | 280 | 64 | 0.230 |
| | July 1, 1985 | all | 18 | 281 | 89 | 0.318 |
| Grand mean | | | | | | 0.217 |

TABLE 2.—Selected $k$ values for various values of $\alpha$ and $\beta$ for one-tailed $t$-tests. Values of $k$ in parentheses are for two-tailed tests.[a]

| | $\alpha$ | | | |
|---|---|---|---|---|
| $\beta$ | 0.20 | 0.10 | 0.05 | 0.01 |
| 0.20 | 5.67 (9.02) | 9.02 (12.37) | 12.37 (15.70) | 20.01 (23.37) |
| 0.10 | 9.02 (13.15) | 13.15 (17.13) | 17.13 (21.02) | 26.00 (29.77) |
| 0.05 | 12.37 (17.13) | 17.13 (21.64) | 21.64 (25.99) | 31.54 (35.63) |

[a] $\alpha$ and $\beta$ are probabilities of type-I and type-II errors, respectively; $k$ relates sample size to sample variability, given $\alpha$ and $\beta$.

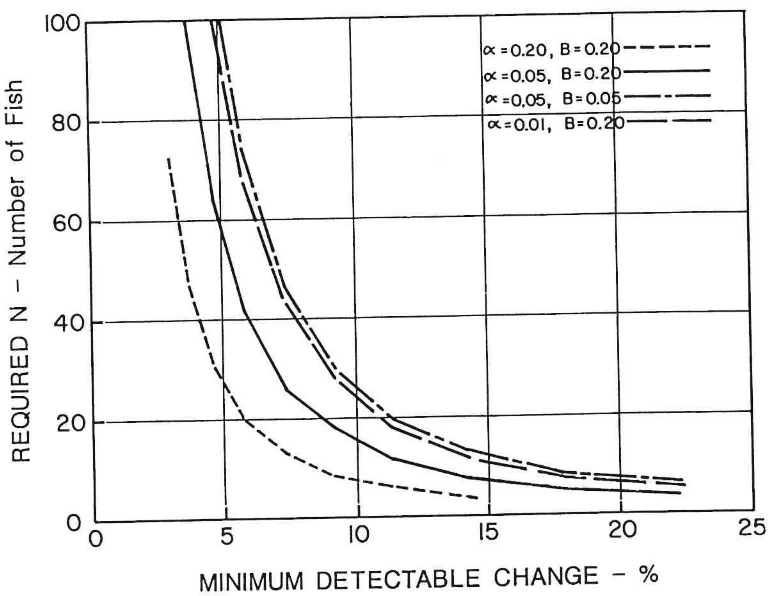


FIGURE 1.—Sample size versus minimum detectable change (as a percentage of the mean, $\bar{X}$) in rainbow trout length at age for a one-tailed $t$-test and four combinations of $\alpha$ and $\beta$ probabilities. All curves are based on the same value of SD/$\bar{X}$ (0.107), the average for eight samples of aged fish from three lakes (Table 3).

*Catch per Unit Effort*

The CPUE index was the ratio of total daily catch to total daily effort for all interviewed anglers sampled on four lakes during two seasons (spring and summer), giving eight sampling strata. In contrast to CPUE for individuals, the distribution of this measure of CPUE was approximately normal and, therefore, does not require transformation to meet the requirements of the $t$-test. The number

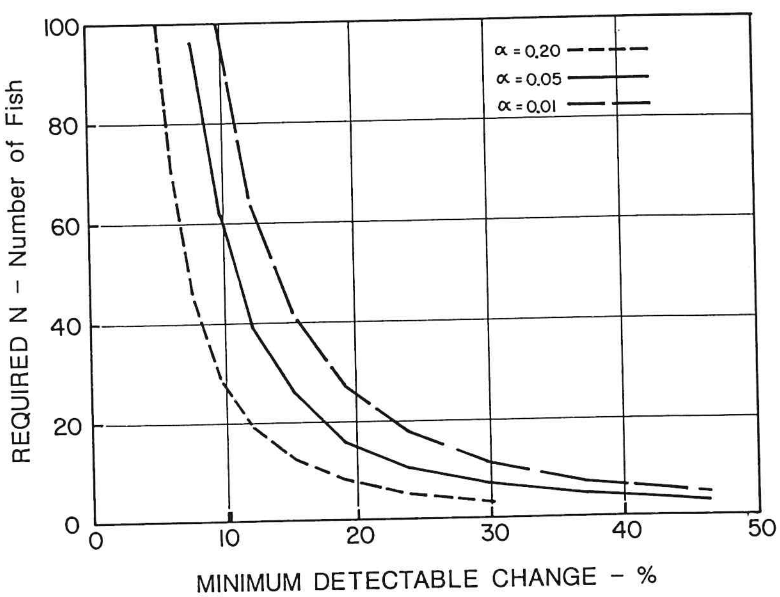


FIGURE 2.—Sample size versus minimum detectable change (as a percentage of the mean, $\bar{X}$) for a one-tailed $t$-test of lengths of all rainbow trout in angler catches from three lakes. All curves are based on the same value of SD/$\bar{X}$ (0.217; Table 3) and $\beta$ probability (0.20).

of census days varied from 9 to 42 per stratum. The total daily effort expended by interviewed anglers in a day ranged from 2.5 to 346 h. The CPUE data were more variable than the length data; the average SD/$\bar{X}$ value for the eight strata was 0.44. As a consequence, detecting a change of 35% requires more than 20 d of sampling (Figure 3). In an attempt to reduce variation, days with less than 50 h of effort were eliminated. The average SD/$\bar{X}$ dropped to 0.34 and the required sample sizes declined by 40% (Figure 3).

*Effort*

Variability in midday boat counts was large even when the data were stratified by season (spring, summer) and day type (weekend, holiday weekend, weekday). Typically, standard deviations were 50–100% of the mean, and SD/$\bar{X}$ tended to be higher for strata with lower mean counts (Figure 4). Average counts on all four lakes usually peaked at midday but counts made at 1000 and 1600 hours were not significantly different from those made at noon (analysis of variance; $P > 0.05$). Aerial counts made between 1100 and 1500 hours were assumed, therefore, to be representative of the daily effort on the group-1 and group-2 lakes identified in Table 1.

Experimental designs for comparing boat counts before and after a management strategy are complicated because of the need to reduce variability through seasonal and day-type stratification. Spatial controls are also desirable to account for confounding factors, such as license fees, that may change during the course of an experiment. To avoid some of these difficulties, we used the difference ($D$) between the count on one lake and the average count on a group of nearby lakes. Instead of asking if the boat count had changed, we decided to ask if the boat count had changed in comparison to boat counts made on nearby lakes on the same day.

Values of $D$ were calculated for the two groups of lakes that were counted on the same day on several occasions during 1979–1985. For each day $i$, the count $C_{ij}$ on lake $j$ was compared with the mean count $\bar{C}_i$ on the remaining lakes, the difference being $D_{ij}$:

$$D_{ij} = C_{ij} - \bar{C}_{i\cdot} \quad (4)$$

In contrast to the raw count data, the distribution of this index was approximately normal and did not require transformation. The variability (SD/$\bar{X}$) in $D_{ij}$ in relation to $\bar{C}_i$ was 0.95 and 0.98 for groups 1 and 2, respectively.

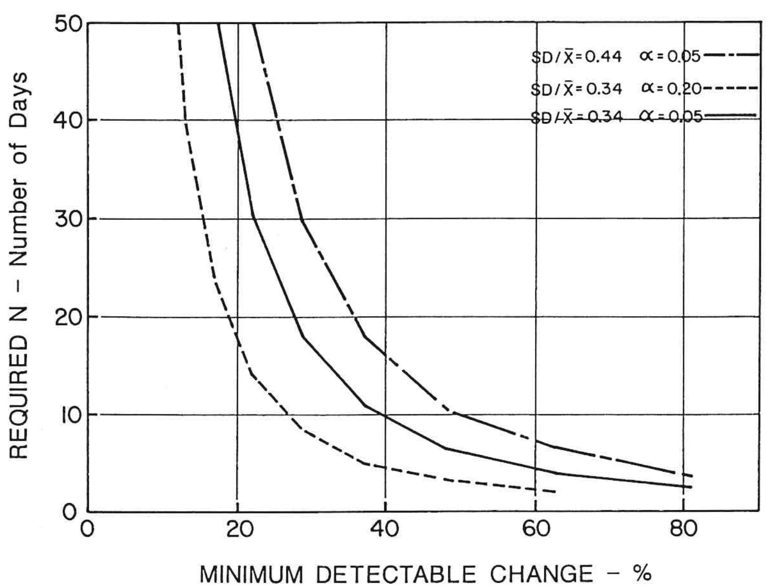


FIGURE 3.—Sample size versus minimum detectable change (as a percentage of the mean, $\bar{X}$) in catch per unit effort (CPUE) for a one-tailed $t$-test and alternative $\alpha$ probabilities. Sample sizes for CPUE are given as the number of personnel-days on which census data were collected. Values of SD/$\bar{X}$ were 0.44 when all days were included and 0.34 when only those days with more than 50 h of effort were included. Probability $\beta = 0.20$ for all curves.

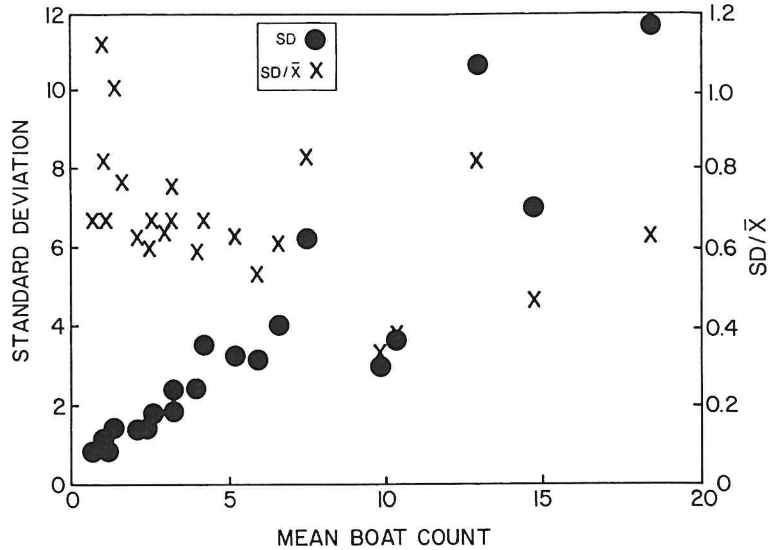


FIGURE 4.—Standard deviation (SD) and SD/$\bar{X}$ values for mean ($\bar{X}$) boat counts on four study lakes on which ground counts were made 5 d/week during the spring–summer fishing season. Sample periods were stratified by day type and season.

Sample sizes required to detect a given change in $D$ (expressed as a percentage of $\bar{C}_i$) are similar for both groups of lakes (Figure 5). The required sample sizes are large, even with a large detectable change—a direct result of the high variability in the raw count data. When the weekdays and autumn counts present in group 2 were deleted, SD/$\bar{X}$ declined to 0.79 and $N$ decreased by 37%.

## *Age Composition*

Changes in the age composition of the catch can be detected with a simple chi-square analysis. If

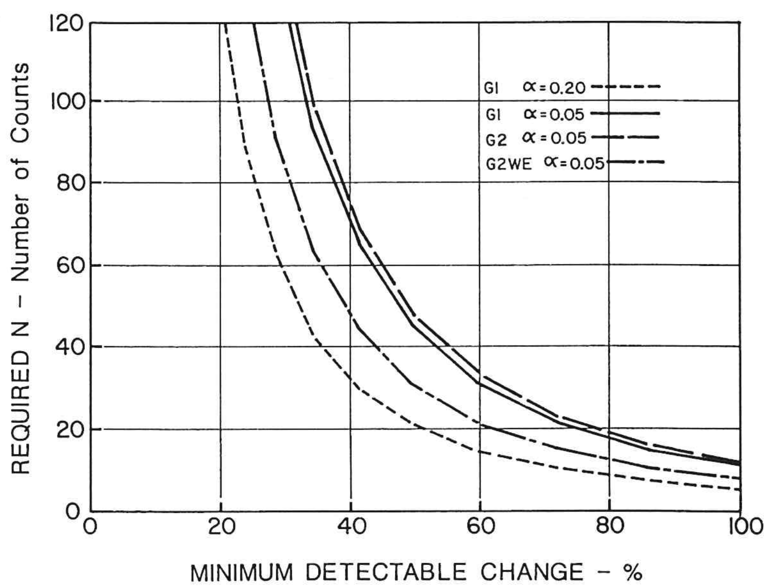


FIGURE 5.—Sample size versus minimum detectable change in the fishing effort index ($D$), expressed as a percentage of the mean boat count on several lakes, for a one-tailed $t$-test for group-1 lakes (G1), group-2 lakes (G2), and weekends only for group-2 lakes (G2WE) in southern British Columbia. The probability $\beta = 0.20$ for all curves. $D$ is the midday boat count on a study lake minus the average midday boat count on a group of nearby lakes.

only two age-classes are used, no preliminary data are needed to calculate the detectable differences for a given sample size, because the variance is a simple function of $r$, the proportion of fish in one age-class. The minimum detectable difference varies with the initial $r$ and reaches a maximum for intermediate values of $r$ (Figure 6). The number of fish required is potentially very large for the size of differences that might interest managers. Changes of less than 0.1 in $r$ (e.g., a 50:50 ratio going to a 60:40 ratio) require sample sizes greater than 160 fish, at an $\alpha$ of 0.05.

## Discussion

Sample size estimation is a subjective process because only one factor ($SD/\bar{X}$) of the three involved can be objectively estimated. Acceptable levels of $\alpha$ and $\beta$ have to be subjectively determined despite the presence of widely accepted conventional values. The choice of $p$ (minimum detectable percentage change) is especially important because halving $p$ quadruples the required sample size, but this choice is difficult because there is no guidance from convention.

Although sample size estimation is useful in reducing the collection of excess data, it is probably more important in avoiding experiments in which biologically significant differences are unlikely to be statistically detectable (e.g., Trautmann et al. 1982). An increase of 20% in stocking may appear to be substantial, but if the minimum expected detectable difference in the highly variable CPUE index is 35%, a 20% increase in stocking is unlikely to produce a detectable response in this index.

The numerical results for each index are strictly applicable only to our situation, but the calculations are easy to make for other situations for which variances can be estimated from preexisting data. If no data are available, our results could be used directly but, because variability is likely to differ among fisheries, this approach may be inappropriate. Data collected in the initial stages of an experiment can be used to estimate sample size requirements and, if necessary, sample sizes could be expanded at this stage. In many cases, however, a small amount of preliminary data will not supply good estimates of variances so that the estimate of the necessary sample sizes will be poor unless extra data are collected specifically for variance estimation.

Sample size calculations are usually specific to individual situations (e.g., Ossiander and Wedemeyer 1973; Hankin 1982; Trautmann et al. 1982). Our methods can only be applied where a 2 × 2 chi-square test or a $t$-test is appropriate, but this covers many situations in which an index is measured before and after a management change. In many cases, a more complicated design is desirable to provide for a spatial as well as a temporal control, but sample size calculations are complex even for simple analysis-of-variance designs. Nonparametric tests such as Wilcoxon's two-sample procedure (Sokal and Rohlf 1981) may have to be used with data that are not normally distributed,

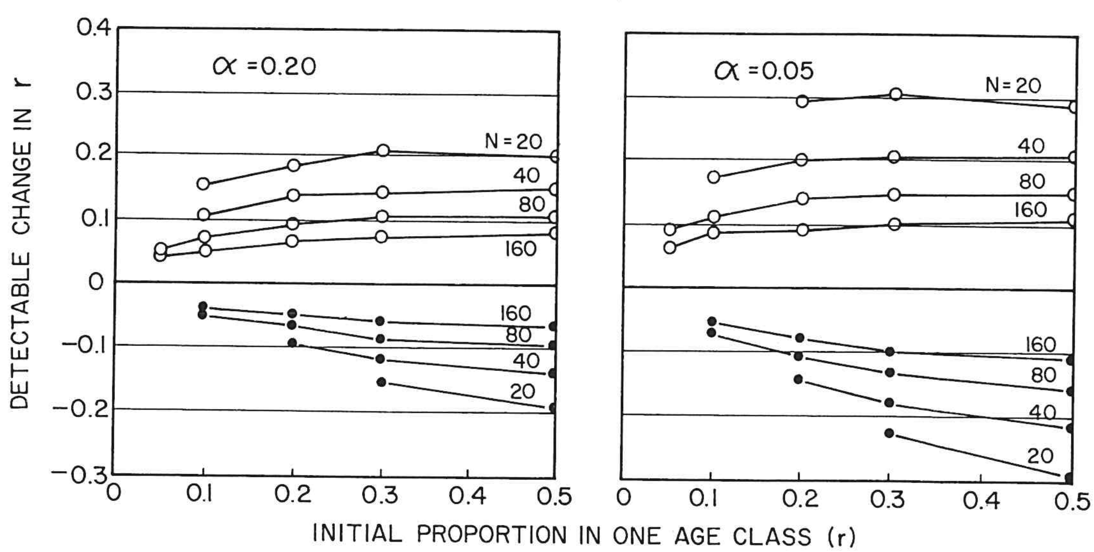


FIGURE 6.—Minimum detectable changes in the proportion of fish in one of two age-classes for various initial proportions and sample sizes at $\alpha$ probabilities of 0.20 and 0.05.

although transformation can usually normalize data that do not have a multimodal distribution. Sample size estimation is complex and poorly understood for nonparametric tests.

Effort is a key index because it represents the anglers' reaction (to fish or not to fish) in response to changes in other indices (CPUE, fish size) that are perceptible. Effort can also be measured by aerial counts, rather than by on-ground clerks. Interviews, therefore, could be eliminated, and a response in terms of effort changes alone could be used to evaluate the effectiveness of management changes on lakes where the expense of creel census cannot be justified by the size of the fishery.

The use of boat counts from a single time stratum (e.g., midday weekends) assumes that the proportion of effort expended in that stratum is similar before and after a mangement change is implemented. There is no way of checking this assumption except to count other strata and thereby greatly increase the amount of sampling required. A similar argument applies to other indices that are measured during a restricted time period or at a certain location.

The CPUE index used here was chosen for convenience, ease of comprehension, and its approximately normal distribution. The choice among alternative CPUE indices may be more critical when a roving creel clerk is to be used in a multispecies situation (Von Geldern and Tomlinson 1973). Hamley and Howley (1985) emphasized the variable nature of CPUE and the need for larger-than-expected samples to make meaningful comparisons among observational units in many fisheries.

Sample size calculations for chi-square tests require no pretreatment data but, in our experience, they are rarely performed even though required sample sizes are surprisingly large relative to those required for characteristics such as length or length at age. Gold (1969) pointed out that substantial differences are often statistically insignificant because of small sample sizes.

The five indices used in this study can be viewed as an integrated data collection system that has internal checks. An increase in effort on experimental lakes relative to control lakes should be accompanied by an increase in either CPUE or fish size. An increase in fish size can be tied to some combination of growth and age at capture. A decrease in age at capture should be accompanied by an increase in either effort or growth. Inconsistencies indicate that either an error has occurred in detecting a change in one of the indices or that factors other than the management change have affected the experiment and benefits cannot be attributed to the management change alone.

Costs of data acquisition appear to be reasonable even on very small lakes. In southern British Columbia, over 20 lakes typically can be counted for each hour of flight time (at Can$100/h) giving a maximum cost for 80 counts ($N$ = 40) of about $400. If the control : experimental lake ratio were 1:1, the total cost of detecting about 50% change in effort ($\alpha = 0.05$; $\beta = 0.2$) would be about $800 spread over 4–8 years. Detecting changes in CPUE is more expensive, because costs are about $125/d for creel surveys. If lakes are reasonably close together, data can be collected from two lakes on the same day giving a total cost of about $1,875 for 30 data points ($N$ = 15), which will be adequate to detect a change of about 30% ($\alpha = 0.05$; $\beta = 0.2$). Large sample sizes of fish should be obtainable at the same time. At a modest CPUE of 0.1 fish/h, 15 d with 50 h effort/d would produce 75 fish, and a change of less than 10% in fish size and growth indices could be detected. The total cost of data collection would be, therefore, about $2,675. Adding 10 d of office time at $100/d spread over 4–8 years gives a total cost of about $3,675 to monitor a typical small lake. On a lake with 1,000 angler-days/year, an increase in effort of only 20% due to improved management would produce benefits of about $4,000/year at $20/angler-day. Meaningful data collection, therefore, seems to be feasible even on small lakes.

Angler behavior will be influenced by the difficulty anglers experience in detecting changes in fishing quality. An increase in quality that is difficult to detect is unlikely to result in an increase in angler effort. To an angler, changes in fish size are easy to detect because the required sample sizes are small and each fish observed constitutes an easily measured sample. In contrast, anglers are likely to have more difficulty detecting changes in CPUE. Sample sizes are larger and each sample involves recording the catch resulting from a substantial amount of effort (over 50 h in this example). As a consequence, we predict that the effort response to changes in fish size should occur more rapidly than the effort response to changes in CPUE.

These characteristics of fishing quality indices also have implications for fisheries management agencies. A noticeable response in effort to a change in CPUE may take place years after the change has occurred. In the British Columbia fishery for steelhead (anadromous rainbow trout), for example, CPUE in some streams has more than doubled in the last 5 years, but the effort response has lagged 2–3 years behind this improvement in angling quality (Billings 1986). Modest improvements in angling quality based on CPUE may be maintained longer than those based on fish size before the effort response induces a higher harvest rate and a subsequent reduction in fishing quality. If a rapid response to CPUE changes is desired, advertising may be necessary. Management experiments that produce changes in CPUE should have a long time frame to allow for a possible lag in effort response.

The use of significance levels other than 0.05 can be justified when the results of an experiment are not to be widely applied and, therefore, the consequences of incorrect conclusions are not serious. With most experiments, the investigator would like to conclude that a treatment has an effect which will be the same in other situations. An example might be an experiment in which primary production increased after phosphorus was added to several lakes. Here, the conclusion is that phosphorus addition raises primary production in a certain class of lakes. In a lake-specific management system, results are only applied to a particular lake. For example, a similar change in stocking rate can result in increased, decreased, or unchanged CPUE in different lakes. A conclusion that an increase in CPUE resulted when the stocking rate on lake A was raised from 100 to 200 yearling/hectare cannot be assumed to be valid on lake B. Each experiment provides information on stocking rates in general but each lake also represents a unique biological and sociological situation.

There is obviously an optimal significance level at which the available sampling effort is allocated among lakes to maximize the number of lakes for which a correct decision will be made. If all the available sampling effort is concentrated on a single lake, the results may be significant at the 0.01 level, but a correct conclusion would have been made on only one lake. If, however, sampling effort is spread over, for example, five lakes, the expected difference may only be detectable at $\alpha$ = 0.2, but the correct conclusion will have been made on four of these lakes. The optimization model is not simple; it includes both the power and significance level, as well as the cost of reallocating sampling effort.

We suggest that values of $\alpha$ other than 0.05 be considered when changes in management of small trout lakes are evaluated. In cases where the consequences of an inappropriate management change are more serious (e.g., a larger lake with several fishing resorts), the use of lower $\alpha$ values (requiring larger sample sizes) should be continued. If, however, a lake is one of many small lakes in a fisheries management district, the use of a higher $\alpha$ should result in a better allocation of sampling effort and a more successful overall management program, even though the error rate for individual lakes will be higher. The controversy over the meaning and interpretation of significance levels is not new (Morrison and Henkel 1970). Specific suggestions for criteria in choosing and interpreting significance levels have been summarized by Labovitz (1968).

Conventional hypothesis-testing theory assumes that type-I errors (detection of a difference that is not real, probability = $\alpha$) are more serious than type-II errors, but the consequences of type-II errors (nondetection of a difference that is real, probability = $\beta$) are often just as serious (Skipper et al. 1967), so the value of $\beta$ should also be considered when the required sample size is estimated. If both $\alpha$ and $\beta$ start at 0.2, for example, the required sample sizes approximately double if $\alpha$ is lowered to 0.05 and double again if $\beta$ is lowered to 0.05 as well (Figure 1). If a higher degree of accuracy is required and sample sizes are limited, lowering $\alpha$ and $\beta$ simultaneously should be considered in cases when not detecting a real difference is just as serious a problem as detecting a spurious difference.

The distinction between one-tailed and two-tailed tests is somewhat arbitrary and needs to be examined in each case (Labovitz 1968). Consider the case in which CPUE is measured before and after an increase in stocking rates. Three outcomes (decrease, no change, or increase in CPUE) are possible. Two of these (decrease, no change), however, would result in the same management decision (restore lower stocking rates). A one-tailed test appears appropriate because we are interested in detecting a change in only one direction. In the more general case, however, we may be interested in the distinction between a decrease and no change in CPUE, because the former suggests that it is possible to overstock, rather than to just saturate, a lake. A two-tailed test is appropriate in this case because a change in either direction is of interest. The advantage of phrasing hypotheses in terms of one-tailed instead of two-tailed tests is a 17–37% reduction in required sample sizes for the 0.05–0.20 range of $\alpha$ and $\beta$.

Although it is a subjective process that can only be performed for simple experimental designs, sample size estimation is important in exploring new approaches (or evaluating old ones) to management. The data needed to estimate SD/$\bar{X}$ for statistics in different fisheries are often available from creel surveys performed for other reasons. Experienced managers can usually provide good guidance on the choice of $\alpha$, $\beta$, and $p$ for particular fisheries. Substantial improvements, therefore, can be made in answering the question of how much management information should be collected, typically without further data collection.


## Acknowledgments

We thank E. J. Mansfield and R. Bencivenga of Pacific Datametrics for their advice in the development of this paper.


## References


Benson, N. G., and R. V. Bulkley. 1963. Equilibrium yield and management of cutthroat trout in Yellowstone Lake. U.S. Bureau of Sport Fisheries and Wildlife Research Report 62.

Billings, S. J. 1986. Steelhead harvest analysis 1984–1985. British Columbia Fisheries Branch, Fisheries Technical Circular 71, Victoria, Canada.

Gold, D. 1969. Statistical tests and substantive significance. American Sociologist 4:42–46.

Hamley, J. M., and T. P. Howley. 1985. Factors affecting variability of trapnet catches. Canadian Journal of Fisheries and Aquatic Sciences 42:1079–1087.

Hankin, D. C. 1982. Estimating escapement of Pacific salmon: marking practices to discriminate wild and hatchery fish. Transactions of the American Fisheries Society 111:286–298.

Hilborn, R. 1987. Living with uncertainty in resource management. North American Journal of Fisheries Management 7:1–5.

Labovitz, S. 1968. Criteria for selecting a significa level: a note on the sacredness of 0.05. Amer Sociologist 3:220–222.

Morrison, D. E., and R. E. Henkel. 1970. The sig icance test controversy. Aldine, Chicago.

Northcote, T. G., and P. A. Larkin. 1956. Indice productivity in British Columbia lakes. Journa the Fisheries Research Board of Canada 13:5 540.

Ossiander, F. J., and G. Wedemeyer. 1973. Comp program for sample sizes required to determine ease incidence in fish populations. Journal of Fisheries Research Board of Canada 30:1383–1

Rieman, B. E., and B. Bowler. 1980. Kokanee tro ecology and limnology in Pend Oreille Lake. Id Department of Fish and Game, Fisheries Bull 1, Boise.

Skipper, J. K., A. L. Guenther, and G. Nass. 1967. sacredness of 0.05: a note concerning the use statistical levels of significance in social scie American Sociologist 1:16–18.

Snedecor, G. W., and W. G. Cochran. 1980. Statist methods, 7th edition. Iowa State University Pr Ames.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry, edition. Freeman, San Francisco.

Tesch, F. W. 1968. Age and growth. Pages 93–12 W. E. Ricker, editor. Methods for assessment of production in fresh water. Blackwell Scientific P lications, Oxford, England.

Trautmann, N. M., C. E. McCulloch, and R. T. Ogle 1982. Statistical determination of data requ ments for assessment of lake restoration progra Canadian Journal of Fisheries and Aquatic Scier 39:607–610.

Von Geldern, C. E., and P. K. Tomlinson. 1973. the analysis of angler catch rate data from wa water reservoirs. California Fish and Game 59:2 292.

Walters, C. 1986. Adaptive management of renewa resources. MacMillan, New York.