TRANSACTIONS of the AMERICAN FISHERIES SOCIETY

July 1964

VOLUME 93
NUMBER 3

Sample Size in Petersen Mark–Recapture Experiments

D. S. ROBSON AND H. A. REGIER

Biometrics Unit, Cornell University

ABSTRACT

The efficient planning of a Petersen-type mark and recapture experiment requires some knowledge of the order of magnitude of the population size N. Sample sizes M and C of the mark and recapture samples, respectively, may then be ascertained on the basis of a guessed value of N to achieve any desired degree of accuracy with any specified degree of confidence. Restrictions on the sample sizes M and C are that MC must exceed 4 times the guessed value of N, and the total costs of M and C must be equal. Graphs and formulas are given defining sample size to attain preassigned levels of accuracy and precision of population estimation. A method of choosing sample sizes such that experimental costs are minimized is described.

INTRODUCTION

One of the major decisions which must be made in planning a Petersen mark–recapture experiment is the amount of resources in terms of men, time, money, etc., which are to be invested in the experiment. The greater the investment, the greater will be the accuracy of the estimate of population size if resources are properly allocated. Our purpose in this paper is to provide a guide to effective planning in the form of formulas and graphs showing the relationship between the size and costs of the experiment and the accuracy and precision of the attendant population estimate.

Determining sample size for any experiment, whether with fish, tomato plants, or vacuum tubes, presents difficulties of the same type as the proverbial problem of lifting oneself by one's bootstraps. In the mark–recapture experiment the precision of estimation associated with any given sample size depends upon the size of the population from which the sample is drawn; consequently, to predict in advance the level of accuracy of an experiment is tantamount to predicting population size—or predicting precisely that unknown quantity which the experiment is designed to estimate.

The answer to this paradox lies in a pilot study, or an educated guess of population size based on other evidence and experience. Some guess, at least as to the order of magnitude of the population, is necessary if the planning of the experiment is not to be done completely in the dark, and the purpose of the experiment may then be regarded as the objective confirmation of or improvement upon this earlier guess. While this approach to planning is admittedly crude, it is presently the only approach available to us, and the logic of the situation suggests that no substantially more refined approach is possible. We shall assume, therefore, that a rough estimate of population size is available from some source. We take up from there the problem of assessing the relation between sample size and accuracy of the resulting estimate, to provide for the experimenter an indication of how he may expect the Petersen estimate to perform at any contemplated sample size. Cost considerations are then introduced and formulas for optimum allocation of resources to different phases of the experiment are developed.

PETERSEN ESTIMATE

A Petersen estimate is obtained by equating the proportion of marked fish observed in a sample to the proportion marked in the population, and then solving the equation for number in the population. Thus

$$\hat{N} = \frac{MC}{R}, \quad (1)$$

where M is the number marked and released,
C is the number subsequently examined for marks,
R is the number of marks found in the sample C,
N is the total (and unknown) number in the population, and
$\hat{N}$ is the Petersen estimate of N.

The Petersen estimate may be used where one or both of the following conditions applies: the marked fish, M, become randomly distributed in the population before the second sample, C, is taken; the second sample, C, is selected at random from the population. Circumstances that might invalidate the above conditions are discussed by Ricker (1958). Here we assume that these conditions are satisfied.

BIAS IN THE PETERSEN ESTIMATE

Various modifications have been suggested to improve the standard Petersen estimator, such as adding 1 to each factor in the numerator and denominator, and Bailey (1951) and Chapman (1951) have shown that under certain conditions the estimator

$$N^* = \frac{(M+1)(C+1)}{R+1} - 1$$

is exactly unbiased. A condition necessary for this unbiasedness, however, is that the total size of the mark and recapture samples $(M + C)$ must equal or exceed the size of the population (N). To express this another way, the size of the recapture sample (C) must equal or exceed the number of unmarked members in the population $(N - M)$. This condition would guarantee that the recapture sample includes at least one marked member $(R \geqslant 1)$. In many sampling situations M and C are small enough to admit the possibility that the recapture sample may not contain any marked members, and when this possibility exists, all forms of the Petersen estimator are biased downward by a factor equal to the probability of getting no recaptures. The estimator N^*, for example, instead of estimating N, is then an estimate of

$$N - (N - M)\ h(0;\ C,\ M+1,\ N)$$
$$\approx N - N\left(1 - \frac{M+1}{N}\right)^{C+1}$$
$$\approx N - Ne^{-\frac{(M+1)(C+1)}{N}},$$

where h denotes the hypergeometric function (Lieberman and Owen, 1961).

For any form of the Petersen estimator such as $\hat{N}$ or N^* the bias is therefore approximately $100e^{-MC/N}$ percent (Chapman, 1951) and thus becomes negligible only when the product of the two sample sizes $(M \times C)$ exceeds the population size (N) by at least a factor of 3 or 4. If $MC \approx N$, then the Petersen estimator underestimates N by approximately 37 percent on the average, while if $MC \approx 3N$ then the bias is reduced to 5 percent and for $MC \approx 4N$ the bias is less than 2 percent. At the other extreme where sample sizes are unreasonably small in relation to population size the bias of Petersen-type estimators approaches 100 percent. This is illustrated by the absurd but simple example where $M = C = 1$; the probability of a recapture $(R = 1)$ is then $1/N$ and the probability of no recapture $(R = 0)$ is $1 - (1/N)$, so the average value of the estimator N^* is

$$\begin{aligned} \text{avg } N^* &= \left(\frac{(M+1)(C+1)}{0+1} - 1\right)\left(1 - \frac{1}{N}\right) \\ &\quad + \left(\frac{(M+1)(C+1)}{1+1}\right)\left(\frac{1}{N}\right) \\ &= 3\left(1 - \frac{1}{N}\right) + 1\left(\frac{1}{N}\right) \\ &= 3 - \frac{2}{N}, \end{aligned}$$

and regardless of the population size N being estimated, the estimator N^* is seen to be an estimate of something less than 3.

The approximation given by

$$\text{Approx. Bias} = 100e^{-\frac{(M+1)(C+1)}{N}} \text{ percent}$$

improves as N gets large, but is of the correct

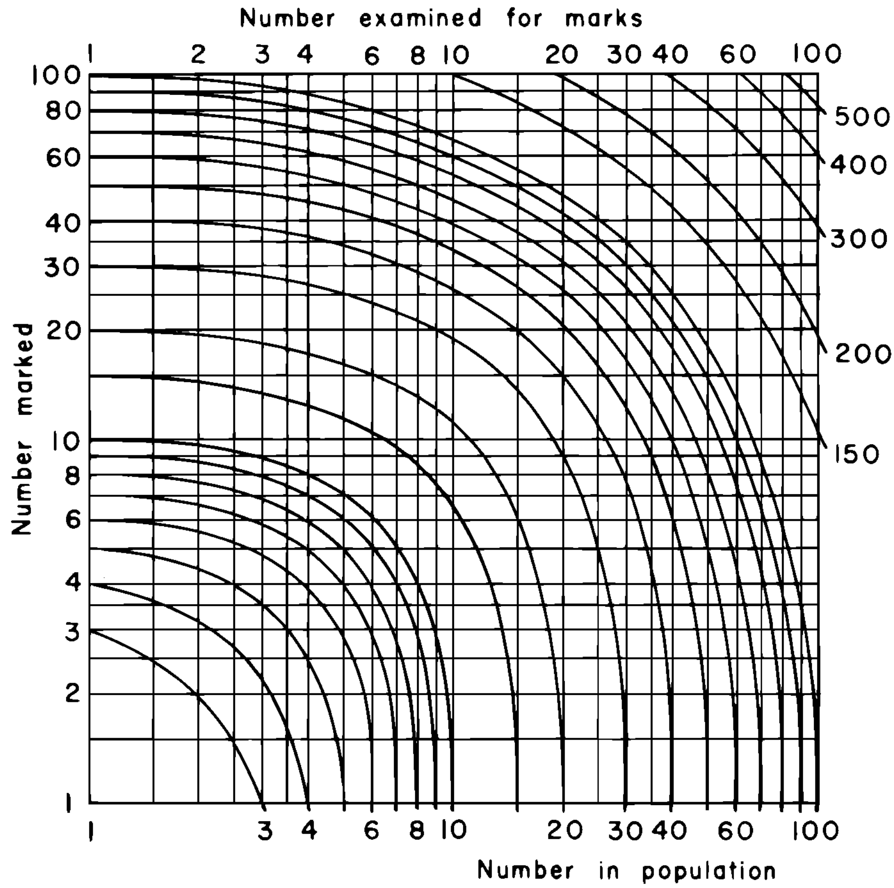

FIGURE 1.—Sample size when $1-\alpha=0.95$ and $p=0.50$; recommended for preliminary studies or management surveys. Data for $N \leqslant 100$ based on tables of hypergeometric distribution.

order of magnitude even for populations as small as $N=100$. Table 1 indicates the error in the approximation at this level of population size for the case of equal-sized samples ($M=C$). We conclude, therefore, that the approximation of bias by the above formula is satisfactory for populations of the larger sizes likely to be dealt with in practice.

The list of conditions necessary for the validity of the Petersen estimate should therefore include the following: *the product of the two sample sizes* M *and* C *must exceed 4 times the population size* N. Since MC/N is both the mean and approximate variance of R (see below), we can use tabulated values of standard normal deviates to determine that *if 7 or more recaptures are made*, then the lower 95 percent confidence limit of MC/N will exceed 4, or *with 95 percent confidence the bias is negligible.* Reference here is only to the theoretical bias of the estimator under random sampling, not to any bias that might result from improper sampling methods.

ACCURACY, PRECISION, AND SAMPLE SIZE

Let $1-\alpha$ be the probability that the population estimate $\hat{N}$ will not differ from the true population size by more than $100p$ percent, *i.e.*, by not more than pN. Here p denotes the level of *accuracy*, and $1-\alpha$ the level of

TABLE 1.—*Comparison of exact and approximate percent bias of* N^* *when* $N=100$ *and* $M=C$

	Sample size, where $M=C$						
Bias	50	25	20	15	10	5	1
Exact	0	0.1	0.3	5.0	26.4	69.3	97.0
Approximate	0.0	0.1	1.2	7.7	29.8	69.8	96.0

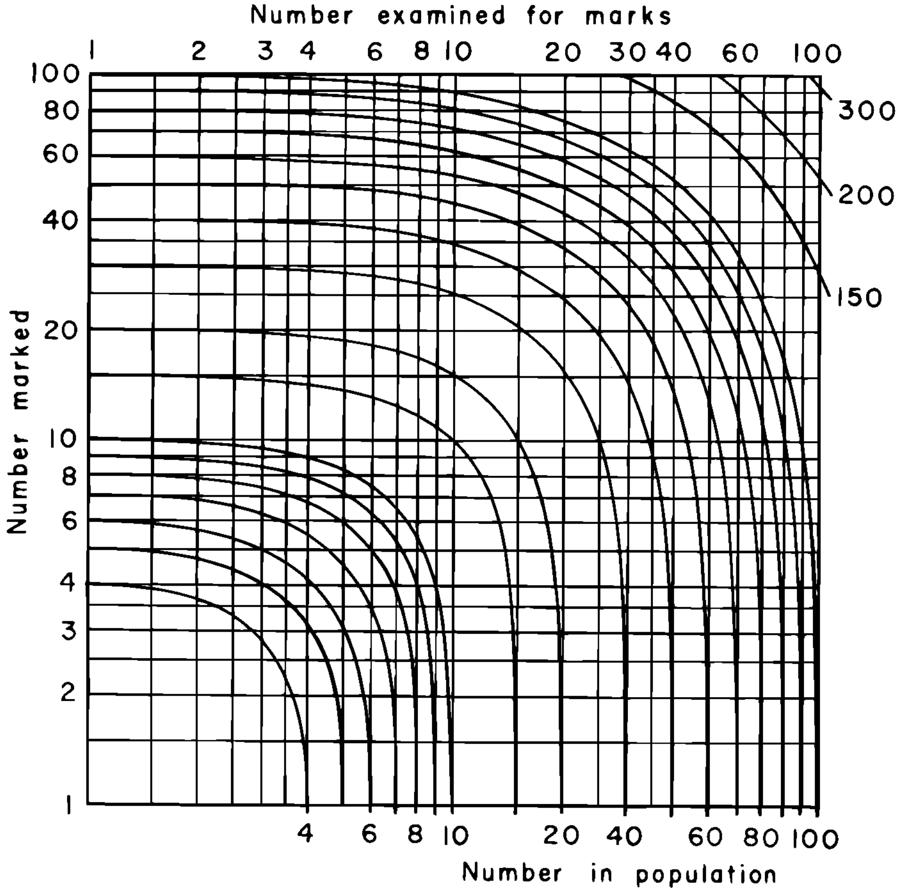

FIGURE 2.—Sample size when $1-\alpha = 0.95$ and $p = 0.25$; recommended for management studies. Data for $N \leqslant 100$ based on tables of hypergeometric distribution.

precision. The relation between α, p, and N as stated above is

$$1-\alpha \leqslant P\left(-p < \frac{\hat{N}-N}{N} < p\right). \quad (2)$$

Sample size in a Petersen mark–recapture experiment is specified by two numbers, either M and C or M and R. Chapman (1952) refers to a census design specifying M and C as "direct" and one specifying M and R as "inverse." We consider only the direct method here though this may require slightly larger sample sizes than would the inverse method. With $p = 0.10$ and $1-\alpha = 0.95$, the practical advantage of the inverse is negligible at all population sizes. With $p = 0.50$ and $1-\alpha = 0.95$, the inverse method requires about 1 percent less effort than the direct method when the population size, N, is 100; with larger N the advantage rapidly becomes negligible (see Chapman, 1952, p. 289).

To determine sample size, substitute equations (1) and (2) and rearrange the resulting equation to obtain (3) which does not contain $\hat{N}$.

$$1-\alpha \leqslant P\left(\frac{MC}{(1+p)N} < R < \frac{MC}{(1-p)N}\right). \quad (3)$$

Designate $MC/[(1+p)N]$ as R_L and $MC/[1-p)N]$ as R_U, thus

$$1-\alpha \leqslant P(R_L < R < R_U). \quad (4)$$

Equation (4) reads thus: the probability of obtaining between R_L ("L" for "lower bound") and R_U ("U" for "upper bound") recaptures when M, C, N, and p are fixed is equal to or greater than $1-\alpha$.

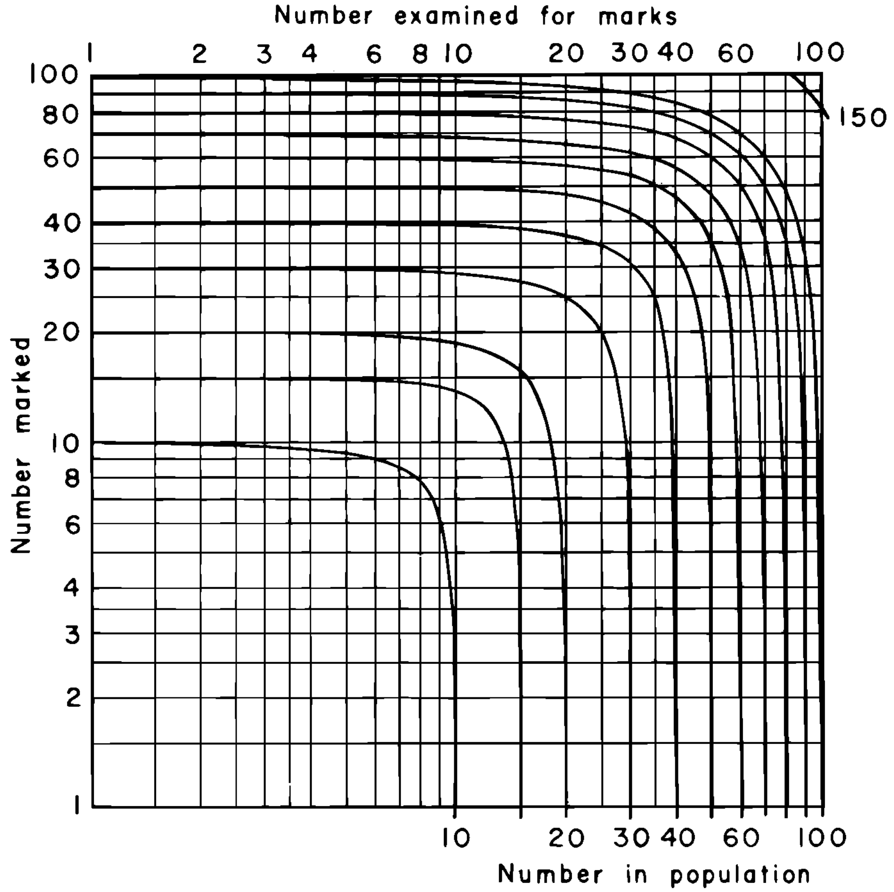

FIGURE 3.—Sample size when $1-\alpha = 0.95$ and $p = 0.10$; recommended for research. Data for $N \leqslant 100$ based on tables of hypergeometric distribution.

The distribution of R in (4) is hypergeometric. Lieberman and Owen (1961) have tabulated the probability function for the hypergeometric distribution for populations to size 100. These tabulations may be used with equations (3) and (4) to determine sample size with the Petersen estimate in the following manner.

Consider population size fixed, say N_1. We intend to find a combination of the sample sizes M and C, say M_1 and C_1, so that (3) will be satisfied and such that sample sizes of $M_1 - 1$ and C_1 or M_1 and $C_1 - 1$ would not satisfy it. For fixed α, p, and N one guesses a combination of M and C, and substitutes all values in equation (3). Generally R_L and R_U will be nonintegers and must be rounded to integers, $R_L{}^*$ and $R_U{}^*$, so that

$$1-\alpha \leqslant P(R_L \leqslant R_L{}^* \leqslant R \leqslant R_U{}^* \leqslant R_U). \quad (5)$$

Having obtained $R_L{}^*$ and $R_U{}^*$, enter the tables and determine if the probability of obtaining an R as in (5) is greater than or equal to $1-\alpha$. Choose a second pair of M and C and repeat the procedure. Continue iterating until the smallest sample sizes M_1 and C_1 are found that satisfy the equation.

Figures 1 to 3 are based on iterations as described above. Figure 1 shows plots of combinations of M and C that satisfy equation (4) for population sizes between 3 and 100 when $p = 0.50$ and $1-\alpha = 0.95$. In other words, Figure 1 summarizes sample sizes required if the experimenter is prepared to accept errors in population size up to 50 percent but requires 95 percent certainty that errors will not exceed 50 percent. Figure 2

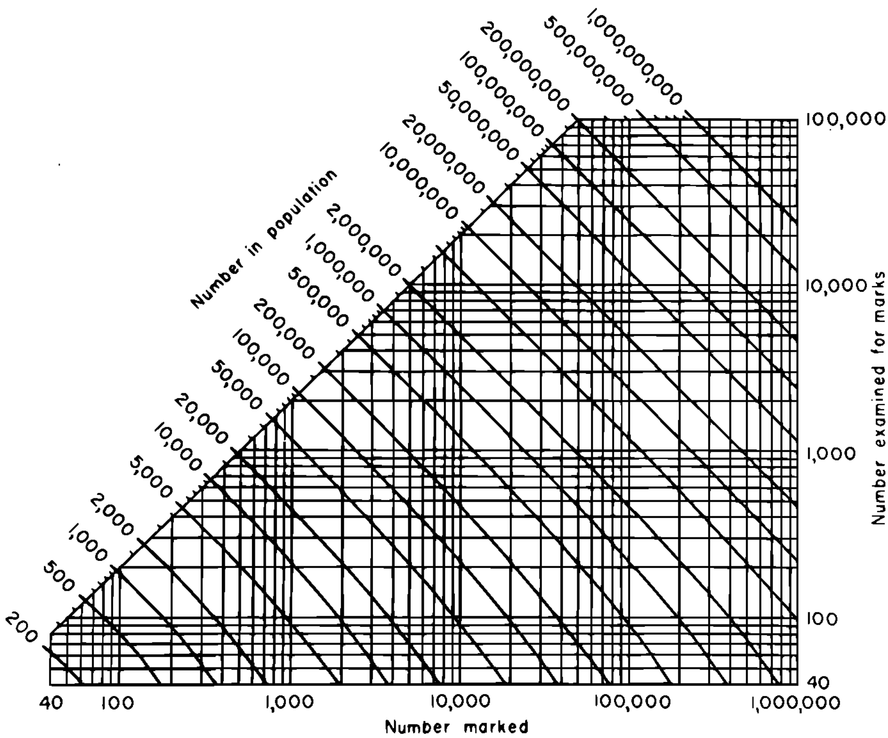

FIGURE 4. – Sample size when $1-\alpha = 0.95$ and $p = 0.50$; recommended for preliminary studies and management surveys. Data based on normal approximation to the hypergeometric distribution.

has similar graphs for 95 percent confidence that errors will not exceed 25 percent, and Figure 3 for 95 percent confidence that errors will not exceed 10 percent.

The curves drawn in the figures denote approximate sample sizes. Because of the rounding necessary in equation (5), it is not always possible to find one pair of numbers M and C such that all larger values satisfy equation (5) and all smaller values do not satisfy it. Thus for $p = 0.50$, $1-\alpha = 0.95$, and $N = 100$, the following combinations of M and C satisfy equation (5): 30, 30; 32, 32; 34, 34; 35, 35; 36, 36; etc. But for the same p, $1-\alpha$, and N the following do not satisfy it: 28, 28; 29, 29; 31, 31; 33, 33. Such interspersion of acceptable and nonacceptable sample sizes is due solely to rounding necessary in equation (5). Little interspersion occurs with $p = 0.10$, more occurs at $p = 0.25$, and still more at $p = 0.50$. The curves drawn in Figures 1 to 3 are a compromise in this respect and exclude about as many acceptable points as they include nonacceptable points. The curves thus include some sample sizes for which $1-\alpha$ is somewhat less than 0.95, but seldom less than 0.92. The band of interspersion thus presents no serious problem.

Tables of the hypergeometric distribution become very bulky at population sizes over 100 and have not yet been tabulated. However, over certain ranges of M, C, and N simpler distributions (binomial, Poisson, and normal) are satisfactorily close approximations of the hypergeometric (see Lieberman and Owen, 1961, for a recent discussion). Figures 4 to 6 are derived from the normal approximation. The method of deriving sample sizes using the normal approximation follows.

Start again with equation (2)

$$1-\alpha \leqslant P\left(-p < \frac{\hat{N}-N}{N} < p\right). \qquad (2)$$

Substitute MC/R for $\hat{N}$ in equation (2), consider N fixed and R varying. Then by rearrangement

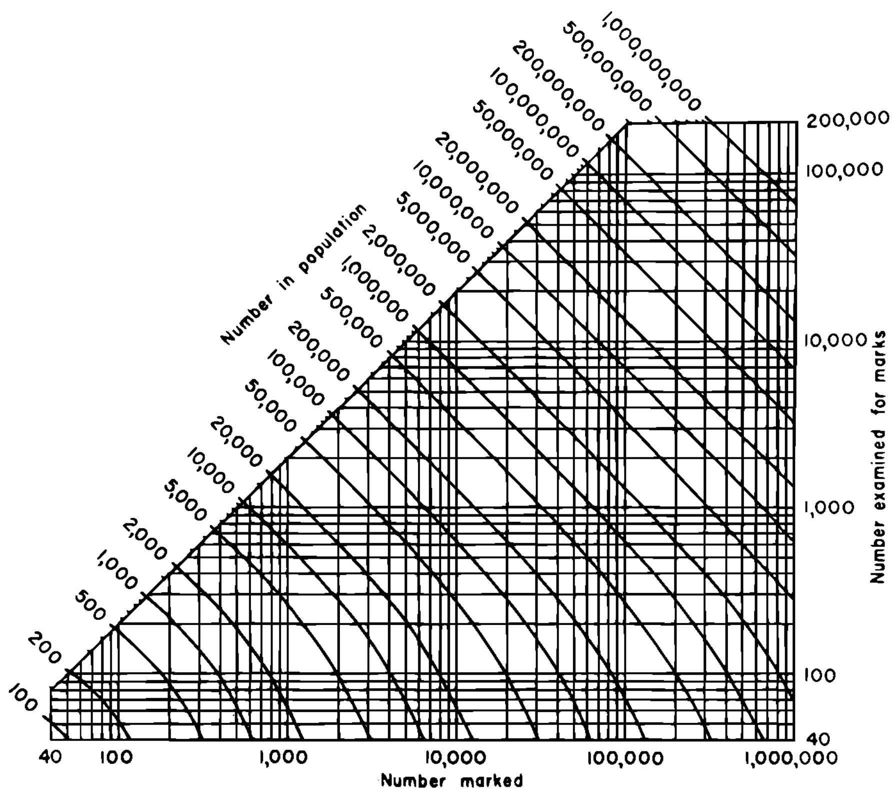

FIGURE 5.—Sample size when $1-\alpha=0.95$ and $p=0.25$; recommended for management studies. Data based on normal approximation to the hypergeometric distribution.

$$1-\alpha=P\left(\frac{MC}{(1+p)N}<R<\frac{MC}{(1-p)N}\right).$$

Classically the normal approximation to the hypergeometric is found by setting

$$\mu=\frac{MC}{N}$$

$$\sigma^2=\frac{M(N-M)C(N-C)}{N^2(N-1)}.$$

After substitution and rearrangement we obtain

$$1-\alpha=\Phi\left(\frac{p}{1-p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right)-\Phi\left(-\frac{p}{1+p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right)\quad(7)$$

where $\Phi(x)$ is the cumulative standard normal probability distribution.

Figures 4 to 6 were obtained from a large number of iterative solutions of equation (7) for $1-\alpha=0.95$, $p=0.50$, 0.25, or 0.10, and N ranging from 100 to 1,000,000,000. The iterations may be done in two stages. First let

$$\frac{MC(N-1)}{(N-M)(N-C)}=D$$

and solve equation (7) for D for a particular combination of p and $1-\alpha$. (Table 2 lists D for various combinations.) In the second step fix $N=N_1$ for a particular $D=D_1$ (in which p and $1-\alpha$ are fixed) and iteratively derive various combinations of M and C, *i.e.*,

$$\frac{MC(N_1-1)}{(N_1-M)(N_1-C)}=D_1.$$

Such combinations of M and C may be plotted as in Figures 4 to 6.

Only parts of the ranges in combinations of M and C for each population size are drawn in Figures 4 to 6 (*cf.* Figures 1 to 3). The relationships are symmetric for M and C, thus the remaining combinations of M and C are

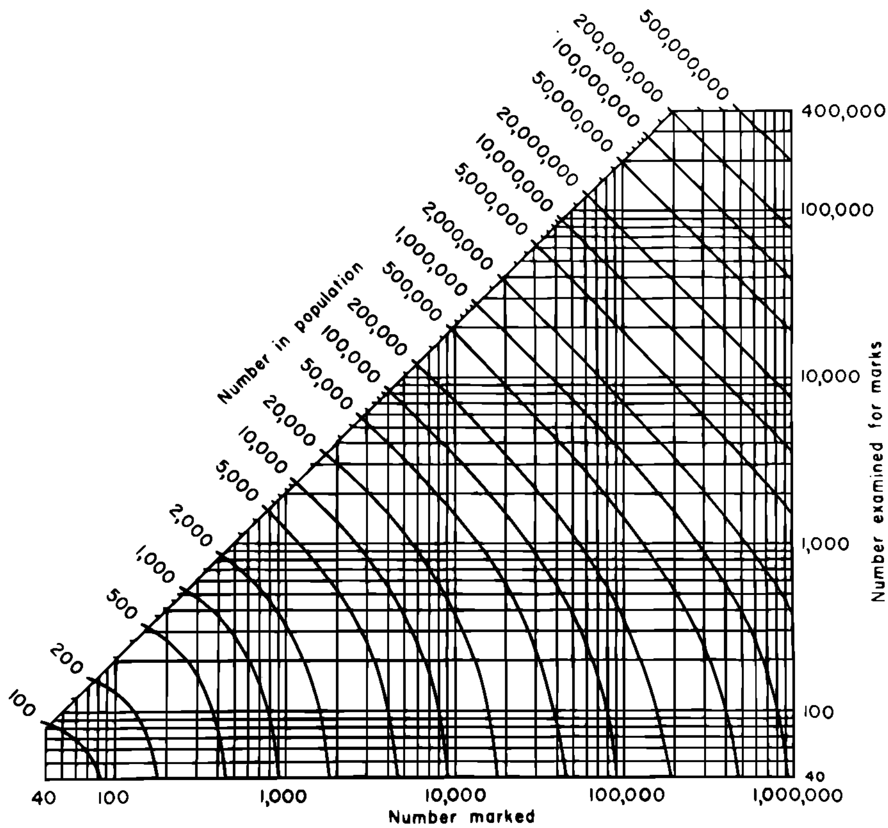

FIGURE 6.—Sample size when $1-\alpha = 0.95$ and $p = 0.10$; recommended for research. Data based on normal approximation to the hypergeometric distribution.

determined simply by mentally reversing the x- and y-axes. Thus sample sizes ($M = 300$, $C = 600$) will yield an estimate as accurate and precise as sample sizes ($M = 600$, $C = 300$).

Sample sizes less than 40 were not graphed since the normal approximation is not sufficiently close at such small sample sizes. The lines for smaller sample sizes could have been derived using either the Poisson or binominal approximations to the hypergeometric but it was not considered worth the effort. Experiments utilizing markedly unequal sample sizes (either large M and very small C or vice versa) would almost certainly be considerably more costly than if more nearly equal-sized samples were used (see below).

The fourth line in Table 2 states that the probability is 5 in 100 that the population estimate will be in error by more than 50 percent if $D = 24.4$. The experimenter may desire to know the probability that the estimate will be in error by more than, say, 25 percent with the same D. The answer is obtained again from equation (7), but now considering D and p fixed. Table 3 lists various combinations of α and p which when combined with the D's from which Figures 4 to 6 were derived satisfy equation (7). As an example, suppose an experimenter decides that he will tolerate an error greater than 50 percent with probability less than $100 - 95 = 5$ percent (Table 3, line 2). It follows that he is also prepared to accept an error greater than 25 percent if its probability is less than $100 - 79 = 21$ percent (line 3). Furthermore, his chances are 62 percent of being in error by more than 10 percent (line 4).

We showed earlier that the Petersen esti-

TABLE 2.—*Values of* D *satisfying equation (7) for selected* α *and* p. *Last column shows sample size, where* M = C, *to satisfy this relationship for* N = *1,000*

$1-\alpha$	p	D	$M=C$
0.75	0.50	4.75	65
0.90	0.50	14.8	108
0.90	0.25	45.5	176
0.95	0.50	24.4	135
0.95	0.25	69.9	210
0.95	0.10	392	385
0.99	0.10	695	455
0.99	0.01	66,300	997

TABLE 3.—*Combinations of* p *and* $1-\alpha$ *that satisfy equation (7) for the* D*'s used in Figures 4 to 6*

D	p	$1-\alpha$
24.4	0.88	0.99
24.4	0.50	0.95
24.4	0.25	0.79
24.4	0.10	0.38
24.4	0	0
69.9	0.39	0.99
69.9	0.25	0.95
69.9	0.10	0.60
69.9	0	0
392	0.14	0.99
392	0.10	0.95
392	0.05	0.75
392	0	0

mate was virtually unbiased if either $M+C > N$, or $MC > 4N$. Trial computations with sample sizes specified in Figures 1 to 3 indicate that these sample sizes generally satisfy the above approximate criteria. Small sections in the top left and bottom right corners of Figure 3 specify sample sizes that do not quite satisfy these criteria, but when these sample sizes are used in the bias equation then the bias is found to be less than 1 percent. Thus the total ranges of sample sizes in Figures 1 to 6 are for practical purposes free of the bias related to the probability of no recaptures.

We have outlined in some detail the steps required to determine various combinations of M and C that will ensure a satisfactorily accurate (p) and precise ($1-\alpha$) estimate of the population size (N) if a fairly good guess of population size was made and if the assumptions on which the Petersen estimate is based were not violated. We now turn to the question of which of the many possible combinations of M and C to use in a particular experiment.

OPTIMUM ALLOCATION OF RESOURCES

The total cost of an experiment consists of fixed plus variable costs. Fixed or overhead costs are independent of sample size and are not considered further here. Variable costs depend on sample size, and we assume here that such costs increase in direct proportion to sample size. Variable costs are then $mM + cC$ where m and c are respectively the per fish costs of catching-marking and catching–examining.

Emphasis in planning may be on achieving highest possible accuracy of estimation with restricted resources, or on achieving a preassigned accuracy with unrestricted resources. Estimates of two parameters are required in each case: the population size and the relative costs of catching-marking and catching–examining. Preliminary estimates of these data can usually be made based on the literature, experience, or a pilot study.

(a) Minimum error with limited funds

If available funds are limited, then optimum resource allocation is here considered to be achieved by maximizing the probability $(1-\alpha)$ that the error of estimate will be less than the preassigned $100p$ percent.

The formula for optimum allocation is achieved by differentiating relation (7) as a function of M and C, subject to the constraint that

$$mM + cC = \text{a constant}, K \ .$$

Since $\Phi(x)$ is a normal probability distribution, this operation will be equivalent to minimizing the variance.

$$1-\alpha = \Phi\left(\frac{p}{1-p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right) - \Phi\left(-\frac{p}{1+p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right) \quad (7)$$
$$= \Phi(x) - \Phi(y) \ .$$

Differentiating by the chain rule and using the functional relation between C and M

$C = \frac{1}{c}(K - mM)$ where K is constant, then

$$\frac{d(1-\alpha)}{dM} = \frac{d\Phi}{dx}\cdot\frac{dx}{dM} - \frac{d\Phi}{dy}\cdot\frac{dy}{dM}$$

$$= \left(\frac{\mathrm{d}}{\mathrm{d}M}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right) \cdot \left(\frac{p}{1-p}\cdot\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \frac{p}{1+p}\cdot\frac{\mathrm{d}\Phi}{\mathrm{d}y}\right).$$

Setting $\mathrm{d}(1-\alpha)/\mathrm{d}M = 0$, it follows that

$$\frac{mM}{cC} = \frac{N-C}{N-M}. \tag{8}$$

Since $mM + cC$ is known (*i.e.*, total resources available) and rough estimates of N and m/c are available, relation (8) can be solved directly for M and C. With large N, both M and C are small relative to N (see Figures 4 to 6) so that relation (8) becomes simply $mM = cC$. Thus *an equal division of resources between catching–marking and catching–examining is optimal.*

(b) Minimum costs with preset accuracy and precision

We have described above the method of determining optimum sample size when resources are limited. We now consider the case where the experimenter desires to estimate the cost of an experiment in which preset levels of p and $1-\alpha$ are to be obtained.

The optimal sample sizes M and C are now constrained to satisfy equation (7), and among all values satisfying (7) we wish to find a pair which minimizes the total cost of experimentation; that is, which minimizes

$$\text{cost} = \text{fixed overhead costs} + mM + cC\,.$$

Introducing a Lagrangian multiplier λ, we may rephrase the problem as one of minimizing

$$F(M, C, \lambda) = mM + cC + \lambda[(1-\alpha) - \Phi(x) + \Phi(y)]$$

where $\Phi(x)$ and $\Phi(y)$ are as defined previously. The roots of the differential equations $\partial F/\partial M = 0$ and $\partial F/\partial C = 0$ will then be seen to satisfy relation (8). With M and C related by (8) the function $\Phi(x) - \Phi(y)$ is monotonic in either M or C, implying that a unique solution exists for this minimum cost problem.

The data required to estimate the cost are m/c, a rough estimate of N and the chosen p and $1-\alpha$. If the chosen p and $1-\alpha$ correspond to one of the three combinations on which Figures 1 to 6 are based, then the relevant figure may be consulted to obtain a large number of pairs of M and C that will satisfy the preset levels of p and $1-\alpha$. But only one combination will minimize costs. This optimum pair is obtained iteratively by testing various combinations by inserting them in the relation

$$\frac{m}{c} = \frac{C(N-C)}{M(N-M)}$$

obtained by rearranging relation (8). The combination of M and C that satisfies the above equation is the pair that minimizes cost at the preset level of p and $1-\alpha$.

The use of the figures as above requires interpolation when the guessed population size is different from those shown. If the experimenter prefers a combination of p and $1-\alpha$ different from those in the figures, then the sample size may be obtained as follows. Consider the three equations previously described:

$$D = \frac{MC(N^*-1)}{(N^*-M)(N^*-C)},$$

$$\frac{m}{c} = \frac{C(N^*-C)}{M(N^*-M)},$$

$$N^* \simeq N\,.$$

D is obtained either from Tables 2 or 3 (if the required combination of p and $1-\alpha$ are tabulated there) or by the method described earlier. Relative costs, c/m, are known, as before, and a rough estimate of N is available. The three equations may then be solved for M and C. The use of D here presupposes that population and sample sizes are such that the normal distribution is a good approximation to the hypergeometric (*i.e.*, $N > 150$, $M > 50$, $C > 50$).

ALTERNATIVE METHODS OF CAPTURE

Two or more alternative methods of capture may be available. A choice of the less costly can readily be made either by calculating and comparing total costs of alternative methods, or by applying relations given below.

Assume overhead (fixed) costs equal in the alternative models. Let M_1, C_1, and M_2, C_2 be optimum sample totals under two different cost systems (m_1, c_1) and (m_2, c_2) at fixed levels of α and p. Since

$$1-\alpha=\Phi\left(\frac{p}{1-p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right)$$
$$-\Phi\left(-\frac{p}{1+p}\sqrt{\frac{MC(N-1)}{(N-M)(N-C)}}\right),$$

it follows that

$$\frac{M_1C_1}{(N-M_1)(N-C_1)}=\frac{M_2C_2}{(N-M_2)(N-C_2)}.$$

Furthermore, since N will normally be large relative to M, C, and MC/N this relation is essentially

$$M_1C_1=M_2C_2 .$$

The corresponding optimality conditions $m_1M_1 = c_1C_1$, $m_2M_2 = c_2C_2$ thus imply that the total sample sizes for the two methods of capture must stand in the ratio

$$\frac{M_2+C_2}{M_1+C_1}=\frac{\sqrt{\dfrac{m_2}{c_2}}+\sqrt{\dfrac{c_2}{m_2}}}{\sqrt{\dfrac{m_1}{c_1}}+\sqrt{\dfrac{c_1}{m_1}}}$$

and the ratio of total cost must be

$$\frac{m_2M_2+c_2C_2}{m_1M_1+c_1C_1}=\frac{m_2M_2}{m_1M_1}=\sqrt{\frac{m_2c_2}{m_1c_1}}.$$

For example, if the second method of capture costs twice as much per captured fish as the first method and both methods utilize the same marking techniques, then we might have

$$m_1=m+c \qquad c_1=c$$
$$m_2=m+2c \qquad c_2=2c$$

so that

$$\sqrt{\frac{m_2c_2}{m_1c_1}}=\sqrt{\frac{2(m+2c)}{m+c}}$$

as the ratio of total costs of the two methods. If the method of capture is the same in both cases but the second method employs a marking technique which costs twice as much as that used in the first, we might have

$$m_1=m+c \qquad c_1=c$$
$$m_2=2m+c \qquad c_2=c$$

giving the ratio of total costs as

$$\sqrt{\frac{m_2c_2}{m_1c_1}}=\sqrt{\frac{2m+c}{m+c}}.$$

DISCUSSION

We have discussed bias, accuracy, and precision of the Petersen mark–recapture experiment as affected by sample size and ultimately by resources allocated to the experiment. Our purpose was to provide as simple a guide as possible to effective planning of such experiments. None of the concepts in this paper are new, but the fact that proper mark–recapture experiments are not easily designed led us to believe that a more detailed statistical guide than presently available (*e.g.*, in Chapman, 1948; Ricker, 1958) would be useful to experimenters and administrators.

It should be strongly emphasized that all relations stated in this paper assume that all practical conditions required for valid use of the Petersen estimate are met. If this is not the case, then sample sizes as specified here will not assure the desired accuracy and precision. In short, intelligent use of the Petersen estimate requires not only estimates of sample sizes but also a considerable knowledge of fish behavior, fish movements, gear bias, etc., as well as skill in obtaining proper samples.

We singled out three combinations of $1-\alpha$ and p for graphing: (0.95, 0.50), (0.95, 0.25), and (0.95, 0.10). *We suggest that these be considered standard levels of precision and accuracy.* The lowest level of accuracy, $p =$ 0.50, might be satisfactory for preliminary studies or in management surveys where only a rough idea of population size and composition is required. For more accurate management work the intermediate level may be acceptable. For careful research into population dynamics, the highest level (0.95, 0.10) should become routine. Population estimates are often separated into age-groups, or are multiplied or divided by other estimates (*e.g.*, multiplied by mean weight to obtain total biomass, divided by another population estimate to obtain survival) and relatively high precision and accuracy are required of each estimate if the variance of the product or ratio of such estimates is to be reasonably small.

ACKNOWLEDGMENTS

Preliminary work on this paper was done while the authors were at Princeton University and with the Ontario Department of Lands and Forests, respectively. We thank Dr. A. W. Eipper of Cornell University for criticism of an earlier paper on this subject which was presented at the Northeast Fish and Wildlife Conference in 1962.

LITERATURE CITED

BAILEY, N. J. J. 1951. On estimating the size of mobile populations from recapture data. Biometrika, 38: 293–306.

CHAPMAN, D. G. 1948. A mathematical study of confidence limits of salmon populations calculated from sample tag ratios. Internat. Pacific Salmon Fisheries Comm. Bull., 2: 69–85.

———. 1951. Some properties of the hypergeometric distribution with applications to zoological censuses. Univ. California Publ. Stat., 1(7): 131–160.

———. 1952. Inverse, multiple and sequential sample censuses. Biometrics, 9(4): 286–306.

LIEBERMAN, G. J., AND D. B. OWEN. 1961. Tables of the hypergeometric probability distribution. Stanford Univ. Press. 726 pp.

RICKER, W. E. 1958. Handbook of computations for biological statistics of fish populations. Fisheries Res. Bd. of Canada Bull., 119. 300 pp.