

## NOTES

### Sample Sizes for Single Mark and Single Recapture Experiments

A. L. JENSEN

*School of Natural Resources, University of Michigan  
Ann Arbor, Michigan 48109*

#### Abstract

Calculation of the sample sizes necessary to attain a desired level of precision should be an important step in the design of a mark-and-recapture experiment. Equations that easily can be applied for this purpose have not been available. In this report, simple equations for calculation of the sample sizes necessary to obtain a desired degree of precision with high certainty are developed for direct single mark and single recapture experiments. If the costs of marking and searching for marks are known in terms of dollars or effort, the equations give the optimum allocation of effort between marking and looking for marks.

Calculation of the sample sizes necessary to obtain a desired level of precision with a high degree of certainty is an important consideration in design of mark-and-recapture experiments. However, equations that easily can be applied for calculation of sample sizes have not been developed for mark-and-recapture studies.

The sample sizes necessary to attain a certain level of precision depend on the design of the mark-and-recapture experiment. Many different approaches are possible but only a few are applied widely. The method of direct single mark and single recapture is applied commonly in fisheries and wildlife studies. The advantages of this approach are its simplicity and the short duration of such experiments; the latter increases the likelihood that assumptions concerning mortality and reproduction will be better realized (Seber 1973).

Robson and Regier (1964) examined the problem of finding the sample sizes necessary to give an error bound of a desired size with a high degree of certainty. They assumed sampling without replacement and applied the hypergeometric distribution to describe the results of a direct single mark and single

recapture experiment. For a population of more than 100 individuals they applied the normal approximation to the hypergeometric. The assumption of sampling without replacement results in equations that are difficult to solve and a computer was applied to calculate sample sizes for a 95% error bound at three levels of precision for several population sizes. These results were summarized in a useful series of graphs. Robson and Regier (1964) also examined the optimum allocation of effort between the first and second sample, but the resulting equations must be solved iteratively. Their charts do not indicate optimum sample sizes.

In this report simple equations are developed for calculation of sample sizes that will give a desired error bound with a high degree of certainty for studies of closed populations by direct single mark and single recapture. If the cost of marking and the cost of looking for marks are known in terms of dollars or effort the equations give sample sizes that result in a minimum variance for a given amount of effort.

#### Direct Single Mark and Single Recapture

To estimate abundance by single mark and single recapture, a sample of  $n_1$  animals is captured, marked, and released. Later a second sample of  $n_2$  animals is captured and the number of marked animals in the sample,  $m_2$ , is determined. To develop a mathematical model for such a survey we assume (1) the population is closed; (2) samples are selected at random (at least the second one); (3) animals do not lose marks; (4) marks on recaptured animals are always detected. When the second sample is taken the animals in the population are of two types—marked and unmarked. Two different models have been applied: the binomial model for sampling with replacement and the hypergeometric model for sampling without replacement.

If in the second sample animals are caught and released in a series of trials, the probability of obtaining  $m_2$  marked animals in  $n_2$  trials is given by the binomial distribution. Application of the method of maximum likelihood, a conventional technique for obtaining estimators of

parameters of probability models (Seber 1973), gives the following estimator of population size ( $N$ ):

$$\hat{N}_1 = n_1 n_2 / m_2; \quad (1)$$

where  $m_2$  is the number of marked animals in the second sample. An estimator is an equation applied to calculate an estimate from sample data, and an estimate is a number calculated from the estimator. The variance ( $V$ ) of the above estimator of  $N$  is approximately (Bailey 1951)

$$V(\hat{N}) = \frac{N^2(N - n_1)}{n_1 n_2}. \quad (2)$$

To estimate the variance from sampling data the following equation is applied:

$$\hat{V}(\hat{N}) = \frac{\hat{N} n_1 (n_2 - m_2)}{m_2^2}. \quad (3)$$

Bailey (1951) showed the above estimator of  $N$  is biased and Robson and Regier (1964) state the estimator is nearly unbiased if  $(n_1 + n_2) > N$  or  $n_1 n_2 > 4N$ .

In practice, sampling usually is done without replacement and the probability of obtaining  $m_2$  marked animals in  $n_2$  trials is given by the hypergeometric distribution. The maximum likelihood estimator of population size for this model was obtained by Chapman (1951). In most practical applications there is little difference between the result obtained with the above two methods (Seber 1973). Chapman (1951) suggested that the binomial distribution is a suitable approximation for the hypergeometric distribution when  $m_2/n_2 > 0.1$ . Either the normal or the Poisson distribution is recommended as an approximation in other situations. But both the Poisson and normal distributions are approximations to the binomial and Cormack (1968) states that the binomial is a suitable theoretical distribution in cases where the Poisson and normal might be applied for practical reasons. Application of the binomial gives simple equations for calculation of sample size.

### Calculation of Sample Size

Development of equations for calculation of sample sizes for mark-and-recapture experiments is more difficult than development of similar equations for survey sampling because there is an additional variable to consider. Two

relations between  $n_1$  and  $n_2$  are necessary in order to develop equations for calculation of  $n_1$  and  $n_2$ . To develop the first relation, the sample sizes will be related to the error bound in terms of the sample sizes necessary to obtain a desired level of precision with high certainty. To develop a second equation relating  $n_1$  and  $n_2$ , a cost equation is developed that gives total cost as a function of the number of animals marked and the number of animals examined for marks. In fisheries work cost most frequently will be measured in terms of time. The cost equation could be applied directly with the equation for the error bound to calculate  $n_1$  and  $n_2$ , but a better approach is to apply the cost equation to find values of  $n_1$  and  $n_2$  that give the most precise estimates for a given effort.

The first relation between  $n_1$  and  $n_2$  is given by the sample size necessary to obtain an error bound of a desired size with a high degree of certainty. The error bound,  $B = 2\sqrt{V(\hat{N})}$ , is an interval estimate that is applied widely in survey sampling for determination of sample size (Mendenhall et al. 1971). In a long series of trials the parameter being estimated will be within the interval described by the error bound ( $\hat{N} - B, \hat{N} + B$ ) at least 75% of the time regardless of the shape of the distribution, provided that the distribution has a finite mean and variance. If the distribution is normal, the 75% error bound is equivalent to a 95.45% confidence interval (Fowler and Hauke 1979). For estimation of population with a single mark and a single recapture survey without replacement, the error bound is

$$B = 2\sqrt{\frac{N^2(N - n_1)}{n_1 n_2}}. \quad (4)$$

The second relation necessary to find  $n_1$  and  $n_2$  is developed from the cost equation. In a mark and recapture survey the sampling effort must be allocated between marking animals and examining animals for marks. A cost equation for such a study is

$$C = C_o + C_1 n_1 + C_2 n_2; \quad (5)$$

where  $C$  = total cost of the survey;

$C_o$  = overhead cost;

$C_1$  = cost per individual of capture and marking;

$C_2$  = cost per individual of capture and examination for marks;

- $n_1$  = number of individuals captured, marked, and released in first sample;  
 $n_2$  = number of individuals captured in second sample and examined for marks.

Cost can be measured in any relevant units such as dollars or the amount of time required to capture and mark animals as compared to the amount of time required to examine animals for marks.

The cost equation could be applied together with the equation for the error bound to calculate  $n_1$  and  $n_2$ . However, the variance of the estimate,  $V(\hat{N})$ , depends on how the available effort is allocated to marking and looking for marks, so a better approach is to find the values of  $n_1$  and  $n_2$  that minimize the variance. If the sample sizes that minimize the variance are sought directly by application of calculus it is found that  $n_1$  and  $n_2$  are infinitely large. To obtain useful results it is necessary to minimize the variance subject to the constraint of the cost equation that places a limit on the available resources. The mathematical method for minimization of a function subject to a linear constraint is termed the Lagrange multiplier method. Application of the Lagrange multiplier method (Appendix 1) to minimize the variance subject to the linear constraint of the cost equation gives

$$n_1 = n_2 \left( \frac{c_2}{c_1} \right) \left( \frac{N}{N - n_1} \right). \quad (6)$$

If  $n_1$  is much less than  $N$ , equation (6) becomes  $n_1 = (c_2/c_1)n_2$ , so if marking is expensive fewer animals should be marked and more animals should be examined for marks. If the cost of marking is low, more animals should be marked and fewer should be examined for marks. If marking and checking for marks require the same effort or cost, sample sizes should be equal. Similar conclusions and an equation similar to equation (6) were obtained by Robson and Regier (1964).

Solution of equations (4) and (6) for  $n_1$  and  $n_2$  gives (Appendix 2)

$$n_1 = \sqrt{\frac{4N^3C_2}{B^2C_1}}; \quad (7)$$

$$n_2 = \sqrt{\frac{4N^3C_1}{B^2C_2}}. \quad (8)$$

In applying equations (7) and (8),  $C_1$  and  $C_2$  are assumed known or equal,  $B$  is the error bound established by the investigator, and  $N$  is a guess or estimate of population size. Population size can be estimated in a pilot study, from a previous study, or simply guessed if no other information is available.

An example will illustrate the ease with which these equations can be applied. Suppose  $N$  is assumed to be 1,000,  $C_1 = C_2$ , and an error bound of 50% of  $N$  is desired with high certainty. The necessary sample sizes are:

$$n_1 = n_2 = \sqrt{\frac{4(1,000)^3}{500^2}} = 126.$$

This example was set up so the results could be compared with those given by Robson and Regier (1964); from their figure with  $1 - \alpha = 0.95$ ,  $n_1 = n_2 \doteq 130$ . Now, suppose  $C_1 = 2C_2$ , that is, marking requires twice the time or cost of looking for marks. Then

$$n_1 = \sqrt{4(1,000)^3/(2)(500)^2} = 89;$$

$$n_2 = \sqrt{(2)(4)(1,000)^3/(500)^2} = 179;$$

In summary, equations (7) and (8) give the necessary sample sizes to obtain an error bound of a desired size with a high degree of certainty. If the cost of marking and searching for marks is known, the equations give the optimum sample size.

### Acknowledgments

I thank Gary Fowler and Richard Patterson for reviewing the manuscript.

### References

- BAILEY, N. T. J. 1951. On estimating the size of mobile populations from recapture data. *Biometrika* 38:293-306.
- BARTLE, R. G. 1964. The elements of real analysis. John Wiley and Sons, New York, New York, USA.
- CHAPMAN, D. G. 1951. Some properties of the hypergeometric distribution with applications to zoological censuses. University of California Publications in Statistics 1:131-160.
- CORMACK, R. M. 1968. The statistics of capture-recapture methods. *Oceanography and Marine Biology an Annual Review* 6:455-506.
- FOWLER, G. W., AND D. HAUKE. 1979. A distribution free method for interval estimation and sample size determination. *Resource Inventory Notes* 19:1-8.
- MENDENHALL, W., L. OTT, AND R. L. SCHEAFFER.

1971. Elementary survey sampling. Wadsworth, Belmont, California, USA.
- ROBSON, D. S., AND H. A. REGIER. 1964. Sample size in Petersen mark-recapture experiments. Trans-
- actions of the American Fisheries Society 93:215-226.
- SEBER, G. A. F. 1973. Estimation of animal abundance. Griffin, London, England.

### Appendix 1. Derivation of Equation (6)

The Lagrange multiplier method is based on the theorem: given that  $V(n_1, n_2)$  and  $G(n_1, n_2)$  (a function defined below) are continuous functions of  $n_1$  and  $n_2$ , that  $G(n_1, n_2) = 0$ , and that  $g$  is a relative maximum or minimum of  $V(n_1, n_2)$ ; then there exists a real number  $\lambda$  and a function  $F = V(n_1, n_2) + \lambda G(n_1, n_2)$  called the Lagrange function, such that at  $g$ ,  $\partial F / \partial n_1 = \partial F / \partial n_2 = 0$ . A more detailed statement of the theorem and a proof can be found in Bartle (1964).

In this case, the function to be minimized as a function of  $n_1$  and  $n_2$  is the variance

$$V(n_1, n_2) = \frac{N^2(N - n_1)}{n_1 n_2};$$

the linear constraint is the equation

$$G(n_1, n_2) = C - C_0 - C_1 n_1 - C_2 n_2 = 0.$$

The Lagrange function,  $F(n_1, n_2)$ , is

$$F(n_1, n_2) = \frac{N^2(N - n_1)}{n_1 n_2} + \lambda(C - C_0 - C_1 n_1 - C_2 n_2).$$

Partial derivatives with respect to  $n_1$  and  $n_2$ , set equal to zero, give

$$\begin{aligned} \frac{\partial F}{\partial n_1} &= \frac{n_1 n_2(-N^2) - N^2(N - n_1)n_2}{(n_1 n_2)^2} - \lambda C_1 = 0; \\ \frac{\partial F}{\partial n_2} &= \frac{-N^2(N - n_1)n_1}{(n_1 n_2)^2} - \lambda C_2 = 0. \end{aligned}$$

Solution of each of these equations for  $\lambda$ , and then substitution for  $\lambda$ , give

$$\frac{N^3 n_2}{C_1 (n_1 n_2)^2} = \frac{N^3 n_1 - N^2 n_1^2}{C_2 (n_1 n_2)^2}.$$

Solution for  $n_1$  gives equation (6).

### Appendix 2. Derivation of Equation (7) and (8)

If equation (4) is squared,  $\frac{B^2}{4} = \frac{N^2(N - n_1)}{n_1 n_2}$ . This squared equation, rearranged, gives  $n_2 = \frac{4N^2(N - n_1)}{n_1 B^2}$ . Substitution of the above equation into equation (6) gives equation (7) and substitution of equation (7) into the above equation gives equation (8).