Yugoslav Journal of Operations Research
8 (1998), Number 2, 215-224

ESTIMATION OF FISH POPULATION SIZE BY SINGLE MARKING AND RECAPTURE METHOD*

Umit DEMIR, Mustafa AKAR

University of Cukurova, Faculty of Fisheries, Turkey

Communicated by Byron Papathanassiou

Abstract: The single marking and recapture method is one of the methods commonly used to estimate fish population size. There are four formulae in this method, two of which are suggested by Bailey (1951), one by Chapman (1952) and the last by Ricker (1975). For this reason, in the current study, firstly, one of Bailey's formulae (1951) was examined mathematically and statistically. Secondly, numerical data obtained from Duzgunes (1985) were used and his results from Mogan Lake were compared. As a result, population size estimation of the carp (*Cyprinus carpio L.*) in this lake was calculated using the four formulae as 9716±360, 9381±3354 (Bailey, 1951), 9409±3218 (Chapman, 1952) and 9410±3364 (Ricker, 1975). It is suggested that as the confidence interval is narrower in the formula proposed by Chapman (1952) than the other formulae, the results obtained using this formula are more reliable.

Keywords: Capture, recapture, fish population size.

1. INTRODUCTION

The single marking and recapture method has been widely used by field biologists and ecologists to investigate the dynamics of biological populations. For a closed population the classical problem for capture-recapture experiments is the estimation of the population size. For this reason, this study was carried out to determine which formula suggested by Bailey (1951), Chapman (1952) and Ricker (1975) is more reliable to estimate fish population size in the single marking and recapture method.

---
* Presented at 4th Balkan Conference on Operational Research

2. MATERIAL AND METHOD

In this study, numerical data obtained from Duzgunes (1985) were used. In addition the single marking and recapture method is given below.

2.1. Single Marking and Recapture Method (Petersen Method)

The Petersen method is the simpliest method known as the Petersen or Lincoln Index method. In this study, we used the following slightly different notation to that used by Ricker (1975). For instance r is the number of individuals in the second sample that are marked (R. Ricker terminology in 1975)

N = Size of population at the time of marking

M = Number of fish marked in the first sample

C = Total number of fish captured for census in the second sample

r = Number of fish in the second sample that are marked

To estimate abudance by the single marking and recapture method, a sample of M animals is captured, marked and released. Later a second sample of C animals is captured and the number of marked animals in the sample, r, is determined. In this situation the probability of catching the marked individuals in the population will be $p = M \ N$.

To develop a mathematical model for such a survey, marking experiments are based on models with assumptions as follows:

1. The population is closed to additions (births or immigrants) and deletions (deaths or emigrants)
2. All animals are equally likely to be captured in each sample
3. Marks are not lost and are not overlooked by the observer
4. Animals do not lose marks between the two sampling periods
5. Marking individuals does not affect their catchability

Other considerations

- Accidental deaths do not affect the first assumption
- The tagged fish are randomly distributed in the population
- All tagged fish are reported

When the second sample is taken the fish 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 sampling is without replacement then the model that will be used is the hypergeometric. In this situation, the function of hypergeometric distribution is formulated as (De Groot, 1986):

$$P(r) = \frac{C_r^{Np} C_{C-r}^{Nq}}{C_C^N}, \quad r = 0,1,2,... \tag{1}$$

But while $N \to \infty$ this distribution approaches the binomial distribution. Thus:

$$\lim_{N \to \infty} \{P(r)\} = C_r^C p^r q^{C-r} \tag{2}$$

At this point, $p = M/N$ and $q = (N-M)/N$. So if N is fairly big then there is no important difference between the sampling without replacement and the replacement sampling. For this reason, population size can be found by the binomial distribution.

The maximum Likelihood Method that is used by Bailey (1951) to estimate population size is written below for the present study:

$$e^L = \binom{C}{r} \left(\frac{M}{N}\right)^r \left(\frac{N-M}{N}\right)^{C-r} \tag{3}$$

This equation can be solved for N. Thus:

$$L = \ln\binom{C}{r} + r\ln\left(\frac{M}{N}\right) + (C-r)\ln\left(\frac{N-M}{N}\right) \tag{4}$$

$$\frac{\partial L}{\partial N} = 0 \Rightarrow$$

$$\frac{\partial L}{\partial N} = r\left(\frac{-M/\hat{N}^2}{M/\hat{N}}\right) + (C-r)\left(\frac{[\hat{N} - (\hat{N}-M)]/\hat{N}^2}{(\hat{N}-M)/\hat{N}}\right)$$

$$\frac{\partial L}{\partial N} = -\frac{r}{\hat{N}} + (C-r)\left(\frac{M}{\hat{N}(\hat{N}-M)}\right) \tag{5}$$

$$\frac{\partial L}{\partial N} = 0$$

$$-r(\hat{N} - M) + M(C-r) = 0$$

$$-r\hat{N} + rM + MC - Mr = 0$$

$$-r\hat{N} = -MC$$

Thus we find that:

$$\hat{N} = MC/r \tag{6}$$

$\hat{N}$ is the Maximum Likelihood estimate of population size. This value is known sometimes as the Lincoln Index (Seber, (1982); cited in Kenneth et al., (1990)). A number of modifications of the basic equation (6) are given in Table 1.

Table 1: Formulae for estimating population size (N) by the Petersen Method.

Reference	Type of sampling	Population size (N)	Variance of ($\hat{N}$)
Bailey (1951)	Direct	$\hat{N} = \frac{MC}{r}$	$\mathrm{var}(\hat{N}) = \frac{M^2C(C-r)}{r^3}$
Bailey (1951)	Direct	$\hat{N}_B = \frac{M(C+1)}{r+1}$	$\mathrm{var}(\hat{N}_B) = \frac{M^2(C+1)(C-r)}{(r+1)^2(r+2)}$
Chapman (1952)	Direct	$\hat{N}_C = \frac{(M+1)(C+1)}{r+1} - 1$	$\mathrm{var}(\hat{N}_C) = \frac{(M+1)(C+1)(M-r)(C-r)}{(r+1)^2(r+2)}$ *
Ricker (1975)	Direct	$\hat{N}_R = \frac{(M+1)(C+1)}{r+1}$	$\mathrm{var}(\hat{N}_R) = \frac{(M+1)^2(C+1)(C-r)}{(r+2)^2(r+2)}$ **

* Variance found by Seber (1982) (Cited in Kenneth et al. (1990))

** Variance found by Chapman (1952).

"Direct" sampling means that sampling is continued until a predetermined sample size (C) is obtained.

If the second differential of Eq. (4) is taken:

$$\frac{\partial^2 L}{\partial N^2} = -\frac{r}{\hat{N}} + \frac{M(2\hat{N}-M)(C-r)}{\hat{N}^2(\hat{N}-M)^2} \tag{7}$$

Here if we put $r = MC/\hat{N}$ in this equation we find;

$$= -\frac{[MC/\hat{N}]}{\hat{N}} + \frac{M(2\hat{N}-M)(C-r)}{\hat{N}^2(\hat{N}-M)^2}$$

$$= -\frac{MC}{\hat{N}^3} + \frac{MC(2\hat{N}-M)(M-\hat{N})}{\hat{N}^2(\hat{N}-M)^2}$$

$$\frac{\partial^2 L}{\partial N^2} = \frac{MC}{\hat{N}^2(\hat{N}-M)}$$

By using the Cramer-Rao Inequality Theory (Hoel, 1971):

$$E\left(\frac{\partial^2 L}{\partial N^2}\right) = -E\left[\left(\frac{\partial L}{\partial N}\right)^2\right] \text{ and } \operatorname{var}(\hat{N}) = \frac{|E(N)|}{E\left(\frac{\partial^2 L}{\partial N^2}\right)}$$

Thus:

$$\operatorname{var}(\hat{N}) = \frac{-1}{-\dfrac{MC}{\hat{N}^2(\hat{N}-M)}}$$

$$\operatorname{var}(\hat{N}) = \frac{\hat{N}^2(\hat{N}-M)}{MC} \qquad (8)$$

Meanwhile if we put $\hat{N} = MC/r$ in Eq. (8), $\operatorname{var}(\hat{N})$ is found as:

$$\operatorname{var}(\hat{N}) = \frac{|(MC/r)^2||(MC/r)-M|}{MC}$$

and thus:

$$\operatorname{var}(\hat{N}) = \frac{M^2C(C-r)}{r^3} \qquad (9)$$

Leslie (1952) (cited in Ricker, 1975) shows $u = r/M$ that the estimate of the rate of exploitation of the population is an unbiased maximum likelihood estimate. Assuming random mixing of marked and unmarked fish, its variance is found from the binomial distribution. According to it, if r is variable, the binomial distribution is:

$$r \sim B[C(M/N), C(M/N)(1-M/N)]$$

$$\operatorname{var}(u) = \frac{\operatorname{var}(r)}{M^2} = \frac{C(M/N)(1-M/N)}{M^2}$$

Thus $\operatorname{var}(u)$ is found as:

$$\operatorname{var}(u) = \frac{C(1-M/N)}{NM} \qquad (10)$$

If the number of marked individuals (r) is small, than r/C is approximate to $M/\hat{N}$. Thus if r/C is written instead of $M/\hat{N}$ in Eq. (10), $\operatorname{var}(u)$;

$$\operatorname{var}(u) = \frac{C}{M\hat{N}}\left(1-\frac{r}{C}\right) \qquad (11)$$

Hence if we put $\hat{N} = MC/r$ value in Eq. (11) then:

$$\mathrm{var}(u) = \frac{r(C-r)}{M^2C} \tag{12}$$

Similarly, an unbiased estimate of the reciprocal of population size can be found by direct proportion below (Ricker, 1975):

$$\frac{1}{\hat{N}} = \frac{u}{C} = \frac{r}{MC} \tag{13}$$

The large-sample sampling variance of $(1/\hat{N})$ is:

$$\mathrm{var}(1/\hat{N}) = \frac{\mathrm{var}(u)}{C^2} = \frac{r(C-r)/M^2C}{C^2}$$

Thus:

$$\mathrm{var}(1/\hat{N}) = \frac{r(C-r)}{M^2C^3} \tag{14}$$

The reciprocal of Eq. (14) is a constant estimate of $\hat{N}$, that is

$$\hat{N} = \frac{MC}{r} = \frac{C}{u} \tag{15}$$

The sampling variance of Eq. (15) is the same with Eq. (9). This is expressed by Bailey (1951). Ricker (1975) found that values of MC/r are not symmetrically distributed whereas values of r/MC are. Thus assuming error that shows a normal distribution, the confidence limit of $1/\hat{N}$ is calculated and the inverted version of the values found is taken and the confidence limits of $\hat{N}$ are found. Moreover, Ricker (1975) suggests using the formulae below to find the confidence limits of r:

$$\%95 \Rightarrow r + 1.92 \pm 1.96\sqrt{r+1} \tag{16}$$

$$\%99 \Rightarrow r + 3.22 \pm 2.576\sqrt{r+1.7} \tag{17}$$

Equation (15) is a consistent estimate of N in that it tends to approximate the correct value as sample size increases. But it is not quite the best estimate.

Bailey (1951) and Chapman (1951) have shown that by ordinary "direct" sampling this is true whether sampling is direct or inverse. Equation (15) tends to over estimate the true population and the above authors proposed modified formulae which give an unbiased estimate in most situations. Chapman's version is as follows:

$$\hat{N}_C = \frac{(M+1)(C+1)}{r+1} - 1 \quad (18)$$

But Ricker (1975) indicates that omitting −1 is of no practical significance. For this reason the proposed equation below is a more practical form of Eq. (18):

$$\hat{N}_R = \frac{(M+1)(C+1)}{r+1} \quad (19)$$

Bailey's (1951) expression is as follows:

$$\hat{N}_B = \frac{M(C+1)}{r+1} \quad (20)$$

Furthermore by using 95% the limits of confidence of $\hat{N}$ can also be found by the equation below (Kenneth et al., 1990);

$$\hat{N} \pm 1.96\sqrt{\mathrm{var}(\hat{N})} \quad (21)$$

3. RESULTS

In this section, the results obtained using the formulae above are compared. Thus we could conclude which formula is more reliable to estimate population size.

Duzgunes (1985) marked 330 carp fish (*Cyprinus carpio L.*) (M) and released them into Mogan Lake. Later 795 carp fish were captured of which 27 fish had been marked. Thus the results of the fish population size according to the single marking and recapture method given are in Table 2.

Table 2: Results of the fish population size according to the single marking and recapture method given in Table 2 in the Duzgunes (1985) study in Morgan Lake.

Reference	Population size (N)	Variance ($\hat{N}$)	Limits of confidence ($\hat{N}$)	
			$\hat{N} \pm 1.96\sqrt{\mathrm{var}(\hat{N})}$	$r+1.92 \pm 1.96\sqrt{r+1}$
Bailey (1951)	9716	3378041	[6114, 13318]	[6727, 13808]
Bailey (1951)	9381	2928115	[6027, 12735]	[6567, 13134]
Chapman (1952)	9409	2696689	[6191, 12627]	[6587, 13173]
Ricker (1975)	9410	2945888	[6046, 12774]	[6587, 13174]

As seen in Table 2, it is suggested that as the confidence interval is narrower in the formula proposed by Chapmen (1952) than in the other formulae, the results obtained using this formula are more reliable in the single marking and recapture method (calculation of results in Table 2 are given in Appendix 1).

4. DISCUSSION

The basic assumption in fact is that N is constant throughout the period in which individuals are being marked and recaptured. It is therefore particularly suited to estimating the fish population size in a pond or lake at a particular time. It is not strictly applicable, however, for estimating fish population size in the sea if the recaptures are made by commercial fishermen over a fairly long period of time.

Estimating fish population size with known methods for commercial fishes is difficult when the very high cost of such investigations is taken into consideration.

The m king technique can be used both for pelagic and demersal fishes. Thus it is pos. le that many parameters belonging to the fish population can be calculated with the single marking and recapture method. But there are some difficulties in practice. For example, as marking small fish causes mortality, this method should be used with caution for small fishes and be caught, or reported. In addition marking individuals can affect their catchability.

REFERENCES

[1] Bailey, N.J., "On estimating the size of mobile populations from recapture data", *Biometrika*, 38 (1951) 293-306.

[2] Chapman, D.G., "Some properties of the hypergeometric distribution with applications to zoological sample censures", *Univ. Calif. Publ. Stat.*, 1 (1951) 131-160.

[3] Chapman, D.G., "Inverse, multiple and sequential sample censuses", *Biometrics*, 8 (1952) 286-306.

[4] De Groot, M.H., *Probability Statistics, Second Edition*, Carnegie-Mellon University, Addison-Wesley Publishing Company, 1986.

[5] Duzgunes, E., "Mogan Gölündeki sazan (Cyprinus carpio L., 1958) Stoklarinin Tahmini Üzerine Bir Arastirma", Doktora Tezi , A. Ü. Zir. Fak. Su Ür. Böl., 1985.

[6] Hoel, G.P., *Introduction to Mathematical Statistics, Fourth Edition*, University of California, Los Angeles, 1971.

[7] Kenneth, H.P., Nichols, J.D., Brownie, C., and Hines, J.E., *Statistical Inference for Capture-Recapture Experiments. WildLife Monographs*, A Publication of the WildLife Society, No 107, 1990.

[8] Leslie, P.H., "The etimation of population parameters from data obtained by means of capture-recapture method II. The estimation of total numbers", *Biometrika*, 38 (1952) 363-388.

[9] Ricker, W.E., *Computation and Interpretation of Biological Statistics of Fish Populations*, Fish. Res: Board Can., 1975.

[10] Seber, G.A.F., *Estimating of Animal Abundance and Related Parameters, Second Edition*, Griffin, London, 1982.

APPENDIX 1.

Calculation of results in Table 2

$$M = 330,\ C = 795,\ r = 27$$

limits of confidence r from Eq. (16):

$$27 + 1.92 \pm 1.96\sqrt{27+1} \qquad [19, 39]$$

Bailey (1951)	$\hat{N} = \frac{MC}{r} = \frac{330 \times 795}{27} = 9716$ $\text{var}(\hat{N}) = \frac{M^2C(C-r)}{r^3} = \frac{330^2 \times 795(795-27)}{27^3} = 3378041$	
	$\hat{N} \pm 1.96\sqrt{\text{var}(\hat{N})}$ $9716 \pm 1.96\sqrt{3378041}$ 9716 ± 3602 [61[illegible]4, 13318]	$\hat{N}_1 = \frac{MC}{r} = \frac{330 \times 795}{19} = 13808$ $\hat{N}_2 = \frac{MC}{r} = \frac{330 \times 795}{39} = 6727$

Bailey (1951)	$\hat{N}_B = \frac{M(C+1)}{r+1} = \frac{330(795+1)}{27+1} = 9381$ $\text{var}(\hat{N}_B) = \frac{M^2(C+1)(C-r)}{(r+1)^2(r+2)} = \frac{330^2(795+1)(795-27)}{(27+1)^2(27+2)} = 2928115$	
	$\hat{N} \pm 1.96\sqrt{\text{var}(\hat{N})}$ $9381 \pm 1.96\sqrt{2928115}$ 9381 ± 3354 [6027, 12735]	$\hat{N}_1 = \frac{M(C+1)}{r+1} = \frac{330(795+1)}{19+1} = 13134$ $\hat{N}_2 = \frac{M(C+1)}{r+1} = \frac{330(795+1)}{39+1} = 6567$

Chapman (1952)	$\hat{N}_C = \frac{(M+1)(C+1)}{r+1} - 1 = \frac{(330+1)(795+1)}{27+1} - 1 = 9409$ $\text{var}(\hat{N}_C) = \frac{(M+1)(C+1)(M-r)(C-r)}{(r+1)^2(r+2)}$ $= \frac{(330+1)(795+1)(330-27)(795-27)}{(27+1)^2(27+2)} = 2696689$	
	$\hat{N} \pm 1.96\sqrt{\text{var}(\hat{N})}$ $9409 \pm 1.96\sqrt{2696689}$ 9409 ± 3218 [6191, 12627]	$\hat{N}_1 = \frac{(M+1)(C+1)}{r+1} - 1 = \frac{(330+1)(795+1)}{19+1} - 1 = 13173$ $\hat{N}_2 = \frac{(M+1)(C+1)}{r+1} - 1 = \frac{(330+1)(795+1)}{39+1} - 1 = 6586$

Ricker (1975)	$\hat{N}_R = \frac{(M+1)(C+1)}{r+1} = \frac{(330+1)(795+1)}{27+1} = 9410$ $\text{var}(\hat{N}_R) = \frac{(M+1)^2(C+1)(C-r)}{(r+1)^2(r+2)} = \frac{(330+1)^2(795+1)(795-27)}{(27+1)^2(27+2)} = 2945888$	
	$\hat{N} \pm 1.96\sqrt{\text{var}(\hat{N})}$ $9410 \pm 1.96\sqrt{2945888}$ 9410 ± 3364 $[6046, 12774]$	$\hat{N}_1 = \frac{(M+1)(C+1)}{r+1} = \frac{(330+1)(795+1)}{19+1} = 13174$ $\hat{N}_2 = \frac{(M+1)(C+1)}{r+1} = \frac{(330+1)(795+1)}{39+1} = 6587$