

SEQUENTIAL ESTIMATION OF POPULATION SIZE*

By

FAURAN SINGH CHAUDHARY¹

J.V. College, Baraut-250611

and

DAROGA SINGH²

Indian Agricultural Statistics Research Institute,

New Delhi-110012

(Received : November, 1978)

SUMMARY

For estimating the size of the plaice populations, the capture-recapture method is quite popular with the biologists. In the present paper an interesting approach has been presented to estimate population and sample sizes alongwith their variances by sequential sampling where units are selected one-by-one randomly. Both with and without replacement procedures have been considered.

INTRODUCTION

Where the object of an enquiry is to estimate the size of a given population, it can be done either by counting all the units of the population or by taking recourse to sampling. There are cases where complete enumeration or census is well-nigh impossible as for example, that of the fish in a lake or some defined sea water. Similarly it is pretty difficult to make a census of the total wild life in a forest or a sanctuary. All that can be done is to arrive at a precise estimate of the population size by a sampling method. Similar situation arises when one is considering the problem of making an estimate of persons having

* This work was facilitated by a grant from the University Grants Commission *vide* Grant No. SR-I.F, 25-3/7585/77.

Present address : 1. Deputy Director (Statistics), Directorate of Research, Haryana Agricultural University, Hissar.
2. Ex Director, IASRI

some rare characteristic. It can be seen that sequential procedures prove useful in tackling such problems.

deLury [4], Baily [2], Chapman [3], Leslie and Chitney [6] and others used sequential methods for the estimation of certain biological populations. Anscombe [1] discussed the fixed-size and sequential methods of estimation at length. Singh [7] has discussed a number of estimators and sketched a new approach to study sequential-methods.

NOTATIONS

The following notations will be used throughout :

N : the size of population being studied

n : the size of sample being drawn

n_1 : the size of sample in with replacement method

n_2 : the size of sample in without replacement method

t : the number of marked individuals in the population

s : the number of marked individuals recovered in the sample

$p \left(= \frac{t}{N} \right)$: the proportion of marked individuals in the population

$d (= \hat{n}_1 - \hat{n}_2)$: difference between the sample size estimates.

Theorem 1.1. If sampling is with replacement, $Pr(n)$, the probability of having to sample n individuals to obtain s marked ones is given by negative binomial distribution, i.e.

$$Pr(n) = \binom{n-1}{s-1} p^s (1-p)^{n-s}$$

where s is fixed and n is a random variate, then

$$\left. \begin{aligned} E(\hat{n}_1) &= \frac{Ns}{t} = \frac{s}{p} \\ V(\hat{n}_1) &= N(N-t)s/t^2 = \frac{s(1-p)}{p^2} \\ v(\hat{n}_1) &= \frac{n(n-s)}{(s+1)} \end{aligned} \right\} \dots(1.1)$$

$$\left. \begin{aligned} \text{or } \hat{N}_1 &= \frac{nt}{n} \text{ with } \hat{V}(N_1) = \frac{N(N-t)}{s} \\ v(\hat{N}_1) &= \frac{nt^2(n-s)}{s^2(s+1)} \end{aligned} \right\} \dots(1.2)$$

Proof is obvious

Theorem 1.2. If sampling occurs without replacement, $Pr(n)$, the probability of having to sample n individuals to obtain s marked one is given by negatively pergeometric distribution, *i.e.*,

$$Pr(n) = \frac{(n-1)! t! (N-t)! (N-n)!}{(s-1)! (t-s)! (n-s)! N! (N-n-t+s)!}$$

where s is fixed and n is a random variate, then

$$\left. \begin{aligned} E(\hat{n}_2) &= \frac{(N+1)s}{(t+1)} \\ V(\hat{n}_2) &= (N+t)(N-t)(t-s+1)s/(t+2)(t+1)^2 \end{aligned} \right\} \dots(1.3)$$

$$\text{or } \left. \hat{N}_2 = \frac{n(t+1)}{s} - 1 \right\} \dots(1.4)$$

$$\text{with } V(\hat{N}_2) = (N+1)(N-t)(t-s+1)/s(t+2)$$

Proof is obvious.

2. EMPIRICAL STUDIES

Corresponding to different values of N , s and p , the differences $d = \hat{n}_1 - \hat{n}_2$ and their relative precisions have been studied.

$$RP = \frac{V_1 - V_2}{V} \times 100$$

where, V_1 denotes the variance of \hat{n}_1 , V_2 denotes the variance of \hat{n}_2 .

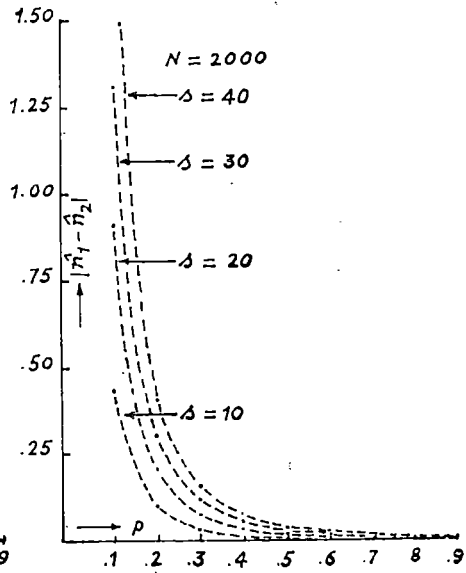
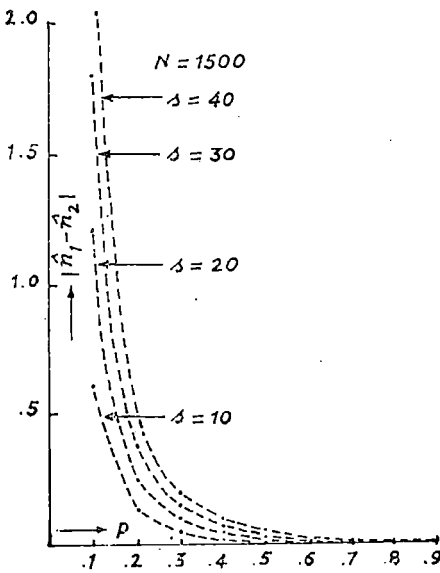
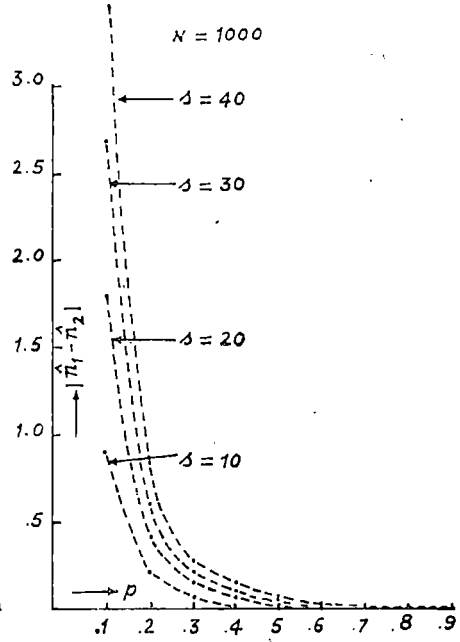
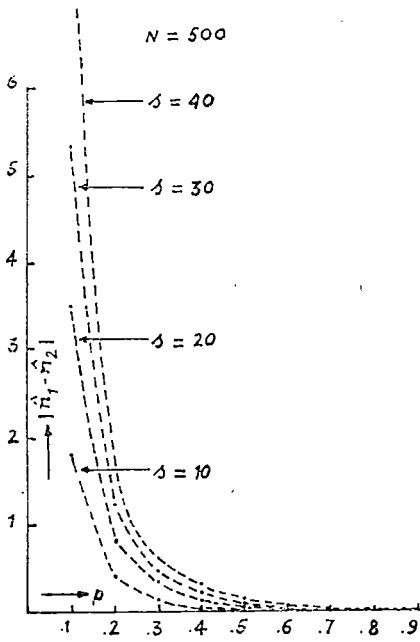
For $N=500, 1000, 1500, 2000$; and $s=10, 20, 30, 40$; the values of the difference of the estimates d alongwith the relative precision (RP) have been calculated for different values of p , *i.e.*, 0.1 to 0.9. An attempt has been made to study sequential sampling without replacement method with the help of with replacement method and a line similar to Dodge-Romig [5] has been developed. The results obtained from the above values have been shown graphically (Graphs I—III) in the form of relations between

(i) p and $d (= \hat{n}_1 - \hat{n}_2)$,

(ii) p and RP , and

(iii) d and RP

The purpose of the present study is to seek some easier procedures for the scientist/biologists for the study of sequential sampling without replacement method, with the help of sequential method with replacement, *i.e.*, to devise a method with minimum possible algebraic relations and computational labour.



The graphs have been plotted by taking different sets of values for N and s , viz.,

- (a) between p and d ($=\hat{n}_1 - \hat{n}_2$) ...Graph I
- (b) between p and RP ...Graph II
- (c) between d and RP ...Graph III

3. METHOD TO ESTIMATE SAMPLE SIZE WHEN N IS KNOWN

When the population size N is known and s is given, the sample size estimates for sequential sampling with and without replacement methods are given by

$$\left. \begin{aligned} \hat{n}_1 &= s/p \\ \hat{n}_2 &= \hat{n}_1 - d \end{aligned} \right\} \dots(3.1)$$

where, the value of d may be obtained from Graph I.

The variance of the estimators in (3.1), using (1.1) and (2.1) will be

$$\left. \begin{aligned} V(\hat{n}_1) &= s(1-p)/p^2 \\ V(\hat{n}_2) &= V(\hat{n}_1) / \left(\frac{1+RP}{100} \right) \end{aligned} \right\} \dots(3.2)$$

For given s and p , the value of RP may be obtained from Graph II. As an illustration, let

$$N=1000, s=20 \text{ and } p=.25 \text{ be given,}$$

and suppose it is required to estimate the sample sizes along with their variances for both the sampling methods

For the values $N=1000$, $p=.25$ and $s=20$, d is obtained from Graph I, which is 0.25. Therefore, using relation (3.1)

$$\hat{n}_1 = 20/.25 = 80, \text{ and } \hat{n}_2 = 80 - 0.25 = 79.75$$

For the same set of p and s , RP is obtained from Graph II, which is 9.5. By relation (3.2)

$$V(\hat{n}_1) = 20 \times .75 / (.25)^2 = 240$$

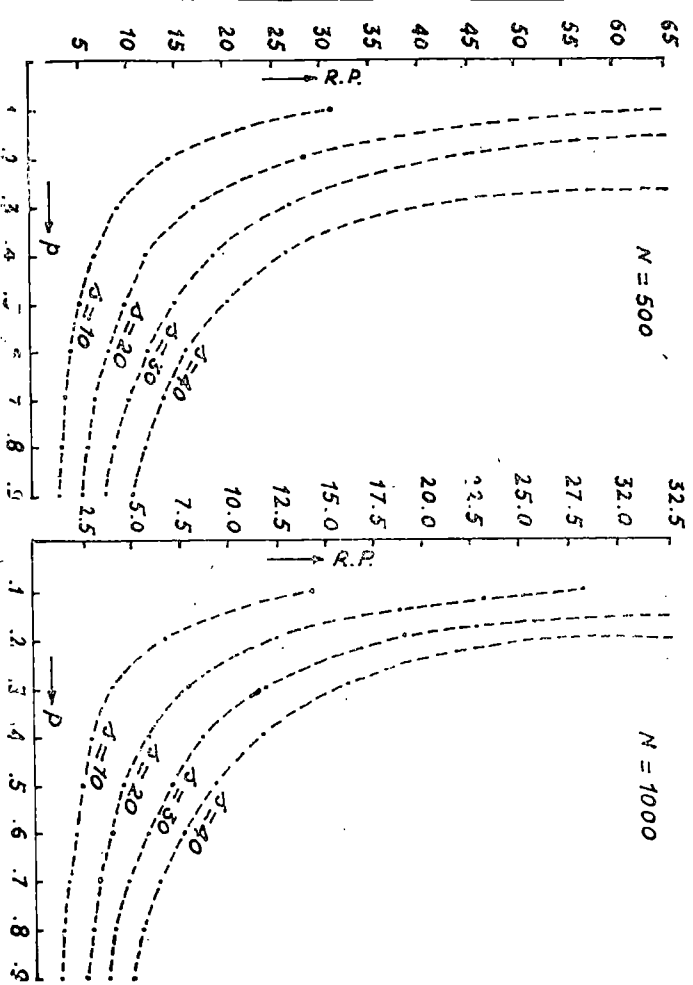
$$V(\hat{n}_2) = 240 / (1 + 9.5/100) = 210$$

For the sake of verification, the values obtained from (1.3)

$$\hat{n}_2 = 79.76$$

$$V(\hat{n}_2) = 216.5$$

which shows that the values obtained by the graphic method are in agreement with those given by the formula,



4. METHOD TO ESTIMATE SAMPLE AND POPULATION SIZES WHEN N IS UNKNOWN

In most of the practical situations, nothing is known except finiteness of population size N , one has to decide for both n and N . According to the requirements of sampling structure D , discussed by Singh [7], the following steps are needed :

(i) stopping rules are to be defined :

(a) number of tagged individuals to be captured, *i.e.* s ;

(b) the risk function, which is expressed generally in terms of specified RP .

and (c) the probability field and the corresponding degree of confidence, *i.e.*, the difference d and the relative error l are to be controlled.

(ii) a choice of suitable graph and decision for N accordingly;

and (iii) to revise them for a proper choice and take the final decision according to the requirements in (i).

The relations used as already mentioned in the theorems are :

$$\left. \begin{aligned} \hat{n}_1 &= d/l = s/p \\ \hat{n}_2 &= \hat{n}_1 - d \end{aligned} \right\} \dots(4.1)$$

$$\left. \begin{aligned} v(\hat{n}_1) &= \hat{n}_1 (\hat{n}_1 - s)/(s + 1) \\ v(\hat{n}_2) &= v(\hat{n}_1)/(1 + RP/100) \end{aligned} \right\} \dots(4.2)$$

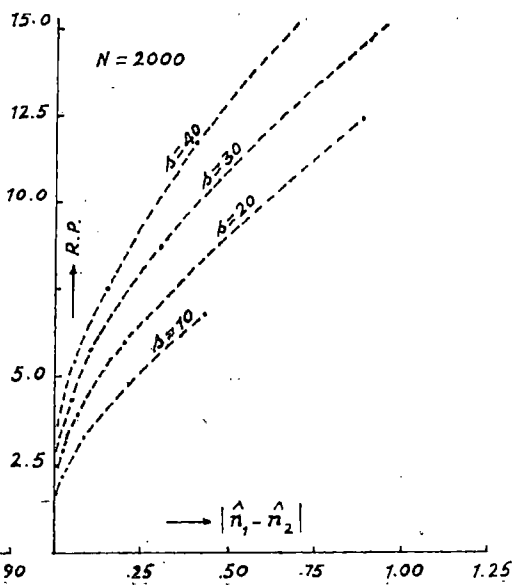
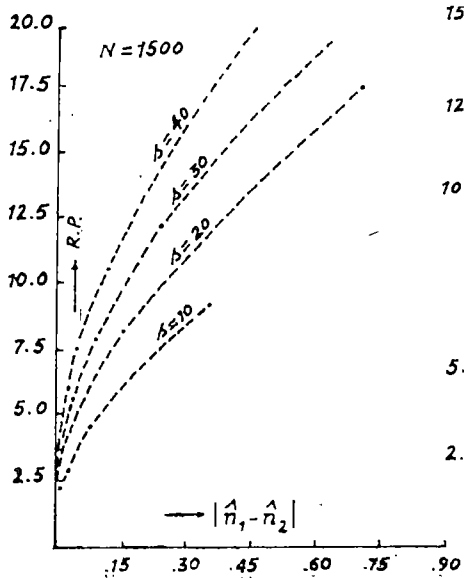
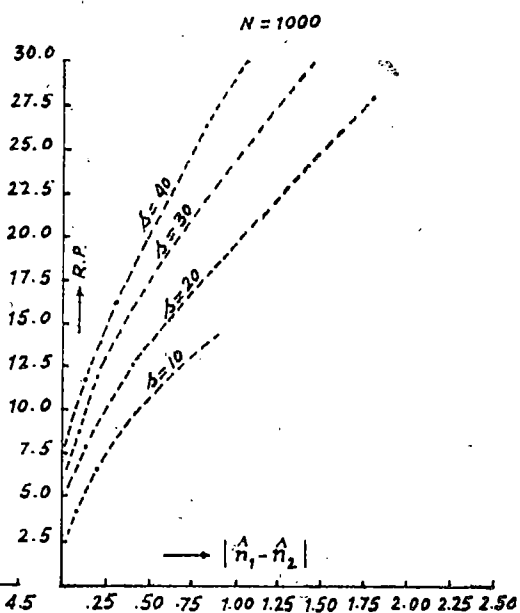
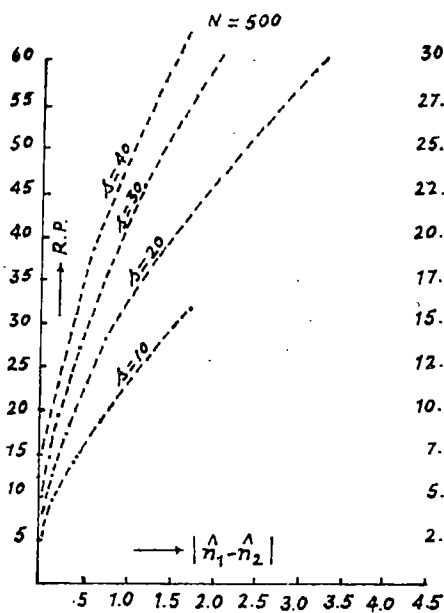
Also

$$\left. \begin{aligned} \hat{N}_1 &= \hat{n}_1 t/s \\ \hat{N}_2 &= \hat{n}_2 (t + 1)/s - 1 \end{aligned} \right\} \dots(4.3)$$

and

$$\left. \begin{aligned} v(\hat{N}_1) &= (\hat{N}_1 p/s)^2 v(\hat{n}_1) \\ v(\hat{N}_2) &= v(\hat{N}_1)/(1 + RP/100) \end{aligned} \right\} \dots(4.4)$$

As an illustration, let $t=100$ and $s=10$ given, and the relative difference be 1 percent and the absolute difference unity



with $RP=.05$. Therefore,

$$\hat{n}_1 = 1/.01 = 100$$

$$\hat{n}_2 = 100 - 1 = 99$$

$$v(\hat{n}_1) = 100(100 - 10)/(10 + 1) = 818$$

$$v(\hat{n}_2) = 818/(1 + 5/100) = 719$$

Also

$$\hat{N}_1 = \hat{n}_1 t/s = 100 \times 100/10 = 1000$$

and therefore p is 0.1.

$$\hat{N}_2 = 99 \times 101/10 - 1 = 998.9$$

$$v(\hat{N}_1) = (100/10)^2 \times 818 = 81800$$

$$v(\hat{N}_2) = 81800/(1 + 5/100) = 77905$$

ACKNOWLEDGEMENT

The authors acknowledge with thanks the valuable comments given by the referees of this paper.

REFERENCES

- [1] Anscombe, F.J. (1953) : Sequential Estimation. *J.R. Statist. Soc. B*, 15, 1-29.
- [2] Bailey, N.T.J. (1951) : On estimating the size of mobile populations from recapture data. *Biometrika*, 38, 292-306.
- [3] Chapman, D.G. (1952) : Inverse, multiple and sequential sample censuses. *Biometrics*, 8, 286-306.
- [4] deLury, D.B. (1947) : On estimation of biological populations. *Biometrics*, 3, 145-67.
- [5] Dodge, H.F. and Romig, H.G. (1959) : *Sampling Inspection Tables* (Single and Double Sampling). Wiley & Sons, New York.
- [6] Leslie, P.H. and Chithey, D. (1951) : The estimation of population parameters from data obtained by capture-recapture method. I. The minimum likelihood equations for estimating the death rate. *Biometrika* 38, 269-92.
- [7] Singh, F. (1977) : A sequential approach to sample surveys. Ph.D. thesis, Meerut University, India.