

Optimum Allocation in Multivariate Double Sampling for Biomass Estimation

JAVED AHMED AND CHARLES D. BONHAM

Double sampling or two-phase sampling involves sampling of any population in 2 phases. The first phase yields data on any desired factor by direct measurements as well as by some indirect method. In the second phase, data are collected by the indirect method only.

The estimated variance in double sampling with regression and ratio estimators are described in detail by Cochran (1963). The first sample is a simple random sample of size n' . The second sample of size n is a random subsample from the first sample, but may be drawn independently if this is more convenient. The first step is to set up the estimate and to determine its variance. The auxiliary variate (x_i) is used to make a regression estimate of y . It is assumed that the population is finite but very large and that the relation between y_i and x_i is linear. In the first (large) sample (size n'), only x_i is measured. In the second (small) sample (size n), both x_i and y_i are measured.

One of the major problems in double sampling is determining the number of samples required in each phase to give the desired accuracy for the maximum economy. The efficiency of double sampling depends on two things: (1) the precision of the mathematical relationship, and (2) the cost of direct measurements compared to indirect estimates.

If too many direct samples are taken, the cost of sampling becomes unnecessarily high, while the use of too few direct samples results in an unreliable mathematical relationship. Thus, it is desirable to estimate the size of the two samples; the large sample (n') and the small sample (n). In its application to estimating crop biomass, the first phase of double sampling involves estimation of the plant biomass ocularly or by capacitance meter. In the second phase, the plant biomass is estimated as in the first phase followed by clipping and weighing of the plants. For a detailed discussion on the statistical aspects of the double sampling, the reader is referred to Schumacher and Chapman (1948), Hansen et al. (1953), National Research Council (1962), and Cochran (1963).

With a single factor under study and for a given sampling procedure, optimum allocation of resources to direct and indirect methods of estimation is well defined. However, a simple procedure for optimum allocation in multivariate double sampling is not available. A technique is described which enables the investigator to find optimum allocation in multivariate double sampling which minimizes cost or variance and also gives achievable variances of the estimated means.

For sampling involving a single independent variable, the procedure for optimum allocation in double sampling is reviewed. According to Cochran (1963), the cost of double sampling is

$$C = nc_n + n'c_{n'} \quad (1)$$

where

- C = total cost of double sampling
- c_n = cost of obtaining one direct sample
- $c_{n'}$ = cost of obtaining one indirect sample
- n = number of direct samples
- n' = number of indirect samples.

Authors are with the Range Science Department, Colorado State University, Fort Collins 80523.

This research was sponsored by a grant from U.S. AID, project No. 391-80183 awarded to Javed Ahmed.

Manuscript received November 24, 1980.

For a fixed cost C , optimum allocation (for minimum variance is obtained when

$$\frac{n'}{n} = \frac{\sqrt{V_{n'}c_n}}{\sqrt{V_n c_{n'}}} \quad (2)$$

$$= \sqrt{\frac{c_n \rho^2}{c_{n'} (1 - \rho^2)}} \quad (3)$$

where

$$V_n = (S_y^2) (1 - \rho^2) \quad (4)$$

$$V_{n'} = \rho^2 S_y^2, \quad (5)$$

and

ρ = coefficient of correlation between y_i and x_i in sample n
 S_y = standard deviation of y_i .

Equations (1) and (2) or (3) determine n and n' . The techniques for allocation in double sampling described by Wilm et al. (1944) and Schumacher and Chapman (1948) are essentially the same as the one described by Cochran (1963), and are, therefore, not given here.

In our study of optimum allocation for multivariate double sampling, the variance of the mean desired for the estimates of each variate was specified, and the optimum allocation was the one which achieved this at a minimum cost. Therefore, the basic approach was to attain specified levels of precision at the minimum cost. First, the variance of the regression and ratio estimator in double sampling is given and then the allocation problem is formulated and its solution described.

Estimated Variances in Double Sampling

The estimate of \bar{y} is

$$\bar{y}_{1r} = \bar{y} + b(\bar{x}' - \bar{x}) \quad (6)$$

where \bar{x}' and \bar{x} are the means of x_i in the indirect and direct samples, respectively, and b is the least squares regression coefficient of y_i on x_i , computed from the direct sample.

The variance of \bar{y}_{1r} , the regression estimate in double sampling, is

$$V(\bar{y}_{1r}) = \frac{V_n}{n} + \frac{V_{n'}}{n'} \quad (7)$$

and a sample estimate of $V(\bar{y}_{1r})$ is

$$\hat{V}(\bar{y}_{1r}) = \frac{s_y^2 \cdot x}{n} + \frac{s_y^2 - s_y^2 \cdot x}{n'} \quad (8)$$

Optimum Allocation

The cost of double sampling is given in equation (1). With k variates, let V_j^0 be the specified variance tolerance for the mean of the j^{th} variate. The precision specifications become

$$V(\bar{y}_j) \leq V_j^0 \quad (9)$$

From the approximate expression of variance in equation (7) and from (9), it follows that

$$\frac{V_{n_j}}{n} + \frac{V_{n'_j}}{n'} \leq V_j^0 \quad (10)$$

and the allocation problem with k variates becomes

$$\text{minimize } C = nc_n + n'c_n', \quad (11)$$

subject to

$$\frac{Vn_j}{n} + \frac{Vn'_j}{n'} \leq V^0_j \quad (12)$$

$$(j = a, \dots, k)$$

$$n' > n > 0.$$

The overhead costs have been neglected because these do not enter into the optimization problems (Kokan 1963).

By obtaining a solution to the above plan, we actually obtain the solution to a series of plans. Let C^0 be the sampling cost

$$C^0 = nc_n + n'c_n', \quad (13)$$

which satisfies

$$\frac{Vn_j}{n} + \frac{Vn'_j}{n'} \leq V^0_j, \quad (14)$$

then the fixed cost (C') sample

$$C' = \frac{C^0}{r} \quad (15)$$

is

$$\frac{n}{r} \text{ and } \frac{n'}{r}, r > 0, \quad (16)$$

which satisfies

$$V(\bar{y}_j) \leq rV^0_j. \quad (17)$$

Choice of Precision Specifications

No definite answer can be given to the question, "what is a desirable level of precision?" The desired precision will depend on the purpose at hand. The quantities V_n and $V_{n'}$ on the left-hand side of the inequalities can be estimated from a preliminary sample in which y_i is measured and x_i is estimated for each variate. The V^0_j , or the righthand side of the inequalities are specified:

1. from the past experience or from values reported in the literature, or
2. by specifying coefficients of variation for \bar{y} of each variate

$$CV(\bar{y}_j) = \frac{SE(\bar{y}_j)}{\bar{y}_j}, \text{ or} \quad (17)$$

3. by specifying a bound (B) on the error of the estimate

$$B = 2SE, \quad (18)$$

which is equivalent to

$$V(\bar{y}_j) = \frac{B^2}{4}. \quad (19)$$

Solution of the Allocation Problem

The purpose of optimization is to find the best possible solution among the many potential combinations of sampling ratios for a given problem in terms of effectiveness or performance criterion. The usual analytical approach for optimization of nonlinear programming problems is to use the calculus and/or Lagrange multipliers. However, the geometric programming approach of Duffin et al. (1967), developed for solving algebraic nonlinear programming problems, was used to obtain solutions to the multivariate double sampling problem. A programming algorithm using the above analytical techniques has been described in detail by Ahmed and Bonham (1980).

Test data for this allocation problem were collected at the Central Plains Experimental Range, administered by the Science and Education Administration-Agricultural Research of the U.S. Department of Agriculture, near Nunn, Colo. Statistical summaries of the double sampling data are given in Table 1.

***** PROGRAM FOR OPTIMUM ALLOCATION IN DOUBLE SAMPLING *****

```

HOW MANY CONSTRAINTS
? 4
COST FUNCTION
? 3 .5
INPUT CONSTRAINT COEFFICIENTS AND RIGHT-HANDSIDE
(SEPARATED BY A COMMA OR A SPACE)
CONSTRAINT 1:
? 5.5 106.8 .24
CONSTRAINT 2:
? 36.8 182.2 3.14
CONSTRAINT 3:
? 5.8 31.8 .28
CONSTRAINT 4:
? .9 5.9 .04
CONSTRAINT 5:
? 3.3 15.2 .056
CONSTRAINT 6:
? 15.8 80.8 .74
    
```

INITIAL OPTIMUM ALLOCATION PLAN

```

THE NUMBER OF DIRECT SAMPLES      =      98.9
THE NUMBER OF INDIRECT SAMPLES    =     671.4
THE COST OF SAMPLING WITH THIS PLAN = $  830.29
    
```

VARIANCES OF THE MEANS

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6
SPECIFIED	.24	3.14	.28	.04	.06	.74
ESTIMATED	.21	.64	.11	.02	.06	.28

```

HOW MANY SAMPLING PLANS DO YOU WANT
? 5
ENTER THE COST OF EACH PLAN
(SEPARATED BY A COMMA OR A SPACE)
? 700 600 500 400 300
    
```

OPTIONAL OPTIMUM ALLOCATION PLANS

	COST DOLLARS	SAMPLES		VARIANCES OF THE MEANS					
		N1	N2	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6
SPECIFIED VARIANCES				.24	3.14	.28	.04	.06	.74
OPTIONAL PLANS									
PLAN 0	830.29	98.9	671.4	.21	.64	.11	.02	.06	.28
PLAN 1	700.00	83.4	566.0	.25	.76	.13	.02	.07	.33
PLAN 2	600.00	71.5	485.2	.30	.89	.15	.02	.08	.39
PLAN 3	500.00	59.6	404.3	.36	1.07	.18	.03	.09	.47
PLAN 4	400.00	47.7	323.4	.45	1.34	.22	.04	.12	.58
PLAN 5	300.00	35.7	242.6	.59	1.78	.29	.05	.15	.78

Fig. 1. Computer printout from program DUBSAM.

Table 1. Statistical analysis of the double sampling data. Regression of clipped green weight against estimated green weight.

Species	Sample size	Clipped green weight (gm)		Standard deviation about the regression line
		Mean	deviation	
<i>Bouteloua gracilis</i> (H.B.K.) Lag. ex Steud.	50	4.90	10.60	2.35
<i>Sporobolus cryptandrus</i> (Torr.) A. Gray	50	17.74	14.80	6.07
<i>Aristida</i> spp.	50	5.32	6.14	2.41
<i>Sphaeralcea coccinea</i> (Pursh) Rydb.	50	2.10	2.61	0.94
<i>Chenopodium</i> spp.	50	2.37	4.30	1.82

The average time needed to clip a plot was 10 times higher than that needed to ocularly estimate a plot. On the average, 15 plots were clipped per day. The daily wage of a skilled person was assumed to be \$75. This gave a cost of \$5 for obtaining one clipped plot and \$.50 for estimating one plot ocularly.

This information was used in formulating the objective function. The quantities V_n and $V_{n'}$ in the left-hand side of the constraints were calculated using equation (8). The right-hand side of the constraints, or the precision specifications, were obtained from the test data using equation (17). A 10% CV for the estimated means to be obtained was specified.

Based on these calculations, the test problem was of the form.

$$\text{Minimize Cost} = 5n + .5n'$$

subject to

$$5.5/n + 106.8/n' \leq .24 \text{ (constraint for } \textit{Bouteloua gracilis})$$

$$36.8/n + 182.2/n' \leq 3.14 \text{ (constraint for } \textit{Sporobolus cryptandrus})}$$

$$5.8/n + 31.8/n' \leq .28 \text{ (constraint for } \textit{Aristida spp.})}$$

$$0.9/n + 5.9/n' \leq .04 \text{ (constraint for } \textit{Sphaeralcea coccinea})}$$

$$3.3/n + 15.2/n' \leq .056 \text{ (constraint for } \textit{Chenopodium spp.})}$$

$$n' > n > 0.$$

Solution to this problem was obtained using the program DUBSAM (Ahmed and Bonham 1980). The computer printout is given in Figure 1.

The most important binding constraint(s) may not be important practically. For a practical solution of the tolerance setting, a series of sampling plans for varying costs and degrees of precision can be worked out. The sampler can then choose the plan which best fits the particular budget and precision requirements.

The program was written for solution of problems with up to 10 constraints (species). It can, however, be modified to handle more than 10.

For the test problem, the number of direct samples (n) was 99 and that of indirect samples (n') was 672. The cost of sampling with this plan, excluding the fixed cost, was \$830.29 (Figure 1). Optional plans for variable costs of \$700, \$600, \$500, \$400, and \$300 were also obtained. With an allocation plan up to the cost of \$400, only the precision requirements for *Bouteloua gracilis* and *Chenopodium* spp. were violated. With a plan cost of \$300, all the precision requirements except that of *Sporobolus cryptandrus* were violated. The optimum ratio of clipped plots to ocularly estimated plots was 6.6 ($672 \div 99$) (Fig. 1).

Conclusions

The optimum allocation problem in multivariate double sampling can be solved by analytical or graphical methods as described by Ahmed and Bonham (1980). However, the computer can be used more conveniently to obtain the solution of the optimum allocation problem and sampling plans for specified costs.

The constraints most binding may not be important practically. For a practical solution of the tolerance setting, develop a series of plans for varying costs and degrees of precision. Then the plan which best fits in terms of the budget and precision requirements can be chosen.

Although the principles and methods are generally applicable, the data obtained and presented here are applicable only to the specific site sampled and for the year. This technique for optimum allocation in multivariate double sampling for biomass estimation is not restricted in its use. The technique will be found useful in all situations of double sampling and in all fields of study where the interest is in finding optimum allocation of resources for taking direct and indirect measurements on one or more variables. This technique will also be found useful in stratified sampling. In stratified double sampling, optimum allocation can be worked out for each stratum, and the information so collected can then be pooled for estimating population parameters.

Literature Cited

- Ahmed, J., and C.D. Bonham 1980. DUBSAM-Algorithm and computer program for optimum allocation in multivariate double sampling for biomass estimation. Colorado State Univ., Range Science Series, No. 33 109 p.
- Cochran, W.G. 1963. Sampling Techniques. John Wiley & Sons, Inc., New York. 413 p.
- Duffin, R.J., E.L. Peterson, and C. Zener. 1967. Geometric Programming: Theory and Application. John Wiley & Sons, Inc., New York. 278 p.
- Hansen, M.H., W.N. Hurwitz, and W.G. Madon. 1953. Sample Survey Methods and Theory. John Wiley & Sons, Inc., New York. 2 vol.
- Kokan, A.R. 1963. Optimum allocation in multivariate surveys. J.R. Statist. Soc. Amer. 126:537-565.
- National Research Council. 1962. Range research: Basic problems and techniques. Nat. Acad. Sci. Nat. Res. Council Pub. No. 890. Washington, D.C. 341 p.
- Schumacher, F.X., and R.A. Chapman. 1948. Sampling methods in forestry and range management. Duke Univ. School of Forest Bull. No. 7. 222 p.
- Wilm, H.G., D.F. Costello, and G.E. Klipple. 1944. Estimating forage yield by the double sampling method. Amer. Soc. Agron. J. 36:194-203.