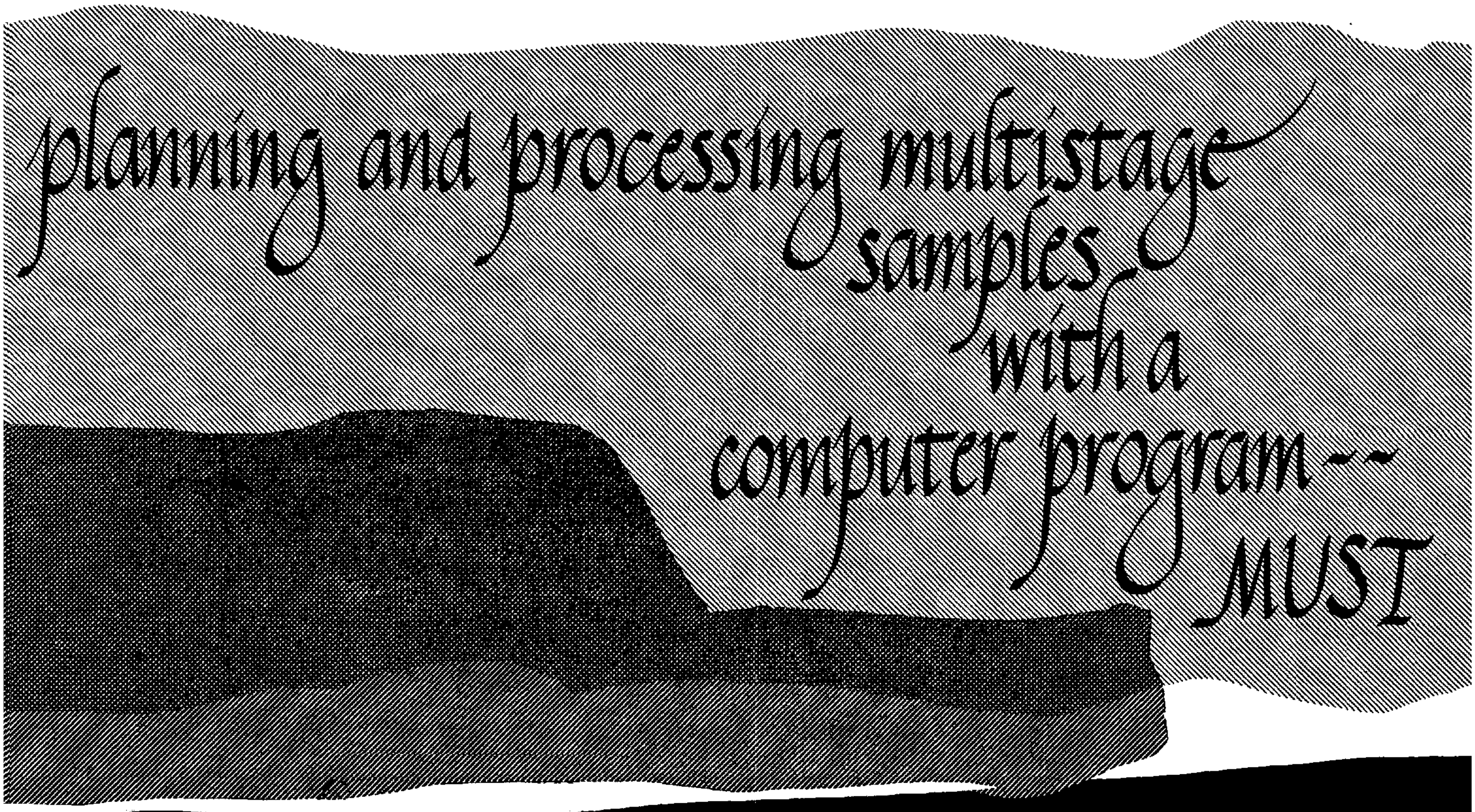
1974
USDA Forest Service General Technical Report PNW-11



planning and processing multistage samples with a computer program--MUST

john w. hazard
larry e. stewart

Pacific Northwest Forest and Range Experiment Station
U.S. Department of Agriculture Forest Service
Portland, Oregon

JOHN W. HAZARD is biometrician with Pacific Northwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

LARRY E. STEWART is mathematician with Bonneville Power Administration, U.S. Department of the Interior.

Abstract

A computer program was written to handle multistage sampling designs in insect populations. It is, however, general enough to be used for any population where the number of stages does not exceed three. The program handles three types of sampling situations, all of which assume equal probability sampling. Option 1 takes estimates of sample variances, costs, and either a specified cost or precision and computes the optimum number of sampling units to select prior to sampling. Option 2 takes the observations, continuous or discrete, from a pilot survey or actual inventory and estimates the mean and variances. It then computes an estimate of the optimum number of units which should have been taken. Option 3 is a special case of the situation in option 2 where some of the observations are lost or for some other reason an unequal number of subunits per primary unit exists.

An explanation of multistage sampling and a processed example are included.

Keywords: Sampling design, computer program, forest entomology.

Introduction

The term multistage sampling refers to a method in which sampling units are sampled. The first stage of sampling occurs when a sample of primary sampling units is selected from the aggregate of such units making up the population. The second stage occurs when samples are selected from each of the primary sampling units from the first stage. Sometimes subsampling units are sampled creating a third stage composed of sub-subsampling units. This process can continue; however, it rarely goes beyond the fourth stage. This type of arrangement of sampling units is illustrated in a forest insect research study designed to estimate the average number of spruce budworm egg masses and the number of larvae per unit of foliage (Carolin 1972). The population was an acreage of infested Douglas-fir trees in eastern Oregon. The infestation occurred in small islands of trees in which Douglas-fir was the predominant tree species. Thus, the primary sampling unit was an area plot of which there were a large number (considered infinite) of such plots. Each plot was sampled by taking subplots systematically within the plot. A subplot involved a portion of an acre around the subplot center. A third stage of sampling occurred when trees were sampled within each subplot.

Studies such as this involving insects, range vegetation, and trees adapt themselves readily to multistage sampling. Multistage sampling can be applied when small elements of the population occur in clusters at one or more stages in the population, it is uneconomical to measure the variable of interest on all the subelements in a cluster, or a list of small elements is prohibitively difficult or expensive to construct.

The theory of multistage sampling is well established in texts (Cochran 1963, Sukhatme 1954, Yates 1949, and others); however, explanations of these techniques are not always easily understood by researchers nor are their applications always apparent. This is particularly the case in biological populations because the sampling units at each stage frequently may be defined in a variety of dimensions and arrangements.

Whenever multistage sampling is used, there is a problem in allocating effort to the various stages of sampling. In other words, what should the sample size be at each stage? One seeks that allocation of resources (i.e., time and money) which produces a given sampling error at least cost or, conversely, will produce the smallest sampling error for a given cost.

Implementing Multistage Sampling

This techniques report introduces a new computer program, MUST, which processes equal probability multistage samples and computes optimum sampling fractions at the various stages of sampling. It also presents to researchers the fundamentals of this particular class of sampling problems (i.e., multi-stage sampling problems), so that they will recognize them when confronted with such situations and be able to plan surveys and studies more efficiently.

One characteristic of biological populations which may exist and complicate the application of multistage sampling is the occurrence of units of unequal size. Drawing a sample from a collection of units of unequal size may be done on an equal or unequal probability basis. Unequal probability multistage sampling is not considered here (Cochran 1963, Langley 1969). Fortunately, most biological populations have large numbers of units at one or more stages of sampling, so the size of samples at subsequent stages may be kept constant and the probability of selection assumed equal.

We mentioned previously that the shapes and dimensions of the units and subunits can frequently be changed by the researcher when they are not fixed by the way they occur in nature. Before one can estimate the optimum allocation of effort to the various stages, he must identify the stages of sampling and specify the sampling units (i.e., define specifically the size and shape of each unit fromwhich observations are to be taken at each stage of sampling). Units which are small in size usually are distributed more evenly over an area than are large units (i.e., for an equal amount of the area sampled), but economic or practical reasons tend to favor larger units--measuring a larger area while on a plot and thus measuring fewer total plots cuts down on costs incurred by travel between plots. The decision of choosing the optimum-size sampling unit becomes one of choosing the alternative that minimizes the variance per unit of cost. This means that if two alternative sampling units are available one should choose the one which provides the smallest estimate of the variance given both samples cost the same. This problem in three-stage sampling is more complicated than this simple explanation because the definition of subsampling units are dependent upon the choice of sampling units at the first stage.

If the problem is extended to determining the optimum size and frequency of sampling units at each stage, then it becomes a mathematical programing problem. Further developments of this aspect of multistage sampling will not be considered in this paper.[1]

Assume that the researcher has identified his stages of sampling and decided upon the definitions of his sampling units by some suitable criteria so that he can progress toward determining the optimum number of units to be selected at each stage. The strategy for determining this optimum will be, as mentioned previously, to satisfy one of two possible objectives. The first is to choose the number of first-, second-, and third-stage units in the sample (n, m, and k) so the total cost of the survey is minimized, subject to achieving an expected level of precision (sampling error) on the estimate of interest. The second is to choose sample sizes to minimize the expected precision subject to achieving a specified cost of the survey.[2]

In three-stage sampling there are three variables (n, m, and k). It should be apparent that the precision can be increased or decreased by changing any one, two, or all three of these sample sizes.

---

[1]John W. Hazard and Lawrence C. Prommitz. Optimality in the design of successive forest surveys. (Submitted to For. Sci.)

[2]The discussion and illustrations from here on will be framed around three-stage sampling, since the program MUST will handle up to three stages.

To calculate an estimate of the optimum allocation of resources to the third and second stages of sampling (i.e., $\hat{k}_{opt}$ and $\hat{m}_{opt}$, respectively), one needs the following sample information:[3]

s_1^2 = the variance among first-stage unit means,

s_2^2 = the variance among second-stage unit means pooled over first-stage units,

s_3^2 = the variance among third-stage units pooled over the second- and first-stage units,

c_i, i = 1–3 are the respective costs per unit associated with taking units at each stage.

Note that these items of information are independent of the population sizes N, M, and K, and the specified precision or total cost. This means that k_{opt} and m_{opt} can be estimated from sample information. The number of sampling units at the first-stage, n, is a function of k_{opt} and m_{opt}, and it also depends upon which objective has been chosen. If the objective specifies a total cost, then an estimate of n results directly from substitution of the known items into the cost function. If, however, the objective involves a specified precision, then one needs to know N unless $\overline{N}$ is very large. In the latter case, n is a function only of the variance at the first stage.

Program Specifications

When an actual sampling problem arises, the researcher or sampler is faced with one of two situations. Either he must conduct a survey with no prior information about his population, or he is fortunate enough to have previous inventories or studies with some estimates of the variances and costs of sampling at each stage.

Perhaps the best approach, in the absence of prior knowledge of the population, is to consider running a pilot study to estimate the inputs to the allocation problem. If the required information (i.e., costs and variances) is present, then the techniques for determining n, m, and k are straightforward.

The program MUST processes three options:

Option 1. Predict *optimum*.--This option predicts the optimum n, m, and k when prior estimates (variances and costs) are available from a pilot study or other sources for input. The optimum is computed for minimizing cost or minimizing variance. Two sets of variance inputs may be used: Either estimates of the true variances, $(\hat{S}_i^2)$, are specified directly, or the sample variances, (s_i^2), are provided (Cochran 1963).

---

[3] **Refer to the appendix for a complete explanation of the calculation of $\hat{k}_{opt}$, $\hat{m}_{opt}$, and $\hat{n}$.**

Option 2. Equal *subsamples*.--This option accepts the individual observations from a pilot or an actual survey with either binomial or continuous random variables. It computes estimates of the population mean and variances, then computes the optimum for the cost or variance restriction from this sample for future surveys of this population or for evaluating the efficiency of this allocation.

Option 3. Unequal *subsamples*.--This option is used when the observations have been taken under the assumption of equal probability sampling and equal-sized units at each stage but, due to laboratory failures, etc., the samples end up unequal. Again, this option will process percentages or continuous variables, and it also computes the new optimum number of samples for future surveys.

Control Cards

card 1

Title card

Columns	Format	Information
1-5	A5	Punch TITLE.
6-75	7A10	Any ident. desired.
76-80	A5	If a Z appears in col. 80. the data will be listed.

As many TITLE cards may be used as desired.

card 2

OPTION 1

Columns	Format	Information
1-15	A5, A10	Punch PREDICT OPTIMUM.
16-19		Blank.
20-24	F5.0	Size of the sampling universe for the first stage (*N*). If *N* = infinity, set to 99999.
25-28	F4.0	Size of the first-stage. sample (n).
29-32	F4.0	Size of the sampling universe for the second-stage (*M*). If *M* = infinity, set to 9999.
33-36	F4.0	Size of the second-stage sample (*m*).
37-40	F4.0	Size of the sampling universe for the third-stage (*K*). If *K* = infinity, set to 9999.
41-44	F4.0	Size of the third-stage sample (*k*).

47-54	F8.I*	Cost of collecting a first-stage unit.
55-62	F8.I*	Cost of collecting a second-stage unit.
63-70	F8.I*	Cost of collecting a third-stage unit.
71-80	F10.I*	Total cost, if specified.

* If the number in the columns does not contain a decimal point, then the field is read with "I" being zero.

Card 3

OPTION 1. Predicting the optimum when the appropriate statistics are available from previous surveys.

Columns	Format	Information
1- 5	F5.1	Relative standard error, (percent/100).
6-15	F10.I	Absolute standard error.
16-25	F10.I	Estimated mean.
26-27	F2.0	Probability level.
28-34	F7.1	t-value.
35-45		Blank
46	A1	Leave blank if variances are unbiased estimates of the true population variance; otherwise, insert S, if variances are computed among means at the various stages (see reference).
47-54	F8.1	Estimated variance at first stage.
55-62	F8.1	Estimated variance at second stage.
63-70	F8.1	Estimated variance at third stage.
71-80	F10.I	$V(\bar{\bar{\bar{y}}})$ specified. If the standard error and the t-value are specified, then $V(\bar{\bar{\bar{y}}})$ will be calculated and this field ignored.

Card 2

OPTION 2. Predicting the optimum when the input must be summarized from a pilot or existing study.

Columns	Format	Information
1-19	5XII	Variable format where II is an integer count of the number of observations per card followed by the format for each observation (i.e., starting in col. 1 (5X10(F6.2)).
20-44		Same as card 2, option 1.
45		"1" if the observations are binomial data (i.e., yes or no). Only a summary of the number of *yes* observations needs to be entered. Blank, if data are continuous.
46		"1" if there are unequal sample sizes. Blank if equal sample sizes.
47-80		Same as card 2, option 1.

A card 3 will not occur with option 2 because the variances will be computed from the data.

Card 2

OPTION 3. Samples are selected under the assumption of equal-sized units (equal probability) but the size of samples ends up unequal due to losses of sampling units.

The card 2 format will be the same as for option 2, except column 46 must have a "1" punch.

card 3

OPTION 3 (optional)

Columns	Format	Information
1- 5	A5	Punch "TWO" if unequal subsamples occur at second stage (left-justified).
6- 9	I3	A count of the number of second-stage samples within the first-stage unit.
10-12	I3	Second-stage unit.
13-80		In three-digit fields, identify the number of subunits sampled in the remaining primary units.

card 4

OPTION 3 (optional)

Columns	Format	Information
1- 5	A5	Punch "THREE" if unequal subsamples occur at third stage.
6- 9	13	Similar to card 3; identify the number of sub-subsampling units at the third stage within each second-stage sampling unit.
10-80		Repeat for all second-stage sampling units in three-digit fields.

Data card

Columns	Format	Information
1- 5	A5	Punch DATA.
6-80	5X(II)	Data are punched in the format specified in card 2.

Data will be ordered $x(i, j, k)$ where k is within j and j within i. Between each job a STOP (cols. 1-4) card is required. The last card will be an end-of-data indicator (6,7,8,9 punched in col. 1). The order must also be consistent with the identification of sample sizes in cards 3 and 4.

Control card specifications, programed error messages, operating instructions, and examples of input data occur on comment cards at the beginning of the source program.

The program MUST is written for the CDC 6400; it will take up to 1,000 measurements over all stages (can be expanded easily), and it will process problems of up to three stages.

An Example of Field Use of MUST

A tussock moth egg survey for virus determination was run without knowledge of the relative variances at each stage. The situation was that a sample of 17 points was distributed over an area (assumed random). At each point, an average of approximately 23 egg masses were collected from a much larger number. Each egg mass had approximately 250 eggs from which an average of approximately 44 larvae were reared. The average number of eggs per egg mass varied from mass to mass but appeared to be constant enough to accept the average. This amounted to roughly 17,204 insects which were reared and examined for virus. The objective was to estimate the percent of insects which possessed this natural virus; thus, the observation $y_{ij\ell}$ is 1 or 0 depending on whether the insect does or does not have the virus. Under the option 2 alternative, the data were processed and the allocation computed. A summary of these data and estimates appears in table 1.

Looking at the lower part of the table, column 1 labels the particular stage of sampling, columns 2–7 present a summary of the statistics (i.e., the inputs to the optimization formulas), and column 8 the optima:

The population and sample sizes for each stage occur in columns 2 and 3.

The 99999 indicates an assumed infinite number of primary units (plots) of which 17 were selected at random.

The 9999 symbolizes an infinite number of egg masses per plot of which an unequal number (average of 22.47) was selected at random from the 17 plots.

There were an average of 250 eggs per egg mass of which an unequal number (approximately 44) was selected per egg mass from the 17 x 22.47 = 382 egg masses selected.

Sampling fractions are not presented for stages 1 and 2 in column 4 because they are infinite in size. The .17596 represents a weighted average sampling fraction for all 382 second-stage samples taken.

The values of s_1^2, s_2^2, and s_3^2 occur in column 5 labeled "SAMPLE VARIANCE," and unbiased estimates of the parameters S_1^2, S_2^2, and S_3^2 occur in column 6, "ESTIMATED POPN. VARIANCE." Either one may be used, but most frequently s_i^2's are available.

Costs in column 7 are assumed proportional to the frequency of sampling units. The cost function thus takes on the form:

$$c = c_i(n + nm + nmk).$$

TABLE 1.--MULTISTAGE ANALYSIS OF SAMPLE COUNTS AYO COST OATA ON SAMPLING
R. MASON'S CALIFORNIA STUDY, VIRUS INSTARS I AND 2, OPTION 2

```
(5X31(F2.0))          99999  179999     250    11
TWO     24 24 24 22 17 25 22 17 23 25 24 24 25 24 30  6 26
THREE 45 45 29 48 46 46 48 50 13 34 40 47 42 47 43 42 31 19 46 48 49
THREE 40 45 49
THREE 53 72 46 41 66 41 48 43 47 46 18 50 45 43 31 49 50 40 43 68 40
TYREE 49 42 21
THRFE 50 42 34 20 27 SO 34 49 58 19 47 25 58 64 50 50 49 5C 24
THREE 38 49 43 35 49
TYRFE 41 50 47 35 41 48 49 32 47 44 11 84 43 50 13 33 24 45 50 38 47
TYRFE 49
THREE 31 39 50 50 C9 49 37 40 33 49 50 49 29 49 53 27 0.3
THREE 50 64 5C 42 79 50 43 50 50 49 48 22 44 54 82 69 50 46 45 07 49
THREE 47 50 50 49
THREE 49 44 07 05 50 11 17 50 28 17 08 25 29 45 03 35 12 46 50 33
THRFE 14 11
THREE 49 48 47 49 14 45 50 47 44 56 47 38 20 87 47 '46 05
THREE 49 46 50 5C 46 49 50 19 50 50 50 50 49 50 49 48 49 55 49 49
THREE 53 50 50
THREE 30 50 49 50 50 48 50 49 26 47 50 49 50 50 50 50 5C 47 50 50 47
THREE 47 27 50 49
THREE 50 27 41 50 46 50 50 34 42 50 50 47 50 50 50 44 48 35 50 59 50
THREE 49 50 48
THREE 50 30 50 24 50 50 48 18 29 11 76 46 50 48 50 49 05 36 50
THREE 49 07 37 12 12
THREE 47 50 47 47 93 50 01 49 47 50 48 48 45 69 63 29 50 49 37 49 50


THREE 49 50 49 50 48
DATA  0311160506130607060507090322081715071604172I...
DATA  07300319020805031412041403110405051606292627260O


CATA  33221010033031364336091116462010O4
OATA  0933153315090202100510080917150IC7080912001208


DATA  0001040406024111270311070706000601182501001904O3
CATA  0105063116020116070618110318090704150302091112120 3
OATA  0239063507030303030000427070000160606010613090607
CATA  2302042204051719020425302316030614180730351505361812060106 20
OATA  144610071933
OATA  160621312120340723151213362116031813353104283536O506
```

```
                 VULTISTAGE                                    SPECIFICATIONS
                  ESTIMATES                                     FOR OPTIMUM
               --------------------                          --------------------
MEAN              2.915253E-01                TOTAL COST

VARIANCE          7.970234E-04                ~ T O. ERROR
               --------------------             ABSOLUTE

                                              PERCENT/100

                                              PROB LEVEL

                                                 T VALUE
* UNEQUAL SAMPLES                                            --------------------
BECAUSE THERE WERE NO COST LIMITS PER STAGE,THE COST OF EACH STAGE WAS CONSIDERED THE SAME
```

The cost values appear in the column labeled "COST" only if costs are entered; otherwise, they are assumed equal.

The optimum sample sizes for minimizing cost subject to meeting the computer variance, .00079702, are $\hat{k}_{opt} = 2.0$; $\hat{m}_{opt} = 2.0$; and $\hat{n} = 17$.

The estimated average or proportion of insects possessing the virus appears under the heading "MULTISTAGE ESTIMATES." The value of this estimate is 29.15253-02. This says that 29.15 percent were virus infected, and the variance of this estimated proportion is 7.970234E-04.

The striking thing about these results is that to satisfy the precision achieved of .00079702 would have required only 2 x 2 x 17 = 68 observations of larvae instead of the 17,204 actually taken. In other words, nothing is gained in the estimated precision of the mean virus occurrence of the population from measuring a large number of eggs per egg mass or a large number of egg masses per plot. It should be realized by now that a particular optimum solution is a characteristic of the population being sampled. If the intensity of infection or occurrence of virus-infected insects increases or decreases, the variances will change thus altering the optimum solution. One must therefore be particularly careful about borrowing input variances from other research studies or surveys. In other words, the variances one chooses for obtaining his optimum solution should be relatively bias-free estimates of his current population of interest.

Optimization problems of this type have a specific, well-defined objective. There are occasions when secondary objectives exist which are of a descriptive nature (e.g., descriptions of relationships of the larvae to their environment). When these occur, it may be desirable to take larger m and k to provide this other type of information, even though it may not be efficient from an estimation point of view.

Appendix

The optimum frequencies n, m, and k are functions of the true variances:

Variance among primary unit means

$$S_1^2 = \sum_{i=1}^{N} (\overline{\overline{Y}}_{i\cdot} - \overline{\overline{Y}})^2/(N - 1)$$

Variance among second-stage unit means

$$S_2^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} (\overline{Y}_{ij} - \overline{\overline{Y}}_{i\cdot})^2/N(M - 1)$$

Variance among third-stage unit means

$$S_3^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{\ell=1}^{K} (y_{ij\ell} - \overline{Y}_{ij})^2/NM(K - 1)$$

where:

$y_{ij\ell}$ is an observation of the ℓ^{th} third-stage unit within the j^{th} second-stage unit within the i^{th} first-stage unit,

$\overline{Y}_{ij}$ is the mean of the ℓ third-stage units within the j^{th} second stage and i^{th} first stage,

$\overline{Y}_{i}$ is the mean over ℓ third-stage units and j second-stage units, within the i^{th} first-stage unit.

$\overline{\overline{Y}}$ is the mean over all stages.

The appropriate sample-based estimates of these parameters are $\overline{y}_{ij}$, $\overline{\overline{y}}_{i}$, and $\overline{\overline{y}}$. When random sampling is assumed at all stages, the sample mean $\overline{\overline{y}}$ and $v(\overline{\overline{y}})$ are unbiased estimators of $\overline{\overline{Y}}$ and $V(\overline{\overline{y}})$. The true variance and sample variance are:

$$V(\overline{\overline{y}}) = (1 - f_1)\frac{S_1^2}{n} + (1 - f_2)\frac{S_2^2}{nm} + (1 - f_3)\frac{S_3^2}{nmk} \qquad (1)$$

$$v(\overline{\overline{y}}) = (1 - f_1)\frac{s_1^2}{n} + f_1(1 - f_2)\frac{s_2^2}{nm} + f_1f_2(1 - f_3)\frac{s_3^2}{nmk} \qquad (2)$$

where $f_1 = n/N$, $f_2 = m/M$, and $f_3 = k/K$ are the sampling fractions at the three stages, which means that N, M, and K must be the number of first-, second-, and third-stage units in the population.

If $k = K$, the problem reduces to a two-stage sampling problem. And logically, if $m = M$ and $k = K$, the problem is one-stage or simple random sampling.

Costs are a necessary consideration in computing estimates of the optimum number of elements at each stage. Let c_i be the average cost of taking an element at the i^{th} stage. Since each unit is composed of subunits, the cost associated with primary units must include all cost components associated with taking a primary unit, but not costs of measuring secondary units. The result is that if we omit travel costs between units and extraneous costs associated with each stage, then the total cost can be expressed as a linear function of n, m, and k; i.e.,

$$c(\text{total cost}) = c_1 n + c_2 nm + c_3 nmk.$$

Then, given these inputs, the optimum n, m, and k are (Cochran 1963):

$$k_{opt} = \left[\frac{S_3}{(S_2^2 - S_3^2/K)^{1/2}}\right]\left[\frac{c_2}{c_3}\right]^{1/2} \tag{3}$$

$$m_{opt} = \left[\frac{(S_2^2 - S_3^2/K)^{1/2}}{(S_1^2 - S_2^2/M)^{1/2}}\right]\left[\frac{c_1}{c_2}\right]^{1/2} \tag{4}$$

The size of n for a specified cost or specified precision are, respectively:

$$n = \frac{(c \text{ specified})}{(c_1 + c_2 m_{opt} + c_3 m_{opt} k_{opt})} \tag{5}$$

$$= \frac{S_1^2 - \frac{(1 - f_2)S_2^2}{m_{opt}} - \frac{(1 - f_3)S_3^2}{m_{opt} k_{opt}}}{(V(\bar{\bar{y}})\text{specified}) + S_1^2/N} \tag{6}$$

Since nothing is usually known about the S_i^2's or the values of N, M, and K, one can estimate n, m_{opt}, and k_{opt} by substituting estimates for the S_i^2's. Unbiased estimates of these true variances exist which are functions of the following sample variances (Sukhatme 1954--expression 56, page 310; 57, page 311):

$$s_1^2 = \sum_{i=1}^{n} (\bar{\bar{y}}_i - \bar{\bar{\bar{y}}})^2/(n - 1)$$

$$s_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (\bar{y}_{ij} - \bar{\bar{y}}_i)^2/n(m - 1)$$

$$s_3^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{\ell=1}^{k} (y_{ij\ell} - \bar{y}_{ij})^2/nm(k - 1)$$

and

$$\hat{S}_1^2 = s_1^2 - (1 - f_2) \frac{s_2^2}{m} - (1 - f_3) \frac{s_3^2}{kM}$$

$$\hat{S}_2^2 = s_2^2 - (1 - f_3) \frac{s_3^2}{k}$$

$$\hat{S}_3^2 = s_3^2$$

Unbiased sample based estimators of equations **(3)** and *(4)* can be obtained by substituting the estimators $\hat{S}_i^2$, $i=1$, 2, or **3** into equations **(3)** and *(4)* and then simplifying. These estimators are:

$$\hat{k}_{opt} = \left[\frac{s_3}{(s_2^2 - s_3^2/k)^{1/2}}\right]\left[\frac{c_2}{c_3}\right]^{1/2} \quad (7)$$

$$\hat{m}_{opt} = \left[\frac{(s_2^2 - s_3^2/k)^{1/2}}{(s_1^2 - s_2^2/m)^{1/2}}\right]\left[\frac{c_1}{c_2}\right]^{1/2} \quad (8)$$

When $\hat{k}_{opt}$ and $\hat{m}_{opt}$ are computed, the values of k and m on the right-hand side of the estimators (7) and (8) come from the sample used to **compute** s_2^2 and s_3^2.

There are two estimators of n--choice depends upon whether the objective specifies a cost or variance restriction. The two forms are:

$$\hat{n}_{(\text{cost})} = \frac{(C \text{ specified})}{c_1 + c_2\hat{m}_{opt} + c_3\hat{m}_{opt}\hat{k}_{opt}} \tag{9}$$

$$\hat{n}_{(\text{variance})} = \frac{\hat{S}_1^{\,2} + \dfrac{(1-f_2)\hat{S}_2^{\,2}}{\hat{m}_{opt}} + \dfrac{(1-f_3)\hat{S}_3^{\,2}}{\hat{m}_{opt}\hat{k}_{opt}}}{(V(\bar{\bar{y}})\text{specified}) + \dfrac{\hat{S}_1^{\,2}}{N}} \tag{10}$$

Note that $\hat{k}_{opt}$, $\hat{m}_{opt}$, and $\hat{n}_{(\text{cost})}$ are functions of sample values. Nothing is required to be known about the population. Also, when M and K are large relative to $\hat{m}_{opt}$ and $\hat{k}_{opt}$, $\hat{n}_{(\text{variance})}$ simplifies to an expression in which only the population size N is required. It should also be apparent that, if N is very large (assumed infinite = 99999), the value of $\hat{n}_{(\text{variance})}$ approaches $s_1^2/V(\bar{\bar{y}})$ (specified) which is the case for simple random sampling from an infinite population. In equation (10), $\hat{n}$(variance) is written as a function of $\hat{S}_i^{\,2}$'s instead of s_i^2's. There is a reason for using (1) instead of (2) for arriving at equation (10). The estimators $\hat{S}_i^{\,2}$'s are unbiased estimators of S_i^2's, and thus their expectations are always constants regardless of the size of n, m, and k. Whereas using (2) requires estimating s_i^2's which are functions of the sample sizes m and k. In other words, if one increases the amount of sampling k, this affects the value of s_2^2 and s_1^2.

This point is not mentioned in Cochran (1963), but it is easy to show that (1) and (2) produce different values of n for the same set of sampling data.

Literature Cited

Carolin, V. M., and W. K. Coulter
19.72. Sampling populations of western spruce budworm and predicting defoliation on Douglas-fir in eastern Oregon. USDA For. Serv. Res. Pap. PNW-149, 38 p., illus. Pac. Northwest For. & Range Exp. Stn., Portland, Oreg.

Cochran, W. G.
1963. Sampling techniques. Ed. 2, 413 p. New York: John Wiley & Sons, Inc.

Langley, Philip G.
1969. New multistage sampling techniques using space and aircraft imagery for forest inventory. Sixth Int. Remote Sensing Environ. Symp. Proc. 2: 1179-1192. Univ. Mich., Ann Arbor.

Sukhatme, P. V.
1954. Sampling theory of surveys with applications. 491 p. Ames: Iowa State Coll. Press.

Yates, F.
1949. Sampling methods for censuses and surveys. 318 p. London, Eng.: Charles Griffin & Co., Inc.

The mission of the PACIFIC NORTHWEST FOREST AND RANGE EXPERIMENT STATION is to provide the knowledge, technology, and alternatives for present and future protection, management, and use of forest, range, and related environments.

Within this overall mission, the Station conducts and stimulates research to facilitate and to accelerate progress toward the following goals:

1. Providing safe and efficient technology for inventory, protection, and use of resources.
2. Development and evaluation of alternative methods and levels of resource management.
3. Achievement of optimum sustained resource productivity consistent with maintaining a high quality forest environment.

The area of research encompasses Oregon, Washington, Alaska, and, in some cases, California, Hawaii, the Western States, and the Nation. Results of the research will be made available promptly. Project headquarters are at:

Fairbanks, Alaska	Portland, Oregon
Juneau, Alaska	Olympia, Washington
Bend, Oregon	Seattle, Washington
Corvallis, Oregon	Wenatchee, Washington
La Grande, Oregon	

Mailing address: Pacific Northwest Forest and Range
Experiment Station
P.O. Box 3141
Portland, Oregon 97208

GPO 989-786

The FOREST SERVICE of the U. S. Department of Agriculture is dedicated to the principle of multiple use management of the Nation's forest resources for sustained yields of wood, water, forage, wildlife, and recreation. Through forestry research, cooperation with the States and private forest owners, and management of the National Forests and National Grasslands, it strives — as directed by Congress -- to provide increasingly greater service to a growing Nation.