# SAMPLE SIZE AND ITS OPTIMUM ALLOCATION WITH CONSTRAINTS ON PRECISION OF STRATUM ESTIMATES—AN EXTENSION OF NEYMAN'S METHOD

By
SHAMBHU DAYAL
*NSSO, Faridabad*

(Received : September, 1979)

SUMMARY

**A procedure of determining sample size and its optimum allocation to various strata under stratified random sampling has been obtained when stratum level estimates are also required. A more general case where stratum level sample sizes are bounded is also discussed and it is observed that in certain situations, the optimum allocation may not be feasible.**

## 1. PROCEDURE OF ALLOCATION

Following Sukhatme and Sukhatme (1970), Neyman's method of allocation consists in choosing $n_h$ so as to minimize

$$V(\bar{Y}_{st})=\sum_{h=1}^{L} (W_h^2\ S_h^2/n_h)-\sum_{h=1}^{L} (W_h^2\ S_h^2/N_h) \qquad (1.1)$$

where $(N_h/N)=W_h$, subject to $\sum_{h=1}^{L} n_h=n$ and its value is given by

$$n_h=n\ W_h\ S_h/\sum W_h\ S_h;\ h=1...L. \qquad (1.2)$$

However, in Neyman's allocation the aim is to minimize the variance of the overall mean $\bar{Y}_{st}$ without any regard for the precision of strata estimates. In many situations, it is desirable to have stratum level estimates with specified precision and, Neyman's method may fail. Suppose the desired precision, in terms of relative variance, for the $h$-th stratum is

$$\{V(\overline{Y}_h)/\overline{Y}_h^2\} \leqslant b_h,\ h=1,\dots,L,$$

where $b_h$'s are given constants. This gives

$$n_h \geqslant a_h,\ h=1,\ .\ .\ ,\ \text{L},$$

where $a_h$'s are constants depending upon $b_h$, $N_h$ and $S_h^2$. Thus the problem is to minimise $V(\overline{Y}_{st})$ subject to constraints

$$n_1+\dots+n_L-n=0 \tag{1.3}$$

and

$$a_h-n_h \leqslant 0,\ h=1,2,\dots,L, \tag{1.4}$$

of course, $n$ should be greater than $\sum a_h$, otherwise the constraints will not be satisfied.

To solve the problem, let us first minimise $V(\overline{Y}_{st})$ subject to the constraint (1.3) following Neyman's method. Let $\underline{n}^*=(n_1^*,\dots,n_L^*)$ be the solution. This solution will either satisfy all the constraints at (1.4) (—case I), or will not satisfy all or some of the constraints at (1.4) (—case II).

*Case (i) :* In this case, $\underline{n}^*$ will be the optimum solution to the problem, since, by Hadley [1], addition of new constraints at (1.4) will not improve the situation.

*Case (ii) :* In this case, we take the equality sign in one of the $L$ constraints given at (1.4), say in $h'$-th constraint and minimise $V(\overline{Y}_{st})$ subject to the constraint $a_h'=n_h'$ as well as (1.3). Following the usual procedures for minimizing

$$\phi=V(\overline{Y}_{st})+\lambda_1(n_1+\dots+n_L-n)+\lambda_2\,(a_h'=n_h') \tag{1.5}$$

let the optimum values of $n_h$ for $h \neq h'$ be

$$\underline{n}^x=(n_1^x,\dots,n_L^x) \tag{1.6}$$

which satisfies $(L-1)$ constraints

$$a_h-n_h<0 \text{ for } h \neq h' \tag{1.7}$$

with strict inequality and also let

$$\lambda_2<0 \tag{1.8}$$

Further, let $V(\bar{Y}_{st})+\sum_{h=1}^{L} u_h\,(a_h-n_h)+u_{L+1}(n_1+\ldots+n_L-n)=\Psi$ (1.9)

where

$$u_j \geqslant 0, \text{ for } j=1,\ldots,(L+1).$$

Then in order that $\underline{n}^x$ be a solution to the problem of minimisation of $V(\bar{Y}_{st})$ subject to (1.3) and (1.4); it is necessary, Kuhn and Tucker [2] that $\underline{n}^x$ and some $\underline{u}^x=(u_1^x,\ldots,u_{L+1}^x)$ satisfy

$$\frac{\partial\Psi}{\partial n_h}=0,\ n_h>0,\ h=1,\ldots,L \qquad \ldots(1.10)$$

and

$$\frac{\partial\Psi}{\partial u_j}\geqslant 0,\ \frac{\partial\Psi}{\partial u_j}u_j=0 \text{ and } u_j\geqslant 0 \qquad \ldots(1.11)$$

for $j=1,\ldots,L+1$, and also $u_j^x=0$ if the $j$-th restriction in (1.4) comes out to be with strict inequality at $\underline{n}^x$. Since the solution (1.6) satisfies (1.7), we get

$$u_j^x=0, \text{ for } j\neq h' \text{ and } (L+1).$$

Then $\Psi$ reduces to

$$\phi=V(\bar{Y}_{st})+u_{h'}\,(a_{h'}-n_{h'})+u_{L+1}\,(n_1+\ldots+n_L-n)$$

if we put

$$u_{h'}=\lambda_2$$

and

$$u_{L+1}=\lambda_1.$$

We note that

$$u_{h'}^x>0 \text{ by } (1.8)$$

$$u_{L+1}^x>0$$

and

$$\frac{\partial\phi}{\partial u_{h'}}=\frac{\partial\phi}{\partial u_{L+1}}=0 \qquad \ldots(1.12)$$

at $\underline{n}^x$, $u_{h'}^x$ and $u_{L+1}^x$. Thus the necessary conditions for $\underline{n}^x$ to be optimum solution to the minimization problem are satisfied. Also since $V(\bar{Y}_{st})$, $\sum n_h$ and $a_h-n_h$ are convex functions of $\underline{n}$, conditions (1.7), (1.8) and (1.12) are also sufficient, Kuhn and Tucker [2]. Thus, $\underline{n}^x$ at (1.6) is the optimum solution to the minimisation problem of $V(\bar{Y}_{st})$

subject to (1.3) and (1.4). It is to be noted that the solution obtained through the above procedure has to satisfy (1.7) and (1.8) and only then it becomes optimal solution. The satisfaction of (1.8) also means

$$\sum_{h \neq h'} (W_h S_h)/(n - a_{h'}) > W_{h'} S_{h'}/a_{h'} \tag{1.13}$$

If the above solution does not give the optimal solution, we make another inequality constraint in (1.4) as equality constraint and proceed as above. If this also does not give the optimal solution, we continue the process until we have tried all the combinations of including just one of the inequality constraints with equality sign. If the optimum is not attained by this, we continue the process making two or more inequality constraints as equality till the optimum is found.

**Note.** It is stated above that $u_j = 0$ for $j \neq h'$, $(L+1)$ if the solution (1.6) satisfies (1.7) with strict inequality. The condition of strict inequality is, however, not necessary for the corresponding $u_j$ to be equal to zero. For, let the solution (1.6) satisfy (1.7) with equality at $h''$-th stratum. This means that

$$n_{h \neq h'} = (n - a_h') W_h S_h / \sum_{h \neq h'} W_h S_h \qquad \ldots(1.14)$$

and

$$a_h'' = n_h'' = (n - a_h') W_h'' S_h'' / \sum_{h \neq h'} W_h S_h \qquad \ldots(1.15)$$

If we add $\lambda_3 (a_h'' - n_h'')$ to (1.5), we get the value of

$$\lambda_3 = \left[ \left\{ \left( \sum_{h \neq h', h''} W_h S_h \right) / (n - a_h' - a_h'') \right\} + W_h'' S_h'' / a_h'' \right] \times$$

$$\left[ \left\{ \left( \sum_{h \neq h', h''} W_h S_h \right) / (n - a' - a_h'') \right\} (-W_h'' S_h'' / a_h'') \right] \qquad \ldots(1.16)$$

The second factor of (1.16) is equal to

$$\left\{ \left( \sum_{h \neq h'} W_h S_h - W_h'' S_h'' \right) / (n - a_h' - a_h'') \right\} - W_h'' S_h'' / a_h''$$

which will be zero if we put the value of $a_h''$ from (1.15). This gives $u_3^x=\lambda_3=0$. Thus if the solution $\underline{n}^{\times}$ satisfies $(L-1)$ constraints

$$a_h - n_h \leqslant 0 \text{ for } h \neq h'$$

and also if $\lambda_2 > 0$, $\underline{n}^{\times}$ is an optimum solution.

## 2. Optimum Sample Size

Here we minimise

$$n = n_1 + \ldots + n_L$$

subject to constraints

$$V(\bar{Y}_{st}) = \sum_{h=1}^{L} (W_h^2 S_h^2 / n_h) - \sum_{h=1}^{L} W_h^2 S_h^2 / N_h$$

$$= V, \text{ a constant} \qquad \ldots(2.1)$$

and

$$a_h - n_h \leqslant 0 \text{ for } h = 1, 2, \ldots, L. \qquad \ldots(2.2)$$

The problem is similar to the one discussed in Section 1 and can be solved accordingly.

**Note 1.** In sections 1–2, we cannot have constraints like.

$$\sum n_h \leqslant n$$

since minimum value of $V(\bar{Y}_{st})$ will be approached when $\sum n_h = n$. Similarly we cannot have $V(\bar{Y}_{st}) \leqslant V$.

**Note 2.** The problem of optimum allocation of sample with given cost and minimum precision to stratum estimates as also the determination of minimum total cost with given precision of $\bar{Y}_{st}$ and given minimum precision of stratum estimates can also be solved in the manner discussed in Section 1-2.

## 3. Optimisation with Some Other Constraints

In some situations instead of constraints $n_h \geqslant a_h$, $h=1, \ldots, L$ as in Section 1, it may be desirable to have

$$n_h \leqslant d_h \text{ for } h = 1, \ldots, L. \qquad \ldots(3.1)$$

where $d'_h$ s are given constants. This problem can also be solved in the manner given in Section I. Of course, there has to be another constraints,

$$n_h > 0, \; h = 1, \ldots, L.$$

A general case will be when we have constraints

$$d_h \geqslant n_h \geqslant a_h \text{ for } h = 1, \ldots, L. \qquad \ldots(3.2)$$

In such a situation, we first solve the problem with constraints

$$n_h \geqslant a_h \text{ or } d_h \geqslant n_h$$

only and at each stage of this solution we see whether the general constraints $d_h \geqslant n_h \geqslant a_h$ are also satisfied.

We stop when this has been achieved and that solution will be the optimum solution. However, in this general case, the solution, sometimes, may not exist even if we have

$$\sum d_h \geqslant n \geqslant \sum a_h \qquad \ldots(3.3)$$

## 4. Application of the Procedure

The procedure given in section I is, in practice, not lengthy as it appears to be. We discuss the cases when one, two or more constraints given at (1.4) are not satisfied.

*Case 1.* Suppose after allocating the sample following Neyman's method *i.e.*, after putting

$$n_h \propto W_h S_h \text{ for } h = 1, \ldots, L \qquad \ldots(4.1)$$

only the first constraint at (1.4) is not satisfied, *i.e.*

$$a_1 > n_1 \qquad \ldots(4.2)$$

and the other constraints are satisfied. We put (4.1) also as

$$n_h = \mathcal{B}\, W_h S_h \text{ for } h = 1, \ldots, L \qquad \ldots(4.3)$$

where $\mathcal{B}$ is a constant. We discuss different situations as follows :

(A) We assume that in strata other than the first, the constraints

$$n_h \geqslant a_h \text{ for } h = 2, 3, \ldots, L$$

are satisfied after putting (4.1) Let us take

$$n_2 - \beta = a_2 \qquad \ldots(4.4)$$

or

$$\mathcal{B}\, W_2 S_2 - \beta = a_2 \qquad \ldots(4.5)$$

from (4.3) for the second stratum where $\beta \geqslant 0$ is a constant. Suppose we allocate $(n_2-\beta)$ units to the second stratum and thereafter, we allocate remaining units following Neyman's method. Further, suppose that this allocation satisfies constraints of (1.4). Now in order that this allocation is optimum, (1.17) also should be satisfied. This means that

$$\sum_{h\neq 2} W_h S_h/(n-a_2) > W_2 S_2/a_2$$

or

$$\left(\sum_{h\neq 2} W_h S_h\Big/\left(\mathcal{L}\sum W_h S_h+\beta\right) > W_2 S_2/\mathcal{L}\, W_2 S_2-\beta\right)$$

or

$$\beta\sum_{h=1} W_h S_h<0 \qquad \ldots(4.6)$$

a condition which is never satisfied.

**(B)** In this situation, in addition to (4.5), we put

$$\mathcal{L}\, W_3 S_3-\delta=a_3 \qquad \ldots(4.7)$$

where $\delta\geqslant 0$ is a constant and allocate $n_2-\delta$ and $n_3-\delta$ units to second and third strata respectively and remaining units following Neyman's method. Suppose that this allocation satisfies constraints of (1.4). Now in order that this allocation is optimum, it should also satisfy

$$\sum_{h\neq 2,\,3} W_h S_h/(n-a_2-a_3)>W_2 S_2/a_2$$

and

$$\sum_{h\neq 2,\,3} W_h S_h/(n-a_2-a_3) > W_3 S_3/a_3$$

which mean that

$$\beta\sum_{h\neq 3} W_h S_h+\delta W_2 S_2 < 0 \qquad \ldots(4.8)$$

and

$$\delta\sum_{h\neq 2} W_h S_h+\beta W_3 S_3 < 0 \qquad \ldots(4.9)$$

the conditions which are never satisfied.

Proceeding as in Situations ($A$) and ($B$), it can be shown that the optimum allocation will not be attained unless we allocate '$a_1$' units to first stratum. This, in general, shows that if after allocating the sample following Neyman's method a particular constraint at (1.4) is not satisfied, the same constraint is to be put with equality sign in order to arrive at an optimum solution. After this, we should allocate the remaining sample to remaining strata following Neyman's method and if this allocation satisfies all the constraints at (1.4), this allocation will be optimum allocation as we prove at situation ($C$) below.

(C). If, after allocating the sample following Neyman's method, (4.2) holds, we allocate $a_1$ units to the first stratum, and $n-a_1$ units to the remaining strata following Neyman's method. Since $a_1 > n_1$, the number of units in the remaining $(L-1)$ strata will now be reduced from what was allocated following (4.1). If this allocation satisfies remaining constraints of (1.4), it will be optimum allocation provided it also satisfies (1.13), which means that

$$\left(\sum_{h\neq 1}^{L} W_h S_h\right) \Big/ \left(\alpha \sum_{h=1}^{L} W_h S_h - a_1\right) > W_1 S_1 / a_1$$

or

$$a_1 > \alpha W_1 S_1 \quad \ldots(4.10)$$

which holds by (4.2). Thus this allocation will be optimum allocation.

If this allocation does not satisfy one of the remaining $(L-1)$ constraints at (1.4), the arguments at ($A$), ($B$) and ($C$) will show that only that constraint which is not satisfied is to be put to equality for finding out an optimum allocation. The procedure can be continued till all the constraints at (1.4) are satisfied.

*Case II:* Suppose after allocating the sample following Neyman's method, the first and second constraints at (1.4) are not satisfied *i.e.*

$$a_1 > n_1 \quad \ldots(4.11)$$

and

$$a_2 > n_2 \quad \ldots(4.12)$$

($A$) Proceeding as in cases $I(A)$ and $I(B)$, it can be shown that if we put equality constraints at (1.4) for strata other than the first and the second, and then allocate the remaining sample to remaining strata following Neyman's method, the conditions like (1.13) will not be satisfied. This type of allocation will not, therefore, give the optimum allocation.

($B$) Next, let us make one of the constraints at (4.11) and (4.12), say at (4.11), as equality constraint and the constraint in one of the remaining strata, say in the third, as equality constraint and allocate the remaining sample to the remaining strata following Neyman's method. Suppose this allocation satisfies all the constraints at (1.4). This allocation will be optimum if it satisfies conditions like (1.17) *i.e.* if

$$\sum_{h\neq 1,3} W_h S_h \Big/ \left( \mathcal{B} \sum_{h=1}^{L} W_h S_h - a_1 - a_3 \right) > W_3 S_3 / a_3 \quad \text{...(4.13)}$$

and

$$\left( \sum_{h\neq 1,3} W_h S_h \right) \Big/ \left( \mathcal{B} \sum_{h=1}^{L} W_h S_h - a_1 - a_3 \right) > W_1 S_1 / a_1 \quad \text{...(4.14)}$$

(4.13) gives

$$a_3 \sum_{h\neq 1,3} W_h S_h > W_3 S_3 \left( \mathcal{B} \sum_{h=1}^{L} W_h S_h - a_1 - a_3 \right)$$

or

$$W_3 S_3 (a_1 - \mathcal{B} W_1 S_1 + a_3 - \mathcal{B} W_3 S_3) > (\mathcal{B} W_3 S_3 - a_3) \sum_{h\neq 1,3} W_h S_h \quad \text{...(4.15)}$$

Since $\mathcal{B} W_3 S_3 > a_3$, RHS of (4.15) is positive. Now, in order that the present allocatian does not satisfy (4.12), it is necessary that

$$\mathcal{B} W_3 S_3 - a_3 > a_1 - \mathcal{B} W_1 S_1 \quad \text{...(4.16)}$$

otherwise the present allocation will reduce the earlier allocated sample to second stratum (*i.e.* $n_2$) and (4.12) will stand. (4.16) will mean that LHS of (4.15) is negative and (4.15) will not be satisfied. This allocation is not, therefore, optimum.

In a similar manner it can be shown that if we take only one of the constraints at (4.11) and (4.12) as equality and also take more than one other constraints at (1.4) as equality, the solution will not be optimal.

(*C*) Further, let us take constraints in the first and second strata as equality and allocate the remaining sample to remaining strata following Neyman's method. In order that this gives an optimum allocation, it should satisfy (1.4) and also conditions like (1.13). The satisfaction of conditions like (1.13) means that

$$\left(\sum_{h\neq 1,2} W_h S_h\right) \Big/ \left( \mathcal{B} \sum_{h=1}^{L} W_h S_h - a_1 - a_2 \right) > W_1 S_1 / a_1 \quad \ldots (4.17)$$

and

$$\left(\sum_{h\neq 1,2} W_h S_h\right) \Big/ \left( \mathcal{B} \sum_{h=1}^{L} W_h S_h - a_1 - a_2 \right) > W_2 S_2 / a_2 \quad \ldots (4.18)$$

(4.17) gives

$$(a_1 - \mathcal{B}\, W_1 S_1) \sum_{h\neq 1,2} W_h S_h + W_1 S_1 (a_2 - \mathcal{B}\, W_2 S_2) > 0 \quad \ldots (4.19)$$

which will be satisfied by (4.11) and (4.12). Similarly (4.18) will also be satisfied. Thus this allocation will be optimum if it satisfies (1.4). If this does not satisfy one or two constraints at (1.4), we have to put the particular constraint/constraints which is/are not satisfied as equality constraint (*s*) and then proceed to see whether this revised allocation satisfies (1.4). The process may continue till we get the optimum allocation.

*Case III :* The procedure given in case II can be extended to cases where, after allocating the sample following Neyman's method, more than two constraints at (1.4) are not satisfied.

*Example 4.1 :* Table 4.1 gives the values of $W_h S_h$ and $a_h$ and the various allocations of the sample of size 1450 to nine strata. Variance of the mean, ignoring f.p.c., has also been given for different allocations. Column 4 gives the values according to Neyman's allocation. After adopting this allocation, constraints (1.4) in stratum numbers 3, 5 and 9 are not satisfied. If we put them as equality constraints and allocate the remaining sample to remaining 6 strata

**TABLE 4.1**

**Allocation of the sample**

| *Stratum number* | $W_h S_h$ | *a* | *Neyman's allocation* | *I* | *II* | *III* | *IV* | *V* | *VI* | *VII* | *VIII* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *1.* | *2.* | *3.* | *4.* | *5.* | *6.* | *7.* | *8.* | *9.* | *10.* | *11.* | *12.* |
| 1 | 36.53 | 63 | 73 | 68 | 62 | 63 | 63 | 69 | 69 | 67 | 65 |
| 2 | 24.62 | 45 | 49 | 46 | 42 | 45 | 45 | 47 | 45 | 49 | 46 |
| 3 | 13.15 | 45 | 26 | 45 | 45 | 45 | 45 | 45 | 45 | 48 | 46 |
| 4 | 314.63 | 600 | 627 | 584 | 600 | 600 | 600 | 600 | 600 | 604 | 615 |
| 5 | 100.37 | 260 | 200 | 260 | 260 | 260 | 260 | 260 | 260 | 264 | 265 |
| 6 | 151.93 | 290 | 303 | 282 | 290 | 290 | 290 | 290 | 290 | 294 | 297 |
| 7 | 57.26 | 79 | 114 | 106 | 96 | 93 | 79 | 79 | 108 | 83 | 81 |
| 8 | 22.17 | 15 | 44 | 41 | 37 | 36 | 50 | 42 | 15 | 19 | 16 |
| 9 | 6.58 | 18 | 13 | 18 | 18 | 18 | 18 | 18 | 18 | 22 | 19 |
| Total | 727.24 | 1415 | 1450 | 1450 | 1450 | 1450 | 1450 | 1450 | 1450 | 1450 | 1450 |
| Variance | — | — | 365.00 | 372.08 | 372.97 | 373.14 | 375.56 | 375.02 | 385.51 | 383.84 | 387.77 |

following Neyman's method, the constraints in stratum numbers 4 and 6 are not satisfied (Col. 5). If we put these also as equality constraints and allocate the remaining sample to remaining 4 strata, constraints in stratum numbers 1 and 2 are not satisfied (Column 6). If we put these also as equality constraints and allocate the remaining sample to remaining 2 strata, all the constraints as well as conditions like (1.13) will be satisfied and this gives the optimum allocation (Column 7). Columns 8 to 12 give examples of allocations where all the constraints at (1.4) are satisfied but these are not the optimum allocations.

## ACKNOWLEDGEMENT

The author is grateful to the referees for their valuable comments.

## REFERENCES

[1] Hadley, G., (1964) : Nonlinear and Dynamic Programming, Chapter 3, Sections 3-6, Addison—Wesley Publishing Company, Japan.

[2] Kuhn, H.W. and Tucker, A.W., (1951) : Nonlinear Programming, Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability, ed. by Neyman, J., University of California Press, Berkeley.

[3] Sukhatme, P.V. and Sukhatme, B.V. (1970) : *Sampling Theory of Surveys with Applications*; Indian society of Agricultural Statistics, New Delhi.