CONTRIBUTIONS TO STRATIFIED SAMPLING AND SOME RELATED PROBLEMS

By

V. K. Sethi

Institute of Social Sciences
Agra University, Agra.

P R E F A C E

In this thesis I have presented some of my results on problems connected with stratified sampling. Most of the results have already been published in various journals or have been submitted for publication. The thesis includes some extension of the published results and a much greater amount of evidence to support some conjectures than could be accomodated in papers submitted to journals.

A major portion of the thesis in contained in Part I. This part deals with the problem of optimum stratification and the use of simple rules to approximate optimum stratification. In this thesis attention has been confined to Normal and Gamma distributions. It is believed that the results can be used for a large variety of populations.

Some of the important conclusions of this part are given below.

i) Stratification that is optimum for minimum variance allocation is almost optimum for equal allocation.

ii) Equal and minimum variance allocations are almost identical for stratification that is optimum for equal allocation.

iii) Equalizing strata totals is a bad rule of stratification for equal or minimum variance allocation.

iv) Equalizing strata ranges is a bad rule (for the populations considered) of stratification for equal or minimum variance allocation.

v) Equalizing the product of stratum range and stratum weight is a good rule for some of the populations (very skew amongst the populations considered) but bad for others (the more symmetrical amongst the populations considered).

vi) Equalizing cumulatives of the square-root of the frequency function gives a close approximation to optimum stratification for both equal and minimum variance allocations.

vii) The optimum points of stratification for proportional allocation for any Gamma distribution divides its aggregate into almost the some proportions into which the optimum points of stratification for proportional allocation for Normal distribution divide its total frequency.

The main results of this part have already been published in the Australian Journal of Statistics (1963).

Part II deals with some other problems of stratified sampling as well as some problems whose scope is not confined to stratified sampling alone. The problems are listed below.

i) Decision about the number of strata.

ii) Optimum method of selecting a pair of units from a stratum and consequent modifications of optimum stratification rules.

iii) Solution of certain problems in programming and its application to the problem of optimum allocation of sample size to strata in multipurpose surveys.

iv) Extension of results derived for equal probability sampling to varying probability sampling.

v) The problem of obtaining unbiased ratio estimates.

Important conclusions from this part are listed below.

i) The decision about the number of strata depends not only on the correlation between the stratification variable and estimation variable but also on the distribution of the stratification variable.

ii) The best method of selecting a pair of units from a population is to arrange the units in a monotonic sequence of the values of the estimation variable and then draw a pair of units equally distant from the two ends.

iii) If pairs of units are to be selected from strata it would help a great deal if the within strata distributions of the estimation variable are made symmetrical and optimum pairs of units drawn from them.

iv) The problem of optimum allocation in a multipurpose survey can be presented as a problem in programming.

v) A large number of problems in programming can be reduced to problems of maximizing a function under equality conditions only.

vi) Selection with varying probabilities can be interpreted as selection from a hypothetical population with equal probabilities. This interpretation enables us to use all the results of equal probability sampling for varying probability sampling also.

Tables given in Part I are abstracted from a much larger set of tables. These tables, as obtained directly from electronic computers, are given in a separate volume as annexure. It is felt that these tables may be used directly for taking decisions about the stratification points and number of strata for a large variety of populations.

I wish to express my gratitude to Dr. Des Raj and Sri D.B.Lahiri who introduced me to the theory and practice of sampling and for suggesting some of the problems included in the thesis. I also wish to express my thanks to Miss Irene Hess and Prof. Leslie Kish of the University of Michigan for the opportunity I have had to work with them on the problem of optimum stratification and for valuable suggestions in that connection. I wish to thank the University of Michigan for allowing me to use the services of the electronic computer which helped in verifying and extending the tables prepared in connection with optimum stratification.

I am indebted to Dr. M.N.Murthy and Mrs. Nanjamma Chenappan for allowing me to include in this thesis parts of the paper 'Some Sampling Systems Providing Unbiased Ratio Estimators' of which they are co-authors with me.

I wish to acknowledge my obligation to Dr.B.P.Adhikari under whose general guidance I have prepared the thesis and who has taken great pains in suggesting modifications in the presentation of the material. Finally I thank Sri Gour Das of the office of the Research and Training of the Indian Statistical Institute for his help in typing the thesis.

Calcutta,
November 1963.

V. K. Sethi

C O N T E N T S

Page

PART I

PART I

PROBLEM OF OPTIMUM STRATIFICATION OF POPULATIONS

CHAPTER I

OPTIMUM STRATIFICATION AND ITS APPROXIMATION

1.1. The Problem: Very often the object of a survey is to estimate the mean of a population from a sample drawn from it. We shall confine our attention to univariate populations. The problem is to find the method of stratification that would reduce the variance of the estimator of the mean based on the selected sample.

The optimum of any operation depends on the way the other operations are planned to be performed. Thus, for example, given the stratification, the optimum allocation of sample size to different strata depends on the method of selection. The allocation that is optimum for equal probability selection may be quite far from it for a varying probability selection. Initially attention will be confined to selection with equal probabilities independently in all strata. Later it will be shown how the results could be applied to varying probability selection and what special techniques should be adopted for dependent selection.

The first problem in stratification is to make strata homogeneous with respect to the estimation variable. This can be done by pooling together similar units. It is customary to think that a vague information about the units is enough for this purpose. Quite often geographic stratification is very good for a multi-purpose socio-economic survey. Many statisticians are of the opinion that only such qualitative data should be used for stratification and the more precise quantitative information reserved for other operations.

We shall start by considering a univariate population only where more precise quantitative information may be very useful. This problem has been discussed in considerable details by Dalenius and Gurney (1950, 1951). They have found the conditions which the optimum points of stratification should satisfy when method of allocation has been specified. The conditions have been derived under the assumption that the distribution of the variable under study is continuous. In practice the distribution is never continuous nor is it known precisely. Thus optimum stratification is an ideal which we can only try to approximate but never hope to achieve.

It has also been demonstrated by Cochran (1953), specifically for two strata, that minor shifts in the strata boundaries near the optimum have negligible effects on increasing variance of the estimator.

We can reasonably assume that the approximation that we can achieve would be quite close to the optimum provided the distribution of a closely related variable is known quite precisely. To achieve this approximation we would find optimum points of stratification for this related variable. For this we will have to search for points for this variable satisfying the conditions given by Dalenius.

The conditions for optimum stratification derived by Dalenius lead to iterative methods of finding the strata boundaries. The iteration becomes quite cumbrous even for simple continuous distributions. Knowing that in practice we can achieve only approximations it seems natural to replace this cumbrous method by some simple rules. A number of alternative simplifications have been proposed by Mahalanobis (1952), Ayoma (1954), Dalenius and Hodges (1957), Ekman (1959), and Durbin (1959).

It is the purpose of this part of the thesis to judge the performance of these simplifications for a large variety of populations. Tables of optimum stratification have been prepared for these populations. Actual stratification may be achieved by referring to these tables. Thus these tables may provide an alternative simplification with a wide use.

Mathematically the problem of optimum stratification for continuous distributions may be presented as follows.

We assume that the estimation variable X has the frequency function f(X). This has to be divided into L strata by (L-1) stratification points, $x_1, x_2, \ldots, x_{L-1}$. The object is to estimate the mean of the distribution with the help of observations on samples selected independently. We shall follow the notation used in the results derived by Dalenius and Gurney.

$$\int_{x_{h-1}}^{x_h} f(x)\,dx = W_h \qquad \ldots \qquad \ldots \qquad (1)$$

$$\int_{x_{h-1}}^{x_h} x f(x)\, dx = W_h \mu_h \qquad \ldots \qquad \ldots \qquad (2)$$

$$\int_{x_{h-1}}^{x_h} (x - \mu_h)^2 f(x)\, dx = W_h \sigma_h^2 \qquad \ldots \qquad (3)$$

and

$$\int_{x_{h-1}}^{x_h} g(x)\, f(x)\, dx = W_h C_h \qquad \ldots \qquad \ldots \qquad (4)$$

Here W_h stands for the stratum weight, μ_h the stratum mean, σ_h^2 the stratum variance, $g(x)$ the cost for investigating a unit with value x and C_h the average cost per unit investigated in the h^{th} stratum.

The estimator of the population mean is

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \qquad \ldots \qquad \ldots \qquad (5)$$

where $\bar{x}_h$ is the sample mean in the h^{th} stratum, the sample size in the h^{th} stratum being n_h.

The problem is to minimize the variance of $\bar{x}_{st}$ with respect to $x_1, x_2, \ldots, x_{L-1}$. As the variance of the estimator depends on the allocations n_h of sample size to different strata also, the solutions of the problem are functions of methods of allocation.

1.2. <u>Conditions for Optimum Stratification</u>: The general conditions were derived by Dalenius and Gurney (1951) of which the following are of particular interest.

i) <u>Proportional Allocation</u>: If $n_h \propto W_h$, the conditions are

$$2\, x_h = (\mu_h + \mu_{h+1}) \quad \ldots \quad \ldots \tag{6}$$

$$h = 1, 2, \ldots, L-1$$

ii) <u>Minimum Variance Allocation</u>: If $n_h \propto W_h\, \sigma_h$, the conditions are

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}, \tag{7}$$

$$h = 1, 2, \ldots, L-1 .$$

iii) <u>Equal Allocation</u>: If the sample sizes in various strata are equal the conditions are

$$W_h\, [\sigma_h^2 + (x_h - \mu_h)^2] = W_{h+1}[\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2], \tag{8}$$

$$h = 1, 2, \ldots, L-1.$$

iv) <u>Optimum Allocation</u>: If $n_h \propto W_h\, \sigma_h\, / C_h^{\frac{1}{2}}$, the conditions are

$$\frac{\sigma_h^2\, g(x_h) + \sigma_h (x_h - \mu_h)^2}{\sigma_h\, C_h^{\frac{1}{2}}} = \frac{\sigma_{h+1}^2\, g(x_h) + \sigma_{h+1}(x_h - \mu_{h+1})^2}{\sigma_{h+1}\, C_{h+1}^{\frac{1}{2}}} \tag{9}$$

$$h = 1, 2, \ldots, L-1 .$$

1.3. The Alternative Simple Rules: As mentioned earlier the exact solutions suggested by Dalenius and Gurney give only the relations that the optimum points of stratification (OPS) should satisfy. To find such points one has to use iterative methods. For proportional allocation and only two strata the method is quite simple. For equal and minimum variance allocations the difficulty starts even at two strata. As the number of strata increases even proportional allocation points become more and more difficult to locate. The following rules have been suggested as approximations to the actual optimum for equal and minimum variance allocation cases.

i) Equalization of Strata Totals: This rule is the one most commonly used in practice. It was observed that the within strata variances increased with an increase in within strata means. As a simplified model of this situation it was assumed that when stratification is optimum the within strata relative variances are nearly equal. This means that equalizing strata totals $W_h\mu_h$ also equalizes the quantities $W_h\sigma_h$ and consequently equal allocation becomes optimum allocation.

Hansen, Hurwitz and Madow (1953) prove formally that, when PSU relative variances are the same and remain about the same on adjusting the strata boundaries, equalization of strata totals gives the optimum stratification for equal allocation of PSU's to all strata.

Though this simple rule was first advocated for equal allocation it should be as good for minimum variance allocation if the observation that, for optimum stratification in that case, $W_h \sigma_h$ tend to be almost equal is true. In a later chapter of the thesis we shall verify if the following observations are true for the type of stratification and the family of distributions under consideration.

(a) σ_h / μ_h are almost equal.

(b) $W_h \sigma_h$ are almost equal.

ii) Equalization of Strata Ranges : This rule is also widely in use. Ayoma (1956) derives it from the conditions for optimality given by Dalenius under the assumption that within a stratum the distribution can be approximated by a rectangular distribution. This assumption is apparently reasonable if the number of strata is large. It will be investigated how this rule performs when the number of strata is small and the assumption is likely to be invalid.

iii) Equal Intervals on Cumulatives of $f^{\frac{1}{2}}(x)$.

Let

$$y = G(x) = \int_a^x f^{\frac{1}{2}}(u)\,du \qquad \dots \qquad \dots \qquad (10)$$

If

$$G(b) = K \qquad \dots \qquad \dots \qquad (11)$$

where a and b are respectively the lower and the upper bounds of the distribution of X, then this rule given by Dalenius and Hodges (1957) defines the stratification points x_h^* by

$$G(x_h^*) = hK/L, \qquad \dots \qquad \dots \qquad (12)$$

$$h = 1, 2, \dots, L-1.$$

A non-rigorous justification given by the authors of this approximation is as follows.

If there are a large number of narrow strata the density at each point within a stratum may be taken to be constant. Then

$$W_h = f(x_h)(x_h - x_{h-1}), \qquad \ldots \qquad (13)$$

$$\sigma_h = (x_h - x_{h-1}) / (12)^{\frac{1}{2}}, \qquad \ldots \qquad (14)$$

and

$$G(x_h) - G(x_{h-1}) = f^{\frac{1}{2}}(x_h)(x_h - x_{h-1}) \qquad \ldots \qquad (15)$$

The standard error of the estimator of the population mean by a stratified sample with minimum variance allocation is proportional to

$$\sum_{h=1}^{L} W_h \sigma_h = \sum_{h=1}^{L} f(x_h)(x_h - x_{h-1})^2$$
$$= \sum_{h=1}^{L} [G(x_h) - G(x_{h-1})]^2 \qquad \ldots \qquad (16)$$

As $G(b) - G(a)$ is fixed (K) the right side of the equation (16) is minimized by making $G(x_h) - G(x_{h-1})$ a constant.

A rigorous proof that the approximation is good involves two assumptions, one that the distribution is bounded and second that the number of strata is large. The first of these assumptions is true for any real survey work. It seems that if the number of strata is very large quite a number of rules may lead to almost the same precision. It may be recalled that this is exactly the same assumption as employed by Ayoma for deriving the rule of equal strata ranges.

Our real interest lies in the performance of the rules for a small number of strata. What makes this rule important is the observation by the authors of the rule that this gives a close approximation to the optimum stratification for as few as two or three strata for certain continuous distributions (1960). Cochran (1961) confirms this observation for a number of actual finite populations.

The non-rigorous demonstration given above with minor modifications (replacing expression (16) by the expression for the variance when allocation is equal) suggests that the rule should be good for equal allocation also.

iv) Equalization of the Product of Strata Weights and Strata Ranges $W_h(x_h - x_{h-1})$: This rule was suggested by Ekman (1959) and derived from the exact conditions by observing that if the distribution is not highly skew

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{4\, W_h(x_h - x_{h-1})}{3 f(x_h)} \quad \ldots \quad (17)$$

This method cannot be applied to unbounded distributions like Normal and Gamma families considered in this thesis. To judge its performance (and that of the rule of equal ranges) we shall use bounded truncated distributions. Cochran (1960) in his studies with some actual discrete populations found its performance to be almost as good as that of equalization of cumulatives of $f^{\frac{1}{2}}(x)$.

v) Equalization of Cumulatives of $\frac{1}{2}[F(x) + f(x)]$: This rule was suggested by Durbin (1959) as a first order correction to the rule of equal ranges by considering simple departures from rectangularity of the distribution. This rule is not being examined in this thesis. Cochran (1961) found its performance good for some and bad for other populations.

CHAPTER II

ITERATIVE METHODS FOR OPTIMUM STRATIFICATION

2.1. Proportional Allocation: Dalenius (1950) proved that the optimum points of stratification for proportional allocation (OPSPA) should satisfy the following relations.

$$2x_h = \mu_h + \mu_{h+1}, \qquad h = 1, 2, \ldots, L-1 \qquad (6)$$

As μ_h and μ_{h+1} are themselves functions of stratification points OPSPA have to be found by some iterative method. Adopting the usual methods of iterative solution, equations (6) may be solved by the following steps

1. Start with some arbitrary set of points

$$x_1' < x_2' < x_3' \ldots < x_{L-1}'$$

x_h' dividing the h^{th} and $(h+1)^{th}$ strata.

2. Calculate W_h and μ_h for the strata defined by x_h'
3. Calculate

$$\Delta_{ph} = x_h' - \frac{(\mu_h + \mu_{h+1})}{2} \qquad \ldots \qquad (18)$$

4. Calculate

$$\frac{\partial \Delta_{ph}}{\partial x_i'} = l_{phi}, \quad i = h-1, h, h+1 \qquad \ldots \qquad (19)$$

5. Solve the following simultaneous linear equations in d_i' s.

$$\Delta_{ph} + \sum_{i=h-1}^{h+1} l_{phi} \, d_i = 0, \quad h=1,2,\ldots, L-1 \qquad (20)$$

6. Replace the initial set by $x_h' + d_h$.

7. Repeat the steps 2 to 6 till two consecutive sets of points are either identical or differ by negligible quantities.

For calculating l_{phi} it may be noted that

$$\frac{\partial \mu_h}{\partial x_h'} = \frac{(x_h' - \mu_h)}{W_h} f(x_h') \quad \ldots \qquad (21)$$

and

$$\frac{\partial \mu_{h+1}}{\partial x_h'} = \frac{(\mu_{h+1} - x_h')}{W_{h+1}} f(x_{h+1}') \quad \ldots \qquad (22)$$

This method becomes extremely cumbrous even for small value of L. The following alternative is much simpler though the number of iterative steps to approach the optimum may increase.

3'. Replace the initial set by

$$x_h'' = \frac{\mu_h + \mu_{h+1}}{2} \quad \ldots \qquad (23)$$

4'. Repeat steps 2 and 3' till two consecutive sets are either identical or differ by negligible quantities. This alternative method can be used directly for actual discrete populations without approximating them by continuous distributions. This is a particularly suitable method if electronic computer facilities are available.

2.2. Equal Allocation: The optimum points of stratification for equal allocation (OPSEA) should satisfy the following conditions.

$$W_h[\sigma_h^2 + (x_h - \mu_h)^2] = W_{h+1}[\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2], \quad \ldots \quad (8)$$

$$h = 1,2,\ldots, L-1.$$

As W_h, μ_h, σ_h^2, W_{h+1}, μ_{h+1} and σ_{h+1}^2 are all functions of the points of stratification, OPSEA have to be found by some iterative method, As before the equations (8) can be solved by the following steps.

1. Start with some arbitrary set of points x_h'.
2. Calculate W_h, μ_h, σ_h^2 for the strata defined by these points.
3. Calculate

$$\Delta_{Eh} = W_h[\sigma_h^2 + (x_h' - \mu_h)^2] - W_{h+1}[\sigma_{h+1}^2 + (x_h' - \mu_{h+1})^2] \quad \ldots \quad (24)$$

4. Calculate

$$\frac{\partial \Delta_{Eh}}{\partial x_i'} = l_{Ehi} \quad , i = h-1, h, h+1 \quad \ldots \quad (25)$$

5. Solve the simultaneous linear equations in d_i's.

$$\Delta_{Eh} + \sum_{i=h-1}^{h+1} l_{Ehi}\, d_i = 0 \quad \ldots \quad (26)$$

$$h = 1,2,\ldots, L-1.$$

6. Replace the initial set by the set of points $x_h' + d_h$.
7. Repeat the steps 2 to 6 till two consecutive sets of points are either identical or differ by negligible quantities.

For these calculations it may be noted that

$$l_{Eh(h-1)} = -(x'_h - x'_{h-1})^2 \, f(x'_{h-1})$$

$$l_{Ehh} = 2[W_{h+1}(\mu_{h+1} - x'_h) + W_h(x'_h - \mu_h)] \qquad \ldots \quad (27)$$

$$l_{Eh(h+1)} = -(x'_{h+1} - x'_h)^2 \, f(x'_{h+1})$$

This cumbersome process can be considerably simplified by replacing the steps 3 to 7 by the following.

3'. Replace the original set of points by the set of solutions of the equations

$$\Delta_{Eh} = 0 \qquad \ldots \qquad \ldots \quad (28)$$

The solutions may be represented as follows.

$$x''_h = \frac{(T_h - T_{h+1}) + \left\{(T_h - T_{h+1})^2 - (S_h - S_{h+1})(W_h - W_{h+1})\right\}^{\frac{1}{2}}}{(W_h - W_{h+1})} \qquad (29)$$

where T_h and S_h represent respectively the sum and sum of squares of the values of the estimation variable in the h^{th} stratum. The positive sign before the curled brackets is used under the assumption that if the original set of points is not extremely bad new x''_h should lie between original μ_h and μ_{h+1}.

4'. Repeat steps 2 and 2' till two consecutive sets of points are either identical or differ by negligible quantities.

This simplified method can be used directly even for discrete distributions. For using the original method an approximation by some continuous distribution is necessary.

2.3. Minimum Variance Allocation: The optimum points of stratification for minimum variance allocation (OPSMVA) must satisfy the following relations.

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}} \quad \ldots \quad (7)$$

The iterative method for minimum variance allocation is similar to the one given for equal allocation. Δ_{Eh} is, however, to be replaced by Δ_{mh}, where

$$\Delta_{mh} = \frac{\sigma_h^2 + (x_h' - \mu_h)^2}{\sigma_h} - \frac{\sigma_{h+1}^2 + (x_h' - \mu_{h+1})^2}{\sigma_{h+1}} \quad \ldots \quad (30)$$

and l_{Ehi} by

$$l_{mhi} = \frac{\partial \Delta_{mh}}{\partial x_i'} \quad \ldots \quad \ldots \quad (31)$$

and so on.

Considerable simplification will be achieved if the second approximations are found by solving the equations

$$\Delta_{mh} = 0 \quad \ldots \quad \ldots \quad (32)$$

These will give

$$x_h'' = \frac{\left(\frac{\mu_h}{\sigma_h} - \frac{\mu_{h+1}}{\sigma_{h+1}}\right) + \left\{\left(\frac{\mu_h}{\sigma_h} - \frac{\mu_{h+1}}{\sigma_{h+1}}\right)^2 - \left(\frac{\sigma_h^2 + \mu_h^2}{\sigma_h} - \frac{\sigma_{h+1}^2 + \mu_{h+1}^2}{\sigma_{h+1}}\right)\left(\frac{1}{\sigma_h} - \frac{1}{\sigma_{h+1}}\right)\right\}^{\frac{1}{2}}}{\frac{1}{\sigma_h} - \frac{1}{\sigma_{h+1}}} \quad (33)$$

Again positive sign before the curled brackets is used under the assumption that x_h'' should lie between μ_h and μ_{h+1}.

This alternative method can be applied directly to actual discrete populations but is still too cumbrous to be used unless services of electronic computer are available.

Another alternative simplification was suggested by Dalenius and Hodges (1959). The rule given by them is as follows.

The new set x_h'' is found by computing

$$x_1'' = x_1' - \frac{\sqrt{3}}{2L}\left[(L-1)\Delta_{m1} + (1-2)\Delta_{m2} + \dots + 2\Delta_{m(L-2)} + \Delta_{m(L-1)}\right]$$

$$x_2'' = x_2' - \frac{\sqrt{3}}{2L}\left[(L-2)\Delta_{m1} + 2(L-2)\Delta_{m2} + \dots + 4\Delta_{m(L-2)} + 2\Delta_{m(L-1)}\right]$$

$$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad (34)$$

and

$$x_{L-1}'' = x_{L-1}' - \frac{\sqrt{3}}{2L}\left[\Delta_{m1} + 2\Delta_{m2} + 3\Delta_{m3} + \dots + (L-2)\Delta_{m(L-2)} + (L-1)\Delta_{m(L-1)}\right].$$

This procedure was derived for the rectangular distribution and it was assumed that the procedure given above will prove reasonably good for other distributions.

These simplified methods (excluding that suggested by Dalenius and Hodges) were used to obtain optimum points of stratification for Normal and Gamma distributions. The procedure is discussed in details in later chapters.

CHAPTER III

CONDITIONS FOR OPTIMUM STRATIFICATION OF DISCRETE POPULATIONS

3.1. Proportional Allocation: The conditions of optimum stratification were derived by Dalenius under the assumption of continuity of distribution. In Chapter II we have seen how iterative methods can be used to find the set of points satisfying these conditions even for discrete distributions. It has not so far been shown that the points satisfying these conditions will stratify the population in the optimum manner. In this chapter we shall see what modifications are introduced due to the discreteness of the distribution.

We shall again consider the units to be arranged in increasing order of values and strata marked off by points dividing them. As the data are discrete there will usually be two points at the boundary. We shall denote the maximum value of the h^{th} stratum by x_h and the minimum by x'_h.

Let us suppose that in h^{th} stratum the number of units is N_h, the mean μ_h and variance S_h^2. If the stratification is optimum an increase in variance would be effected by keeping all other strata intact but including x_h in the $(h+1)^{th}$ stratum or x'_{h+1} in the h^{th} stratum. This is the general method of deriving conditions of optimality.

For proportional allocation this method leads to the following conditions concerning x_h and x'_{h+1}.

$$(N_h+1)\,\frac{(N_h-1)S_h^2+N_h\mu_h^2+x'^2_{h+1}-(N_h\mu_h+x'_{h+1})^2/(N_h+1)}{N_h}$$

$$+\;(N_{h+1}-1)\,\frac{(N_{h+1}-1)S_{h+1}^2+N_{h+1}\mu_{h+1}^2-x'^2_{h+1}-(N_{h+1}\mu_{h+1}-x'_{h+1})^2/(N_{h+1}-1)}{N_{h+1}-2}$$

$$>\;N_h\,S_h^2+N_{h+1}\,S_{h+1}^2$$

or $$\frac{S_h^2(N_h^2-1)}{N_h}+\mu_h^2-2\mu_h x'_{h+1}+S_{h+1}^2\,\frac{(N_{h+1}-1)^2}{N_{h+1}-2}-\mu_{h+1}^2\,\left(1+\frac{2}{N_{h+1}-2}\right)$$

$$+\,2\mu_{h+1}\,x'_{h+1}\left[1+\frac{2}{N_{h+1}-2}\right]-x'^2_{h+1}\left[1+\frac{2}{N_{h+1}-2}\right]>N_hS_h^2+N_{h+1}\,S_{h+1}^2$$

or $$(\mu_h-\mu_{h+1})(\mu_h+\mu_{h+1}-2x'_{h+1})+\left[\frac{S_{h+1}^2}{N_{h+1}-2}-\frac{S_h^2}{N_h}-\frac{2}{N_{h+1}-2}(\mu_{h+1}-x'_{h+1})^2\right]>0 \qquad (35)$$

similarly $$(\mu_{h+1}-\mu_h)(\mu_h+\mu_{h+1}-2x_h)+\left[\frac{S_h^2}{N_h-2}-\frac{S_{h+1}^2}{N_{h+1}}-\frac{2}{N_h-2}(x_h-\mu_h)^2\right]>0 \qquad (36)$$

If N_h and N_{h+1} are large so that we can ignore the quantities involving their reciprocals we get the simple conditions

$$x_h<\frac{\mu_h+\mu_{h+1}}{2} \qquad \ldots \qquad \ldots \qquad (37)$$

and

$$x'_{h+1}>\frac{\mu_h+\mu_{h+1}}{2} \qquad \ldots \qquad \ldots \qquad (38)$$

Clearly the points satisfying Dalenius conditions will also satisfy conditions (37) and (38).

It may be remembered that, for extremely skew distributions, the number of units in the tail end strata is likely to become small though the strata variances remain large. (See Annexure). In this case the set of points satisfying Dalenius conditions may not lead to exact stratification. After reaching these points it may be worth-while to examine if the boundary of the last stratum is optimum.

3.2. Equal Allocation: The conditions of optimum stratification for discrete distributions for equal allocation are derived in the same way as in the case of proportional allocation case. They are as follows.

$$[\frac{L}{n}(N_{h+1}+1)+\frac{1}{N_{h+1}}]S^2_{h+1}+[\frac{L}{n}(N_{h+1}+1)-1](x_h-\mu_{h+1})^2$$

$$> [\frac{L}{n}(N_h-1)+\frac{1-L/n}{N_h-2}]S^2_h+[\frac{L}{n}(N_h-1)-1]\frac{N_h}{N_h-2}(x_h-\mu_h)^2, \qquad (39)$$

$$h = 1, 2, \ldots, L-1.$$

and

$$[\frac{L}{n}(N_h-1)+\frac{1}{N_h}]S^2_h+[\frac{L}{n}(N_h+1)-1](x'_{h+1}-\mu_h)^2$$

$$> [\frac{L}{n}(N_{h+1}-1)+\frac{1-L/n}{N_{h+1}-2}]S^2_{h+1}+[\frac{L}{n}(N_{h+1}-1)-1]\frac{N_{h+1}}{N_{h+1}-2}(x'_{h+1}-\mu_{h+1})^2, \qquad (40)$$

$$h = 1, 2, \ldots, L-1.$$

For large values of N_h and N_{h+1} they can be approximated by

$$N_{h+1}[S^2_{h+1}+(x_h-\mu_{h+1})^2] > N_h[S^2_h+(x_h-\mu_h)^2], \qquad \ldots \qquad (41)$$

$$h = 1, 2, \ldots, L-1,$$

and

$$N_{h+1}[S_{h+1}^2 + (x'_{h+1} - \mu_{h+1})^2] < N_h[S_h^2 + (x'_{h+1} - \mu_h)^2] \qquad (42)$$

$$h = 1, 2, \ldots, (L-1).$$

As in the case of proportional allocation the conditions given by Dalenius may be enough except for the last stratum boundary for every skew distributions.

3.3. Minimum Variance Allocation: The conditions for optimum stratification under the assumption of large number of units in each stratum reduce to

$$\frac{S_{h+1}^2 + (x_h - \mu_{h+1})^2}{S_{h+1}} > \frac{S_h^2 + (x_h - \mu_h)^2}{S_h} \quad \ldots \qquad (43)$$

$$h = 1, 2, \ldots, (L-1),$$

and

$$\frac{S_{h+1}^2 + (x'_{h+1} - \mu_{h+1})^2}{S_{h+1}} < \frac{S_h^2 + (x'_{h+1} - \mu_h)^2}{S_h}, \quad \ldots \qquad (44)$$

$$h = 1, 2, \ldots, (L-1).$$

We may rely on these approximations except perhaps for the last stratum boundary of a very skew population.

CHAPTER IV

OPTIMUM STRATIFICATION FOR PROPORTIONAL ALLOCATION FOR NORMAL AND GAMMA DISTRIBUTIONS

4.1. It was mentioned that even the simplified iterative method for optimum stratification is quite time consuming. As there are no simpler rules available for proportional allocation case, it was decided to prepare tables of OPSPA for a variety of distributions. The Normal and Chi-square distributions were selected for this purpose.

It is easy to see that if $x_1', x_2', \ldots, x_{L-1}'$ are the OPSPA for the distribution of X then $y_1' = ax_1' + b$, $y_2' = ax_2' + b$, ..., $y_{L-1}' = ax_{L-1}' + b$ are the OPSPA for the distribution of $Y = aX + b$. Thus the OPSPA found for the selected distributions in terms of percentage points of the distribution could also be used for all distributions resembling them in shape (distributions of linear functions of the variables).

Tables have been formed to give the following information.

i) The OPSPA in terms of the value of the distribution function at these points.

ii) Within strata variances for OPSPA. This is to verify the observation that proportional allocation becomes optimum for such a stratification. If this conjecture be true then

$$N_h \propto N_h \sigma_h$$

or

$$\sigma_h = \text{Constant.}$$

Thus for the conjecture to be true the variances in different strata should be equal.

iii) Values of $W_h \sigma_h^2$ in different strata. This is to verify a conflicting conjecture that the contribution from each stratum to the overall variance of an estimator of the population mean is the same.

iv) Variance of the estimate of mean based on a stratified proportional allocation design as a ratio of the variance for a simple random sample of the same size. This may prove useful in determining the sample size needed for a given precision.

The detailed tables for different distributions are given in Annexure. Summary tables derived from them are included in this chapter for easy reference. The formulae used for calculating the means and variances for different theoretical distributions are given below. These are given in a form which can utilize the standrad tables of these distributions.

4.2. Strata Means and Variances for Standardized Normal Distribution:

$$\mu_h = \frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} x\, e^{-\frac{1}{2}x^2}\, dx \Big/ \frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} e^{-\frac{1}{2}x^2}\, dx$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_{h-1}^2} - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_h^2} \Big/ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_h} e^{-\frac{1}{2}x^2} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_{h-1}} e^{-\frac{1}{2}x^2} dx$$

$$= \left[Z(x_{h-1}) - Z(x_h)\right] / \left[F(x_h) - F(x_{h-1})\right] \quad \dots \quad \dots \quad (45)$$

$$\sigma_h^2 = \frac{\frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} x^2\, e^{-\frac{1}{2}x^2}\, dx}{\frac{1}{\sqrt{2\pi}} \int_{x_{h-1}}^{x_h} e^{-\frac{1}{2}x^2}\, dx} - \mu_h^2$$

$$= 1 - \frac{x_h\, Z(x_h) - x_{h-1}\, Z(x_{h-1})}{F(x_h) - F(x_{h-1})} - \mu_h^2 \quad \dots \quad \dots \quad (46)$$

where $z(x)$ and $F(x)$ are respectively the ordinate and the distribution function at x for a $N(0, 1)$ distribution.

4.3. Strata Means and Variances for Gamma Distributions: A Gamma variable with parameters α and p is denoted by $G(\alpha, p)$ whose density function is given by

$$f(x) = \frac{\alpha^p}{\lceil(p)} e^{-\alpha p} x^{p-1} \qquad \text{when } x > 0$$

$$= 0 \qquad \text{when } x \leq 0$$

For this distribution

$$\mu_h = \frac{\frac{\alpha^p}{\Gamma(p)} \int_{x_{h-1}}^{x_h} e^{-\alpha x} x^p \, dx}{\frac{\alpha^p}{\Gamma(p)} \int_{x_{h-1}}^{x_h} e^{-\alpha x} x^{p-1} \, dx} = \frac{p}{\alpha} \; \frac{G(\alpha, p+1)_{x_h} - G(\alpha, p+1)_{x_{h-1}}}{G(\alpha, p)_{x_h} - G(\alpha, p)_{x_{h-1}}} \qquad (47)$$

and

$$\sigma_h^2 = \frac{p(p+1)}{\alpha^2} \; \frac{G(\alpha, p+2)_{x_h} - G(\alpha, p+2)_{x_{h-1}}}{G(\alpha, p)_{x_h} - G(\alpha, p)_{x_{h-1}}} - \mu_h^2 \qquad \ldots \qquad (48)$$

where $G(\alpha, p)_x$ is the value of the distribution function of $G(\alpha, p)$ variable at x.

For obtaining the OPSPA the simplified iterative method mentioned in Chapter II was used. Of the Gamma distributions only those of the type $G(\frac{1}{2}, n/2)$ were used. They are the same as Chi-square distributions. As any Gamma variable can be expressed as a linear function of some Chi-square distribution the OPSPA tables for Chi-square distributions in terms of the values of the distribution function at the stratification points will cover the entire family of Gamma distributions also.

4.4. The Tables: The tables of OPSPA given on the next few pages are based on Biometrika Tables Vol. I. The tables do not give the exact optima but a very close approximation to the same. This was inspite of the fact that interpolated values of the functions tabulated in Biometrika Tables were used to get a closer approximation.

TABLE I

OPSPA FOR NORMAL AND CHI-SQUARE DISTRIBUTIONS

Number of Strata	h	Distribution Function at OPSPA for N(0,1)	χ^2_1	χ^2_2	χ^2_3	χ^2_4	χ^2_6
2	1	.500	.862	.798	.759	.733	.697
3	1	.271	.746	.632	.576	.537	.494
	2	.729	.952	.926	.906	.893	.871
4	1	.161	.657	.528	.448	.408	.350
	2	.500	.879	.817	.779	.751	.715
	3	.839	.976	.961	.950	.941	.931
5	1	.107	.561	.423	.363	.337	.269
	2	.352	.808	.713	.666	.620	.557
	3	.648	.926	.889	.861	.841	.815
	4	.893	.986	.976	.971	.964	.957
6	1	.078	.520	.362	.294	.264	.217
	2	.261	.746	.632	.559	.522	.482
	3	.500	.879	.817	.778	.751	.715
	4	.739	.954	.929	.910	.897	.879
	5	.922	.991	.985	.981	.977	.972

TABLE I - Continued

Number of Strata	h	Distribution Function at OPSPA for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	1	.674	.658	.631	.614	.603	.594
3	1	.464	.439	.405	.360	.353	.346
	2	.857	.849	.832	.778	.768	.768
4	1	.330	.313	.276	.247	.243	.233
	2	.690	.673	.644	.579	.570	.564
	3	.923	.916	.904	.865	.858	.855
5	1	.242	.228	.200	.185	.180	.169
	2	.547	.533	.497	.478	.463	.456
	3	.798	.787	.763	.748	.740	.731
	4	.951	.948	.939	.935	.931	.928
6	1	.191	.176	.150	.140	.130	.126
	2	.452	.430	.398	.380	.364	.366
	3	.690	.673	.644	.626	.613	.604
	4	.866	.856	.839	.827	.811	.814
	5	.969	.966	.960	.957	.954	.952

TABLE II

WITHIN STRATA VARIANCES FOR THE SCHEMES GIVEN IN TABLE I

Number of Strata	h	N(0,1)	Within Strata Variances for χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
2	1	.363	0.33	0.75	1.23	1.74	2.81
	2	.363	2.99	4.00	4.97	5.91	7.74
3	1	.118	0.13	0.32	0.56	0.83	1.46
	2	.252	0.49	0.75	0.97	1.21	1.64
	3	.118	0.21	4.00	4.76	5.49	6.97
4	1	.078	0.07	0.18	0.32	0.50	0.88
	2	.200	0.18	0.29	0.43	0.54	0.83
	3	.200	0.53	0.71	0.87	1.04	1.42
	4	.078	3.31	4.01	4.67	5.33	6.56
5	1	.047	0.03	0.10	0.21	0.36	0.62
	2	.059	0.10	0.16	0.24	0.27	0.47
	3	.173	0.18	0.30	0.36	0.46	0.64
	4	.059	0.57	0.71	0.91	1.04	1.33
	5	.047	3.37	3.99	4.62	5.26	6.32
6	1	.034	0.02	0.07	0.14	0.25	0.48
	2	.048	0.05	0.10	0.14	0.19	0.33
	3	.156	0.10	0.16	0.24	0.30	0.40
	4	.156	0.20	0.29	0.36	0.43	0.61
	5	.048	0.57	0.72	0.86	1.03	1.29
	6	.034	3.43	4.02	4.61	5.13	6.18

TABLE II - Continued

Number of Strata	h	Within Strata Variances for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	1	3.93	5.11	8.11	11.25	14.41	17.62
	2	9.53	11.33	15.59	19.84	23.94	27.98
3	1	2.13	2.78	4.62	6.54	8.69	10.67
	2	2.05	2.62	3.87	5.11	6.15	7.30
	3	8.32	9.58	12.81	16.07	18.90	22.43
4	1	1.37	1.88	3.19	4.62	6.11	7.82
	2	1.05	1.31	2.05	2.83	3.54	4.29
	3	1.78	2.12	3.00	3.89	4.67	5.12
	4	7.73	8.87	11.74	14.46	16.72	19.64
5	1	0.97	1.38	2.45	3.63	4.87	2.53
	2	0.64	0.84	1.36	1.80	2.31	5.89
	3	0.83	1.02	1.42	1.94	2.46	2.92
	4	1.47	1.99	2.76	3.46	4.07	4.92
	5	8.09	8.35	10.87	13.25	15.65	17.35
6	1	0.77	1.10	1.95	3.00	3.98	5.27
	2	0.48	0.59	1.01	1.37	1.84	2.46
	3	0.52	0.66	0.95	1.35	1.69	2.01
	4	0.78	0.92	1.35	1.77	1.87	2.57
	5	1.49	1.84	2.51	3.12	3.71	4.12
	6	7.15	7.94	10.35	12.33	14.61	16.49

TABLE III

CONTRIBUTION TO OVERALL VARIANCE FROM DIFFERENT STRATA FOR SCHEMES IN TABLE I

Number of Strata	h	Contribution to Variance ($W_h \sigma_h^2$) from strata for N(0,1)	χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
2	1	.182	.283	.602	.933	1.271	1.957
	2	.182	.412	.808	1.196	1.581	2.344
3	1	.068	.097	.201	.326	.448	.723
	2	.054	.101	.222	.321	.431	.618
	3	.068	.155	.297	.446	.590	.892
4	1	.032	.043	.096	.144	.203	.307
	2	.026	.039	.083	.142	.187	.301
	3	.026	.052	.103	.149	.197	.308
	4	.032	.079	.156	.235	.313	.456
5	1	.019	.017	.043	.075	.122	.168
	2	.014	.024	.046	.071	.075	.145
	3	.014	.021	.052	.070	.103	.151
	4	.014	.034	.062	.100	.128	.189
	5	.019	.048	.094	.135	.188	.272
6	1	.012	.011	.024	.041	.065	.105
	2	.009	.012	.027	.038	.049	.087
	3	.008	.013	.030	.053	.069	.092
	4	.008	.015	.032	.047	.062	.100
	5	.009	.021	.040	.061	.083	.121
	6	.012	.031	.060	.090	.115	.171

TABLE III - Continued

Number of Strata	h	χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
		Contribution to Variance ($W_h \sigma^2_h$) from strata for					
2	1	2.650	3.360	5.116	6.908	8.686	10.471
	2	3.103	3.879	5.755	7.653	9.512	11.357
3	1	0.986	1.223	1.874	2.526	3.265	3.911
	2	0.805	1.073	1.650	2.215	2.675	3.195
	3	1.188	1.451	2.157	2.894	3.573	4.389
4	1	0.452	0.587	0.881	1.196	1.515	1.892
	2	0.377	0.471	0.753	1.037	1.293	1.550
	3	0.415	0.517	0.781	1.056	1.302	1.417
	4	0.594	0.744	1.127	1.488	1.804	2.346
5	1	0.235	0.316	0.489	0.673	0.875	0.427
	2	0.195	0.254	0.404	0.527	0.655	1.690
	3	0.209	0.260	0.378	0.524	0.680	0.801
	4	0.225	0.320	0.487	0.644	0.777	0.974
	5	0.393	0.433	0.658	0.868	1.077	1.242
6	1	0.147	0.194	0.293	0.422	0.516	0.662
	2	0.125	0.150	0.250	0.328	0.432	0.593
	3	0.125	0.161	0.234	0.332	0.419	0.476
	4	0.136	0.169	0.263	0.356	0.370	0.539
	5	0.153	0.201	0.304	0.404	0.530	0.512
	6	0.223	0.273	0.418	0.534	0.676	0.788

TABLE IV

VARIANCE OF THE ESTIMATOR OF POPULATION MEAN BASED ON OPTIMALLY STRATIFIED PROPORTIONAL ALLOCATION DESIGN AS A RATIO OF THE VARIANCE FOR A SIMPLE RANDOM SAMPLE OF THE SAME SIZE

Number of Strata	Variance Ratio V_{prop} / V_{ran} N(0,1)	χ^2_1	χ^2_2	χ^2_3	χ^2_4	χ^2_6
2	.363	.348	.353	.355	.357	.359
3	.190	.177	.179	.182	.184	.186
4	.117	.107	.110	.112	.113	.114
5	.080	.072	.075	.076	.077	.077
6	.058	.052	.054	.055	.056	.056

TABLE IV - Continued

Number of Strata	Variance Ratio V_{prop} / V_{ran} χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	.359	.362	.363	.364	.364	.364
3	.186	.187	.189	.191	.191	.191
4	.115	.116	.118	.119	.119	.120
5	.078	.079	.080	.081	.082	.085
6	.057	.058	.058	.059	.059	.059

4.5. <u>Conclusions from the Tables</u>: On the basis of the tables given in section 4.4 and Annexure we come to the following conclusions.

i) If the OPSPA for Normal distribution are its $P_1, P_2, \ldots, P_{L-1}$ percent points, then, for any value of n, OPSPA for Chi-square distribution with n degrees of freedom are closely approximated by the $P_1, P_2, \ldots, P_{L-1}$ percent points of Chi-square distribution with n+2 degrees of freedom. (See Annexure). It may be mentioned that this fact was first observed for some of the Chi-square distributions. For the rest of them the initial set of stratification points was selected using this conjecture. In hardly any of them was it necessary to proceed to the next set.

ii) The within strata variances are widely different for stratification defined by OPSPA. This means that proportional allocation is not optimum even for this scheme. This suggests large gains of optimum allocation over the proportional. The two allocations are compared in the accompanying graph. (See Table II, Annexure ~~and graph~~).

iii) The contributions to the overall variance of the estimator of the population mean from different strata are not equal. (See Table III).

iv) The amount of reduction due to stratification is almost the same for all the distributions. (See Table IV).

v) For five strata or more the last stratum contains a very small proportion of the total population. Such small strata may become meaningless for proportional allocation unless the total sample size is large. It may be mentioned that actual discrete distributions are unlikely to resemble continuous distributions in such small right hand tail area. This is specially true for very skew distributions.

It may be mentioned that it is not necessary to fit a Chi-square distribution to the population before utilising the conclusion (i). The simplest procedure would be to divide the cumulatives of the values (after adjusting the origin) in the observed constant proportions $P_1, P_2, \ldots, P_L$.

If an approximating Chi-square distribution has been found, it may be better to divide the cumulative squares of the values in the proportions into which the Chi-square distribution with four more degrees of freedom is divided by the OPSPA. This suggestion does not follow from the tables but from the conjecture that the tail area of a skew discrete distribution will be better taken care of by including consideration of higher powers.

CHAPTER V

OPTIMUM STRATIFICATION FOR NORMAL AND GAMMA DISTRIBUTIONS FOR EQUAL ALLOCATION

5.1. Although simplifications are available for OPSEA (optimum points of stratification for equal allocation) it was decided to prepare tables of OPSEA for Normal and Chi-square distributions. This may help in discriminating between various approximating rules. As before the knowledge of OPSEA for Chi-square distributions amounts to knowledge of OPSEA for all Gamma distributions. OPSEA found for these distributions in terms of their percentage points could be utilized directly for stratifying any population resembling them in shape.

Tables have been formed to provide the following information.

i) The OPSEA in terms of the value of the distribution function at these points.

ii) Values of $W_h \sigma_h$ for different strata. This is to verify if equal allocation is also the minimum variance allocation when the stratification is optimum. It also checks the hypothesis that the strata contribute equally to the overall variance. The tables will enable us to verify this conjecture.

iii) The points of stratification for some distributions derived by equalising the cumulatives of $f^{\frac{2}{3}}(x)$.

iv) Strata totals $W_h \mu_h$ for different strata defined by OPSEA.

v) Points of stratification derived by equalizing strata totals for some distributions.

vi) Comparison of the variances of the estimator of population means of some truncated distributions for different approximations to OPSEA.

TABLE V

OPSEA FOR NORMAL AND CHI-SQUARE DISTRIBUTIONS

Number of Strata	h	Distribution Function at OPSEA for					
		N(0,1)	χ^2_1	χ^2_2	χ^2_3	χ^2_4	χ^2_6
2	1	.500	.808	.740	.704	.680	.341
3	1	.285	.657	.551	.487	.475	.430
	2	.715	.911	.878	.855	.841	.821
4	1	.184	.561	.423	.385	.337	.310
	2	.500	.808	.740	.704	.680	.641
	3	.816	.949	.926	.914	.905	.895
5	1	.128	.473	.362	.294	.264	.230
	2	.367	.706	.632	.576	.552	.518
	3	.633	.879	.826	.796	.777	.754
	4	.872	.968	.953	.945	.939	.928
6	1	.095	.416	.295	.247	.228	.191
	2	.277	.657	.528	.487	.459	.430
	3	.500	.808	.740	.704	.680	.641
	4	.723	.911	.878	.861	.847	.821
	5	.905	.976	.967	.958	.952	.946

TABLE V - Continued

Number of Strata	h	Distribution function at OPSEA for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ_{30}
2	1	.623	.610	.591	.579	.570	.564
3	1	.397	.400	.375	.360	.353	.346
	2	.809	.798	.791	.778	.768	.768
4	1	.286	.275	.247	.247	.243	.233
	2	.623	.610	.591	.579	.570	.564
	3	.884	.882	.869	.866	.858	.855
5	1	.211	.202	.187	.175	.171	.165
	2	.495	.487	.467	.452	.440	.430
	3	.742	.734	.716	.706	.701	.694
	4	.926	.921	.913	.910	.904	.905
6	1	.181	.168	.150	.141	.137	.119
	2	.409	.391	.367	.460	.347	.325
	3	.623	.610	.591	.579	.570	.559
	4	.814	.803	.791	.791	.778	.771
	5	.943	.940	.933	.930	.926	.928

TABLE VI

VALUES OF $W_h\sigma_h$ FOR DIFFERENT STRATA DEFINED BY CPSEA

Number of Strata		Values of $W_h\sigma_h$ for Different Strata for χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
2	1	.371	.553	.694	.813	.980
	2	.327	.520	.665	.786	1.016
3	1	.168	.250	.303	.384	.466
	2	.140	.235	.319	.348	.450
	3	.156	.245	.318	.377	.481
4	1	.097	.133	.186	.203	.268
	2	.077	.145	.174	.226	.257
	3	.082	.130	.171	.202	.277
	4	.091	.148	.187	.223	.273
5	1	.055	.094	.110	.131	.166
	2	.047	.085	.113	.141	.184
	3	.064	.084	.113	.130	.169
	4	.054	.092	.123	.149	.181
	5	.058	.095	.119	.141	.184
6	1	.037	.060	.079	.101	.124
	2	.041	.053	.077	.087	.123
	3	.034	.074	.090	.105	.109
	4	.036	.060	.082	.096	.120
	5	.040	.064	.073	.086	.123
	6	.044	.067	.091	.109	.136

TABLE VI - Continued

Number of Strata	h	Values of $W_h \sigma_h$ for Different Strata for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	1	1.146	1.293	1.601	1.867	2.093	2.298
	2	1.184	1.333	1.643	1.911	2.139	2.343
3	1	.521	.631	.773	.892	1.010	1.107
	2	.557	.573	.754	.871	.947	1.073
	3	.563	.644	.767	.912	1.042	1.119
4	1	.308	.352	.419	.520	.596	.639
	2	.308	.348	.452	.505	.554	.626
	3	.316	.368	.447	.529	.582	.636
	4	.329	.361	.459	.521	.602	.657
5	1	.194	.223	.285	.325	.373	.411
	2	.214	.244	.307	.360	.393	.418
	3	.200	.221	.270	.318	.385	.418
	4	.218	.241	.298	.349	.373	.450
	5	.208	.234	.298	.336	.390	.414
6	1	.154	.172	.210	.244	.278	.270
	2	.138	.152	.196	.242	.260	.284
	3	.131	.155	.196	.228	.269	.310
	4	.149	.162	.204	.243	.286	.308
	5	.140	.156	.196	.243	.246	.308
	6	.157	.179	.223	.255	.289	.298

TABLE VII

POINTS OF STRATIFICATION
DERIVED BY EQUALIZING CUMULATIVES OF $f^{\frac{1}{2}}(x)$

Number of Strata	h	Distribution Function at Points Dividing Cumulatives of $f^{\frac{1}{2}}(x)$ into Equal Proportions for N(0,1)	χ^2_2	χ^2_4	χ^2_6	χ^2_8	χ^2_{10}	χ^2_{20}	χ^2_{30}
2	1	.500	.750	.684	.650	.632	.618	.583	.568
3	1	.271	.551	.475	.430	.397	.371	.347	.325
	2	.729	.889	.853	.837	.830	.808	.799	.792
4	1	.171	.451	.337	.296	.264	.256	.235	.228
	2	.500	.750	.684	.650	.632	.618	.583	.568
	3	.829	.939	.915	.905	.895	.888	.875	.874
5	1	.117	.330	.264	.217	.221	.185	.170	.146
	2	.359	.632	.552	.531	.485	.487	.439	.430
	3	.641	.835	.785	.762	.748	.741	.711	.702
	4	.883	.959	.944	.912	.928	.930	.912	.911
6	1	.085	.295	.209	.179	.161	.152	.132	.112
	2	.271	.551	.475	.430	.397	.371	.347	.325
	3	.500	.750	.684	.650	.632	.618	.583	.568
	4	.729	.889	.853	.837	.830	.808	.799	.792
	5	.915	.973	.963	.938	.951	.951	.945	.945

TABLE VIII

PROPORTIONS OF AGGREGATE $\Sigma W_h \mu_h$ IN DIFFERENT STRATA DEFINED BY OPSEA

Number of Strata	h	$W_h \mu_h / \Sigma W_h \mu_h$ for χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
2	1	.363	.391	.407	.418	.420
	2	.637	.609	.593	.582	.580
3	1	.175	.191	.194	.217	.221
	2	.417	.429	.437	.424	.428
	3	.408	.380	.369	.359	.351
4	1	.104	.106	.124	.121	.134
	2	.260	.285	.283	.297	.286
	3	.352	.342	.341	.336	.349
	4	.284	.267	.252	.246	.232
5	1	.060	.075	.076	.080	.086
	2	.163	.189	.194	.203	.211
	3	.283	.258	.264	.259	.260
	4	.290	.286	.286	.284	.273
	5	.204	.192	.180	.174	.170
6	1	.040	.049	.055	.063	.066
	2	.135	.125	.139	.141	.156
	3	.189	.217	.213	.213	.198
	4	.229	.230	.235	.233	.229
	5	.243	.233	.216	.207	.216
	6	.165	.147	.146	.143	.134

TABLE VIII - Continued

Number of Strata	h	$W_h \mu_h / \Sigma W_h \mu_h$ for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	1	.430	.436	.448	.455	.459	.463
	2	.570	.564	.552	.545	.541	.537
3	1	.219	.239	.244	.248	.253	.255
	2	.439	.430	.433	.430	.424	.430
	3	.342	.341	.323	.322	.323	.315
4	1	.137	.142	.143	.146	.161	.159
	2	.292	.294	.304	.299	.298	.303
	3	.341	.344	.335	.334	.330	.329
	4	.230	.220	.218	.211	.211	.209
5	1	.090	.094	.101	.102	.106	.107
	2	.213	.220	.225	.228	.225	.224
	3	.265	.264	.260	.262	.268	.269
	4	.272	.264	.261	.260	.251	.257
	5	.160	.158	.153	.148	.150	.143
6	1	.072	.074	.077	.079	.082	.073
	2	.156	.157	.161	.169	.166	.163
	3	.201	.206	.210	.206	.211	.221
	4	.236	.230	.230	.224	.230	.232
	5	.206	.208	.200	.204	.200	.200
	6	.129	.125	.122	.118	.111	.111

TABLE IX

POINTS DIVIDING THE POPULATION INTO EQUAL SIZED STRATA

Number of Strata	h	Values of Distribution Function at Points Dividing the Cumulatives of $xf(x)$ into Equal Proportions for χ_1^2	χ_6^2	χ_{30}^2
2	1	.879	.715	.585
3	1	.794	.554	.430
	2	.935	.832	.751
4	1	.727	.456	.325
	2	.879	.715	.585
	3	.957	.885	.823
5	1	.683	.404	.275
	2	.832	.620	.483
	3	.917	.790	.678
	4	.968	.912	.863
6	1	.629	.350	.228
	2	.794	.554	.430
	3	.879	.715	.585
	4	.935	.832	.757
	5	.976	.929	.895

TABLE X

WITHIN STRATA COEFFICIENTS OF VARIATION
FOR STRATA DEFINED BY OPSEA

Number of Strata	h	Within Strata Coefficients of Variation for χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
2	1	1.021	.708	.569	.487	.389
	2	.514	.427	.373	.337	.292
3	1	.963	.655	.520	.443	.351
	2	.335	.274	.243	.205	.175
	3	.383	.323	.288	.262	.228
4	1	.940	.630	.500	.420	.333
	2	.296	.255	.205	.190	.150
	3	.231	.190	.167	.149	.133
	4	.322	.278	.247	.227	.197
5	1	.925	.621	.486	.409	.321
	2	.286	.225	.195	.174	.145
	3	.224	.162	.143	.126	.105
	4	.187	.161	.143	.131	.111
	5	.283	.248	.221	.203	.180
6	1	.919	.611	.478	.403	.315
	2	.308	.214	.186	.155	.132
	3	.182	.171	.141	.123	.091
	4	.156	.130	.117	.103	.087
	5	.164	.138	.115	.104	.095
	6	.265	.227	.207	.192	.169

TABLE X - Continued

Number of Strata	h	Within Strata Coefficients of Variation for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	1	.333	.296	.238	.205	.182	.166
	2	.259	.237	.198	.175	.158	.145
3	1	.297	.264	.211	.180	.160	.145
	2	.159	.136	.116	.101	.089	.083
	3	.205	.189	.158	.142	.129	.118
4	1	.281	.247	.195	.167	.148	.134
	2	.132	.118	.099	.085	.074	.069
	3	.116	.107	.089	.079	.071	.064
	4	.179	.164	.140	.123	.114	.105
5	1	.270	.237	.188	.159	.141	.128
	2	.125	.111	.091	.079	.070	.062
	3	.094	.084	.069	.061	.057	.052
	4	.100	.091	.076	.067	.059	.058
	5	.162	.149	.128	.113	.104	.097
6	1	.265	.232	.182	.155	.136	.123
	2	.110	.097	.081	.072	.063	.058
	3	.082	.075	.062	.055	.051	.047
	4	.079	.071	.059	.054	.049	.044
	5	.085	.075	.065	.060	.051	.051
	6	.153	.143	.121	.108	.098	.089

TABLE XI

PRODUCT OF STRATUM RANGE AND STRATUM WEIGHT
FOR OPSEA

Number of Strata	h	χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_6^2
		Values of $W_h(x_h - x_{h-1})$ for				
3	1	.591	.882	1.120	1.520	2.064
	2	.508	.850	1.141	1.244	1.599
4	1	.337	.465	.693	.809	1.210
	2	.271	.507	.606	.789	.894
	3	.296	.462	.609	.717	.991
5	1	.189	.326	.412	.528	.759
	2	.163	.297	.395	.488	.636
	3	.225	.291	.396	.450	.566
	4	.196	.330	.444	.535	.644
6	1	.125	.207	.296	.410	.573
	2	.145	.186	.264	.300	.430
	3	.120	.256	.304	.355	.378
	4	.125	.205	.283	.334	.414
	5	.143	.231	.262	.304	.441

TABLE XI - Continued

Number of Strata	h	Values of $W_h(x_h - x_{h-1})$ for χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
3	1	2.541	3.320	4.762	6.192	7.695	9.134
	2	1.978	2.025	2.662	3.051	3.361	3.756
4	1	1.544	1.925	2.717	3.804	4.811	5.615
	2	1.078	1.210	1.578	1.726	1.902	2.158
	3	1.127	1.301	1.557	1.830	2.016	2.212
5	1	0.992	1.252	1.889	2.467	3.129	3.712
	2	0.738	0.855	1.064	1.246	1.345	1.452
	3	0.689	0.766	0.921	1.092	1.279	1.431
	4	0.589	0.841	1.049	1.224	1.299	1.540
6	1	0.796	0.974	1.425	1.889	2.398	2.523
	2	0.479	0.533	0.673	0.832	0.882	0.989
	3	0.449	0.528	0.672	0.741	0.870	1.058
	4	0.518	0.560	0.700	0.780	0.940	1.039
	5	0.486	0.575	0.682	0.836	0.853	1.036

TABLE XII

COMPARISON OF SOME SCHEMES OF STRATIFICATION
FOR SOME TRUNCATED DISTRIBUTIONS

Number of Strata	Truncated at	Equal Aggregate of $f^{\frac{1}{2}}(x)$	Equal Aggregate of x	Equal Range
		Distribution χ_1^2		
2	4.2	1.00	1.19	1.19
3	5.2	1.00	1.37	1.22
4	6.1	1.00	1.37	1.37
5	6.8	1.00	1.04	1.32
		Distribution χ_6^2		
2	8.9	1.00	1.08	1.13
3	10.5	1.00	1.19	1.07
4	11.6	1.00	1.18	1.12
5	12.4	1.00	1.21	1.16
		Distribution χ_{30}^2		
2	35.3	1.00	1.02	2.70
3	38.2	1.00	1.11	2.25
4	40.7	1.00	1.15	2.19
5	42.0	1.00	1.07	2.24

5.2. Conclusions from the Tables: Before describing the conclusions drawn from the tables given in this chapter it should be mentioned how the OPSEA were arrived at. The iterative method mentioned in Chapter II was used. At the time the work on this problem started the approximation by equalizing the cumulatives of $f^{\frac{1}{2}}(x)$ was not known. For proportional allocation it was found that the OPSPA for chi-square distribution divided the cumulative distribution of another Chi-square variable with two more degrees of freedom into more or less the same proportions as the OPSPA of a Normal distribution divide it. It was conjectured that some such relation may hold for the equal allocation case also. For two strata it was found that the OPSEA divide the Chi-square distribution with 1.3 more degrees of freedom*. It was assumed that the OPSEA for larger number of strata would divide the Chi-square distribution with 1.3 more degrees of freedom in the same proportions as the OPSEA of a Normal distribution divide it. So the initial set of points to start the iterative process was selected on the basis of this assumption. It was found that the initial set was very close to the second approximation. So no iteration was actually used. As before it must be realized that what is represented as OPSEA in the tables is only a very close approximation to it.

i) The OPSEA of Gamma distributions divide the cumulatives of $x^{.65}$ into almost the same proportions as the OPSEA of Normal distribution divide it.

*Conventionally Chi-square distributions are supposed to have on integral number of degrees of freedom. But considered as the parameter p in G $(\frac{1}{2},p)$ it could have any positive value. The points for fractional degrees of freedom were obtained by simple interpolation.

ii) For strata defined by OPSEA the values of $W_h \sigma_h$ are fairly constant. (Table VI) This implies that OPSEA are also close approximation to the optimum points of stratification for minimum-variance allocation (OPSMVA), for this stratification equal allocation is a close approximation to minimum variance allocation and the two allocations lead to very nearly equal variances of the estimates of the population means.

iii) The points of stratification derived by equalizing cumulatives of $f^{\frac{1}{2}}(x)$ are very close to the OPSEA. (In some cases they were found to be even closer than the approximation used in the tables. This was discovered while comparing the variances for various stratification schemes. The improvement, however, was minor as was expected). (Tables V and VII).

iv) The strata totals $W_h \mu_h$ are not approximately equal. This suggests that equalizing strata totals may not be a good approximation to OPSEA. (Table VIII).

v) The same conclusion is arrived at by comparing the points of stratification derived by equalizing strata totals with OPSEA. (Table IX).

vi) For strata defined by the OPSEA, the within strata coefficients of variation are far from equal. This shows that for such populations the assumption under which equal sized strata were proved to be optimum is not valid. (Table X).

vii) The product of the stratum range and stratum weight is almost a constant for Chi-square distributions with fewer degrees of freedom. For higher degrees of freedom it is far from constant. This cannot be interpreted in terms of skewness because it is absolutely constant for a rectangular distribution which is symmetric. (Table XI).

viii) A comparison of the methods of stratification shows that equalizing cumulatives of $f^{\frac{1}{2}}(x)$ is always superior to equal strata sizes or equal strata ranges. The amount of gain of one over the other depends on the population and the number of strata. No more definite conclusion should be drawn from Table XII because the strata derived from these methods equalized the respective quantities only roughly.

CHAPTER VI

ON EQUALIZATION OF STRATA TOTALS

6.1. Hansen, Hurwitz and Madow (1953) proved that under certain conditions OPSEA are achieved by equalizing strata totals. The conditions stated by them are as follows.

''A rule of thumb which will provide a rough guide to the optimum sizes of strata when a constant number of psu's are selected from each stratum is to make the strata equal in terms of X_h when the psu rel - variances are about the same and remain about the same on adjusting the sizes of strata."

This rule has been widely practiced and claimed to be good. For the type of stratification discussed in this part of the thesis, however, the conditions are unlikely to hold exactly for any population. The proof given for the optimality of the rule assumes the equality of strata coefficients of variation exactly and their remaining invariant under change of boundary. Not only were these conditions not satisfied for any one of the populations considered and the rule found unsatisfactory but we expect that they will never be satisfied and we can always improve upon the rule.

A shift in one of the points of stratification will generally shift the means in the same direction (the two means will either both increase or both decrease) but variances in opposite directions, one of the variances will increase and the other decrease. Thus if the rel-variances were initially equal this shift will make them unequal. It is possible that the conditions will hold when stratification is based on geographical contiguity or similar qualitative data. But for the type of stratification discussed here they may not hold for any population.

Though the above reasoning goes against the assumptions on which the proof of the goodness of the rule was based it does not imply that the rule is not good.

6.2. It can be demonstrated that the chance of coming across a distribution where this rule gives an actual optimum is very small. Let there be a variable X for which the rule gives the OPSEA. The

$$W_h \quad \sigma_h^2 + (x_h - \mu_h)^2 = W_{h+1} \quad \sigma_{h+1}^2 + (x_h - \mu_{h+1})^2 \qquad (8)$$

$$h = 1, 2, \ldots, L-1$$

for strata given by this rule.

Let us consider a variable $Y = X + b$. We shall describe the parameters of its distribution with a subscript y. In the optimum strata defined for X

$$\sigma_{hy}^2 = \sigma_h^2 \qquad \ldots \qquad \ldots \qquad (49)$$

$$(y_h - \mu_{hy})^2 = (x_h - \mu_h)^2 ; (y_h - \mu_{(h+1)y})^2 = (x_h - \mu_{h+1})^2 \qquad (50)$$

Thus

Thus

$$W_h \left\{ \sigma^2_{hy} + (y_h-\mu_{hy})^2 \right\} = W_{h+1} \left\{ \sigma^2_{(h+1)y} + (y_h-\mu_{(h+1)y})^2 \right\} \qquad (51)$$

and the original stratification is also optimum for Y. But the strata totals of the variable Y are

$$W_h\mu_{hy} = W_h \left[\mu_h + b\right] \qquad \ldots \qquad (52)$$

The X totals $W_h \mu_h$ were all equal. So the Y totals are equal only if W_h's are all equal. But the equality of both W_h and $W_h \mu_h$ implies the equality of μ_h's. This is not possible for the type of stratification and is bad in general.

Thus, if at all, the rule can give optimum stratification for only one of the family of variables X+b. That in actual practice we come across a large number of such exceptional cases seems unlikely.

6.3. Another way of demonstrating that the rule is bad is to show that it achieves very little gain over a corresponding unstratified sampling. The demonstration given by Sethi (1960) is repeated below with slight modifications in notation.

<u>Notation</u>: The number of units in the h^{th} stratum is N_h, the i^{th} population value in the h^{th} stratum is denoted by Y_{ih} and the i^{th} sample value by y_{ih}.

$$\sum_{h=1}^{L} N_h = N \qquad \ldots \qquad \ldots \qquad (53)$$

$$\sum_{i=1}^{N_h} Y_{ih} = Y_h \qquad \ldots \qquad \ldots \qquad (54)$$

$$\sum_{h=1}^{L} Y_h = Y \qquad \ldots \qquad \ldots \qquad (55)$$

<u>Lemma</u>: A stratified sampling procedure with L strata is always better than an unstratified sampling design where the probability of selection of a unit at a particular draw is 1/L times the probability of its selection in the stratified design. Here we shall assume that the sample is drawn with replacement.

$$\bar{y}_{st} = \frac{1}{nN} \sum_{h=1}^{L} \sum_{i=1}^{n} \frac{y_{ih}}{p_{ih}} \qquad \ldots \qquad (56)$$

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} \frac{1}{n} \left[\sum_{i=1}^{N_h} \frac{Y_{ih}^2}{p_{ih}} - Y_h^2 \right] \qquad \ldots \qquad (57)$$

$$\bar{y}_{ran} = \frac{1}{nNL} \sum_{h=1} \sum_{i} \frac{y_{ih}}{p_{ih}/L} \qquad \ldots \qquad \ldots \qquad (58)$$

$$V(\bar{y}_{ran}) = \frac{1}{nN^2 L} \left[\sum_{h=1}^{L} \sum_{i=1}^{N_h} \frac{Y_{ih}^2}{p_{ih}/L} - Y^2 \right] \qquad \ldots \qquad (59)$$

$$\therefore \; V(\bar{y}_{ran}) - V(\bar{y}_{st}) = \frac{1}{nN^2} \sum_{h=1}^{L} \left(Y_h - \frac{Y}{L}\right)^2 \qquad \ldots \qquad (60)$$

Thus, given the number L of strata and the number n to be selected from each stratum the gain due to stratification is proportional to the variance between strata totals. If there are no differences between strata totals the unstratified design is as good as the stratified one for estimating population mean or total. For estimating the variance of the estimator of mean, the unstratified design may be superior because of an increase in the degrees of freedom.

Suppose a scheme of stratification is given as also the probabilities of selection of different units. We can easily find the probabilities of selection of the units for the corresponding unstratified design by dividing the original probabilities by the number of strata. These probabilities could then be regrouped so that total in each group is the same. If originally there were L strata and finally there are L' groups one can consider a stratification scheme with L' strata where the probabilities of selection of the units are L'/L times their probabilities in the original stratification scheme. Let Y_h denote the stratum total for the h^{th} stratum in the original scheme and Y'_k for the k^{th} stratum in the modified scheme.

The modified scheme is better than the original scheme provided

$$\sum_{k=1}^{L'} L'(Y'_k - \frac{Y}{L'})^2 > \sum_{h=1}^{L} L(Y_h - \frac{Y}{L})^2 \qquad (61)$$

Thus we may gain over the scheme of equalization of strata totals even if the number of strata is reduced.

6.4. Illustrations:

a) There is a population of 15 units with the following values of the character.

1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 8, 12.

The initial scheme is to divide the population into two strata so that the totals are equal. After that 2 units are to be selected from each stratum at random with equal probability with replacement.

The two strata are

Stratum	Values of Units	Total
I	1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3	24
II	4 8 12	24

The probabilities of selection for the units in the first stratum are 1/12 and for the second stratum 1/3 each. The variance of an estimate of the total in this case is 96.

The modification suggested is to retain the same probability as before for each unit but to change the stratification to the following.

Stratum	Values of the Units	Total
I	1, 1, 1, 1, 2, 2, 2, 2, 4	16
II	3, 3, 3, 3, 8, 12	32

The variance of the estimate of population total for this scheme is only 32. This remarkable reduction in the variance is mainly due to the differences between the strata sizes.

b) The following illustration is to show how deliberately making the strata totals unequal may result in reduced variance even if the number of strata is reduced.

Let the original stratification of population of 17 units be

Stratum			Total
I	Values of Units	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	55
	Sizes	1, 1, 1, 2, 2, 3, 3, 4, 4, 5	26
II	Values of Units	10, 13, 14, 16	53
	Sizes	4, 6, 8, 8	26
III	Values of Units	17, 18, 20	55
	Sizes	8, 9, 9	26

The scheme is to draw a sample of two units from each stratum with probabilities proportional to sizes with replacement. The variance of the estimate of the total is: $59 \frac{47}{72}$.

It will be better to use the following stratification.

Stratum			Total
I	Values of Units	3, 4, 5, 7, 9, 10, 13, 17, 20	88
	Sizes	1, 2, 2, 3, 4, 4, 6, 8, 9	39
II	Values of Units	1, 2, 6, 8, 10, 14, 16, 18	75
	Sizes	1, 2, 3, 4, 5, 8, 8, 9	39

The variance of the estimate of total for this modified scheme is only $32 \frac{59}{72}$.

6.5. The rule of equalization of strata totals has also been seriously considered for minimum variance allocation. As before it is easy to see that if a stratification is optimum for a variable X it is also optimum for all linear functions of X. The rule can, if at all, give optimum stratification for only one of the family of variables X+b. It seems unlikely that one should come across many such variables.

In practice the strata are equalized for the number of secondary stage units in a varying probability two-stage design. This helps in achieving self-weighting and equal work-load. If the mean per secondary stage unit is a monotonic function of the number of secondary stage units in the psu this may also help in increasing the differences between strata total . This fact alone may help in achieving substantial gains due to stratification. Thus we are likely to gain rather than lose by having to approximate this rule in the absence of the knowledge about the estimation variable.

PART II

MISCELLANEOUS PROBLEMS CONNECTED WITH STRATIFIED SAMPLING

CHAPTER VII

DECISION ABOUT THE NUMBER OF STRATA

7.1. From the tables in the previous chapters it is obvious that a larger number of strata is bound to reduce the variance of the estimator for a given mode of allocation. It is also otherwise easy to see that when we are optimizing the strata boundaries additional boundary points should reduce the error. This may mean that one should use as many strata as possible consistent with the total sample size. In an extreme case one should take only one unit per stratum. This will work out very well if the observation that equal allocation is optimum for optimum stratification continues to hold for larger number of strata than have been covered in this study.

The advantage of continued stratification diminishes quite rapidly. The fact that there are other sources of error besides the sampling makes the reduction beyond a certain limit superfluous. Cochran (1963) came to this conclusion on the basis of studies connected with some real populations. The present study of stratification of some theoretical distributions lead to the same conclusions qualitatively. There is, however, some minor quantitative difference .

7.2. Law of Reducing Variance with Increasing Number of Strata:

Dalenius (1957) suggested that variance varies inversely as the square of the number of strata. This is an approximation, exact for the rectangular distribution and quite close for a number of skew distributions studied by Cochran. He expected to discover some relationship between the rate of reduction of variance and skewness of the distributions. The study failed to reveal any such relationship. He remarks that in general the eight distributions which he had studied followed the approximation quite closely. There are, however, quite remarkable differences between some of them. For example, two of them give the following variances compared to unstratified random sampling variance.

Number of strata	V_{min}/V_{ran} Variable 3 Ind. Loans	Variable 5 Coll. Stu.
2	.288	.178
3	.124	.076
4	.067	.047

This suggests that although the effect of stratification may not be related to skewness it varies from one population to another. It will be worthwhile, for fixing the sample size in a stratified sample, to take into account this variation. The study of the class of Chi-square distributions is likely to prove both simple and useful in this connection. If we have the variance ratios for various Chi-square distributions we can guess the appropriate sample size for a population which resembles any one of them in shape. Tables are given for the strata defined for proportional as well as equal (which happen to be the same as minimum variance) allocation.

TABLE XIII

VARIANCE OF THE ESTIMATOR OF POPULATION MEAN BASED ON OPSPA DEFINED STRATIFIED PROPORTIONAL ALLOCATION DESIGN

Number of	Variance Ratio V_{prop}/V_{ran}				
Strata	χ^2_1	χ^2_2	χ^2_3	χ^2_4	χ^2_6
2	.348	.353	.355	.358	.359
3	.177	.175	.182	.185	.186
4	.107	.110	.112	.114	.115
5	.072	.075	.076	.078	.077
6	.052	.054	.055	.056	.056

TABLE XIII - Continued

	χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	.360	.362	.363	.364	.364	.364
3	.186	.186	.188	.191	.191	.192
4	.115	.116	.118	.120	.120	.120
5	.079	.079	.081	.081	.082	.085
6	.057	.058	.059	.060	.060	.061

TABLE XIV

VARIANCE OF THE ESTIMATOR OF POPULATION MEAN BASED ON OPSEA DEFINED STRATIFIED EQUAL ALLOCATION DESIGN

Number of Strata	Variance Ratio V_{eq}/V_{ran}				
	χ^2_1	χ^2_2	χ^2_3	χ^2_4	χ^2_6
2	.244	.288	.308	.320	.332
3	.108	.134	.148	.154	.163
4	.061	.078	.086	.091	.097
5	.039	.051	.056	.060	.065
6	.027	.036	.041	.043	.045

TABLE XIV - Continued

	χ^2_8	χ^2_{10}	χ^2_{15}	χ^2_{20}	χ^2_{25}	χ^2_{30}
2	.339	.345	.351	.357	.358	.358
3	.168	.171	.178	.179	.180	.182
4	.100	.102	.105	.108	.109	.109
5	.067	.068	.070	.071	.074	.075
6	.047	.048	.050	.053	.053	.053

It is found that for both proportional and equal (minimum variance) allocation cases and for all the gamma populations the variance ratio can be expressed as the inverse of a quadratic function in the number of strata. The constants of the quadratic functions were calculated on the basis of the variances for two and three strata. Then the fitted values of the variances were compared with the observed variances. The two agreed so closely that any difference could be assigned to the fact that the observed values refer to a close approximation to the optimum stratification and not to the exact optimum. Thus one could correct the Tables XIII and XIV on the basis of these fitted values.

As the number of strata increases it is the quadratic term which dominates the other two in the quadratic expression. Thus Dalenius' conjecture (1957) that the ratio of the variance for (L-1) and L strata is nearly $L^2/(L-1)^2$ is nearer the truth than the absolute statement that the variance is inversely proportional to the square of the square of the number of strata. In this form the approximation becomes closer as the number of strata increases.

It was considered desirable to find how the constants of the quadratic expression varied with the degrees of freedom of the Chi-square distribution. It was found that except for the quadratic term the others behaved rather erratically. Assuming that the variation in the constants should be smooth with variation in the degrees of freedom and knowing that we are fitting the quadratic to approximate variance ratio values, it seems appropriate to assume some model which forces the variation to be smooth. It is clear that the constant term of the quadratic is the least important. Thus if we keep it the same for all degrees of freedom it may not affect the fit adversely. In case of proportional allocation it was found that the constant could be taken as zero. For the equal allocation case the value was arbitrarily selected as some sort of an average of the fitted values when no assumption was made about its variation with the degrees of freedom of the distributions. Tables XV and XVI show the comparison of the fits with and without the assumption. The closeness of the two justifies the simplification of the model.

7.3. Fitting the Model : The model used was

$$R_L = V_{ran}/V_{stL} = aL^2 + bL + c \quad \ldots \quad (7.1)$$

where L is the number of strata, V_{ran} the variance for an unstratified sample and V_{stL} the variance for a stratified sample of the same size and L strata. For L=1 the stratified and unstratified selections are identical. This gives a condition that

$$a + b + c = 1 \quad \ldots \quad (7.2)$$

The other two equations used were derived by putting L=2 and L=3 in the model (7.1). This gives the following expressions for a, b and c.

$$a = 0.5 R_3 - 1.0 R_2 + 0.5 \quad \ldots \quad (7.3)$$

$$b = 4.0\ R_2 - 1.5\ R_3 - 2.5 \qquad \ldots \qquad (7.4)$$

$$c = 1.0\ R_3 - 3.0\ R_2 + 3.0 \qquad \ldots \qquad (7.5)$$

When the value of c was assumed (either zero or otherwise) only two equations were required. They were derived by putting $L=1$ and $L=2$ in the model (7.1). These gave the following expressions for a and b.

$$a = 0.5\ R_2 - 1 + 0.5\ c \qquad \ldots \qquad (7.6)$$

$$b = 2.0 - 0.5\ R_2 - 1.5\ c \qquad \ldots \qquad (7.7)$$

The values of the expression (7.1) were then found for 4, 5 and 6 strata and compared with the observed values.

TABLE XV

COMPARISON OF THE GOODNESS OF FIT WITH AND WITHOUT THE ASSUMPTION THAT C = O : PROPORTIONAL ALLOCATION CASE

Distribution	Number of Strata	Values of R_L According to: Observation	Model: $R_L = aL^2 + bL + c$	Model: $R_L = aL^2 + bL$
χ^2_1	4	9.34	9.37	9.29
	5	13.89	13.98	13.78
	6	19.60	19.50	19.16
χ^2_6	4	8.70	8.76	8.74
	5	12.83	12.92	12.89
	6	17.86	17.92	17.84
χ^2_{25}	4	8.33	8.51	8.48
	5	12.20	12.52	12.47
	6	16.67	17.29	17.21

TABLE XVI

COMPARISON OF THE GOODNESS OF FIT WITH AND WITHOUT THE ASSUMPTION THAT c IS THE SAME FOR ALL DISTRIBUTION: EQUAL ALLOCATION CASE

Distribution	Number of Strata	Values of R_L According to		
		Observation	Model	
			$R_1=aL^2+bL+c$	$R_L=aL^2+bL+.0580$
χ^2_1	4	8.33	8.33	8.47
	5	12.99	12.99	13.33
	6	18.52	18.86	19.23
χ^2_6	4	5.18	5.21	5.21
	5	7.75	7.94	7.94
	6	11.11	11.24	11.11
χ^2_{25}	4	4.59	4.65	4.55
	5	6.80	6.99	6.80
	6	9.52	9.80	9.43

Once c is given some fixed value the only constant to be evaluated is a, b is calculated from the condition (7.2). The values of a for $c = 0.0580$ are tabulated below. These are continuously decreasing at a speedily diminishing rate.

TABLE XVII

RELATIONSHIP BETWEEN a AND DEGREES OF FREEDOM

n	1	2	3	4	6	8	10	15	20	25	30
a	.556	.396	.341	.311	.282	.266	.254	.242	.229	.227	.225

7.4. <u>Stratification With Respect to A Correlated Variable</u>: So far we have considered the case where the stratification variable was the same as the estimation variable. In that case increasing the number of strata to twice always reduced the variance considerably and justifies as many strata as possible. In practice stratification is always effected with respect to another variable about which a detailed knowledge exists. As noted earlier if the stratification variable is linearly related to the estimation variable one can still get the optimum stratification and get as much advantage out of it as noted in the tables. The relationship is, however, unlikely to be perfectly linear. At best we could assume a linear regression with high correlation. This model proposed by Cochran (1963) is the starting point of the search for real appreciation of the gains of stratification in quantitative terms.

The model is as follows.

$$y_x = a + bx + e$$

where y is the estimation variable, y_x denotes a typical value of y for a given value x of the stratification variable X. e assumed to have mean zero and variance S_e^2 for all values of X.

This model implies that

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h (a + b\bar{x}_h + \bar{e}_h) \qquad \ldots \qquad (7.8)$$

$$S_y^2 = b^2 S_x^2 + S_e^2 \qquad \ldots \qquad (7.9)$$

and

$$V(\bar{y}_{st}) = b^2\ V(\bar{x}_{st}) + S_e^2 \sum_{h=1}^{L} W_h^2/n_h$$

$$= r^2(S_y^2/S_x^2)\ V(\bar{x}_{st}) + (1-r^2)S_y^2 \sum_{h=1}^{L} W_h^2/n_h$$

Thus

$$V(\bar{y}_{st})/V(\bar{y}_{ran}) = r^2\ V(\bar{x}_{st})/V(\bar{x}_{ran}) + (1-r^2) \sum_{h=1}^{L} W_h^2/w_h \quad (7.10)$$

In case the allocation is proportional, the second term of equation (7.10) on the right hand side becomes $(1-r^2)$. For equal allocation the same term becomes $(1-r^2)L \sum_{h=1}^{L} W_h^2$. We shall not discuss the minimum variance allocation which will be different for the variables X and Y. We are already in possession of tables of the ratios $V(\bar{x}_{st})/V(\bar{x}_{ran})$ for various Chi-square distributions for strata defined by OPSPA and OPSEA for proportional and equal allocations respectively. We also know from the tables in Annexure what proportions of population are in various strata. Thus for these cases the gains due to stratification can be readily computed for different values of r.

TABLE XVII

$V(\bar{y}_{st})/V(\bar{y}_{ran})$ AS A FUNCTION OF L AND r FOR PROPORTIONAL ALLOCATION FOR STRATA DEFINED BY OPSPA FOR VARIOUS CHI-SQUARE DISTRIBUTIONS

Distribution	L	r = .99	.95	.90	.85
χ^2_1	2	.360	.412	.472	.529
	3	.193	.258	.333	.406
	4	.125	.194	.276	.355
	5	.091	.163	.249	.330
	6	.071	.145	.232	.315
χ^2_2	2	.366	.416	.476	.533
	3	.196	.260	.336	.408
	4	.127	.197	.279	.359
	5	.093	.165	.251	.332
	6	.073	.147	.234	.317
χ^2_6	2	.372	.422	.481	.537
	3	.203	.266	.251	.413
	4	.132	.202	.283	.361
	5	.096	.168	.253	.334
	6	.075	.149	.236	.318
χ^2_{10}	2	.375	.425	.483	.540
	3	.204	.267	.342	.414
	4	.134	.203	.284	.362
	5	.098	.170	.254	.335
	6	.077	.150	.237	.320
χ^2_{15}	2	.395	.425	.484	.540
	3	.206	.269	.344	.415
	4	.136	.204	.286	.364
	5	.099	.171	.255	.336
	6	.073	.151	.238	.320
χ^2_{20}	2	.377	.427	.485	.541
	3	.207	.271	.345	.416
	4	.137	.206	.287	.365
	5	.100	.171	.256	.337
	6	.079	.152	.238	.321
χ^2_{30}	2	.377	.427	.485	.541
	3	.208	.271	.345	.417
	4	.138	.207	.287	.365
	5	.104	.175	.260	.340
	6	.080	.153	.239	.322
Minimum		.020	.098	.190	.278

TABLE XVIII

$V(\bar{y}_{st})/V(\bar{y}_{ran})$ AS A FUNCTION OF L AND r FOR EQUAL ALLOCATION FOR STRATA DEFINED BY OPSEA FOR VARIOUS CHI-SQUARE DISTRIBUTIONS

Distri-bution	L	r .99	.95	.90	.85
χ^2_1	2	.266	.354	.459	.558
	3	.135	.244	.374	.497
	4	.091	.210	.351	.485
	5	.069	.189	.332	.468
	6	.059	.182	.328	.468
χ^2_2	2	.306	.380	.467	.549
	3	.156	.245	.350	.451
	4	.101	.194	.305	.410
	5	.076	.173	.288	.398
	6	.061	.157	.273	.382
χ^2_6	2	.347	.405	.469	.540
	3	.182	.255	.342	.425
	4	.117	.196	.292	.382
	5	.086	.169	.273	.362
	6	.067	.151	.252	.347
χ^2_{10}	2	.358	.414	.479	.541
	3	.188	.259	.342	.421
	4	.122	.200	.291	.380
	5	.089	.171	.268	.360
	6	.070	.152	.251	.345
χ^2_{15}	2	.364	.418	.480	.540
	3	.183	.263	.546	.424
	4	.125	.202	.293	.379
	5	.091	.171	.268	.359
	6	.071	.153	.251	.343
χ^2_{20}	2	.370	.422	.484	.542
	3	.197	.215	.346	.424
	4	.127	.202	.294	.379
	5	.092	.172	.268	.359
	6	.074	.155	.252	.343
χ^2_{30}	2	.372	.423	.484	.542
	3	.200	.268	.348	.425
	4	.128	.204	.294	.378
	5	.095	.174	.269	.359
	6	.074	.157	.256	.349
Minimum Variance		.020	.098	.190	.278

Though the observation by Cochran that unless the correlation is very high the gains due to stratification start fading out is true yet even at six strata one seems to be very far from achieving the maximum possible gains of stratification. This is specially true for proportional allocation. The rate of decrease, however, does become so slow that one would try other methods of controlling error if available rather than use the same variable for further stratification. One obvious solution is to use some other variable for deep stratification. This observation helps in dealing with multipurpose surveys. Different variables may be used for stratification to help reducing the errors of estimates of poorly correlated variables.

There are one or two interesting results that come out of these tables. Proportional allocation is uniformly worse than equal allocation for perfect correlation between the estimation and stratification variables. But as the corfelation becomes smaller proportional allocation becomes better. This suggests that one should use proportional allocation unless an extremely good stratification variable is available.

It also appears that though the reduction in variance ratio is larger for the distributions with smaller degrees of freedom when there is perfect correlation (equal allocation case) the position starts changing as the correlation becomes smaller and the number of strata increase. This observation may be useful in selecting the stratification variable.

CHAPTER VIII

SOME CONSEQUENCES OF AN INTERPRETATION OF VARYING PROBABILITY SAMPLING

8.1. Instead of treating simple random sampling as a special case of sampling with probabilities proportional to sizes (pps) where sizes of all the units are equal, this chapter proposes to treat sampling with probabilities proportional to sizes as simple random sampling from a hypothetical population. This approach was considered by Ecimovic (1956). The form of presentation here is, however, slightly different. Ecimovic suggests a method of assigning values to the units of the hypothetical population which is simple and convenient. In the present chapter it is shown that for the particular case of single-stage unstratified sampling that method of assigning values is also the one which reduces the variance of the estimator to a minimum. The method can, however, be improved upon in case of stratified and systematic sampling. This possible improvement is explained through examples.

For stratified varying probability sampling, two allocations have been considered which correspond to the proportional and optimum allocations in case of simple random sampling. Using the known results for simple random sampling it is possible to obtain directly the expression for the variance of the estimator and prove that stratified sampling with proportional allocation is better than simple unstratified case. The solutions of the problem of optimum stratification as given by Dalenius and Gurney (1951) for simple random sampling can again be directly used for selection with probability proportional to size (pps).

The problem of pps with replacement sampling is considered in detail. It is shown, however, that simple random sampling without replacement from the hypothetical population improves upon the pps with replacement sampling. This improvement is substantial only if the sizes are all small. This leads to the suggestion of rounding off the sizes.

The method mentioned above is midway between pps with replacement sampling and pps without replacement sampling. A method of estimation corresponding to the one suggested by Des Raj for pps without replacement sampling is also given. This is likely to reduce the variance considerably though the estimator is slightly more complicated.

8.2. Let there be N units in the population $U_1, U_2, \ldots, U_N$. Let the i^{th} unit U_i have a size X_i $(i = 1, 2, \ldots, N)$. Corresponding to U_i can be conceived a group of X_i sub-units U_{ij}, $(j = 1, 2, \ldots, X_i)$. To these hypothetical sub-units one may assign values for the character under study which depend on the values for the same character in the real unit U_i. If Y_i be the value of the character Y for U_i, one may assign values $Y_{ij} = r_{ij} Y_i$ for U_{ij} such that

$$\sum_{j=1}^{X_i} r_{ij} = 1. \qquad (8.1)$$

Thus the total of the values of the hypothetical sub-units is the same as the total in the original population. An estimate of the total for the hypothetical population based on a sample of sub-units is also an estimate of the total for the original real population of N Units. It is obvious that the values of r_{ij}'s must be fixed in advance before the sample is drawn, and for knowing the value of U_{ij} one has to observe U_i.

Selection of units with pps. Selection of sub-units with equal probabilities leads to the probabilities of observing the original units proportional to their sizes. The cumulative total method of selecting units with pps can easily be interpreted as a method of selecting sub-units with equal probabilities.

Estimation. The estimator of the total

$$Y = \sum_{i=1}^{N} Y_i = \sum_{i=1}^{N} \sum_{j=1}^{X_i} Y_{ij}$$

is clearly

$$Y = \frac{X}{n} \sum_{ij=1}^{n} y_{ij} \qquad \ldots \qquad (8.2)$$

where X is the sum of the sizes (or the total number of sub-units) and n is the total sample size.

Variance of the estimator of total. The variance of the estimator Y is

$$V(Y) = \frac{X}{n} \sum_{i=1}^{N} \sum_{j=1}^{X_i} (Y_{ij} - \frac{Y}{X})^2$$

$$= \frac{X}{n} \sum_{i=1}^{N} \sum_{j=1}^{X_i} (Y_{ij} - \frac{Y_i}{X_i}) + (\frac{Y_i}{X_i} - \frac{Y}{X})^2$$

$$= \frac{X}{n} [\sum_{i=1}^{N} \sum_{j=1}^{X_i} (Y_{ij} - \frac{Y_i}{X_i})^2 + \sum_{i=1}^{N} (\frac{Y_i}{X_i} - \frac{Y}{X})^2 X_i] \qquad (8.3)$$

Optimum values of r_{ij}'s. It is obvious from the expression given above that the variance of the estimator Y will be minimum if all r_{ij}'s are equal for any given i. In this case the values of all Y_{ij}'s are equal to Y_i / X_i.

The estimator

$$Y = \frac{X}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} \quad \ldots \qquad (8.4)$$

and

$$V(Y) = \frac{X}{n} \sum_{i=1}^{N} \left(\frac{Y_i}{X_i} - \frac{Y}{X}\right)^2 X_i. \quad \ldots \qquad (8.5)$$

These optimum values of r_{ij}'s are, however, for the case when we are considering unstratified sampling or sampling from within a stratum, the strata being already defined. For the general stratified sampling equalising all r_{ij}'s may not be the best. This problem will be considered in a separate section.

Estimation of the variance. An unbiased estimator of the expression (8.3) for variance is

$$V(Y) = \frac{X^2}{n(n-1)} \sum_{i,j=1}^{n} \left(y_{ij} - \frac{Y}{X}\right)^2 . \qquad (8.6)$$

For the optimum case it becomes

$$V(Y) = \frac{X^2}{n(n-1)} \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \frac{Y}{X}\right)^2. \quad \ldots \qquad (8.7)$$

Allocations to strata in pps sampling. Proportional allocation for the stratified population of sub-units will mean allocation proportional to the sum of the sizes of the units, i.e.,

$$n_s = n \frac{\sum_{i=1}^{N_s} X_{is}}{X} = n \frac{X_s}{X} . \qquad \ldots \qquad (8.8)$$

Where n_s is the allocation to the s-th stratum, N_s is the number of units in the s-th stratum and X_{is} is the size of the i-th unit of the s-th stratum. The variance of $\hat{Y}$ for proportional allocation is

$$V_{prop} = \frac{X}{n} \left[\sum_{s=1}^{k} \sum_{i=1}^{N_s} \frac{Y_{is}^2}{X_{is}} - \sum_{s=1}^{k} \frac{Y_s^2}{X_s^2} \right] \ldots \qquad (8.9)$$

This is smaller than the variance for unstratified sampling for all populations. The reduction in variance is

$$V_{ran} - V_{prop} = \frac{X}{n} \sum_{s=1}^{k} X_s \left(\frac{Y_s}{X_s} - \frac{Y}{X}\right)^2 . \qquad \ldots \qquad (8.10)$$

Optimum allocation for pps sampling will mean

$$n_s \propto X_s \sigma'_s . \qquad \ldots \qquad \ldots \qquad (8.11)$$

Where σ'_s is the standard deviation in the s-th stratum of the population of sub-units.

$$\sigma'_s = \sqrt{\sum_{i=1}^{N_s} \left(\frac{Y_{is}}{X_{is}} - \frac{Y_s}{X_s}\right)^2 \frac{X_{is}}{X_s}} . \qquad \ldots \qquad (8.12)$$

The reduction in variance of the estimator of total due to using optimum allocation instead of proportional allocation is

$$V_{prop} - V_{opt} = \frac{X}{n} \sum_{s=1}^{k} X_s (\sigma'_s - \bar{\sigma}')^2 \qquad \ldots \qquad (8.13)$$

where

$$\bar{\sigma}' = \sum_{s=1}^{k} \frac{X_s}{X} \sigma'_s . \qquad \ldots \qquad \ldots \qquad (8.14)$$

8.3. The principle used to stratify a population when simple random sampling is employed is to make the strata homogeneous with respect to the value of the character in question. The same principle applied to the sub-unit population means that the original population should be stratified in such a manner that strata become homogeneous with respect to the value of character per unit size.

The method will consist in arranging the units in a monotone sequence of the value of character per unit size $[\frac{Y_i}{X_i}]$ and marking off strata at suitable points depending on the allocation proposed. It is better to consider the hypothetical universe of the sub-units for this stratification.

For proportional allocation Dalenius (1950) has proved that the optimum point of stratification is that where the value of character is the mean of the means of two adjacent strata. The mean of the s-th stratum is $\frac{Y_s}{X_s}$ and that of the (s+1)-th stratum is $\frac{Y_{s+1}}{X_{s+1}}$. For optimum stratification, the value of the character for sub-units at the point of stratification should be $\frac{1}{2}[\frac{Y_s}{X_s} + \frac{Y_{s+1}}{X_{s+1}}]$. A similar principle may be applied to optimum or other allocations.

The approximations discussed in Chapter I can all be applied for pps sampling. For example the rule given by Dalenius and Hodges is reduced to the following:

If the sampling is done with probability proportional to X, a very good approximation to optimum stratification points (for equal allocation) is given by that set of points which divides the cumulatives of $X^{\frac{1}{2}}$ into equal parts.

Here one may note that some of the sub-units corresponding to a unit may go to one stratum and the rest to another. In that case it is not necessarily the best policy to assign equal values to all the sub-units. The possibility of reducing the variance of an estimator of total (or mean) by assigning unequal values to the sub-units is illustrated by a simple example given below:

serial no. of unit i	Y_i	X_i	$\frac{Y_i}{X_i}$
1	4	2	2.00
2	7	3	2.33
3	8	4	2.00
4	10	4	2.50
5	39	6	6.50
6	45	3	15.00
7	58	4	14.50
8	60	4	15.00
9	74	5	14.80

Suppose two strata are to be formed. As the strata are to be made homogeneous with respect to the ratio $\frac{Y_i}{X_i}$ obviously the first four units must go in one stratum and the last four in another. The problem is only about the 5-th unit where the ratio is very different from those in these two groups.

For allocations of one unit per stratum the variances for the estimate of the total for the two schemes of stratification are given below

sl. no.	scheme of stratification	variance of estimate of total
1	stratum 1 : first five units stratum 2 : last four units	1444.03
2	stratum 1 : first four units stratum 2 : last five units	6855.74

An entirely different scheme may be as follows :

Four out of the six sub-units corresponding to the fifth unit belong to the first stratum each having a value 2.25. The other two sub-units belong to the second stratum, each having a value 15.00.

In this case the fifth unit overlaps over the two strata. The variance corresponding to this modified scheme is only 23.52.

<u>Without replacement sampling of sub-units</u>. Instead of taking a simple random sample of sub-units with replacements one may take such a sample without replacement.

The estimation will remain

$$Y = \frac{X}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} \quad \ldots \quad \ldots \quad (8.15)$$

but the variance will be slightly reduced

$$V(Y) = X \left[\frac{1}{n} - \frac{1}{X} \right] \sum_{i=1}^{N} \left(\frac{Y_i}{X_i} - \frac{Y}{X} \right)^2 X_i . \quad \ldots \quad (8.16)$$

The slight reduction will not be worthwhile if $\frac{n}{X}$ is very small. If, for example, the sizes X_i's are all three digited numbers and the sample size n is some number less than 20, one would gain little by this modification. If however the X_i's are rounded off to the nearest hundred, the gain may be substantial.

If X_i's provide good sizes for estimating the total Y, it is very likely that the rounded off values of X_i's will be as good. This rounding off may moreover make the estimation simpler because calculation of multipliers $\frac{X}{nX_i}$ will be easier if X_i's are small numbers.

The simple random sampling of sub-units without replacement is an approach whose position lies between pps with replacement sampling and pps without replacement sampling. Units with size I cannot be repeated in the sample. The maximum number of times a unit can be repeated is equal to its size.

<u>Systematic sampling.</u> If a systematic sample of sub-units is taken then also the estimator of total is

$$\hat{Y} = \frac{X}{n} \sum_{i,j=1}^{n} y_{ij}. \quad \ldots \quad \ldots \quad (8.17)$$

In this case the best values of r_{ij}'s are not necessarily all equal. If, however, it is decided to take all $r_{ij} = \frac{1}{X_i}$ for $j = 1,2,\ldots, X_i$, $i = 1,2,\ldots, X_i$, $i = 1,2,\ldots, N$, then the estimator in this case also is

$$\hat{Y} = \frac{X}{n} \sum_{i=1}^{n} \frac{y_i}{x_i} \quad \ldots \quad \ldots \quad (8.18)$$

If all the sub-units correcponding to a unit are kept together in the arrangement of sub-units for systematic sampling, this amount to the method of selection known as pps systematic sampling. The interpretation of pps systematic sampling as a systematic selection of sub-units with equal probabilities provides the simplest proof of the unbiasedness of the estimator (8.18). It also suggests the possibility of improving the estimator by assigning unequal values to the sub-units.

Example: For illustrating the improvement by assigning unequal values to different sub-units, a particularly simple example is given below where the population consists of four units and the sample of two units.

sl. no. of unit i	value of unit Y_i	size of unit X_i	ratio $\frac{Y_i}{X_i}$
1	4	2	2
2	12	3	4
3	9	3	3
4	20	4	5

Variance for an estimator of total Y using pps systematic sampling when sub-units are assigned equal sizes is 125.33.

If, however, the 12 sub-units are assigned the following values, the variance becomes 45.00.

sl. no. of unit i	sl. no. of sub-units	value assigned to sub-units
1	1	2
1	2	2
2	3	4
2	4	4
2	5	4
3	6	1
3	7	4
3	8	4
4	9	4
4	10	4
4	11	5
4	12	7

8.4. Overlapping Units :

One may think of a two-stage sampling when some of the second stage units belong to more than one first-stage unit. In this case it is not apparent as to what method of estimation for total or for the variance of the estimator of total should be adopted.

Such cases are not imaginary and do arise in practice. For example, the problem may be to select a sample of households of factory workers. In the absence of a frame of such households, they may be approached in two stages. In the first stage a sample of factories may be selected. In the second stage an appropriate number of workers may be selected using the lists of workers of the selected factories. The households of these selected workers may be approached for investigation.

In such a method of selection, those households which have more than one workers in the same or different factories will have higher probabilities of selection. It may be quite difficult to calculate these probabilities for the purpose of estimation. It may even be impossible in the absence of any knowledge about the number of factory workers in all the factories. The following is a simple device to avoid this calculation and still get unbiased estimators.

If the value of some character under study for the i-th household is Y_i and there are n_i factory workers in the household, one may consider a group of n_i sub-units corresponding to this household each having a value $\frac{Y_i}{n_i}$. Although households overlap the first-stage units, i.e., factories, the corresponding sub-units do not and the estimation procedure becomes apparent.

Here again there is a possibility of improving the estimator by assigning unequal values to the sub-units. The principle of assigning these values may differ for different characters.

8.5. PPS Sampling Without Replacement

If the values Y_i of certain units U_i, $(i=1,2,\ldots, n_1)$ are already known, the total Y may be estimated by taking a sub-unit at random from the sub-units corresponding to the rest of the units. If a sub-unit corresponding to U_{n+1} is selected, an unbiased estimator is

$$\hat{Y}_{n_1+1} = \sum_{i=1}^{n_1} y_i + \frac{y_{n_1+1}}{x_{n_1+1}} (X - \sum_{i=1}^{n_1} x_i). \quad \ldots \quad (8.19)$$

In sampling without replacement, as the values of the selected units are known in the order of selection, one gets a set of such estimators with $n_1 = 0,1,2,\ldots,(n-1)$. Des Raj has shown that these estimators are uncorrelated. Because of this remarkable property the variance of the simple average of these estimators is $\frac{1}{n^2}$ times of sum of the individual variances.

On considerations of sampling of sub-units with equal probabilities, it can be easily proved that $\hat{Y}_{n_1+1}$ has a smaller variance than $\hat{Y}_{n_1}$ for $n_1 = 1, 2, \ldots, (n-1)$.

Hence

$$V[\frac{1}{n}\sum_{n_1=0}^{n-1} \hat{Y}_{n_1+1}] = \frac{1}{n^2}\sum_{n_1=0}^{n-1} V(\hat{Y}_{n_1+1}) \leq \frac{1}{n^2}\cdot nV(\hat{Y}_1) = \frac{1}{n} V(\hat{Y}_1). \quad (8.20)$$

But $\frac{1}{n} V(\hat{Y}_1)$ is the variance of an estimator based on a sample of size n drawn with pps with replacement. This proves that the estimator proposed by Des Raj for without replacement sampling is superior to the usual estimator for a sample drawn with replacement*.

It may be noted that in cases where a unit overlaps two or more strata as suggested in section 3, and if a sub-unit corresponding to such a unit is selected in one stratum, the values of the other sub-units in other strata are also known. If instead of ignoring this information one utilises it, the variance will be still further reduced.

---

* Ecimovic also considers the case of without replacement sampling. But the estimator he arrives at is a biased one. As is well known, the unbiased estimator used for pps with replacement sampling is biased for without replacement sampling. This bias is, however, unlikely to be large whenever the sample size is small compared to the number of units in the population.

CHAPTER IX

ON OPTIMUM PAIRING OF UNITS

9.1. Introduction: Cluster sampling is one of the well known and widely used sampling procedures. The populations is divided into a number of clusters of units and one or more of them are selected at random. By a cluster we generally understand a set of units such that for any proper subset there is atleast one more unit of the set which is adjacent to it in location. In this chapter, however, the word 'cluster' will be used for any set of units whether it satisfies the above mentioned property or not. The problem considered is that of defining clusters so as to minimize the risk involved in estimating the population total of some characters from a sample of clusters. An exact solution is obtained for clusters of two units. This solution is extended to the case of selecting two units with varying probabilities.

The last mentioned problem was viewed as one of linear programming by Des Raj (1956) and solved by the Simplex Method. It is shown here that removing some of the restrictions imposed on the solution, one can improve upon the optimum described by Des Raj. The magnitude of the improvement is illustrated by considering the same populations as were discussed by Des Raj in his paper.

The method suggests a new approach to the problem of stratification, namely, by dividing the population ~~into a number of symmetric subpopulation~~ into a number of symmetric sub-populations and then selecting clusters in the optimum manner. Repeated use of the method for different characters, or its combinations with stratified sampling with probabilities proportional to sizes may provide a satisfactory solution for the problem of multi-character surveys.

No simple solution exists for optimum formation of clusters of more than two units. Some methods are described which are nearly optimum. The problem of forming clusters of three or four units has been considered in some detail. A number of near optimum methods are compared for some populations and a thumb rule is suggested for populations in general.

The optimum method of pairing of units indicates how a class of linear programming problems can be solved by a method simpler than the Simplex method. It has been illustrated by a small example in Section 9.10.

9.2. <u>Optimum Pairing for Populations of Four Units:</u> We start by considering a population with only four units (Ui) i = 1,2,3,4. By X_i we denote the value of the character X for Ui. The subscripts are so assigned that

$$X_1 \leq X_2 \leq X_3 \leq X_4 \quad \ldots \quad (9.1)$$

and let

$$\sum_{i=1}^{4} X_i = X \quad \ldots \quad (9.2)$$

By enumerating all the possible pairings we can compare the errors of the estimates of X to which they lead. There are only three possible pairings:

i) (U_1, U_2) and (U_3, U_4) ... Pairing P_1

The absolute error of an estimate of total based on any one of the pairs of P_1 is

$$E_1 = |(X_3 + X_4) - (X_1 + X_2)| \quad \ldots \quad (9.3)$$

ii) (U_1, U_3) and (U_2, U_4) ... Pairing P_2

The absolute error of an estimate based on any one of the pairs of P_2 is

$$E_2 = |(X_2 + X_4) - (X_1 + X_3)| \quad \ldots \quad (9.4)$$

iii) (U_1, U_4) and (U_2, U_3) ... Pairing P_3

The absolute error of an estimate of total based on any one of the pairs of P_3 is

$$E_3 = \left| (X_1 + X_4) - (X_2 + X_3) \right| \quad (9.5)$$

Now it is easy to see that E_3 is the minimum of the three absolute errors enumerated above.

<u>Proof</u>: $E_1 = (X_3 - X_1) + (X_4 - X_2)$

and

$$E_3 = \left|(X_4 - X_2) - (X_3 - X_1)\right|.$$

As $(X_3 - X_1)$ and $(X_4 - X_2)$ are both non-negative, it is clear that $E_3 \leq E_1$.

Similarly

$$E_2 = (X_2 - X_1) + (X_4 - X_3)$$

and

$$E_3 = \left|(X_4 - X_3) - (X_2 - X_1)\right|$$

As $(X_2 - X_1)$ and $(X_4 - X_3)$ are both non-negative, it is clear that $E_3 \leq E_2$. In the same manner it can be proved that $E_2 \leq E_1$.

Thus whatever be the values X_i satisfying (9.1), P_3 is the best pairing. Let the loss due to an absolute error E be denoted by $L(E)$. We may reasonably assume that $L(E)$ is a monotonically increasing function of E. Hence

$$L(E_3) \leq L(E_2) \leq L(E_1) \qquad \ldots \qquad (9.6)$$

If we denote the expected value of $L(E)$ over the pairs of P_i by $R(P_i)$ then, for a population of four units

$$R(P_i) = L(E_i) \qquad i = 1, 2, 3, \qquad (9.7)$$

and the risk is the minimum for pairing P_3. In particular, if $L(E) = KE^2$, the risk is proportional to the variance of the estimator based on pairs of P_i.

9.3. <u>Optimum Pairing for Populations of an Even Number of Units</u>: Let there be 2N units in the population and X_i be the value of the character x for the unit U_i, i = 1, 2, ..., 2N. The sub-scripts are given in such a way that

$$X_1 \leq X_2 \leq \cdots \leq X_{2N} \qquad \ldots \qquad (9.8)$$

and

$$\sum_{i=1}^{2N} X_i = X. \qquad \ldots \qquad (9.9)$$

Let us consider a pairing P where two of the pairs are (U_1, U_i) and (U_j, U_{2N}). Corresponding to P we can find a pairing P' which has pairs (U_1, U_{2N}), (U_i, U_j) and the rest of the pairs the same as in P. Then for comparing the risks for the two pairings we have to consider the losses corresponding to the pairs formed by the units U_1, U_{2N}, U_i and U_j only.

If the absolute error corresponding to a pair (U_i, U_k) be denoted by E_{ik}, for all (i, k), then

$$E_{1i} = |N(X_1+X_i) - X| = \left| \frac{N}{2}(X_1+X_i+X_j+X_{2N}) - X - \frac{N}{2}\{(X_j+X_{2N}) - (X_1+X_i)\} \right| \qquad (9.10)$$

$$= |\Lambda - E_1|$$

where

$$\Lambda = \frac{N}{2}(X_1 + X_i + X_j + X_{2N}) - X \qquad \ldots \qquad (9.11)$$

$$E_1 = \frac{N}{2}\{(X_j + X_{2N}) - (X_1 + X_i)\} \qquad \ldots \qquad (9.12)$$

$$E_2 = \frac{N}{2} \{(X_1 + X_{2N}) - (X_i + X_j) \quad \ldots \quad (9.13)$$

$$E_{j(2N)} = | \Lambda + E_1 | \quad \ldots \quad (9.14)$$

$$E_{1(2N)} = | \Lambda + E_2 | \quad \ldots \quad (9.15)$$

and

$$E_{ij} = | \Lambda - E_2 | \quad \ldots \quad (9.16)$$

It is easy to see that $| E_2 | \leq | E_1 |$. Thus if the loss function L be such that for all values of Λ and $| E_2 | < | E_1 |$

$$L(|\Lambda + E_2|) + L(|\Lambda - E_2|) < L(|\Lambda + E_1|) + L(|\Lambda - E_1|) \quad (9.17)$$

then it is clear that the risk corresponding to P' is not greater than that corresponding to P. Thus we may be sure that for optimum pairing one of the pairs ~~must~~ be (U_1, U_{2N}). By following the same argument repeatedly, it is easy to see that it is not possible for the risk to be less than that corresponding to the pairing $P_o : (U_i, U_{2N+1-i})$, $i = 1,2,\ldots, N$. P_o can thus be termed as optimum pairing. If the loss function L is monotonically increasing with the absolute error and is either linear with absolute error or is convex downwords then it can be shown that conditions (9.17) are satisfied.

<u>Proof</u>: Without loss of generality we may take Λ, E_1, and E_2 to be all positive.

a) If $\Lambda \leq E_2 \leq E_1$

$$L(|\Lambda - E_1|) + L(|\Lambda + E_1|) = L(E_1 - \Lambda) + L(E_1 + \Lambda)$$

and

$$L(|\Lambda - E_2|) + L(|\Lambda + E_2|) = L(E_2 - \Lambda) + L(E_2 + \Lambda)$$

Now if L is a non-decreasing function of the argument,

$$L(E_2 - \Lambda) \leq L(E_1 - \Lambda) \quad \text{and} \quad L(E_2 + \Lambda) \leq L(E_1 + \Lambda).$$

Thus the condition (9.17) is satisfied.

b) If $E_2 \leq \Lambda \leq E_1$

$$L(|\Lambda - E_1|) + L(|\Lambda + E_1|) = L(E_1 - \Lambda) + L(E_1 + \Lambda)$$

and

$$L(|\Lambda - E_2|) + L(|\Lambda + E_2|) = L(\Lambda - E_2) + L(\Lambda + E_2)$$

If $(E_1 - \Lambda) \geq (\Lambda - E_2)$, (9.17) clearly holds. If, however, $(E_1 - \Lambda) < (\Lambda - E_2)$, and if L is convex downwards

$$L(\Lambda + E_1) - L(\Lambda + E_2) \geq L(\Lambda - E_2) - L(0) \geq L(\Lambda - E_2) - L(E_1 - \Lambda)$$

because $\Lambda + E > 0$ and $(E_1 - E_2) \geq (\Lambda - E_2)$. So for convex downwards non-decreasing function L, (9.17) is satisfied.

c) $E_2 \leq E_1 \leq \Lambda$

$$L(|\Lambda - E_1|) + L(|\Lambda + E_1|) = L(\Lambda - E_1) + L(\Lambda + E_1)$$

$$L(|\Lambda - E_2|) + L(|\Lambda + E_2|) = L(\Lambda - E_2) + L(\Lambda + E_2)$$

If L be a convex downwards function (9.17) is satisfied even in this case. Thus (9.17) is satisfied for all populations for any non-decreasing convex-downwards loss function. In particular $L(E) = kE^2$ is such a function and the risk corresponding to it is proportional to the variance of the estimator of the total. It also follows that for some populations (9.17) may not hold if the loss function is convex upwards. In particular it will not hold for (c).

9.4. Selection of a Pair of Units with Varying Probabilities: Let there be N units in the population: U_i $i = 1, 2, \ldots, N$. One has to select a pair of them with some probability system. Let the probability of selection of the pair (U_i, U_j) be denoted by p_{ij}. We have to find a set of values of p_{ij}'s satisfying the N restrictions,

$$\sum_{i \neq j} p_{ij} + 2p_{ii} = p_i \qquad i = 1,2,\ldots, N \tag{9.18}$$

where p_i's are given, in such a way that an estimator of the population total of the type

$$\hat{X} = \frac{X_i}{p_i} + \frac{X_j}{p_j} \qquad \ldots \tag{9.19}$$

has the smallest risk associated with it; p_i here denotes the expected number of times the i-th unit occurs in the selected pair.

It will be assumed that the probabilities p_{ij} are rational numbers. This restriction is reasonable because no known physical method can assure any given irrational probability of selection.

Let $p_{ij} = n_{ij}/d_{ij}$, $i, j = 1, 2, \ldots, N$ be a given set of p_{ij}'s, and L be a number divisible by all the d_{ij}'s. Let us associate Lp_i sub-units with U_i, $i = 1, 2, \ldots, N$. If X_i is the value of the character x for U_i we shall impute a value X_i/Lp_i to each of the sub-units corresponding to U_i. Thus the total of the imputed values of the sub-units is also $X = \sum_{i=1}^{2N} X_i$. The total number of these sub-units is 2L. The total X can thus be estimated from a pair of sub-units selected with equal probability from L pairs into which the sub-units can be divided. The estimator will be of the type (9.19) and the expected number of sub-units corresponding to U_i in a pair selected at random is p_i.

If the loss function be non-decreasing convex downwards clearly the best method is to arrange the sub-units in increasing or decreasing order of their magnitudes i.e., of the $\frac{X_i}{Lp_i}$'s and take the optimum pairing as found in the previous section. The whole method can be summarised as follows:

1. Order the units in increasing order of $\frac{X_i}{p_i}$

2. Find the cumulative totals $S_i = L \sum_{j=1}^{i} p_j$ where Lp_i is an integer for all values of i. Put $S_o = 0$.

3. Select a number r at random from 1 to L with equal probability.

4. If $S_i - 1 < r \leq S_i$ select the i-th unit

5. If $S_j - 1 < 2L+1 - r \leq S_j$ select the j-th unit.

This method of selection shall be called <u>Balanced Systematic Sampling</u> (of size two). The probability p_{ij} resulting from this procedure will form an optimum set for given p_i's.

The case of an odd number of units can be dealt with by taking $L = N$, all the p_i's being equal to $\frac{2}{N}$. Similarly the case of simple random sample of two units can be compared with the balanced systematic sample of two units by assigning $(N-1)$ sub-units to each of the units.

In considering the problem in this section we have allowed the possibility of having either one or two units in a cluster. With this relaxation of conditions, it may be possible to improve upon the optimum clustering derived in Section 2. One such method of improvement will be discussed in a subsequent section.

9.5. <u>Optimum Set of Probabilities - a Problem in Linear Programming:</u>

Let

$$Y_i = a + bX_i, \quad i = 1,2,\ldots, N \tag{9.20}$$

where X_i's are known but a and b are not. A pair of units is to be selected so that the expected number of times the i-th unit is included in the sample is proportional to X_i. The problem is to assign the probabilities p_{ij} to pairs of units (U_i, U_j) which minimizes the variance of the estimator

$$Y = \frac{X}{2}\left(\frac{Y_i}{X_i} + \frac{Y_j}{X_j}\right) \quad \ldots \tag{9.21}$$

where

$$\sum_{i=1}^{N} X_i = X \qquad \ldots \qquad (9.22)$$

In this case

$$p_i = \frac{2X_i}{X} \qquad \ldots \qquad (9.23)$$

and

$$\frac{Y_i}{p_i} = \frac{X}{2}\left(b + \frac{a}{X_i}\right) \qquad \ldots \qquad (9.24)$$

Hence the problem can be solved by arranging the units in a monotonic sequence of X_i's and drawing a balanced systematic sample of units with pps. The p_{ij}'s corresponding to this sampling procedure obviously form the required set.

Des Raj (1956) has viewed the problem from a different angle. He has shown that, given X_i's, the variance of $\hat{Y}$ is a linear function of p_{ij}'s. The problem is to minimize $V(\hat{Y})$ with respect to p_{ij}'s under the conditions

$$p_{ij} > 0, \quad i,j = 1,2,\ldots, N \qquad \ldots \qquad (9.25)$$

and

$$\sum_{j \neq i} p_{ij} + 2p_{ii} = p_i, \quad i = 1,2,\ldots, N \qquad (9.26)$$

As the conditions (9.25) and (9.26) are linear and $V(\hat{Y})$ is also linear in p_{ij}'s, the problem is one of linear programming. Des Raj proposes to solve it by Simplex Method. But he has also imposed another set of restrictions, namely,

$$p_{ii} = 0, \qquad i = 1, 2, \ldots, N \qquad \ldots \qquad (9.27)$$

Obviously the optimum obtained under added restrictions cannot give smaller variance. As these restrictions are not essential it is better to dispose them off. In that case the Balanced Systematic Sampling is an improvement over Des Raj's optimum solution. It is to be observed that Des Raj's solution assumes a strictly linear relation between Y and X. This restriction is too severe compared to that of order only involved in Balanced Systematic Sampling. There it is assumed that the order of Y_i/X_i's is known. If Y^* refers to the value of the same character in some recently past period we may not err much in taking this order to be the same as that of Y_i^*/X_i's. In a number of population Y/X increases or decreases monotonically with X. In all such cases the units can be arranged in a monotonic sequence of the values of X. Above all this Balanced Systematic Sampling is a very simple method for solving problems of this nature.

The three populations given by Yates and Grundy (1953) and used by Des Raj for illustration are given below. The variances corresponding to the three methods of selection are also given. The method given by Yates and Grundy will be denoted by YG, that due to Des Raj by DR and the Balanced Systematic Sampling (ordering according to p_i's) by BSS.

The populations are

		A		B		C	
Unit	p	y	y/p	y	y/p	y	y/p
1	.1	0.5	5	0.8	8	0.2	2
2	.2	1.2	6	1.4	7	0.6	3
3	.3	2.1	7	1.8	6	0.9	3
4	.4	3.2	8	2.0	5	0.8	2

Comparison of Sampling Variances

Population[‡]	V_{YG}	V_{DR}	V_{BSS}	% Reduction by DR over YG	% Reduction by BSS over DR
A	.323	.200	.100	38.1	50.0
B	.269	.200	.100	25.7	50.0
C	.057	.100	.100	-75.4	00.00

For populations A and B the orders are already optimum hence no further reduction is possible. For C BSS would have given an estimate without any error if the correct order were known.

9.6. Problems of Symmetrization: It may be observed that for any symmetrical population, BSS gives an estimate without any error. If the populations are nearly symmetrical the error of estimate is likely to be very small. This fact may be made use of in a number of ways.

a) Addition of Dummy Units: Once we allow that the sample may contain one or two units of the population, we may add dummy units to the population to achieve greater symmetry and assign arbitrary values to them. This method may be an improvement over the optimum pairing described above where there was a restriction that every pair contains two different units of the population. A simple illustration is given below.

Let the population consist of three units with values 1, 2 and 20. By balanced systematic sample we get estimates 6 and 63/2 for the total with probabilities 1/3 and 2/3 respectively, whereas the actual total is 23. Suppose we add one dummy unit and assign a value X to it. We shall assume that we know the order of magnitude of the four units. If a balanced sample of these units leads to estimates which do not differ from 23 by more than 17/2 we definitely gain over the balanced sample from three units alone. The conditions for this are that either $-51/2 \leq x \leq -17/2$ or $21/2 \leq x \leq 59/2$.

As the range of values of x for which there is a gain is quite wide, we may use this method with advantage with a little bit of intelligent guessing, e.g., with the help of the distribution of the same character in some recently past period. Large populations do not necessarily make the problem more complicated because the values of the character may be put in a small number of equal intervals.

b) <u>Arbitrary Increments in the Values of Some of the Units</u>: The same population of three units given above can be used to illustrate this method. The value of the middle unit may be increased by an arbitrary quantity x. The two estimates that result will be $\underline{6} + 2x$ and $63/2 - x$, with probabilities 1/3 and 2/3 respectively. Any value of x between 0 and 17 will give better results than an ordinary balanced systematic sample without this modification. For larger populations, the increments may be decided upon after the population is represented by a histogram with a small number of class intervals of equal width. For unimodal populations a suggestion is to add $(X_{min} - X_{mod})$ to every value greater than $2X_{mod} - X_{min}$ if the distribution is positively skew. For negatively skew distributions, the corresponding suggestion is to add $X_{max} - X_{mod}$ to every value less than $2X_{mod} - X_{max}$. Here X_{max}, X_{min}, and

X_{mod} stand for the maximum, the minimum and the modal values of the character respectively.

It is clear that if one adds a quantity $\bar{X} - X_i$ to X_i for all values of i, the result is an estimator without any error. As the values of x are not known in advance one has to depend on some related variable x' for this modification. The resulting estimator is the well-known difference estimator. The balancing in this case becomes superfluous. It is very unlikely that after applying such increments one can still arrange the units in increasing values of x. The estimator even without balancing may be very good.

c) Unequal Assignment of Values to the Sub-units : It may be possible to make the population of sub-units symmetrical by assigning unequal values to them. The following example illustrates the method.

Unit No. i	Y_i	X_i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	Y_{i6}	Y_{i7}
1	335	5	67	67	67	67	67		
2	217	7	28	28	29	29	29	37	37
3	203	5	41	41	41	40	40		
4	175	3	59	58	58				
5	125	4	30	30	32	33			
6	114	6	44	12	13	22	26	37	
7	98	5	4	13	22	25	34		
8	98	2	49	49					
9	93	3	44	45	44				
10	92	4	4	4	42	42			

This division of Y_i's into Y_{ij}'s leads to the following distribution of the cluster totals.

Total	Frequency
69	1
70	10
71	11

This is much better than what one would get by assigning equal values to all the sub-units corresponding to a unit. The equal assignment leads to the following distribution.

Total	Frequency
60 1/5	1
62	3
62 1/4	4
62 3/5	4
68 3/5	2
77 14/15	2
86	5

In practice such a symmetrization may not be adopted because it involves a great deal of calculation. As this can be achieved with respect to a related variable only, this labour may be a waste. Very large divergence from symmetry may, however, be reduced by this method with profit.

9.7. Possibility of a New Criterion for Stratification: If the problem is to estimate the population total by dividing the population into a number of strata and selecting an even number of units from each stratum, balanced systematic sampling from the strata would give better results than would be otherwise achieved. Once it is decided to select samples balanced systematically it would be worthwhile to so construct the strata that they are nearly symmetrical. A small example is given below to illustrate this method. The population consists of the units with the following values.

1, 1, 1, 2, 3, 3, 4, 5, 6, 6, 6, 6, 7, 8, 9, 10, 11,
11, 11, 14, 16, 19, 19, 20, 24, 24.

The following two schemes of stratification are suggested.

i) S_1 : 1, 1, 1, 2, 3, 3, 4, 5, 6, 6
S_2 : 6, 6, 7, 8, 9, 10, 11, 11
S_3 : 11, 14, 16, 19, 19, 20, 24, 24

ii) S_1 : 1, 1, 5, 6, 6, 9, 11, 14, 16, 19, 19, 20, 24, 24
S_2 : 1, 2, 3, 3, 4, 6, 6, 7, 8, 10, 11, 11

Taking only one pair from each stratum leads to a variance 138 for an estimator of the total for the first scheme. For the second scheme with only two strata the variance is 184. This means that for samples of equal sizes the second method is slightly better though, considering the individual values the strata in this scheme are very heterogenous.

In case of pps sampling, unequal assignment of values to the sub-units may help in achieving symmetry within strata. It is not necessary that all the units corresponding to a unit should belong to the same stratum. They may spread over two or more strata if it helps in achieving either homogeneity or symmetry.

9.8. Use of Optimum Pairing for Multipurpose Surveys : No completely satisfactory sampling design exists for multipurpose surveys. The design most suitable for one character may not be good for another. Several methods for controlling error exist and they may be used in an optimum manner for different variables. One may use the best method of stratification for one character, the most suitable sizes for deciding the probabilities of selection for a second character and optimum pairing for the third.

In deep stratification, the strata homogeneous with respect to one character are further divided to make the sub-strata homogeneous with respect to another character. A similar technique may be used in pair formation. Initially balanced pairs may be formed with respect to one variable, then optimum pairs of these pairs may be formed for another variable and so on. An example is given below where such a method was applied.

The population considered was a group of 128 villages of Tehsil Basauli in District Badaun. The characters considered were (i) Area (ii) Population (iii) Number of Literates. The information was taken from the Census District Handbook for Badaun. There were two orders of balancing tried. The first was area, population and number of literates and the other was area, number of literates and again number of literates. The variances of the estimators of the population means are denoted by V_{APL} and V_{ALL} respectively for these two orders. The variance for the simple random sampling of eight villages is also given and denoted by V_{SRS}.

Variable	Total	V_{SRS}	V_{APL}	V_{ALL}
Area in Acres	76253	20068	8290	4792
Population	79373	28218	1705	8273
No. of Literates	3638	236.88	176.44	133.28

The two schemes of clustering selected out of the 27 possible ones are not perhaps the two best and a comparison of two schemes is not easy. But both of them give, for all the three variables, more precise estimators than the simple random sampling. There is, however, a limitation of this method. This may be applied with ease only if the number of units is a multiple of 2^n where n is the number of variables under consideration.

Another method may be to arrange the units in increasing order of one variable. For other variables, the same order may be made increasing or decreasing by adding suitable quantities. This makes the same pairing optimum for all characters. For large populations the additions may be made for groups of units put into class intervals. If the units have been arranged in increasing order of a variable x and one has to make this arrangement increasing for x' also, a simple method may be to add $(X_i - X_i')$ to X_i' for $i = 1,2,\ldots, N$.

As x, x' are not known in advance, the increments will have to be made with the help of the most recent figures of the two variables available.

9.9. Some Suggestions for Formation of Larger Clusters: From the following example it would appear that the knowledge of order is not sufficient for optimum formation of larger clusters.

In the following examples, two populations each of six units are to be divided into two clusters of three units each. The purpose is to select one of the clusters with equal probabilities to estimate the population total. The units are arranged in increasing order of values of both the characters under study.

	X_1	X_2	X_3	X_4	X_5	X_6
Population I	1	2	4	8	16	32
Population II	1	2	3	4	5	6

The optimum clustering for Population I is (U_1, U_2, U_6) and (U_3, U_4, U_5). This clustering is not the optimum for the second population. Thus considerations other than order must enter optimum formation of clusters of size larger than two.

Suppose there are Nn units in the population which are divided into N clusters of n units each. Let two of the clusters have the following values of the character

$$\text{Cluster } C_i \qquad X_{i1} < X_{i2} < \dots < X_{in} \tag{9.28}$$

$$\text{Cluster } C_{i'} \qquad X_{i'1} < X_{i'2} < \dots < X_{i'n}$$

If

$$X_{ij} < X_{i'j} \qquad j = 1, 2, \dots, n \tag{9.29}$$

then by exchange of any one U_{ij} by $U_{i'j}$ in the two clusters there would obviously be a reduction in risk. Thus (9.29) cannot hold for any two clusters if the clustering is optimum. A clustering which does not satisfy (9.29) for any pair of clusters will be called a Nearly Optimum Clustering:

A simple method of getting a nearly optimum clustering is the following.

i) Arrange the units in increasing order of values of the character under study.

ii) Divide them into n homogeneous strata of N units each. Thus the first N units in the arrangement from the first stratum, The next N units the second stratum, and so on.

iii) Take the strata in pairs. The suggested clustering has clusters which include one optimum pair of units from each of the pairs of strata. Depending on the pairing of strata, there are a large number of nearly optimum clusterings. Two of them are given below which depend entirely on order.

a) Cluster C_i is $\{U_{2kN+i}, U_{2(k+1)N+1-i}\}$,

k taking all the values for which a unit exists in the population.

b) Cluster C_i is $\{U_{4kN+i}, U_{(4k+2)N-i+1},$

$U_{(4k+3)N-i+1}, U_{(4k+3)N+i}\}$

for all integral values of k for which the units exist.

The formation of clusters of three or four units has been studied more thoroughly because of comparative simplicity. A number of populations with varying degrees of skewness were taken for finding the best among a number of methods of cluster formation. These populations were derived from various Chi-square distributions and will be so denoted in what follows. A large number of real populations resemble the Chi-square distributions in shape. Thus the results derived from the study may prove useful guiding rules in practice.

For clusters of three units the various schemes considered are

S_1 : U_{2i+1} , U_{2i+2}, U_{3N-i} $\quad i = 0, 1, 2, \ldots, N-1$

S_2 : Balancing between the first two strata, and then balancing the pair so derived with the units of the third stratum.

S_3 : Balancing between the second and the third strata and then balancing the pair so derived with the units of the first stratum.

S_4 : Balancing between the first and the third strata and then balancing the pairs so derived with the units of second stratum.

For clusters of four units the various schemes compared were

S_1' : U_{3i+1}, U_{3i+2}, U_{3i+3}, U_{4N-i}, $\quad i = 0, 1, 2, \ldots, (N-1)$

S_2' : Balancing the balanced pairs formed from the entire population.

S_3' : Balancing the balanced pairs from the first two and the last two strata.

Taking the value of the distribution function at the mean as the measure of skewness, the following rules were derived from this study as guides for cluster formation.

i) For clusters of three units use S_4 except if

$$0.34 < F(\bar{x}) < 0.43 \quad \text{when } S_1 \text{ should be used}$$

and

$$0.57 < F(\bar{x}) < 0.64 \quad \text{when } S_1 \text{ should be used}$$

after reversing the arrangement of the units.

ii) For clusters of four units

If $F(\bar{x}) < 0.39$ use S'_1

$0.39 < F(\bar{x}) < 0.41$ use S'_3

$0.41 < F(\bar{x}) < 0.59$ use S'_2

$0.59 < F(\bar{x}) < 0.61$ use S'_3 ~~after reversing the arrangement of units~~

$F(\bar{x}) > 0.61$ use S'_1 after reversing the arrangement of units.

9.10. <u>A Class of Linear Programming Problems</u>: It has been shown in section 9.5 that a particular problem in linear programming may be solved by a simple method. It is suggested here that a whole class of linear programming problems may be solved by this method. These problems may be stated as follows.

Problem : To minimize $(C_1, C_2, C_3, \ldots, C_n)((b_{ij}x_{ij}))\begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}$ (9.30)

with respect to x_{ij}'s (where C_i's and b_{ij}'s are all positive known numbers and the matrix $((b_{ij}\, x_{ij}))$ is symmetrical) under the restrictions of given row-totals and $x_{ij} \geq 0$, i.e., to minimize $\sum_i \sum_j C_i C_j\, b_{ij}\, x_{ij}$ under the restrictions

$$\sum_j b_{ij}\, x_{ij} = X_i \qquad i = 1, 2, \ldots, n \qquad (9.31)$$

This problem may be solved by transforming it into one of sample selection by the following method.

Put $b_{ij}\, x_{ij} = \frac{X}{2}\, p_{ij}$, $\quad b_{ii} = X\, p_{ii}$

$$\sum_i X_i = X$$

Put $p_i = 2X_i / X \qquad i = 1, 2, \ldots, n \qquad (9.32)$

So one has only to arrange the n units in increasing order of C_i's and find out the probabilities of selection for different pairs (i, j) for balanced systematic sampling. From these calculated values of p_{ij}'s and known values of b_{ij}'s one can calculate x_{ij}'s.

A small example is given below to illustrate the method.

Example : To minimize $(5, 6, 7, 8)((b_{ij}\, x_{ij}))\begin{pmatrix} 5 \\ 6 \\ 7 \\ 8 \end{pmatrix}$

with respect to x_{ij}'s where b_{ij}'s are positive and known, under the restrictions

$$x_{ij} \geq 0 \qquad\qquad i,j = 1,2,3,4$$

$$\sum_i b_{1i}\, x_{1i} = 1 = X_1$$

$$\sum_i b_{2i}\, x_{2i} = 2 = X_2$$

$$\sum_i b_{3i}\, x_{3i} = 3 = X_3$$

$$\sum_i b_{4i}\, x_{4i} = 4 = X_4$$

For the balanced systematic sample with conditions

$$2X_1 / X = p_1 = 0.2$$

$$2X_2 / X = p_2 = 0.4$$

$$2X_3 / X = p_3 = 0.6$$

$$2X_4 / X = p_4 = 0.8$$

$$p_{14} = 0.2, \quad p_{24} = 0.4, \quad p_{34} = 0.2, \quad p_{33} = 0.2$$

and all other p_{ij}'s are zero.

Hence $x_{14} = X\, p_{14}/2b_{14} = 1/b_{14}, \quad x_{41} = 1/b_{41}$

$$x_{24} = 2/b_{24}\,, \quad x_{42} = 2/b_{42}$$

$$x_{34} = 1/b_{34}, \quad x_{43} = 1/b_{43}$$

$$x_{33} = 2/b_{33}$$

It may be mentioned that if a function $\sum_i \sum_j f_{ij} (c_1, \ldots, c_n) b_{ij} x_{ij}$ has to be minimized under the restrictions (9.31) where $f_{ij} (c_1, \ldots, c_n)$ is a monotonic increasing function of

$$\left| C_i + C_j + 2 \sum_{k=1}^{n} \frac{C_k x_k}{x} \right| ,$$

the same method as above may be used without any modification.

CHAPTER X

SOLUTION OF A CLASS OF PROGRAMMING PROBLEMS

10.1. Introduction: Most of the problems in statistics are concerned with maximizing or minimising some function under one or more equality or inequality restrictions. When the restrictions are equalities the problems can very often be solved by the method of Lagrange's multipliers. If some of the restrictions are inequalities we call it a Programming Problems and it is generally more difficult to solve. The particular case when the restrictions are linear has been extensively studied under the title of Linear Programming. It is the purpose of the present chapter to show that a much more general class of programming problems can be reduced to the solution of problems which involve only equality restrictions.

10.2. Programming Problems 1: Find the minima of $f(\underline{x})$ with respect to $\underline{x}$ given that $\underline{x} \in A = \bigcap_{i=0}^{m} A_i$ ... (10.1)

where

$$A_0 = \{\underline{x} \mid \underline{x} \geq 0\} \quad \ldots \quad (10.2)$$

$$A_i = \{\underline{x} \mid g_i(x) \leq b_i\}, \quad (i=1,2,\ldots, m) \quad \ldots \quad (10.3)$$

and

$$\underline{x} = (x_1, x_2, \ldots, x_n) \quad \ldots \quad (10.4)$$

In the following an 'interval' shall always means an open n-dimensional interval.

Definition: A point P in A will be called a minimum of $f(\underline{x})$ in A if there exists an interval I containing P such that $f(P) < f(\underline{x})$ for all $\underline{x} \in [(I \cap A) - (P)]$.

The particular class of programming problems considered here imposes some restrictions on the functions $f(\underline{x})$, $g_1(\underline{x})$, $g_2(\underline{x})$ ... and $g_n(\underline{x})$.

a) Throughout the region A, $f(\underline{x})$, $g_1(\underline{x})$, $g_2(\underline{x})$, ... and $g_n(\underline{x})$ are continuous.

b) $f(\underline{x})$ is bounded below in A and is strictly decreasing in at least one of the co-ordinates of $\underline{x}$.

10.3. Lemma (1): If there is any solution $\underline{x}'$ of problem (1) it must satisfy the following conditions:

$$\underline{x}' \in B = A \cap \left(\bigcup_{i=1}^{m} B_i \right) \qquad \ldots \qquad (10.5)$$

where

$$B_i = \{ \underline{x} \mid g_i(\underline{x}) = b_i \} \qquad \ldots \qquad (10.6)$$

Proof: Let us consider a point $\underline{x} \in (A-B)$, i.e., $\underline{x}$ satisfy the conditions

$$\underline{x} \geq 0 \qquad \ldots \qquad (10.7)$$

and

$$g(\underline{x}) < b_i \qquad (i = 1,2,\ldots, m) \qquad \ldots \qquad (10.8)$$

Let x_j be the co-ordinate for which $f(\underline{x})$ is strictly decreasing in A. Keeping all the co-ordinates fixed we can increase x_j slightly such that conditions (10.8) continue to be satisfied. Call this new point $\underline{x}'$. As $f(\underline{x})$ is strictly decreasing in x_j

$$f(\underline{x}) > f(\underline{x}') \qquad \ldots \qquad (10.9)$$

Thus no point can be a solution of problem (1) unless it belongs to B.

__Note:__ In the proof above it is not necessary to assume continuity of all the functions $f(\underline{x})$, $g_1(\underline{x})$, $g_2(\underline{x})$, ..., $g_m(\underline{x})$. It is enough if all these are continuous to the right in x_j.

10.4. __Lemma (2):__ If $\underline{x}'$ is a minimum of $f(\underline{x})$ in a region C_1 such that

$$\underline{x}' \in C_2 \subset C_1$$

then $\underline{x}'$ is also a minimum of $f(\underline{x})$ in C_2.

The proof is obvious.

__Corollary:__ Let $\underline{x}$ denote a point where $f(\underline{x})$ is minimized in the region $A_o \bigcap\limits_{i \in w} B_i$. If $\underline{x}$ belongs to A, it is also a minimum of $f(\underline{x})$ in A.

10.5. __Lemma__ (3): If P is a minimum of $f(\underline{x})$ in A and it belongs to some region $A_o \bigcap\limits_{i \in w} B_i$, then either P is a minimum of $f(\underline{x})$ in $A_o \bigcap\limits_{i \in w} B_i$ or it also belongs to some B_j, $j \notin w$.

__Proof:__ (i) Let there be some interval I' containing P such that

$$I' \bigcap A_o \bigcap_{i \in w} A_i = I' \bigcap A.$$

As P is a minimum of $f(\underline{x})$ in A, there is some interval I containing P such that

$$f(\underline{x}) > f(P) \quad \text{for all} \quad \underline{x} \in [(I \cap A) - (P)].$$

So $f(\underline{x}) > f(P)$ for all $\underline{x} \in [(I \cap I' \cap A) - (P)] = [(I \cap I' \cap A_o \bigcap_{i\in w} A_i) - (P)]$.

Hence P is a minimum of $A_o \bigcap_{i\in w} A_i$ and so, by Lemma 2, of $A_o \bigcap_{i\in w} B_i$.

(ii) Let there be no interval I' containing P for which

$$I' \cap A_o \bigcap_{i\in w} A_i = I' \cap A.$$

In this case every neighbourhood of P must have at least one point belonging to the complement of $\bigcap_{i\notin w} A_i$. Thus in every neighbourhood of P, at least one of the functions $g_1(\underline{x})$, $j \notin w$ must assume a value greater than b_j at some point.

As $P \in A_j$, $g_j(P) \leq b_j$. If $g_j(P) < b_j$ and $g_j(\underline{x})$ is continuous at P, then we can find a neighbourhood of P throught which $g_j(\underline{x}) < b_j$. As this is not the case, $g_j(P) = b_j$. Thus P belongs to B_j, $j \notin w$.

10.6. <u>Solution of Problem 1</u>: We can solve Problem (I) by the following steps:

<u>step (1)</u>: Find the minima of $f(\underline{x})$ on $A_o \cap B_j$, $(j=1,2,\ldots, m)$. If some of them belong to A, they are also minima of $f(\underline{x})$ in A.

<u>step (2)</u>: Find the minima of $f(\underline{x})$ on $A_o \cap B_j \cap B_{j'}$, $(j \neq j'=1,2,..,m)$. If some of these points belong to A they are also minima of $f(\underline{x})$ in A.

Proceed step by step increasing the number of restrictions. In this way all the minima of $f(\underline{x})$ in A are obtained. The absolute minimum, if required, can be found by comparing the values of $f(\underline{x})$ at these various minima.

Note (1): If some A_i is defined by $\{\underline{x} \mid g_i(\underline{x}) \geq b_i\}$, it can be redefined by $\{\underline{x} \mid g_i'(\underline{x}) \leq b_i'\}$, where $g_i'(\underline{x}) = -g_i(\underline{x})$ and $b_i' = -b_i$. Thus the definition of A_i as given in Problem (1) is quite general.

Note (2): Let some of the restrictions, say 1 to k, in Problem (1) be equalities, This means that

$$\underline{x} \in A = A_o \bigcap_{i=1} B_i \bigcap_{j=k+1} A_j$$

The solution of the problem will then remain essentially the same. Only the first step will be to find the minima of $f(\underline{x})$ on the surface $A_o \bigcap_{i=1} B_i$ and in the subsequent steps we take additional B_j's from $j = k+1, k+2, \ldots, m$ only.

10.7. Problem 2: Find the maxima of $f(\underline{x})$ with respect to $\underline{x}$ given that

$$\underline{x} \in A = \bigcap_{i=0} A_i$$

which A_i's are defined as in Problem (1) and the restriction (a) of Problem (1) is satisfied. Restrictions (b) is, however, changed to (b') $f(\underline{x})$ is bounded above in A and is strictly increasing in at least one of the co-ordinates of $\underline{x}$.

The solution of Problem (2) can be found by the same method as outlined in 10.5. The word 'Minima' has to be replaced by 'maxima'. The proof requires minor modifications in the lemmas proved above. These are obvious and need not be discussed in details.

10.8. Thus we find that the solution of programming problems reduces to the solution of a number of problems of the type given below.

<u>Problem 3:</u> Minimize a function $f(\underline{x})$ with respect to $\underline{x}$ given that

$$\underline{x} \in B = \bigcap_{i=1} B_i \qquad \ldots \qquad (10.10)$$

where

$$B_i = \{ \underline{x} \mid g_i(x) = b_i \},$$

$$i = (1, 2, \ldots, m) \qquad \ldots \qquad (10.6)$$

and

$$\underline{x} = (x_1, x_2, \ldots, x_n) \qquad \ldots \qquad (10.4)$$

If $f(\underline{x})$, $g_1(\underline{x})$, $g_2(\underline{x}), \ldots, g_m(\underline{x})$ are all partially differentiable with respect to $x_1, x_2, \ldots,$ and x_n and these partial derivatives are continuous in B the problem can generally be solved with the help of Lagrange's multipliers.

<u>Solution:</u> Let us define

$$Q = f(\underline{x}) + \underline{k} \{ \underline{g}(\underline{x}) - \underline{b} \} \qquad \ldots \qquad (10.11)$$

where

$$\underline{k} = (k_1, k_2, \ldots, k_m) \qquad \ldots \qquad (10.12)$$

and

$$[\underline{g}(\underline{x}) - b] = [\; g_1(\underline{x}) - b_1 \;, \ldots, \; g_m(\underline{x}) - b_m \;] \qquad (10.13)$$

The $(\underline{x}, \underline{k})$ which minimizes Q unconditionally also minimizes $f(\underline{x})$ in B. Differentiating Q with respect to x_j and equating to zero, we get the following equations

$$f^j(\underline{x}) + \underline{k}\,[\underline{g}^j(\underline{x})] = 0, \quad (j = 1,2,\ldots, n) \quad \ldots \quad (10.14)$$

where

$$f^j(x) = \frac{\partial f(\underline{x})}{\partial x_j} \quad \ldots \quad (10.15)$$

$$g_i^j\,(\underline{x}) = \frac{\partial g_i(\underline{x})}{\partial x_j} \quad \ldots \quad (10.16)$$

and

$$\underline{g}^j(x) = g_1^j\,(x), g_2^j\,(x), \ldots, g_m^j\,(x) \quad \ldots \quad (10.17)$$

The n equations (10.14) together with the m equations (10.10) give a solution to problem 3. The explicit solution of these equations may not be simple. If $f^j(\underline{x})$ are linear and $g_i^j\,(\underline{x})$ are constants for all i and j, the problem reduces to the solution of (m + n) simultaneous linear equations in $(\underline{x}, \underline{k})$. They can be solved by sweep-out method.

If the second order partial derivatives of $f(\underline{x})$ exist and are continuous, an iterative method starting with an approximate solution may be obtained.

Let $\underline{x}_o$ be an approximate solution.

$$f^i(\underline{x}) = f^i(\underline{x}_o + d) \doteq f^i(\underline{x}_o) + \sum_{j=1}^{n} d_j\, f^{ij}(\underline{x}_o) \quad \ldots \quad (10.18)$$

and

$$g_i(\underline{x}) \doteq g_i(\underline{x}_o) + \sum_{j=1}^{n} d_j\, g_i^j\,(\underline{x}_o) \quad \ldots \quad (10.19)$$

With these approximations the problem again reduces to the solution of (m+n) simultaneous linear equations. We consider below two special cases of problem (1) which are of practical interest in sampling.

10.9. Problem 4: Minimize

$$f(\underline{x}) = \sum_{i=1}^{n} a_i / x_i \qquad \text{given that}$$

$$g_i(\underline{x}) = \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \quad (i = 1,2,\ldots, m) \qquad \ldots \qquad (10.20)$$

and

$$\underline{x} \geq 0 \qquad \ldots \qquad (10.21)$$

where a_i and a_{ij} and b_i are all positive for all i and j.

As $f(\underline{x})$ is strictly decreasing in all the x's and g's $g_1(\underline{x})$, $g_2(\underline{x})$, ..., $g_m(\underline{x})$ and $f(\underline{x})$ are all continuous in x_1, x_2, and x_n in the region $\underline{x} > 0$, the optimum lies on $\bigcup_{i=1}^{m} B_i$, where

$$B_i = \{\underline{x} \mid g_i(\underline{x}) = b_i\}, \quad (i = 1,2,\ldots, m) \qquad \ldots \qquad (10.60)$$

Minimum value of $f(\underline{x})$ on B_i is obtained at $\underline{x}_i$, where

$$\underline{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \qquad \ldots \qquad (10.22)$$

and

$$x_{ij} = b_i (a_j / a_{ij})^{\frac{1}{2}} / \sum_{j=1}^{n} (a_j a_{ij})^{\frac{1}{2}} \qquad \ldots \qquad (10.23)$$

If any one of the $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_m$ satisfy the conditions of the problem it is a solution. If none of them satisfies the conditions, we have to consider the problem of minimizing $f(\underline{x})$ on $B_i \bigcap B_j$. The equations involving Lagrange's multipliers are

$$(a_h / x_h^2) = k_i a_{ih} + k_j a_{jh} \qquad (h=1,2,\ldots,n) \qquad \ldots \qquad (10.24)$$

or

$$x_h = [a_h / (k_i a_{ih} + k_j a_{jh})]^{\frac{1}{2}}, \ (h=1,2,\ldots, n) \qquad \ldots \qquad (10.25)$$

and

$$\underline{x} \in B_i \cap B_j \qquad \ldots \qquad (10.26)$$

Eliminating $x_1, x_2, \ldots, x_n$ from (10.25) and (10.26)

$$b_i = \sum_h a_{ih} (a_h / (k_i a_{ih} + k_j a_{jh})]^{\frac{1}{2}} \qquad \ldots \qquad (10.27)$$

and

$$b_j = \sum_h a_{jh} [a_h / (k_i a_{ih} + k_j a_{jh})]^{\frac{1}{2}} \qquad \ldots \qquad (10.28)$$

If one could solve these two equations in k_i and k_j, one could find the $\underline{x}$ where $f(\underline{x})$ is minimized on $B_i \cap B_j$. It seems however that the solution of these equations is not simple.

An approximation to the solution can be taken as

$$C \underline{x}_o = \underline{x}' , \quad \text{where}$$

$$\underline{x}_o = (x_{01}, x_{02}, \ldots, x_{0n}) \qquad \ldots \qquad (10.29)$$

and

$$x_{0i} = \max [x_{1i}, x_{2i}, \ldots, x_{mi}] \qquad \ldots \qquad (10.30)$$

If

$$g_i(\underline{x}_o) = b_{i0}, \quad (i = 1,2,\ldots, m) \qquad \ldots \qquad (10.31)$$

then

$$C = \min [b_1 / b_{10}, \ldots, b_m / b_{m0}] \qquad \ldots \qquad (10.32)$$

Taking this approximation, $f(\underline{x})$ can be approximated by the following relation.

$$f(\underline{x}) \doteq \sum_{i=1}^{n} a_i / x_i' - \sum_{i=1}^{n} a_i \; d_i / x_i'^2 + \sum_{i=1}^{n} a_i d_i^2 / x_i'^3 \qquad (10.33)$$

where

$$d_i = (x_i - x_i') \qquad \ldots \qquad (10.34)$$

$g_i(\underline{x})$ expressed in terms of the new variables $\underline{d}$ can be written as

$$g_i(\underline{x}) = g_i(\underline{x}' + \underline{d}) = g_i(\underline{x}') + g_i(\underline{d}) \qquad \ldots \qquad (10.35)$$

To obtain the minimum of $f(\underline{x})$ on $\bigcap_i B_i$, we have to solve the following set of simultaneous linear equations.

$$2a_i \, d_i / x_i'^3 = a_i / x_i'^2 + \sum_{j=1}^{m} k_j \, a_{ji}, \quad (i = 1,2,\ldots, n) \qquad (10.36)$$

and

$$\sum_{j=1}^{n} a_{ij} d_j = b_i - \sum_{j=1}^{n} a_{ij} \, x_j', \qquad (i = 1,2,\ldots, m) \qquad (10.37)$$

If $\underline{d}'$ is a solution to the equations (10.36) and (10.37), then $(\underline{x}' + \underline{d})$ is a second approximation of problem 4.

10.10. <u>Problem 5</u>: Minimize

$$f(\underline{x}) = \sum_{i=1}^{n} a_i / x_i + \sum_{i=1}^{n} a_i' / (x_i)^{\frac{1}{2}} \quad \text{given that}$$

$$\underline{x} \in \Lambda = \bigcap_{i=a}^{m} \Lambda_i \qquad \ldots \qquad (10.1)$$

where

$$\Lambda_0 = \{ \underline{x} \mid \underline{x} \geq 0 \} \qquad \ldots \qquad (10.2)$$

and

$$\Lambda_i = \{ \underline{x} \mid \sum_{j=1}^{n} a_{ij} x_j \leq b_i \}, \quad (i=1,2,\ldots, m) \qquad \ldots \qquad (10.3)$$

Here $f(\underline{x}) = f(\underline{x}' + \underline{d})$

$$\doteq \sum_{i=1}^{n} \left\{ \left| \frac{a_i}{x_i'} + \frac{a_i'}{\sqrt{x_i'}} \right| - d_i \left(\frac{a_i}{x_i'^2} + \frac{1}{2} \frac{a_i'}{x_i'^{3/2}} \right) + d_i^2 \left(\frac{a_i}{x_i'^3} + \frac{3}{8} \frac{a_i'}{x_i'^{5/2}} \right) \right\} \qquad \ldots \qquad (10.38)$$

Thus the (m+n) simultaneous linear equations in $(\underline{d}, \underline{k})$ will be

$$2\left(\frac{a_i}{x_i'^3} + \frac{3}{8} \frac{a_i'}{x_i'^{5/2}}\right) d_i = \left(\frac{a_i}{x_i'^2} + \frac{1}{2} \frac{a_i'}{x_i'^{3/2}}\right) + \sum_{j=1}^{n} k_j a_{ji},$$

$$(i = 1,2,\ldots, n) \qquad \ldots \qquad (10.39)$$

and

$$\sum_{j=1}^{n} a_{ij} d_j = b_i - \sum_{j=1}^{n} a_{ij} x_j', \quad (i=1,2,\ldots, m) \qquad \ldots \qquad (10.40)$$

It may be mentioned that problems 4 and 5 have special relevance to the problem of allocation of sample sizes to various strata in a multi-purpose survey when the upper bounds to the variances of certain estimators are given and the task is to minimize the overall cost.

Let the allocations to the L strata be denoted by $n_1, n_2, \ldots, n_L$, and the variance of the i-th variate in the h-th stratum be S_{ih}. Then the variance of the estimate of the population total is of the form

$$V(Y_i) = A_i + \sum_{j=1}^{L} S_{ij}/n_j \qquad \ldots \qquad (10.41)$$

The condition on the upper bound of the variance amounts to

$$A_i + \sum_{j=1}^{L} S_{ij}/n_j \leq B_i \qquad \ldots \qquad (10.42)$$

If there are m estimators for which the upper bounds of the variances are specified. There are m linear restrictions in $(1/n_1, 1/n_2, \ldots, 1/n_L)$. If the cost per unit of investigation be c_i in the i^{th} stratum, c_o be the cost of planning and there is no cost of travelling between units, the total cost of investigation is

$$C = c_o + \sum_{i=1}^{L} c_i n_i = c_o + \sum_{i=1}^{L} c_i/(1/n_i) \qquad \ldots \qquad (10.43)$$

Thus the problem of minimization of cost is identical with problem 4. If the cost of travelling in the i-th stratum is $c_i'/(n_i)^{\frac{1}{2}}$ the problem reduces to problem 5.

CHAPTER XI*

UNBIASED RATIO ESTIMATORS

11.1. Introduction: As the relationship between two characteristics is usually of much interest, estimation of ratios of certain population parameters has become quite important in a large number of surveys. The method of ratio estimation is also being used to estimate population totals, since a ratio estimator is more efficient than the conventional unbiased estimator under certain circumstances not uncommon in actual practice. The usual procedure of using the ratio method in estimating any population ratio or total has been to take the ratio of unbiased estimators of the numerator and the denominator and in the latter case multiply it by the population total of the supplementary variate taken in the denominator. A disadvantage of this method is that the estimator so obtained is biased for many of the selection procedures commonly adopted in surveys. Further a completely satisfactory (at least to the present authors) treatment of the errors and biases of a ratio estimator is not yet available. For small samples, at least, the bias is not likely to be small.

In recent years attempts have been made to give selection and estimation procedures which provide unbiased ratio estimators. Lahiri (1951) has given a method of selecting a sample with probability proportional to its total size (pps) (sum of the sizes of the units in the sample) which is essentially similar to his method of selecting a unit with pps,

* This chapter consists of parts of the paper 'Some Sampling Systems Providing Unbiased Ratio Estimators' by Nanjamma, Murthy and Sethi, Sankhyā (1959).

namely, of selecting a unit with equal probability and including that unit in the sample if a number chosen at random from one to an upper bound of the units is less than or equal to the size of the selected unit. By 'size' here is meant the value of the supplementary variate under consideration. Obviously this method avoids the need for completely enumerating all possible samples and finding their total sizes and the cumulated sizes. Once a sample is chosen with pps it is easy to obtain an unbiased ratio estimator. The disadvantage of the selection procedure given by Lahiri is that it involves rejection of some draws.

Midzuno (1952) and Sen (1952) have independently given a simple procedure for obtaining a sample with pps. Their method consists in selecting one unit with pps and the rest with equal probability without replacement from the remaining units of the population. It may be observed that Lahiri's method of selecting one unit with pps could profitably be used in the selection procedure given by Midzuno and Sen.

In the case of stratified sampling Lahiri has pointed out that his method could be applied to select a sample with probability proportional to $\sum_{s=1}^{k} N_s \bar{x}_s$ where k is the number of strata, N_s the number of units in the s-th stratum and $\bar{x}_s$ the s-th stratum sample mean of the supplementary variate under consideration with a view to get an unbiased ratio estimator. Des Raj (1954) has given the expressions for the variance and an unbiased variance estimator of the ratio estimator in the case of a multi-stage design where the sample of first stage units is selected with pps.

For many of the situations commonly met with in practice it will be shown that the modification of a given sampling scheme which provides unbiased ratio estimator consists essentially in first selecting one unit with probability proportional to its value of the variate occurring in the denominator of the ratio and then the remaining units in the sample according to the original scheme of sampling. It might be expected that in large samples the bias of the conventional ratio estimator is unlikely to be large, since the form of the ratio estimator is the same in the case of the biased and the unbiased ratio estimators and the sample based on the original sampling scheme and that on the modified scheme could be made the same but for a difference of one unit at the most.

In this chapter the estimator and its variance estimator have been given in the case of estimating the ratio $R = \frac{Y}{X}$ where Y and X are the population totals for two characters. The ratio estimator and its variance estimator for estimating Y can be obtained by multiplying the corresponding estimators in the case of estimation of R by X and X^2 respectively.

11.2. Generalised Estimation Procedure: In this section a generalised procedure for estimating unbiasedly certain types of parameters applicable to a large number of sampling designs is given. For the sake of generality it has become necessary to use some notations which are explained below with suitable examples.

Let [illegible] denote a population of finite number of units, say, a universe of N units $u_1\, u_2 \ldots u_N$ and Λ the class of sets α whose elements belong to [illegible]. In such a set the same unit may or may not occur more than once. The class of all point sets and the class of all pairs of units belonging to [illegible] are examples of the class Λ.

Let the population parameter F be expressible as

$$F = \sum_{\alpha \in \Lambda} f(\alpha) \qquad \ldots \qquad (11.1)$$

where $f(\alpha)$ is a single-valued set function defined over the class Λ and $\sum_{\alpha \in \Lambda}$ stands for summation over all sets α belonging to the class Λ. For example, the population total Y can be expressed as F in (7.1) with α as a point set (u_i) and $f(\alpha)$ as y_i the value of the i-th unit for the character y, and Y^2 can be expressed as F in (11.1) with α as a set with two units $(u_i,\, u_j)$* and with $f(\alpha)$ defined as

$$\begin{aligned} f(\alpha) &= 2y_i y_j \qquad \alpha = (u_i,\, u_j), \quad i \neq j = 1,2,\ldots, N \\ &= y_i^2 \qquad\quad \alpha = (u_i,\, u_j), \quad i = 1,\, 2,\ldots,\, N \end{aligned}$$

Let a sample [illegible] be drawn from the population [illegible] with probability P([illegible]). This [illegible] again is a set whose elements belong to [illegible]. It may be noted that the same unit may or may not occur more than once in The class of all such sets will be denoted by Ω which will be the total sample space.

---

* The curled brackets [illegible] are used to denote unordered sets, that is, $u_i,\, u_j$ and $u_j,\, u_i$ are the same.

It will be possible to estimate the population parameter F from the sample ω only if each ω contains at least one set α and each set α is contained in at least one .

An estimator of the parameter F is given by

$$\hat{F} = \frac{\sum_{\alpha \subset \omega} f(\alpha)\phi(\omega, \alpha)}{P(\omega)} \qquad (11.2)$$

where $\sum_{\alpha \subset \omega}$ stands for the summation over all sets α contained in the sample ω and $\phi(\omega, \alpha)$ is a function of ω and α. This estimator will be unbiased if

$$\sum_{\omega \supset \alpha} \phi(\omega, \alpha) = 1$$

where $\sum_{\omega \supset \alpha}$ stands for the summation over all samples ω which contain α, since

$$E(\hat{F}) = \sum_{\omega \in \varepsilon} \sum_{\alpha \subset \omega} f(\alpha)\phi(\omega, \alpha)$$

$$= \sum_{\alpha \in A} f(\alpha) \sum_{\omega \supset \alpha} \phi(\omega, \alpha)$$

An unbiased estimator of the variance of $\hat{F}$ is given by

$$\hat{V}(\hat{F}) = \hat{F}^2 - \frac{\sum_{\alpha, \alpha' \subset \omega} f(\alpha)f(\alpha')\psi(\omega, \alpha, \alpha')}{P(\omega)} \qquad (11.3)$$

where $\sum_{\alpha, \alpha' \subset \omega}$ stands for the summation over all pairs α, α' contained in the sample and $\psi(\omega, \alpha, \alpha')$ is a function of and the pair (α, α') such that

$$\sum_{\omega \supset \alpha,\alpha'} \phi(\omega, \alpha,\alpha') = 1$$

where $\sum_{\omega \supset \alpha,\alpha'}$ stands for the summation over all the samples containing the pair of sets (α,α'), since

$$E\left[\frac{\sum_{\alpha,\alpha' \subset \omega} f(\alpha)f(\alpha')\, \phi(\omega, \alpha,\alpha')}{P(\omega)}\right]$$

$$= \sum_{\alpha,\alpha' \in \Lambda} f(\alpha)f(\alpha') \sum_{\omega \supset \alpha,\alpha'} \phi(\omega, \alpha,\alpha') = F^2$$

The case where $\phi(\omega, \alpha)$ is taken as $P(\omega/\alpha)$, the conditional probability of getting the sample ω given that the set α has been selected first is of interest as in that case it is possible to verify that for many of the designs in general use, this estimator

$$\hat{F} = \frac{\sum_{\alpha \subset \omega} f(\alpha)\, P(\omega/\alpha)}{P(\omega)} \tag{11.4}$$

reduces to the usual estimators of the parameter. An unbiased estimator of its variance is given by

$$\hat{V}(\hat{F}) = \hat{F}^2 - \frac{\sum_{\alpha,\alpha' \subset \omega} f(\alpha)f(\alpha')P(\omega/\alpha \cup \alpha')}{P(\omega)} \tag{11.5}$$

where, $P(\omega/\alpha \cup \alpha')$ is the conditional probability of getting the sample ω given that the units in the union of the two sets α and α' have been selected first. The above variance estimator may take negative values.

An estimator of the variance is possible only if every set

$(\alpha \cup \alpha')$ is contained in at least one ω and every ω contains at least one set $(\alpha \cup \alpha')$.

11.3. Unbiased Ratio Estimator : The above estimation procedure, an estimator for the ratio R of two parameters F and G which can be expressed as

$$F = \sum_{\alpha \in A} f(\alpha)$$

$$G = \sum_{\alpha \in A} g(\alpha)$$

where $g(\alpha)$ is another single valued set function defined over the class A is given by

$$\hat{R} = \frac{\sum_{\alpha \subset \omega} f(\alpha)\phi(\omega, \alpha)}{\sum_{\alpha \subset \omega} g(\alpha)\phi(\omega, \alpha)} . \qquad (11.6)$$

This estimator will be unbiased if

$$P(\omega) = \frac{\sum_{\alpha \subset \omega} g(\alpha)\phi(\omega, \alpha)}{\sum_{\alpha \in A} g(\alpha)} \qquad (11.7)$$

since
$$E(\hat{R}) = \sum_{\omega \in \Omega} \frac{\sum_{\alpha \subset \omega} f(\alpha)\phi(\omega, \alpha)}{\sum_{\alpha \subset \omega} g(\alpha)\phi(\omega, \alpha)} P(\omega).$$

If $g(\alpha)$'s are either all positive or all negative the above form of $P(\omega)$ can be obtained by first selecting a set α with probability proportional to $g(\alpha)$ and then drawing the rest of the units with some probability scheme. In this case the probability of getting ω is

$$P(\omega) = \frac{\sum_{\alpha \subset \omega} g(\alpha) P(\omega/\alpha)}{\sum_{\alpha \in A} g(\alpha)}$$

This shows that if in the general case $\phi(\omega, \alpha)$ is taken as $P(\omega/\alpha)$, the estimator given in (11.6) becomes unbiased for the ratio

$$\hat{R} = \frac{\sum_{\alpha \subset \omega} f(\alpha) P(\omega/\alpha)}{\sum_{\alpha \subset \omega} g(\alpha) P(\omega/\alpha)} \qquad (11.8)$$

An unbiased ratio estimator of F is given by $\hat{F} = \hat{R}.G.$ (11.9)

If $P(\omega/\alpha)$ is independent of the set α, the estimator becomes

$$\hat{R} = \frac{\sum_{\alpha \subset \omega} f(\alpha)}{\sum_{\alpha \subset \omega} g(\alpha)} \qquad (11.10)$$

and if $\frac{P(\omega/\alpha)}{h(\alpha)}$ is independent of the set α,

$$R = \frac{\sum_{\alpha \subset \omega} f(\alpha) h(\alpha)}{\sum_{\alpha \subset \omega} g(\alpha) h(\alpha)}. \qquad (11.11)$$

An unbiased estimator of the variance of $\hat{R}$ is given by

$$\hat{V}(\hat{R}) = \hat{R}^2 - \frac{\sum_{\alpha, \alpha' \subset \omega} f(\alpha) f(\alpha') P(\omega/\alpha \cup \alpha')}{G \sum_{\alpha \subset \omega} g(\alpha) P(\omega/\alpha)} \qquad (11.12)$$

It is possible that F and G can be expressed as sums of set functions defined over more than one class of sets, that is,

$$F = \sum_{\alpha \in \Lambda} f_1(\alpha) = \sum_{\alpha' \in \Lambda'} f_2(\alpha') = \ldots$$

$$G = \sum_{\alpha \in \Lambda} g_1(\alpha) = \sum_{\alpha' \in \Lambda'} g_2(\alpha') = \ldots$$

For each such expression we can give a sampling procedure providing an unbiased ratio estimator. From the point of view of operational convenience, it is preferable to take that class of sets which contains the smaller sets. The size of a set is judged by the number of units it contains. Two examples are given to illustrate the point.

a) The population total X can be expressed in the following two ways

$$X = \sum_{i=1}^{N} X_i = S \frac{\sum_{i=1}^{n} x_i}{\binom{N-1}{n-1}}$$

where S stands for the summation over all sets of n distinct units. In this case the former is to be preferred to the latter because in the former only one unit is to be selected with ppx whereas in the latter case n units are to be drawn with probability proportional to their total size.

b) The population variance $\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{X})^2$ where $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ can again be expressed in the following two ways.

$$\sigma_2 = \frac{1}{N} \cdot \frac{N-1}{n-1} \cdot \frac{1}{\binom{N}{n}} \cdot S \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{N^2} \cdot S'(x_i - x_j)^2.$$

Where S stands for the summation over all sets of n units, $\bar{x}$ is the sample mean and S' stands for the summation over all sets of two units. Here the latter is to be preferred to the former because in the latter case only two units as compared to n units in the former case are to be selected with probability proportional to their measure of size.

The general procedure explained above will be illustrated by applying it to the case of stratified sampling.

11.4. Stratified Sampling : Let k be the number of strata and N_i and n_i be the number of units in the population and the sample respectively for the i-th stratum. For stratified simple random sampling without replacement the modification in the selection procedure for getting an unbiased ratio estimator consists of selecting one unit (say the j-th unit in the i-th stratum) from the whole population with ppx, $(n_i - 1)$ units from the remaining $(N_i - 1)$ units in the i-th stratum and $n_{i'}$ units from $N_{i'}$ units of the i'-th stratum $(i' \neq i)$ with equal probability without replacement. The probability of getting a particular sample s is given by

$$P(s) = \frac{\sum_{i=}^{k} N_i \bar{x}_i}{X \prod_{i=1}^{k} \binom{N_i}{n_i}} \tag{11.13}$$

where $\bar{x}_i$ is the sample mean in the i-th stratum for the variate x. With this procedure an unbiased estimator of the ratio R is given by

$$\hat{R} = \frac{\sum_{i=1}^{k} N_i \bar{y}_i}{\sum_{i=1}^{k} N_i \bar{x}_i} \qquad (11.14)$$

An unbiased estimator of the variance of R is given by

$$\hat{V}(\hat{R}) = \hat{R}^2 - \frac{1}{X\left(\sum_{i=1}^{k} N_i \bar{x}_i\right)} \left[\sum_{i=1}^{k} \frac{N_i}{n_i} \sum_{j=1}^{n} y_{ij}^2 + \sum_{i=1}^{k} \frac{N_i(N_i-1)}{n_i(n_i-1)} \sum_{\substack{j, j' \\ j \neq j'}}^{n} y_{ij} y_{ij'} + \right.$$

$$\left. + \sum_{i \neq i'}^{n} \frac{N_i \, N_{i'}}{n_i \, n_{i'}} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} y_{ij} \, y_{i'j'} \right] \qquad (11.15)$$

It may be noted that $\hat{R}$ resembles the biased combined ratio estimator of R. The modifications of sampling schemes with other types of designs in the strata can be given on similar lines with a view to getting unbiased ratio estimators.

BIBLIOGRAPHY

(Stratified sampling and related subjects)

1. Anderson, P.H. (1942) : 'Distributions in stratified sampling' Ann. Math. Stat. 13, pp. 42-52.

2. Anderson, R.L., Finkner, A.L. and Sen A.R. (1954) : 'A comparison of stratified two-state sampling systems', Jour. Am. Stat. Ass., 49, pp. 539-558.

3. Aoyama, H. (1954) : 'A study of the stratified random sampling', Ann. Inst. Stat. Math. 6, pp. 1-36.

4. Armitage, P. (1947) : 'A comparison of stratified with unrestricted random sampling from a finite population', Biometrika, 34, pp. 273-280.

5. Bryans, E.C., Hartley, H.O. and Jessen, R.J. (1960) : 'Design and estimation in two-way stratification'. Jour Am. Stat. Ass. 55, pp. 105-124.

6. Chapman, D.G. and Junge, C.D. (Jr.) (1956) : 'Estimation of the size of a stratified animal population', Ann. Math. Stat. 27, pp. 375-389.

7. Church, B.M. (1954) : 'Problems of sample allocation and estimation in an agricultural survey', Jour Roy. Stat. Soc.(B) 16, pp. 223-235.

8. Cochran, W.G. (1946) : 'Relative accuracy of systematic and stratified random samples for a certain class of populations', Ann. Math. Stat. 17, pp. 164-177.

9. Cochran, W.G. (1953) : 'Sampling techniques', John Wiley and Sons, Inc., New York.

10. Cochran, W.G. (1963) : 'Sampling techniques', Second edition, John Wiley and Sons Inc., New York.

11. Cornell, F.G. (1947) : 'A stratified random sample of a small finite population', Jour. Am. Stat. Ass. 42, pp. 523-532.

12. Dalenius, T. (1950) : 'The problem of optimum stratification', Skand. Akt. 33, pp. 203-213.

13. Dalenius, T. and Gurney, M. (1951) : 'The problem of optimum stratification II', Skand. Akt. 34, pp. 133-148.

14. Dalenius, T. (1952) : 'The problem of stratification in a special type of sampling design', Skand. Akt. 35, pp. 61-70.

15. Dalenius, T. (1953) : The economics of one-stage stratified sampling', Sankhyā 12, pp. 351-356.

16. Dalenius, T. and Hodges, Jr. (1957) : 'The choice of stratification points', Skand. Akt. 40, pp. 198-203.

17. Dalenius, T. (1957) : 'Sampling in Sweden - contributions to the methods and theories of sample survey practice', Almquist and Wiksell, Stockholm.

18. Dalenius, T, and Hodges, Jr. (1959) : 'Minimum variance stratification', Jour. Am. Stat. Ass. 54, pp. 88-101.

19. Deming, W.E., Hurwitz, W.N. and Tepping, B.J. (1943) : 'On the efficiency of deep stratification in block sampling', Jour. Am. Stat. Ass. 38, pp. 93-100.

20. Deming, W.E. (1950) : 'Some theory of sampling', John Wiley and Sons Inc., New York.

21. Deming, W.E. (1960) : 'Sample design in business research', John Wiley and Sons Inc. London.

22. Des Raj (1954) : 'Ratio estimation in sampling with equal and unequal probabilities', Jour. Ind. Soc. Agr. Stat. 6, pp. 127-138.

23. Des Raj (1956) : 'Some estimators in sampling with varying probabilities without replacement', Jour. Am. Stat. Ass. 51, pp. 269-284.

24. Des Raj (1956) : 'A note on the determination of optimum probabilities in sampling without replacement', Sankhyā 17, pp. 197-200.

25. Des Raj (1957) : 'On estimating parametric functions in stratified sampling designs', Sankhyā 17, pp. 361-366.

26. Des Raj (1958) : 'On the relative accuracy of some sampling techniques', Jour. Am. Stat. Ass. 53, pp. 98-101.

27. Ecimovic, J.P. (1956) : 'Three stage sampling with varying probabilities of selection', Jour. Ind. Soc. Agr. Stat. 8, pp. 14-44.

28. Ekman, G. (1959) : 'A limit theorem in connection with stratified sampling', Skand. Akt. 42, pp. 208-223.

29. Ekman, G. (1959) : 'An approximation useful in univariate stratification', Ann. Math. Stat. 30, pp. 219-229.

30. Ekman, G. (1959) : 'Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution', Ann. Math. Stat. 30, pp. 1131-1135.

31. Evans, W.D. (1951) : 'On stratification and optimum allocation', Jour. Am. Stat. Ass. 46, pp. 95-104.

32. Frankel, L.R. and Stock, J.S. (1939) : 'The allocation of samples among several strata', Ann. Math. Stat. 10, pp. 288-293.

33. Ghosh, B. (1947) : 'Bias introduced by changing the system of stratification', Cal. Stat. Ass. Bull. 1, pp. 43-45.

34. Ghosh, M.N. (1958) : 'A note on stratified random sampling with multiple characters', Cal. Stat. Ass. Bull. 8, pp. 81-90.

35. Goodman, L.A. and Hartley, H.O. (1958) : 'The precision of unbiased ratio-type estimators', Jour. Am. Stat. Ass. 53, pp. 491-508.

36. Goodman, R. (1948) : 'Collapsed strata', Am. Stat. 2, pp. 22.

37. Goodman, R. and Kish, L. (1950) : 'Controlled selection - a technique in probability sampling', Jour. Am. Stat. Ass. 45, pp. 350-372.

38. Hagood, M.J. and Bernert, E.H. (1945) : 'Component indexes as a basis for stratification', Jour. Am. Stat. Ass. 40, pp. 330-341.

39. Hansen, M.H. and Hurwitz, W.N. (1943) : 'On the theory of sampling from finite populations', Ann. Math. Stat. 14, pp. 333-362.

40. Hansen, M.H., Hurwitz, W.G. and Madow, W.G. (1953) : 'Sample Survey method and theory. Vol I - Methods and applications, Vol. II - theory, John Wiley and Sons Inc., New York.

41. Hendricks, W.A. (1948): 'The mathematical theory of sampling', Baily Bros. and Swinfen, London.

42. Jensen, E.L. (1959): 'Optimum stratification of the logarithmic normal distribution - a comment', Skand. Akt. 42, pp. 144-147.

43. Johnson, P.O. and Rao, M.S. (1959): 'Modern sampling methods, theory, experimentation and application', University of Minnesota Press, Minneapolis, Oxford University Press, London.

44. Kitagawa, T. (1955): 'Some contributions to the design of sample surveys, parts I, II and III', Sankhyā, 14, pp. 317-362.

45. Lahiri, D.B. (1951): 'A method of sample selection providing unbiased ratio estimates', Bull. Int. Stat. Inst. 33, pp.133-140.

46. Lahiri, D.B. (1954): 'Technical paper on some aspects of the development of the sample design, NSS No. 5', Department of Economic Affairs, Ministry of Finance, Government of India.

47. Mahalanobis, P.C. (1952): 'Some aspects of the design of sample surveys', Sankhyā 12, pp. 1-7.

48. Maruyama, E., Hayashi, C. and Isida, M.D. (1950): 'On some criteria for stratification', Ann. Inst. Stat. Math. 2, pp. 77-86.

49. Midzuno, H. (1950): 'An outline of the theory of sampling systems', Ann. Inst. Stat. Math. 1, pp. 149-156.

50. Midzuno, H. (1952): 'On the sampling system with probability proportional to sum of sizes', Ann. Inst. Stat. Math. 3, pp. 99-107.

51. Murthy, M.N., Nanjamma, N.S. and Sethi, V.K. (1959): 'Some sampling systems providing unbiased ratio estimators', Sankhyā 21, pp. 299-314.

52. Murthy, M.N., and Nanjamma, N.S. (1959): 'Almost unbiased ratio estimates based on interpenetrating sub-sample estimates', Sankhyā 21, pp. 381-392.

53. Neyman, J. (1934): 'On the two different aspects of representative method. The method of stratified sampling and the method of purposive selection', Jour. Roy., Stat., Soc., 97, pp. 558-625.

54. Nordbotten, S. (1956) : 'Allocation in stratified sampling by means of linear programming', Skand. Akt. 39, pp. 1-6.

55. Nordbotten, S. (1957) : 'On errors and optimal allocation in a census', Skand. Akt. 40, pp. 1-10.

56. Pascual, J.N. (1961) : 'Unbiased ratio estimators in stratified sampling', Jour. Am. Stat. Ass. 56, pp. 70-87.

57. Pearson, E.S. and Hartley, H.O. (1954) : 'Biometrika tables for statisticians', Vol. I, Cambridge University Press.

58. Quensel, C.E. (1958) : 'Some sampling problems when a stratification variable follows a logarithmic normal distribution', Skand. Akt. 41, 177-184.

59. Robson, D.S. and Vithayasai, C. (1959) : 'Unbiased componentwise ratio estimation', Am. Stat. Ass. Proceedings, pp. 155-159.

60. Ross, A and Hartley, H.A. (1954) : 'Unbiased ratio-estimators', Nature, 174, pp. 270-271.

61. Sen, A.R. (1955) : 'On the selection of 'n' primary sampling units from a stratum structure $(n \geq 2)$', Ann. Math. Stat., 26, pp. 744-751.

62. Sethi, V.K. (1960) : 'On the possibility of improving upon the principle of equalization of strata totals', Jour. Soc. Sciences 2, pp. 47-51.

63. Sethi, V.K. (1963) : 'A note on optimum stratification of populations of populations for estimating the population means', Aust. Jour. Stat. 5, pp. 20-33.

64. Showel, M. (1951) : 'How much stratification (on the basis of community size)', Int. Jour. Op. and Att. Res. 5, pp. 229-240.

65. Stephan, F.F. (1941) : 'Stratification in representative sampling', Jour. of Marketing 6, pp. 38-46.

66. Stemens, W.L. (1952) : 'Samples with the same number in each stratum', Biometrika, 39, pp. 414-417.

67. Sukhatme, P.V. (1953) : 'Sampling theory of surveys with applications', Ames. Iowa, U.S.A. and Indian Society of Agricultural Statistics, New Delhi.

68. Taga, Yasushi (1953) : 'On optimum balancing between sample size and number of strata in sub-sampling', Ann. Inst. Stat. Math. 4, pp. 95-102.

69. Yates, F. (1953) : 'Sampling methods for censuses and surveys', (second edition), Hafner, New York.

70. Yates, F. and Grundy, P.M. (1953) : 'Selections without replacement from within strata with probability proportional to size', Jour. Roy. Stat. Soc. (B), 15, pp. 253-261.

71. Zarkovic, S.S. (1956) : 'An illustration of some characteristic situations in the application of the difference estimate', Rev. Int. Stat. Inst. 24, pp. 52-63.