SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited By : P. C. MAHALANOBIS

Vol. 14 | Part 4 | 1955

SOME CONTRIBUTIONS TO THE DESIGN OF SAMPLE SURVEYS

By TOSIO KITAGAWA

Kyusyu University, Fukuoka
and
Indian Statistical Institute, Calcutta

1. Introduction

In spite of some remarkable developments both on the theoretical and the practical sides, certain aspects of the designing of sample surveys seem to remain unexplored. Indeed the last twenty five years may be recognised as a pioneering period in which experts like Hubback (1927), Fisher, Yates (1949) and their school in Rothamsted, Mahalanobis (1940, 1944, 1946, 1952) and his colleagues in India, Neyman (1934) in Poland, Hansen and Hurwitz (1943) and King and Jessen (1945) in U.S.A. and others have contributed to the epoch-making developments of sampling techniques, as shown in various excellent monographs such as those of Yates (1949), Deming (1950) and Cochran (1953). Nevertheless there still remain a number of problems which have been solved only empirically because of the lack of any theory concerning them. In a recent paper, Mahalanobis (1952) pointed out some aspects of the design of sample surveys which should deserve theoretical consideration.

The purpose of the present paper is to take into consideration the real situations dealt with in some types of sample surveys, to analyse them into more fundamental conditions, to give some suitable formulation for each of them, and then to suggest some methods of solving these problems formulated mathematically. Although some of the results which are given in this paper were obtained by the author in connection with his designs of sample surveys executed in Japan, the greater part of this paper was prepared during his stay at the Indian Statistical Institute. In preparing this paper it was a great stimulus to read the latest paper of Mahalanobis (1952) as well as to receive his kind suggestions and encouragement for which the author wishes to express his sincere thanks to Professor P. C. Mahalanobis and his colleagues at the Institute.

The subjects dealt with in this paper are given below. Part I is devoted to the discussion of optimum allocation in sample designs under the assumption of complete information about the population parameters which enter into the designs. Such complete information is not realized and hence Part I really deals with preparatory considerations upon which the results in Parts II and III are based. Nevertheless our emphasis is on multi-purpose estimation which seems to us to be our real aim in sample surveys. Certain unified principles will be followed for finding optimum designs and for the mapping problem. We shall suggest a linearization procedure which will make our designs practicable in sample surveys.

In Part II we shall introduce the notion of the relative efficiency of some sampling designs in which our allocations to each stratum are defined in terms of approximate values of the population parameters involved. The relative efficiency is derived formally from the inequality of Schwartz or Holder so far as the system of stratification or of multi-stage sampling is kept invariant. The generalised cost and variance functions introduced by Mahalanobis (1944) can be suitably transformed into this general formulation and the loss of efficiency studied.

Part III is devoted to a discussion in which a certain system of stratification is used in successive surveys and in which our aim is to accumulate information concerning within-strata variances so as to increase our relative efficiency of a sequence of sample surveys. Here we shall discuss various possible procedures and compare their respective merits. Thus the problems treated in Part III are more realistic approaches as regards efficiency and feed-back principle in comparison with those of Parts I and II, and belong to a category of sequential designs or of historical designs in the language of Mahalanobis (1952). The guiding principle in designing a sequence of sample surveys is the idea of relative minimax solutions which have some connection with Wald's decision functions. This principle seems to us to be an answer to the proposal suggested by Mahalanobis (1952). The detailed analysis will be too complicated to be performed under weak conditions as to the structure of the parent population. We consider it more practical to introduce solutions of the zero order and the first order so that our analysis might be so simplified as to enable us to deal with concrete observations without any essential loss of generality so far as large-scale surveys are concerned. We should turn to a more realistic formulation of our problems in which some improvement of stratification may be effected as we go from one design to the next in the sequence of designs. For this purpose various preparatory considerations should be taken into account. Consequently we consider it necessary to establish some exact sampling distributions of our statistics in order to obtain a theory of inference and a scheme of analysis of variance specially suited to finite populations.

Further problems connected with designs raised by Mahalanobis (1952) are considered in a subsequent paper.

SOME CONTRIBUTIONS TO THE DESIGN OF SAMPLE SURVEYS

PART I. DESIGNS AIMING AT ESTIMATION OF ONE OR MORE PARAMETERS UNDER COMPLETE INFORMATION ABOUT THE PARAMETERS INVOLVED IN THE DESIGNS

1. Introductory

Current theories of designs of sample surveys are usually devoted to the estimation of a single population parameter, specially total or mean of the population concerned. Nevertheless it should be pointed out that the uses of information obtained from a sample survey are sometimes much more than the mere estimation of parameters originally aimed at in the sample design. Moreover, once we are in a position to make a sequence of similar sample surveys in which accumulation of information may serve to improve our designs, it becomes not only natural but also indispensable to have many aims of inference in any one survey. Now we shall give some of these aims.

(a) *Stratified Random Sampling:*

(1°) To estimate a set of certain prescribed linear combinations of stratum-means.

(2°) To estimate the set of within-stratum variances individually and/or their total.

(3°) To estimate the size of each stratum.

(4°) To estimate the between strata variance.

(5°) To estimate the cost function within each stratum.

(6°) To estimate the maximum of stratum-means.

(7°) To give a map of the field showing the dimensional distribution of the characteristic over the field (mapping problem).

(8°) To supply estimates which would be useful in another sampling system different from the one adopted.

(b) *Multi-stage Sampling:* Sub-sampling procedures can be mainly divided into two classes according to whether they are involved with stratification or not. In case of stratification there may arise problems corresponding to (1°)—(8°) in (a), while in both the cases our aims may be concerned with the analysis of variance and cost both between and within primary sampling units (p.s.u.).

(c) *Regression estimates and ratio estimates:* The current formulae concerning variance of ratio estimates are not useful unless there are some accurate estimates of coefficients of variation and of correlation coefficients. A correct approach to the problem of obtaining certain regression and ratio estimates should be the one which makes use of the method of successive designs. Therefore, in the actual survey what we should estimate is not only the population mean but also all or some of these parameters.

In this Part, we shall be mainly concerned with (a) and shall be content merely with giving some indications about (b) and (c). Our main object is to formulate our aims of surveys more precisely so as to be able to define our optimum allocation of resources more suitably. We shall here make use of variances as a loss function in order that our mathematical analysis and allied calculations may be so simplified as to be applicable to actual problems. In this Part, we are concerned with the situations in which complete information about various parameters involved in the design is available.

2. Simultaneous estimation of several parameters which are linear combinations of stratum means

Let us denote by Π the whole parent population, and let it be assumed that Π is divided into k strata $\{\Pi_i\}$ $(i = 1, 2, ..., k)$, where any two strata Π_i and Π_j are assumed to be mutually exclusive. Let the size of Π_i be denoted by $N_i (i = 1, 2, ..., k)$, and let the size of Π be N, so that $N = N_1 + ... + N_k$.

Let us start with the case when we are interested in only one characteristic say x. Let $x_{ij} (i = 1, 2, ..., k; j = 1, 2, ..., N_i)$ be the j-th observation in the i-th stratum. Then the i-th stratum mean $\bar{x}_i (i = 1, 2, ..., k)$ and the general mean $\bar{x}$ are defined by

$$\bar{x}_i = \sum_{j=1}^{N_i} x_{ij}/N_i, \qquad \dots \quad (2.01)$$

$$\bar{x} = \sum_{i=1}^{k} N_i \bar{x}_i/N. \qquad \dots \quad (2.02)$$

Let c_i be the cost per unit in the i-th stratum and let σ_i^2 be the variance within the i-th stratum. In stratified random sampling we draw random samples of size n_i from the i-th stratum. Let C be the prescribed total cost, and let us assume $C = c_1 n_1 + c_2 n_2 + ... + c_k n_k$.

Let us here consider the simultaneous estimation of h linear combinations of these stratum means, viz.,

$$L_i(\bar{x}) = \sum_{j=1}^{k} a_{ij} \bar{x}_j \qquad (i = 1, 2, ..., h) \qquad \dots \quad (2.03)$$

where (a_{ij}) is a given matrix of constants. The first type of information in the sense of Mahalanobis (1952) and the enumerative study in the sense of Deming (1950) may be observed to correspond to the case when $h = 1$ and $a_{ij} = N_j$ $(j = 1, 2, ..., k)$. The case $h > 1$ is a more complicated one. As an estimator for $L_i(\bar{x})$ we may make use of $L_i(\bar{\bar{x}})$ where $\bar{\bar{x}} = (\bar{\bar{x}}_1, \bar{\bar{x}}_2, ..., \bar{\bar{x}}_k)$, $\bar{\bar{x}}_i$ being the sample mean in the i-th stratum. The variance of this estimate is easily seen to be

$$V(L_i(\bar{\bar{x}})) = \sum_{j=1}^{k} a_{ij}^2 \frac{N_j - n_j}{N_j n_j} \sigma_j^2 . \qquad \dots \quad (2.04)$$

When we are concerned with simultaneous estimation, we must lay down certain rules by which relative weights concerning these variances should be determined. The simplest case will be that in which a system of constants $\{\lambda_i\}$ is prescribed and the error variance concerning our problem is defined simply by

$$V = \sum_{i=1}^{h} \lambda_i V\{L_i(\bar{x})\} . \qquad \text{... (2.05)}$$

Theorem 1.1 : *The optimum allocation of sample sizes* $\{n_i\}$ *to each stratum which minimizes* (2.05) *is given by*

$$n_i = CB_i\sigma_i c_i^{-1}\Big\{\sum_{j=1}^{k} B_j\sigma_j c_j^{\,1}\Big\}^{-1} \qquad \text{... (2.06)}$$

where

$$B_i = \sum_{l=1}^{h} \lambda_l^2 a_{li}^2 . \qquad \text{... (2.07)}$$

The proof can be readily obtained by the method of Lagrange's undetermined multipliers.

Another case is that in which a certain positive quadratic form is defined thus :

$$Q(\bar{x}) \equiv \sum_{i=1}^{h}\sum_{j=1}^{h} \lambda_{ij}\{L_i(\bar{x})-L_i(\bar{\bar{x}})\}\{L_j(\bar{x})-L_j(\bar{\bar{x}})\} . \qquad \text{... (2.08)}$$

Here the optimum allocation should minimise the mean value of $Q(\bar{x})$. This case can be treated in a way similar to the previous one, because we have

$$E\{Q(\bar{x})\} = \sum_{p=1}^{k}\sum_{i=1}^{h}\sum_{j=1}^{h} \lambda_{ij}a_{ip}a_{jp}V(\bar{x}_p)$$

$$= \sum_{p=1}^{k} d_p^2\, \frac{N_p-n_p}{N_pn_p}\sigma_p^2, \text{ say.} \qquad \text{... (2.09)}$$

Example 1.1: Let us consider the case when we are concerned with the following k parameters : the general mean $\bar{x}$ and the $k-1$ successive differences of stratum-means $\bar{x}_i-\bar{x}_{i-1}$ $(i = 2, ..., k)$ with weights on $\bar{x}$ and $\bar{x}_i-\bar{x}_{i-1}$ being μ_1 and $\mu_i(i = 2, ..., k)$ respectively. Then the expression for B's in (2.07) are given by

$$B_1 = \mu_1\left(\frac{N_1}{N}\right)^2+\mu_2,$$

$$B_i = \mu_1\left(\frac{N_i}{N}\right)^2+\mu_i+\mu_{i+1} \qquad (i = 2, ..., k-1),$$

$$B_k = \mu_1\left(\frac{N_k}{N}\right)^2+\mu_k .$$

Example 1.2 : Let us consider a two-way classification of the population into the strata $\{\Pi_{ij}\}$ $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$, with respect to two characters. Let $\bar{x}_{ij}$ be the stratum mean of the (i, j)-th stratum.

Then we have

$$\bar{x}_{i\cdot} = \sum_{j=1}^{b} \frac{N_{ij}\bar{x}_{ij}}{N_{i\cdot}} \qquad (i = 1, 2, \ldots, a), \qquad \ldots \quad (2.10)$$

$$\bar{x}_{\cdot j} = \sum_{i=1}^{a} \frac{N_{ij}\bar{x}_{ij}}{N_{\cdot j}} \qquad (j = 1, 2, \ldots, b). \qquad \ldots \quad (2.11)$$

In some situations we may have to obtain the estimated values for $\{\bar{x}_{i\cdot}\}$ and $\{\bar{x}_{\cdot j}\}$ the weights on $\bar{x}, \bar{x}_{i\cdot}$, and $\bar{x}_{\cdot j}$ being λ, μ_i and ν_j respectively. In such circumstances, we have

$$B_{ij}^2 = \lambda\left(\frac{N_{ij}}{N}\right)^2 + \mu_i\left(\frac{N_{ij}}{N_{i\cdot}}\right)^2 + \nu_j\left(\frac{N_{ij}}{N_{\cdot j}}\right)^2, \qquad \ldots \quad (2.12)$$

where B_{ij} occurs in the expression of the sample size n_{ij} similar to (2.06).

We shall give examples on Theorem 1.1.

Example 1.3 : Let us consider a demand and supply plan concerning the trees in a forest. In such a situation, we must not only estimate the total volume of timber stands in the whole forest, but we must also estimate with greater accuracy the volume of the parts of our forest which are to be cut in a certain cutting period. If we stratify our whole forest according to some intervals of ages of say, ten years, then our strata belonging to or nearly belonging to the cutting period become important and their total volume must be estimated with higher accuracy than the volume of parts not belonging to these strata. There are now two linear combinations of stratum means. The first is the total sum, and the other is the sum of the strata belonging to the cutting period.

Consequently we may have in (2.03), $h = 2$ and $a_{1j} = N_j$ $(j = 1, 2, \ldots, k)$ while $a_{2j} = N_j$ if Π_j belongs to cutting period, and $=0$ otherwise. For the sake of brevity let $\{\Pi_j\}$ $(j = 1, 2, \ldots, l)$ be the set of strata belonging to the cutting period. Then we have

$$L_1(\bar{x}) = \sum_{j=1}^{k} N_j\bar{x}_j, \qquad \ldots \quad (2.13)$$

$$L_2(\bar{x}) = \sum_{j=1}^{l} N_j\bar{x}_j. \qquad \ldots \quad (2.14)$$

What we must minimize will then be

$$\lambda_1 V\{L_1(\bar{x})\} + \lambda_2 V\{L_2(\bar{x})\} \qquad \ldots \quad (2.15)$$

and we shall obtain the optimum allocation after assigning weights λ_1 and λ_2 which can only be determined from practical considerations. In forest management the problem is to estimate the whole volume of stands in our forest and then to find some areas which may be expected to supply us a certain assigned fraction of the estimated total volume. From the practical point of view there will be a certain order according to which each stand may be cut down. For the moment, let it be assumed that this is the order in which the strata are numbered.

Then our problem is concerned with inferring the value of l from the following condition :

$$\sum_{i=1}^{l-1} T_i < pT \leqslant \sum_{i=1}^{l} T_i \qquad \ldots \quad (2.16)$$

where T and T_i mean the volumes of the total and of the i-th stratum respectively, p being assumed to be given. We must of course replace T and $\{T_i\}$ by their estimated values respectively. Then it becomes clear that the parameter to be inferred is really the number l, whose estimated value $\hat{l}$ should be considered as a stochastic variable.

To find out the distribution of $\hat{l}$ and to give an optimum stratified random sampling procedure in which $\hat{l}$ should have the least variance is a rather difficult mathematical problem. The following one is suggested as an approximate procedure which may be useful for practical purposes.

(1) Suppose there is some *a priori* knowledge that l belongs to the interval $l_1 < l < l_2$.

(2) Let us consider $h = l_2 - l_1$ linear combinations

$$L_i(\bar{x}) = (1-p) \sum_{j=1}^{l_1+i-1} T_j - p\sum_{r=l_1+i}^{k} T_r \qquad \ldots \quad (2.17)$$

for $i = 1, 2, \ldots, h$.

(3) Let us design a stratified random sampling procedure in such a way as to minimise

$$\sum_{i=1}^{h} V\{L_i(\bar{x})\} \qquad \ldots \quad (2.18)$$

where $$L_i(\bar{x}) = (1-p) \sum_{j=1}^{l_1+i-1} N_j\bar{x}_j - p\sum_{r=l_1+i}^{k} N_r\bar{x}_r. \qquad \ldots \quad (2.19)$$

(4) Let us plot the values $(i, L_i(\bar{x}))$ in a graph, and let us determine the abscissa l for which the first positive value of $L_i(\bar{x})$ is attained.

Example 1.4 : There are many occasions when stratification is mainly adopted either for the purpose of increasing the precision of our estimates and/or for administrative convenience in carrying out large-scale surveys. In such situations it will be found frequently that some of our strata are not large enough to give us any reliable estimates concerning the respective stratum means. But it may not be possible to group our strata in some districts so as to avoid the existence of overlapping groups. We think it better to make use of grouping of strata in which some overlapping may be admitted, so far as group totals or group means are our objects of statistical estimation. The estimates of the group means are then not independent, but it does not matter, because their covariances can also be easily calculated from the sample design. The present author had made use of such grouping when he was connected with the sample survey of labour forces in Fukuoka Prefecture (Japan).

3. Simultaneous estimation of linear combinations of within-stratum variances or standard deviations in stratified random sampling

Let $\{x_{jl}\}$ $(l = 1, 2, \ldots, n; j = 1, 2, \ldots, h)$ be a sample from the i-th stratum and let us define for each stratum sample means $\bar{x}_{j\cdot}$ and sample variances s_j^2 given by

$$s_j^2 = \frac{1}{n_j-1} \sum_{l=1}^{n_j} (x_{jl}-\bar{x}_{j\cdot})^2, \qquad (j = 1, 2, \ldots, h). \qquad \ldots \quad (3.01)$$

In addition to the assumptions in § 2, let us assume the following :—

(1°) Let there be a set of functions $\{f_j(\sigma_j^2)\}$, $\{g_j(s_j^2, n_j)\}$ which satisfies the following conditions:

$$E\{g_j(s_j^2, n_j)\} = f_j(\sigma_j^2), \qquad \ldots \quad (3.02)$$

$$V\{g_j(s_j^2, n_j)\} = \phi_j(n_j, \sigma_j^2, \theta_j), \qquad \ldots \quad (3.03)$$

where ϕ_j is a function depending upon the sample size n_j, the within-stratum variance σ_j^2, and the vector θ_j of a certain set of the population parameters.

(2°) Let our object be the estimation of a certain set of linear combinations

$$L_i(\sigma^2) = \sum_{j=1}^{k} b_{ij} f_j(\sigma_j^2) \qquad \ldots \quad (3.04)$$

for $i = 1, 2, \ldots, h$ and let our estimators be defined by

$$L_i(s^2) = \sum_{j=1}^{k} b_{ij} g_j(s_j^2, n_j) \qquad \ldots \quad (3.05)$$

for $i = 1, 2, \ldots, h$ respectively.

(3°) Let there be given a set of weights $\{\lambda_i\}$ such that our consolidated variance may be defined

$$V(L) = \sum_{i=1}^{h} \lambda_i V\{L_i(s^2)\}. \qquad \ldots \quad (3.06)$$

Then our problem of the optimum allocation of the sample sizes $\{n_j\}$ under a given total cost will be reduced to the one in which we should minimize (3.06).

The solution of this minimizing problem belongs to elementary calculus but there are two kinds of difficulties to be pointed out. The first one is that $\phi_j(n_j, \sigma_j^2, \theta_j)$ may contain the unknown parameter σ_j^2 which at least formally lead to a reasoning in a circle. The second one is that $\phi_j(n_j, \sigma_j^2, \theta_j)$ may contain the higher moments of the j-th stratum (i.e. beyond the first and the second moments), which usually show more fluctuations than the first two moments. In spite of these apparent difficulties which prevent us from following arguments similar to those of § 2, it may be possible in some cases to overcome these difficulties. Indeed, one of the merits of successive or sequential process of statistical inference is that we can deal with such circular reasoning. Further there may be cases in which $\sum_{i=1}^{h} \lambda_i b_{ij}^2 = B_j \sigma_j^{-2}$ and B_j are independent of σ_j, and in which we may assume *a priori* some skewness and/or kurtosis.

Example 1.5 : Let us consider the case when the size of each stratum is so large that the approximation by normal theory may yield the probability density function of s_j/σ_j as

$$p\left(\frac{s_j}{\sigma_j}\right) = \frac{(n_j-1)^{\frac{n_j-1}{2}}}{2^{\frac{n_j-3}{2}}\Gamma\left(\frac{n_j-1}{2}\right)}\left(\frac{s_j}{\sigma_j}\right)^{n_j-2} \exp\left\{-\frac{(n_j-1)s_j^2}{2\sigma_j^2}\right\}. \qquad \dots \quad (3.07)$$

Thus the statistic defined by

$$\hat{s}_j = \frac{(n_j-1)^{\frac{1}{2}}\,\Gamma\left(\frac{n_j-1}{2}\right)}{2^{\frac{1}{2}}\,\Gamma\left(\frac{n_j}{2}\right)} s_j \qquad \dots \quad (3.08)$$

gives us an unbiased estimate of σ_j with the variance

$$V(\hat{s}_j) = \left\{\frac{(n_j-1)\,\Gamma^2\left(\frac{n_j-1}{2}\right)}{2\,\Gamma^2\left(\frac{n_j}{2}\right)} - 1\right\}\sigma_j^2. \qquad \dots \quad (3.09)$$

Furthermore let our weights be such that, for $j = 1, 2, \dots, k$,

$$\sum_{i=1}^{h} \lambda_i b_{ij}^2 = B_j \sigma_j^2, \qquad \dots \quad (3.10)$$

where B_j are independent of σ_j^2. Then our problem is to minimize

$$\sum^{k} B_j \left\{\frac{(n_j-1)\,\Gamma^2\left(\frac{n_j-1}{2}\right)}{2\,\Gamma^2\left(\frac{n_j}{2}\right)} - 1\right\} \qquad \dots \quad (3.11)$$

under the condition $\sum_{j=1}^{k} c_j n_j = C$, c_j being the cost per unit in the j-th stratum. This problem can be solved at least approximately.

4. Estimation of a set of linear combinations of multiple stratum-means

Let our whole population Π be divided into the sum of hk mutually exclusive subpopulations $\{\Pi_{ij}\}$ $(i = 1, 2, ..., h; j = 1, 2, ..., k)$. Let us denote by $\{n_{ijl}\}$ $(i = 1, 2, ..., h; j = 1, 2, ..., k; l = 1, 2, ..., N_{ij})$, the values belonging to each Π_{ij}, and $\bar{x}_{ij}$ and σ_{ij}^2 the (i, j)th stratum-mean and the (i, j)th within-stratum variances respectively. Furthermore we shall consider the stratification systems $\{\Pi_{i\cdot}\}$ $(i = 1, 2, ..., h)$ and $\{\Pi_{\cdot j}\}$ $(j = 1, 2, ..., k)$, which are defined by

$$\Pi_{i\cdot} = \Pi_{i1}+\Pi_{i2}+...+\Pi_{ik}, \quad ... \quad (4.01)$$

$$\Pi_{\cdot j} = \Pi_{1j}+\Pi_{2j}+...+\Pi_{hj}, \quad ... \quad (4.02)$$

respectively and whose stratum-means and within-stratum variances will be defined by

$$\bar{x}_{i\cdot} = N_{i\cdot}^{-1} \sum_{j=1}^{k} N_{ij}\bar{x}_{ij}, \quad ... \quad (4.03)$$

$$\bar{x}_{\cdot j} = N_{\cdot j}^{-1} \sum_{i=1}^{h} N_{ij}\bar{x}_{ij}, \quad ... \quad (4.04)$$

and

$$\sigma_{i\cdot}^2 = (N_{i\cdot}-1)^{-1} \sum_{j=1}^{k} N_{ij}(\bar{x}_{ij}-\bar{x}_{i\cdot})^2+(N_{i\cdot}-1)^{-1} \sum_{j=1}^{k} (N_{ij}-1)\sigma_{ij}^2 \quad ... \quad (4.05)$$

$$\sigma_{\cdot j}^2 = (N_{\cdot j}-1)^{-1} \sum_{i=1}^{h} N_{ij}(\bar{x}_{ij}-\bar{x}_{\cdot j})^2+(N_{\cdot j}-1)^{-1} \sum_{j=1}^{h} (N_{ij}-1)\sigma_{ij}^2 \quad ... \quad (4.06)$$

where we have put

$$N_{i\cdot} = \sum_{j=1}^{k} N_{ij}, \qquad N_{\cdot j} = \sum_{i=1}^{h} N_{ij}. \quad ... \quad (4.07)$$

Let us now consider the situations in which we have introduced the one-way stratification, that is to say, the i-th system and in which we have allocated random samples of certain sizes to the i-system strata. There are many occasions in which our aims of estimation are not only for a set of linear combinations of the stratum-means of the i-system but also for that corresponding to the j-system. Thus a life insurance company will make sample surveys of policy holders and make use of stratification by the sums assured. After they have obtained the data by stratified random sampling, they find it necessary to analyse the data by age-classes and by districts where these persons have insured. In sequential sample surveys we are able to make use of the information, which gradually accumulates, to improve the sampling design of subsequent surveys. The drastic change of one system of stratification into another is not rare. The problem will be discussed in greater detail on another occasion, but here our object will be to show some effect of coordinating our data from another system of stratification different from the one used in actual sampling. Some aspects of changing the system of stratification were discussed by B. Ghosh (1947) to show that biases which would be introduced are too large to advocate use of estimates based on a change of stratification. But his arguments are based upon the assumption that the sizes $\{N_{ij}\}$ should be estimated from the present samples themselves. If, however, this is not so, there is some advantage in changing the stratification.

(a) Let us consider a set of g linear combinations of the stratum-means $\{\bar{x}_{ij}\}$

$$L_p(\bar{x}) = \sum_{i=1}^{h} \sum_{j=1}^{k} b_{pij}\bar{x}_{ij} \qquad (p = 1, 2, \ldots, g). \qquad \ldots \quad (4.08)$$

where $\{b_{pij}\}$ are assigned constants. and also a set of their respective unbiased estimates

$$L_p(\bar{x}) = \sum_{i=1}^{h} \sum_{j=1}^{k} b_{pij}\bar{x}_{ij} \qquad (p = 1, 2, \ldots, g). \qquad \ldots \quad (4.09)$$

Let us denote by $\{\dot{n}_{pq}\}$ $(q = 1, 2, \ldots, k)$ the sizes of the sample which belong to the (p, q) stratum for each sample of the p-th stratum of the i-system. These k non-negative integers may be considered as stochastic variables under the restriction that $\dot{n}_{i1}+\dot{n}_{i2}+\ldots+\dot{n}_{ik} = n_i$, for $i = 1, 2, \ldots, h$. The unbiasedness of $L_p(\bar{x})$ as an estimator of $L_p(\bar{x})$ is evident and its variance may be readily shown to be

$$V\{L_p(x)\} = \sum_{i=1}^{h} \sum_{j=1}^{k} a_{pij}^2 \sigma_{ij}^2 E\left(\frac{1}{n_{ij}}\right) - \sum_{i=1}^{h} \sum_{j=1}^{k} a_{pij}^2 \sigma_{ij}^2 \frac{1}{N_{ij}} \qquad \ldots \quad (4.10)$$

In view of Stephan (1945) we have now

$$E\left\{\frac{1}{n_{ij}}\right\} = \frac{N_i}{n_i N_{ij}} - \frac{1}{n_i^2}\left[\left(\frac{N_i}{N_{ij}}\right) - \left(\frac{N_i}{N_{ij}}\right)^2\right] \qquad \ldots \quad (4.11)$$

The optimum allocation may sometimes be rather different from the ordinary one, which is one of the simplified cases. For instance, let us consider the case when $p = 1$ and when (4.10) may be written approximately

$$V\{L(x)\} = \sum_{i=1}^{h} \frac{N_i}{n_i} \sum_{j=1}^{k} \frac{a_{ij}^2}{N_{ij}} \sigma_{ij}^2 \equiv \sum_{i=1}^{h} \frac{A_i^2}{n_i}, \text{ say,} \qquad \ldots \quad (4.12)$$

which shows that the optimum allocation under an assigned total cost C shall be given by

$$n_i = CA_i c_i^{-\frac{1}{2}} \left(\sum_{i=1}^{h} A_i c_i^{\frac{1}{2}}\right)^{-1}. \qquad \ldots \quad (4.13)$$

(b) Next let us consider a set of linear combinations of the (i, j) stratum-sums

$$H(\bar{x}) = \sum_{j=1}^{k} \frac{B_j}{N_{\cdot j}} \sum_{i=1}^{h} N_{ij}\bar{x}_{ij} \qquad \ldots \quad (4.14)$$

and let its estimator be defined by

$$\hat{H}(x) = \sum_{j=1}^{k} \frac{B_j}{n_{\cdot j}} \sum_{i=1}^{h} n_{ij} x_{ij}. \quad \ldots \quad (4.15)$$

Then its mean value and variance may be readily observed to be equal to

$$E\{\hat{H}(\bar{x})\} = \sum_{j=1}^{k} B_j \sum_{i=1}^{h} E\left\{\frac{n_{ij}}{n_{1j}+\ldots+n_{hj}}\right\} \bar{x}_{ij} \quad \ldots \quad (4.16)$$

and

$$V\{\hat{H}(\bar{x})\} = E\left[\left\{\sum_{j=1}^{k} B_j \sum_{i=1}^{h} \left(\frac{n_{ij}}{n_{1j}+\ldots+n_{hj}} - p_{ij}\right)\bar{x}_j\right\}^2\right]$$

$$+E\left[\left\{\sum_{j=1}^{k} B_j \sum_{i=1}^{h} \frac{n_{ij}}{n_{1j}+\ldots+n_{hj}} \left(\bar{x}_{ij}-\bar{x}_{ij}\right)\right\}^2\right]. \quad \ldots \quad (4.17)$$

5. Problems concerning maximum values in sample surveys

In order that we may discuss these problems in detail, it seems convenient to introduce certain mathematical models different from those of a finite population lacking any notion of dimensional arrangement. Thus sampling from processes depending upon one or several continuous parameters may be naturally taken into consideration.

(i) Let $f(t)$ be a single-valued continuous function defined over a certain interval $a \leq t \leq b$. Let us draw a set of n independent points $\{t_i\}$ $(i = 1, \ldots, n)$ according to a certain sampling distribution. Then it can be readily seen that the stochastic variable defined by

$$m_n = \max\{f(t_1), \ldots, f(t_n)\} \quad \ldots \quad (5.01)$$

is an estimator for the maximum of the function

$$\mu = \max_{a \leq t \leq b} f(t).$$

Its probability density function may be readily seen to be

$$\Pr\{u < m_n < u+du\} = n\left\{\int_{\nu}^{u} \Phi(t)dt\right\}^{n-1}\Phi(u)du \quad \ldots \quad (5.02)$$

where $\Phi(u)$ denotes the distribution of the stochastic variable $u = f(X)$, X being a stochastic variable which is distributed according to a certain assigned probability law and ν is the minimum of the function $f(X)$ in the interval.

(ii) Now let us consider a stratified random sampling procedure such that an i-th stratum Π_i corresponds to the sub-interval $a_i \leq t \leq a_{i+1}$ where we assume $a_0 = a < a_1 < \ldots < a_n < a_{n+1} = b$, and let a random sample of $\{t_{ij}\}$ $(j = 1, 2, \ldots, n_i)$ be drawn independently from each stratum according to their respective probability distributions. Let us define

$$m_i = \max\{(f(t_{i1}), f(t_{i2}) \ldots, f(t_{in_i})\}, \quad \ldots \quad (5.03)$$

$$m = \max(m_1, m_2, \ldots, m_k). \quad \ldots \quad (5.04)$$

The statistic m may be recognised as an estimator of μ and indeed we have

$$\Pr\{u < m < u+du\} = \prod_{i=1}^{k} F_i(u) \sum_{j=1}^{k} \frac{F'_j(u)}{F_j(u)} du \quad \ldots \quad (5.05)$$

in which

$$F'_i(u) = n_i\left\{\int_{a_i}^{u} \Phi_i(t)dt\right\}^{n_i-1}\Phi_i(u), \quad \ldots \quad (5.06)$$

$\Phi_i(u)$ being the density function associated with $f(t)$ in the interval $a_i < t < a_{i+1}$.

Here comes also the problem of the optimum allocation to each stratum. The problem of stratification will be discussed later. In this Part we intend to discuss the question of optimum allocation under certain simplified situations.

(iii) There is another method of seeking the maximum by making use of sampling. Under the assumptions in (i) we may define the sum such that

$$m_p(f) \equiv \frac{\sum_{i=1}^{n} f(t_i)^{p+1}}{\sum_{i=1}^{n} f(t_i)^p} \quad \ldots \quad (5.07)$$

where p is an assigned large positive number. The merits of such statistics as estimators of the maximum of the function lie in the fact that for each fixed set of $\{t_i\}$ and n we shall have

$$\lim_{p\to\infty} m_p(f) = \max\{f(t_1), \ldots, f(t_n)\}. \quad \ldots \quad (5.08)$$

Thus it is our main idea that corresponding to each size of p we should be able to choose a sample size n such that both of the following conditions may be nearly satisfied.

(1°) In some sense $m_p(f)$ gives a sufficiently good approximate value of

$$I_p(f) = \frac{\int_a^b f(t)^{p+1}dt}{\int_a^b f(t)^p dt}. \quad \ldots \quad (5.09)$$

(2°) The power p in (5.09) is so large that we may expect that $I_p(f)$ gives us an approximate value of the maximum of the function.

(iv) The generalisation of the method proposed in (iii) may be easily performed in the case of stratified random sampling. That is to say, we may take into consideration

$$m_p(f) = \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} f(t_{ij})^{p+1} \right\} \left\{ \sum_{i=1}^{k} \sum_{j=1}^{n_i} f(t_{ij})^{p} \right\}^{-1} \quad \ldots \quad (5.10)$$

There will be some combination of these proposed procedures.

6. Some aspects of mapping problems concerning stratified random sampling designs

In order to discuss the mapping problems in some detail it may be better to introduce certain processes depending upon continuous parameters, as we shall do later. Nevertheless here we shall first discuss some aspects of mapping problems concerning stratified random sampling without any assumption of continuous parameters. The method which will be developed in this section may be considered as an application of that adopted in a previous paper by the author [1953(c)]. The problem treated in this section may also be regarded as a smoothing process namely the substitution of the detailed map by the average values within strata consisting of the adjacent elements.

For the sake of brevity, let us consider the case when our parent population shows one-dimensional structure concerning x-values. We may assume now that all elements of our population are arranged on the straight line t-axis in the same order as the numbering of the strata. Let us denote the central position of the j-th stratum by t_j. Our situation may be described as such that our function $f(t)$ takes the value $f(t_j) = \bar{x}_j$ at the point t_j with the weight $w_j = N_j/N$ $(j = 1, 2, \ldots, h)$. The theory of orthogonal functions will show us that there is at least one system of functions $\{\phi_u(t_j; k, w)\}$ $(u = 0, 1, \ldots, k-1;\ j = 1, 2, \ldots, k)$ satisfying the following conditions:

$$\sum_{j=1}^{k} w_j \phi_\mu(t_j; k, w) \phi_\nu(t_j; k, w) = \delta_{\mu\nu} \quad \ldots \quad (6.01)$$

where $\delta_{\mu\nu}$ denotes Kronecker's delta.

Let us now consider for $i = 1, 2, \ldots, k$

$$\bar{g}_m(t_i) \equiv \sum_{\mu=0}^{m} \left\{ \sum_{j=0}^{k} w_j \bar{x}_j \phi_\mu(t_j; k, w) \right\} \phi_\mu(t_i; k, w) \quad \ldots \quad (6.02)$$

where m is to be suitably chosen according to our convenience so that $0 \leqslant m \leqslant k-1$. What we want to assert here is that in some of the designs of stratified random sampling our object of obtaining a smooth picture of the parent population may be satisfied by giving estimated values to (6.02). The estimators to (6.02) will be defined by

$$\bar{g}_m(t_i) = \sum_{k}^{m} \left\{ \sum_{j=1}^{k} w_j \bar{x}_j \phi_\mu(t_j; k, w) \right\} \phi_\mu(t_i; k, w) \quad \ldots \quad (6.03)$$

since we have assumed that $\{w_j\}$ and $\{t_j\}$ are already known to us. Each $\bar{g}_m(t_i)$ is obviously an unbiased estimator for each value of t_i. From the standpoint of fitting curves, however, the circumstances are more complicated. Indeed unless $m = k-1$, $\bar{g}_m(t_i)$ may be a biased estimator for $\bar{x}_i$ $(i = 1, 2, \ldots, k)$.

It will be natural to introduce the norm of $\bar{g}_m(t)-f(t)$ and that of $\bar{g}_m(t)-\tilde{g}_m(t)$ defined as follows:

$$\| \bar{g}_m - f \| \equiv \sum_{i=1}^{k} w_i(\bar{g}_m(t_i)-f(t_i))^2, \quad \ldots \quad (6.04)$$

$$\| \bar{g}_m - \tilde{g}_m \| \equiv \sum_{i=1}^{k} w_i(\bar{g}_m(t_i)-\tilde{g}_m(t_i))^2. \quad \ldots \quad (6.05)$$

For the sake of convenience, let us introduce

$$L(t_i, t_j, m; w) \equiv \sum_{s=0}^{m} \phi_s(t_i; k, w)\phi_s(t_j; k, w). \quad \ldots \quad (6.06)$$

Now let us introduce the following definitions :

The variance and the mean square error of the empirical function $\bar{g}_m$ are defined by

$$M.S.E.\{\bar{g}_m\} \equiv E\{ \| \bar{g}_m - f \| \}, \quad \ldots \quad (6.07)$$

$$V\{\bar{g}_m\} \equiv E\{ \| \bar{g}_m - \tilde{g}_m \| \}. \quad \ldots \quad (6.08)$$

We can readily observe

$$M.S.E.\{\bar{g}_m\} = V\{\bar{g}_m\} + \| \bar{g}_m - f \| \quad \ldots \quad (6.09)$$

and

$$V\{g_m\} = \sum_{j=1}^{k} w_j V\{\bar{x}_j\} \sum_{i=1}^{k} w_i\{L(t_i, t_j; m, w)\}^2$$

$$= \sum_{j=1}^{k} G_j^{(m)} \frac{N_j-n_j}{N_j n_j} \sigma_j^2 \quad \ldots \quad (6.10)$$

where

$$G_j^{(m)} = w_j \sum_{i=1}^{k} w_i\{L(t_i, t_j; m, w)\}^2. \quad \ldots \quad (6.11)$$

So far as we do not know our population values, we shall rather choose $m = k-1$. Nevertheless when k is reasonably large, this may mean a great amount of labour. Unless our function $f(t_i)$ $(i = 1, 2, \ldots, k)$ is rather complicated it may often be better to choose more simplified curves, that is, to select a small m. It is evident that unless m is equal to $k-1$, there is a danger of using a biased estimate. Nevertheless it may be noted that the smaller the m the smaller will be the variance $V(\bar{y}_m)$.

Here is the fundamental problem of choosing the degree m, and the solution to this problem seems to be given only after discussing our whole situation from the standpoint of successive process of statistical inference. The solution will be obtained easily if we endeavour to minimise the variance $V(\bar{y}_m)$ for a certain chosen value of m and for an assigned total cost C. Indeed for an assigned approximate degree of m and an assigned total cost C the allocation which will minimise the variance $V(\bar{y}_m)$ is determined by

$$n_j = CG_j^{(m)\frac{1}{2}}\sigma_j c_j^{-\frac{1}{2}}\{\sum_{i=1}^{k} G_i^{(m)\frac{1}{2}}\sigma_i c_i^{\frac{1}{2}}\}^{-1} \quad \ldots \quad (6.12)$$

for $j = 1, 2, \ldots, k$.

7. Optimum allocation under generalised variance and cost functions introduced by Mahalanobis

In this section, our topics are concerned with a stratified random sample survey in which sample units are grids in the terminology of Mahalanobis (1944) of a constant size a. Let A_i be the area of the i-th zone, that is, the i-th stratum, and let w_i be the density or number of grids per unit area in the i-th zone. The following forms of equations were advocated by Mahalanobis for the total cost and the variance V of the estimated total

$$V = \sum_{i=1}^{k} A_i b_i / w_i a_i^g, \quad \ldots \quad (7.01)$$

$$T = \sum_{i=1}^{k} A_i(c_0 + c_1 w_i + c_2 w_i a_i + c_3 w_i^{\frac{1}{2}}), \quad \ldots \quad (7.02)$$

where c_0, c_1, c_2 and c_3 are cost parameters supposed to be constant over the different zones, b_k is a zonal constant, and g is a pooled constant. To get the optimum size-density distribution at any given cost T, Mahalanobis gave a graphical method of solution which is practical and he also discussed the uncertainty in the estimated optimum size, density and variance due to errors in the parameters obtained by graduation in the graphs. From our general viewpoints of the principles of designing sample surveys, we think it adequate to bring these generalised cost and variance functions to the realm of simplified schemes where certain fundamental principles will be relevantly applied. To this purpose we shall devote this section and also §4 of Part II.

Let us now start with the more generalised functional forms

$$V = \sum_{i=1}^{k} A_i b_i \psi_i(a_i, w_i), \qquad \dots \quad (7.03)$$

$$T = \sum_{i=1}^{k} A_i \phi_i(a_i, w_i). \qquad \dots \quad (7.04)$$

Our aim is to find a certain set of sufficient conditions for the functional forms in which some of our principles are exactly or nearly valid. Let λ be an undetermined multiplier and after differentiating $V+\lambda T$ with respect to a_i and w_i, we have the following relation which the optimum solution should satisfy

$$\frac{\delta(\psi_i, \phi_i)}{\delta(a_i, w_i)} = \begin{vmatrix} \frac{\delta\psi_i}{\delta a_i} & \frac{\delta\phi_i}{\delta a_i} \\ \frac{\delta\psi_i}{\delta w_i} & \frac{\delta\phi_i}{\delta w_i} \end{vmatrix} = 0 \qquad \dots \quad (7.05)$$

for $i=1, 2, \dots, k$. It is to be noted that (7.05) will determine certain functional relations between the optimum values of a_i and w_i, say

$$a_i = l_i(w_i) \qquad \dots \quad (7.06)$$

and that (7.05) and hence (7.06) are independent of $\{A_i\}$ and $\{b_i\}$. This may be said to be the principle of localisation. Now putting (7.06) into ψ_i and ϕ_i, let us introduce a new parameter n_i by means of which we shall have

$$\psi_i(l_i(w_i), w_i) = 1/n_i, \qquad \dots \quad (7.07)$$

$$\phi_i(l_i(w_i), w_i) = h_i(n_i). \qquad \dots \quad (7.08)$$

Some conditions have to be assumed in securing the possibilities of introducing new parameters such as certain monotonicities of our functions. Anyhow under our assumptions the optimum solution must satisfy

$$n_i^2 \frac{dh_i(n_i)}{dn_i} = \frac{b_i}{\lambda} \qquad \dots \quad (7.09)$$

for $i = 1, 2, \dots, k$.

There is an important condition under which we may be able to establish some simple formulae for the optimum allocation, whereas without this condition our analysis would be too complicated to be of any practical use. The condition is that there should be a constant ρ common to all k strata and a set of positive constants $\{d_i\}$ for $i = 1, 2, \ldots, k$, such that the following equalities hold true

$$\left(n_i^2 \frac{dh_i(n_i)}{dn_i}\right)^\rho = d_i h_i(n_i) \qquad \ldots \quad (7.10)$$

at least approximately for all $i = 1, 2, \ldots, k$.

Indeed under this condition we can determine the undetermined multiplier λ and the possible optimum allocation $\{n_i\}$ as follows :

$$\lambda = \left\{T^{-1} \sum_{i=1}^{k} A_i b_i^\rho d_i^{-1}\right\}^{1/\rho}, \qquad \ldots \quad (7.11)$$

$$n_i = h_i^{-1}\left(\frac{b_i^\rho d_i^{-1} T}{\sum_{i=1}^{k} A_i b_i^\rho d_i^{-1}}\right), \qquad \ldots \quad (7.12)$$

for $i = 1, 2, \ldots, k$. This will give us at least a stationary value of V :

$$V = \sum_{i=1}^{k} \frac{A_i b_i}{h_i^{-1}\left(\frac{b_i^\rho d_i^{-1} T}{\sum_{j=1}^{k} A_j b_j^\rho d_j^{-1}}\right)} . \qquad \ldots \quad (7.13)$$

This can be examined in detail for each special case as to whether it is a minimum or not.

It is to be noted that this form of (7.13) can be transformed to a more convenient one by expanding the inverse function $h^{-1}(y)$ into the power series of y. The solution of the differential equation under the initial condition of being zero at the origin gives us

$$h_i^{-1}(y) = \frac{1-\rho}{\rho} \, d_i^{\frac{1}{\rho}} \, y^{\frac{1-\rho}{\rho}} = e_i \, y^{\frac{1-\rho}{\rho}} \ , \quad \text{say}, \qquad \ldots \quad (7.14)$$

which may be recognised as an approximate expression of the true function, because we have assumed the approximate validity of (7.10).

Putting (7.14) into (7.13) we get

$$V = \frac{\left\{\sum_{i=1}^{k} A_i \, (b_i e_i^{-1})^\rho\right\}^{\frac{1}{\rho}}}{T^{\frac{1-\rho}{\rho}}}. \qquad \ldots \quad (7.15)$$

This relation corresponds to the case which will be discussed in §3 of Part II. This correspondence is in fact a formal one, but it is important for further development to establish this formal unity through which our general theory may be evolved. To observe the meaning of our constant ρ in relation to the results of section 3 of Part II, let us put $\rho = p/(1+p)$. From the general consideration there is the restriction that $0 < p < 1$ which leads to $0 < \rho < \frac{1}{2}$.

Now let us apply our transformation to the case of Mahalanobis' functions (7.01) and (7.02). Here (7.05) and hence (7.06) gives us, as Mahalanobis remarked, the local relations

$$a_i = l_i(w_i) = (c_1 + 2c_2 w_i) g/(1-g) c_2 \qquad \dots \quad (7.16)$$

for $i = 1, 2, \dots, k$, which yield a new parameter n_i such that

$$n_i = \frac{1}{\psi_i(l_i(w_i), w_i)} = \left(\frac{g}{(1-g)c_2}\right)^{\rho} w_i (c_1 + 2c_2 w_i)^{\rho} \qquad \dots \quad (7.17)$$

$$h_i(n_i) = \phi_i(l_i(w_i), w_i) = \frac{1}{1-g} w_i \{c_1 + c_2(1+g) w_i\}. \qquad \dots \quad (7.18)$$

These two relations now lead us to

$$n_i^2 \frac{dh_i(n_i)}{dn_i} = \left(\frac{g}{(1-g)c_2}\right)^{\rho} \frac{1}{1-g} w_i^2 (c_1 + 2c_2 w_i)^{\rho+1}. \qquad \dots \quad (7.19)$$

Consequently, using the approximation.

$$(c_1 + 2c_2 w_i)^{(\rho+1)/2} = c_1^{(\rho-1)/2} (c_1 + c_2(g+1) w_i) \qquad \dots \quad (7.20)$$

from the differential equation (7.19) in which we shall put $\rho = \frac{1}{2}$, $d_i = \{G(1-g)c_1^{\rho-1}\}^{\frac{1}{2}}$ where $G = g^{\rho}/(1-g)^{\rho} c_2^{\rho}$, we have $e_i = d_i^2$ for $i = 1, 2, \dots, k$. Thus in this important case the minimum variance (7.13) will be given by

$$V = \frac{\left\{ \sum_{i=1}^{k} A_i b_i^{\frac{1}{2}} d_i^{-1} \right\}^2}{T} \qquad \dots \quad (7.21)$$

which may be simple enough to be adopted in practice, and this is also a sufficiently accurate approximation if the conditions (7.20) are nearly valid.

It is to be noted that in our argument there is no need of an existence of constant g; it may change with i. Also that when $g=0$, our result is coincident with the optimum allocation of Deming. The validity of (7.20) may be readily observed specially for the data which Mahalanobis (1944) obtained.

PART II. LOSS OF EFFICIENCY DUE TO APPROXIMATE VALUES OF SIZES, VARIANCES AND UNIT COSTS IN THE STRATA

Under the general assumptions in §2 of Part I, let us consider the estimation of the general mean (2.02) as defined there. The optimum allocation obtained there can be actually obtained only when we have complete information about the parameters $\{N_i\}$, $\{\sigma_i\}$ and $\{c_i\}$. In this section we shall discuss the loss of efficiency which may be expected when we make use of approximate values of these parameters. First let us consider the following theorem.

Theorem 2.1 : *Let* $\{N'_i\}$, $\{\sigma'^2_i\}$ *and* $\{c'_i\}$ *be certain sets of approximate values of* $\{N_i\}$, $\{\sigma^2_i\}$ *and* $\{c_i\}$ *respectively. Let C be the total cost provided for our sample survey. Let us assume that our cost function is a linear function as enunciated in Part I. We shall now determine the allocation* $\{n'_i\}$ *to each of our strata by means of these values such that*

$$n'_i = CN'_i\sigma'_i c'^{-\frac{1}{2}}_i\Big(\sum_{j=1}^{k} N'_j\sigma'_j c'^{\frac{1}{2}}_j\Big)^{-1}, \qquad \ldots \quad (1.01)$$

for $i = 1, 2, \ldots, k$. *Let us make use of the estimator* $\bar{x}'$ *defined by*

$$\bar{x}' = \frac{N'_1\bar{x}_1+N'_2\bar{x}_2+\ldots+N'_k\bar{x}_k}{N'}. \qquad \ldots \quad (1.02)$$

Then we have

$$E(\bar{x}') = \bar{x} - \sum_{i=1}^{k}\Big(\frac{N_i}{N}-\frac{N'_i}{N'}\Big)\bar{x}_i, \qquad \ldots \quad (1.03)$$

$$E(\bar{x}'-\bar{x})^2 = \sigma^2(\bar{x}') + \Big(\sum_{i=1}^{k}\Big(\frac{N_i}{N}-\frac{N'_i}{N'}\Big)\bar{x}_i\Big)^2, \qquad \ldots \quad (1.04)$$

where

$$\sigma^2(\bar{x}') = \frac{1}{N'^2}\sum_{i=1}^{k} N'^2_i\frac{N_i-n'_i}{N_i n'_i}\sigma^2_i. \qquad \ldots \quad (1.05)$$

Then

$$E(\bar{x}'-\bar{x})^2 = \frac{1}{CN'^2}\sum_{i=1}^{k} N'_i c'^{\frac{1}{2}}_i\frac{\sigma^2_i}{\sigma'_i}\sum_{j=1}^{k} N'_j c'^{\frac{1}{2}}_j\sigma'_j - \frac{1}{N'^2}\sum_{i=1}^{k} N'^2_i N^{-1}_i\sigma^2_i.$$

The proof is immediate. The bias component in (1.04) which is independent of the size n suggests to us some sort of double sampling procedure by which we may obtain more accurate values of $\{N_i'\}$. Another point, apart from the bias, is that the real total cost will be

$$C' = \sum_{i=1}^{k} c_i n_i' = C \frac{\sum_{i=1}^{k} N_i' \sigma_i' c_i c_i'^{-\frac{1}{2}}}{\sum_{i=1}^{k} N_i' \sigma_i' c_i'^{\frac{1}{2}}} \quad \dots \quad (1.06)$$

Comparison of the variance above with the variance of the estimator $\bar{x}'$ under optimum allocation with the total cost C', omitting the *fpc* (finite population correction), will give the concept of relative efficiency.

Definition 2.1. The relative efficiency of a set of approximate values of $\{N_i'\}$, $\{\sigma_i'^2\}$ and $\{c_i'\}$ compared with that of $\{N_i'\}$ and the true values $\{\sigma_i^2\}$ and $\{c_i\}$ is defined by

$$e_1(\sigma', c'; \sigma, c) = \frac{\left(\sum_{i=1}^{k} N_i' \sigma_i c_i^{\frac{1}{2}}\right)^2}{\left(\sum_{i=1}^{k} N_i' \sigma_i' \frac{c_i}{c_i'^{\frac{1}{2}}}\right)\left(\sum_{i=1}^{k} N_i' \frac{\sigma_i^2}{\sigma_i'} c_i'^{\frac{1}{2}}\right)} \quad \dots \quad (1.07)$$

It is to be noticed that this definition is only attached to a common system of stratification, and that the value of $e_1(\sigma', c', \sigma, c)$ is always lying between 0 and 1. The value is equal to 1 when and only when $\sigma_i'^2 c_i / \sigma_i^2 c_i' =$ const., as Schwartz's inequality shows us.

2. Loss of efficiency due to the use of approximate values of the variance components in designing multi-stage sampling procedures

Let us consider a finite population of $N_1 N_2 \dots N_k$ elements where each element can be denoted by $y_{i_1 i_2 \dots i_k}$ ($i_1 = 1, 2, \dots, N_1$; $i_2 = 1, 2, \dots, N_2$; ...; $i_k = 1, 2, \dots, N_k$).

Let us define $\{\bar{y}_{i_1 i_2 \dots i_{k-1} \cdot}\}$, $\{\bar{y}_{i_1 i_2 \dots i_{k-2} \cdot \cdot}\}, \dots, \{\bar{y}_{i_1 \cdot \cdot \dots \cdot}\}$ and $\{\bar{y}_{\cdot \cdot \dots \cdot}\}$ as usual and define $S_k^2, S_{k-1}^2, \dots, S_3^2, S_2^2$ and S_1^2 by the recurrence formulae

$$(1^\circ)\ S_k^2 = \frac{1}{N_1 N_2 \dots N_{k-1}(N_k - 1)} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_k=1}^{N_k} (y_{i_1 i_2 \dots i_k} - \bar{y}_{i_1 i_2 \dots i_{k-1} \cdot})^2$$

$$(2^\circ)\ S_k^2 + N_k S_{k-1}^2 + \dots + N_j S_j^2 \quad \dots \quad (2.01)$$

$$= \frac{1}{N_1 N_2 \dots N_j (N_j - 1)} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_j=1}^{N_j} (\bar{y}_{i_1 i_2 \dots i_j} - \bar{y}_{i_1 i_2 \dots i_{j-1} \cdot})^2$$

for $j = k-1, k-2, \dots, 1$.

Let us consider a multi-stage sampling design which will yield us as an unbiased estimate of the grand mean

$$\bar{y}_{\cdot \cdot \cdot \dots \cdot} = \frac{1}{n_1 n_2 \dots n_k} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_k=1}^{n_k} y_{i_1 i_2 \dots i_k}. \quad \dots \quad (2.02)$$

The problem of the optimum allocation under an assigned total cost C will be quite similar to that for stratified random sampling, in view of the variance function.

$$V(\bar{y}_{\cdots\cdots}) = \sum_{h=1}^{k} \frac{S_h^2}{n_1 n_2 \ldots n_h} - \sum_{h=1}^{k} \frac{S_h^2}{N_1 N_2 \ldots N_h} \qquad \ldots \quad (2.03)$$

when we assume the following cost function

$$C = \sum_{j=1}^{k} c_j n_1 n_2 \ldots n_j \qquad \ldots \quad (2.04)$$

c_j being constants independent of n's.

Adopting Definition 2.2, the relative efficiency of a set of approximate values of the variance components $\{S'_i\}$ and cost per i-th stage unit $\{c'_i\}$ in a multistage sampling procedure is defined by

$$e_2(S', c'; S, c) = \frac{\left(\sum\limits_{h=1}^{k} c_h^{\frac{1}{2}} S_h\right)^2}{\left(\sum\limits_{h=1}^{k} c_h c_h'^{-\frac{1}{2}} S'_h\right)\left(\sum\limits_{h=1}^{k} c_h'^{\frac{1}{2}} S_h^2 S_h'^{-1}\right)} \qquad \ldots \quad (2.05)$$

3. General considerations

From the mathematical point of view and also from the practical one, it will be worth while to discuss the case where our job is to minimize the variance function

$$V(n_1, n_2, \ldots, n_k) = \sum_{j=1}^{k} \frac{A_j}{n_j} \qquad \ldots \quad (3.01)$$

under an assigned total cost C, provided that the cost function is now given by

$$C(n_1, n_2, \ldots, n_k) = \sum_{j=1}^{k} c_j n_j^p \qquad \ldots \quad (3.02)$$

where $\{c_j\}$ and p are constants independent of $\{n_j\}$. It can be readily seen that under the assumption of complete knowledge regarding $\{A_j\}$ and $\{c_j\}$ the optimum allocation will be given by

$$n_j = \frac{C^{1/p} A_j^{1/(1+p)} c_j^{-1/(1+p)}}{\left(\sum\limits_{j=1}^{k} c_j^{1/(1+p)} A_j^{p/(1+p)}\right)^{1/p}}$$

for $j = 1, 2, \ldots, k$.

Similarly, we may define the relative efficiency by making use of a set of approximate values $\{A'_j\}$ and $\{c'_j\}$ by the following ratio which from the Holder inequality lies between 0 and 1.

$$e_3(A', c'; A, c) = \frac{\left(\sum\limits_{j=1}^{k} A_j^{p/(1+p)} c_j^{1/(1+p)}\right)^{\frac{1+p}{p}}}{\left(\sum\limits_{j=1}^{k} A_j'^{\,p/(1+p)} c_j c_j'^{-p/(1+p)}\right)^{1/p}\left(\sum\limits_{j=1}^{k} A_j A_j'^{-1/(1+p)} c_j'^{1/(1+p)}\right)} \qquad \ldots \quad (3.03)$$

Concerning a cost function, the problem whether it is a linear function or a form such as defined in (3.02) should be carefully discussed. It will also be necessary to discuss the loss involved due to the use of an approximate value p' instead of a true value p. Indeed in such situations, an argument similar to that just developed will lead us to the relative efficiency

$$e_4(p'; p) \equiv \frac{\left(\sum_{j=1}^{k} A_j^{p/(1+p)} c_j^{1/(1+p)}\right)^{\frac{1+p}{p}}}{\left(\sum_{j=1}^{k} A_j^{p'/(p'+1)} c_j^{1/(1+p')}\right)\left(\sum_{j=1}^{k} A_j^{p/(p'+1)} c_j^{\frac{p'+1-p}{p'+1}}\right)^{1/p}} \qquad \dots \quad (3.04)$$

Thus for instance, the assumption of a linear cost function means an efficiency $e_4(1; p)$. The points considered in this section are important, because there may be some real circumstances where certain rather complicated cost and loss functions are to be adopted which may be reducible to the forms presented here, as we have shown in §7 of Part I, and which we will verify in the next section.

4. Loss of efficiency under generalised cost and variance functions of Mahalanobis

The optimum allocation under complete information was given in §7 of Part I, where the cost and variance functions were of certain generalised forms some of which were advocated by Mahalanobis (1944). However, in actual situations we have to make use of incomplete information about some of the parameters. Thus let $\{b'_i\}$ and $\{d'_i\}$ be sets of approximate values of $\{b_i\}$ and $\{d_i\}$ respectively. It is also to be noted that we may assume ρ to be exactly known. This assumption seems reasonable, so far as we assume the exact functional forms of $\{\psi_i\}$ and $\{\phi_i\}$.

Calculations similar to those stated in the previous two sections lead to the following deviation formula

$$V = \frac{\left(\sum_{i=1}^{k} A_i (b'_i e'^{-1}_i)^{\rho}\, b_i b'^{-1}_i\right)\left(\sum_{i=1}^{k} A_i (b'_i e'^{-1}_i)^{\rho}\right)^{\frac{1-\rho}{\rho}}}{T^{(1-\rho)/\rho}} \qquad \dots \quad (4.01)$$

which also yields us the following definition of relative efficiency

$$e_4(b', e'; b, e) = \frac{\left(\sum_{i=1}^{k} A_i (b_i e_i^{-1})^{\rho}\right)^{1/\rho}}{\left(\sum_{i=1}^{k} A_i (b'_i e'^{-1}_i)^{\rho} b_i b'^{-1}_i\right)\left(\sum_{i=1}^{k} A_i (b'_i e'^{-1}_i)^{\rho} (e'_i e_i^{-1})^{\frac{\rho}{1-\rho}}\right)^{\frac{1-\rho}{\rho}}} \qquad \dots \quad (4.02)$$

which of course lies between 0 and 1, as may be verified by Hölder's inequality. Here we have put

$$e_i = \frac{1-\rho}{\rho}\, d_i^{1/\rho}, \quad e'_i = \frac{1-\rho}{\rho}\, d'^{1/\rho}_i \qquad \dots \quad (4.03)$$

for $i = 1, 2, \dots, k$.

PART III. SEQUENTIAL DESIGNS OF SAMPLE SURVEYS AIMING TO IMPROVE ALLOCATIONS TO STRATA

1. INTRODUCTORY

In this Part we shall consider a sequence of stratified random sample surveys in which a certain prescribed stratification will be continuously adopted. Usually the accumulation of information regarding our populations may be useful in order to improve later designs and the whole theory is better appreciated in a realm of succesive designs of sample surveys. In this Part we shall start with the more elementary considerations, which however may readily be recognised as indispensable. It is possible to suggest various procedures of successive designs but it will turn out that some of them can only be justified under restrictions. Thus in this Part we shall give some examples of designs and compare their relative merits.

2. ALLOCATION OF OUR RESOURCES TO A PRELIMINARY SURVEY

Let our whole population Π be stratified into the strata $\{\Pi_i\}$ $(i = 1, 2, ..., k)$, whose sizes $\{N_i\}$ are assumed to be known to us. On the other hand suppose there are no reliable estimates of the within-stratum variances sufficient to determine our allocation of sample sizes to each stratum. For the sake of simplicity let it be assumed that cost per unit in each stratum is a common constant $\bar{c}$ throughout all strata. Let C be a given total cost. We shall now consider the following procedure :

(1^0) Draw a sample of size n_0 from each stratum and let us put $kn_0 = aC/\bar{c}$ where $a < 1$.

(2^0) Let $\hat{\sigma}_i$ be an unbiased estimate of the within-stratum standard deviation σ_i based on the samples in (1^0) for $i = 1, 2, ..., k$.

(3^0) Let us allocate our remaining amount $(1-a)C$ to a further sample of size $(1-a)C/\bar{c}$ distributed amongst the strata according to the following formula

$$n_i = \frac{(1-a)CN_i\hat{\sigma}_i}{\bar{c}(N_1\hat{\sigma}_1+N_2\hat{\sigma}_2+...+N_k\hat{\sigma}_k)} \qquad ... \quad (2.01)$$

for $i = 1, 2, ..., k$. The stratified random sampling according to this allocation will yield us an estimate $\bar{x}$ for the grand mean with the variance

$$V(\bar{x}) = \frac{\bar{c}}{(1-a)C}\sum_{i=1}^{k} N_i\sigma_i\frac{\sigma_i}{\hat{\sigma}_i}\sum_{j=1}^{k} N_j\sigma_j\frac{\hat{\sigma}_j}{\sigma_j}, \qquad ... \quad (2.02)$$

as may be seen from the argument developed in Part II.

Broadly speaking our intrinsic problem is to determine the fraction a which minimizes (2.02). To find a solution to this problem, we must take into consideration two factors. The first one is concerned with the sampling distribution of $\{\hat{\sigma}_i\}$ and the second one with the *a priori* knowledge regarding $\{\sigma_i\}$.

For this we shall introduce a certain relative loss function defined as follows :—

Definition 3.1 : Let us denote $V(\hat{z})$ in (2.02) by $W(\sigma; \hat{\sigma}, a, C)$. Let $\phi_a(\hat{\sigma}, \sigma, C)$ be the density function of the joint distribution of $(\hat{\sigma}_1, \hat{\sigma}_2, ..., \hat{\sigma}_k)$ when the population standard deviations are $\{\sigma_i\}$ and our allocation is aC and let us denote the expected value of $W(\sigma; \hat{\sigma}, a, C)$ by $W(\sigma; a, C)$, that is,

$$W(\sigma; a, C) = \int W(\sigma; \hat{\sigma}, a, C)\phi_a(\hat{\sigma}, \sigma, C)d\hat{\sigma}. \qquad ... \quad (2.03)$$

Then the relative loss function of our procedure in which aC is devoted to the preliminary survey and $(1-a)C$ to the aimed survey is defined by

$$W_r(\sigma; a, C) \equiv \frac{W(\sigma; a, C)}{\frac{\bar{c}}{C}\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2}. \qquad ... \quad (2.04)$$

It will now be obvious that our strategy is concerned with the choice of the fraction a while "nature" may assign any system of the real constants $\{\sigma_i\}$. One possible principle in such circumstances in view of the minimax principle due to Wald (1950) is to determine a fraction a which will minimize

$$\sup_{\sigma\epsilon\Omega} W_r(\sigma; a, C) \qquad ... \quad (2.05)$$

where Ω denotes the set of all the possible points $\sigma = (\sigma_1, \sigma_2, ..., \sigma_k)$. Such a fraction a will be called hereafter a relative minimax solution. The existence and uniqueness of such a fraction a may be proved under somewhat general conditions. As a first approach, however, we may be concerned with the cases which satisfy the following conditions :

Assumption 1 : Ω is the k-dimensional set of all the points $\sigma = (\sigma_1, \sigma_2, ..., \sigma_k)$ such that every $\sigma_i > 0$.

Assumption 2 :

$$E\left\{\frac{\sigma_i}{\hat{\sigma}_i}\right\} = 1 + \frac{\gamma}{n_0} + O\left(\frac{1}{n_0^2}\right), \qquad ... \quad (2.06)$$

where γ is a positive constant independent of both n_0 and i.

When we omit the third term of the right-hand, we call it the first term approximation.

Now our solution will be enunciated in the following :

Theorem 3.1 : *Under the Assumptions* (1) *and* (2) *in addition to the general assumption stated in this section, we have only one relative minimax solution* $\hat{a}$ *upto the first term approximation of* (2.06) *which will be given by*

$$\hat{a} = (g_1 + g_1^2)^{\frac{1}{2}} - g_1, \qquad ... \quad (2.07)$$

where

$$g_1 = \frac{\gamma(k-1)\bar{c}}{C}, \qquad ... \quad (2.08)$$

and which yields us

$$\min_{0<a<1} \sup_{\Omega} W_r(\sigma; a, C) = (1+g_1)\left\{1+\left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\right\}^2 . \qquad ... \quad (2.09)$$

Proof : We shall have

$$\int W(\sigma; \hat{\sigma}, a, C)\phi_s(\hat{\sigma}; \sigma, C)d\hat{\sigma}$$

$$= \frac{\bar{c}}{(1-a)C}\left[\sum N_i^2\sigma_i^2 + \sum_{i\neq j}\sum N_i\sigma_i N_j\sigma_j E\left\{\frac{\sigma_i}{\hat{\sigma}_i}\right\}\right]$$

$$= \frac{\bar{c}}{(1-a)C}\left[\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2 + \sum_{i\neq j}\sum N_i\sigma_i N_j\sigma_j \frac{\gamma}{n_0}\right]$$

$$+ \frac{\bar{c}}{(1-a)C} O\left(\frac{1}{n_0^2}\right)\sum_{i\neq j}\sum N_i\sigma_i N_j\sigma_j. \qquad \ldots \quad (2.10)$$

In view of the fact that

$$\max_{\sigma\epsilon\Omega} \frac{\sum\sum_{i\neq j} N_i\sigma_i N_j\sigma_j}{\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2} = \frac{k-1}{k}, \qquad \ldots \quad (2.11)$$

we shall have

$$\sup_{\sigma\epsilon\Omega} W_i(\sigma; a, C) = \frac{1}{1-a}\left[1 + \frac{\gamma(k-1)\bar{c}}{C}\frac{1}{a}\right] + \frac{k-1}{a} O\left(\left(\frac{\bar{c}}{aC}\right)^2\right), \qquad \ldots \quad (2.12)$$

which will yield us the results to be proved, provided that the second term on the right hand side in (2.12) is omitted, under the assumption of the 1st term approximation.

Remark : It is to be noted that up to the first term approximation we have

$$\min_{0<a<1} \sup_{\Omega} W_i(\sigma; a, c) = \sup_{\Omega} \min_{0<a<1} W_i(\sigma; a, c), \qquad \ldots \quad (2.13)$$

which means to be strictly determined in the terminology of the theory of games.

Example 3.1 : There are various circumstances in which we may reasonably assume that sample variances multiplied by their respective degrees of freedom from each stratum are distributed approximately according to the chi-square distribution and that

$$= \frac{(n_i-1)^{\frac{1}{2}}}{2^{\frac{1}{2}}} \frac{\Gamma\left(\frac{n_i-1}{2}\right)}{\Gamma\left(\frac{n_i}{2}\right)} s_i \qquad (2.14)$$

will give us an unbiased estimate of the i-th within-stratum standard deviation where s_i^2 are the unbiased estimates of variance. In such circumstances, we have

$$E\left\{\frac{\sigma_i}{\hat{\sigma}_i}\right\} = \frac{\Gamma\left(\frac{n_i}{2}\right)\Gamma\left(\frac{n_i}{2}-1\right)}{\left\{\Gamma\left(\frac{n_i-1}{2}\right)\right\}^2} = 1+\frac{1}{2n_i}+O\left(\frac{1}{n_i^2}\right) \qquad \dots \quad (2.15)$$

that is, $\gamma = \frac{1}{2}$ in this case.

Sukhatme (1935) discussed this case with the purpose of comparing stratified sampling designs with the purely random ones, but he did not consider the problem of how the total cost may be divided in the case of subsampling procedures, ratio-estimates and regression estimates.

Example 3.2 : It is to be noted here that our method adopted in Theorem 3.1 may be applicable to the following problem of Robbins (1952). Let us now deal with two normally distributed populations with unknown means μ_1, μ_2 and variances σ_1^2 and σ_2^2 and let our problem be to divide the given total of n observations into two stages, where in the first stage, preliminary samples of size m from each of the two populations and in the second stage, the remainder of the $n-2m$ observations will be allocated to the two populations in accordance with the sample values obtained in the first stage for σ_1/σ_2. The solution is obtained from Theorem 3.1 which gives

$$\frac{\hat{a}n}{2} = \frac{n}{2}\left\{\left(\frac{1}{2n}+\frac{1}{4n^2}\right)^{1/2} - \frac{1}{2n}\right\} \qquad \dots . \quad (2.16)$$

since $g_1 = \frac{1}{2n}$ provided that we have restricted ourselves within $\sigma_1+\sigma_2 =$ constant and that our approximate approach is sufficient in this case. A simple rule will be

$$m = \frac{1}{2}\left(\frac{n}{2}\right)^{1/2}. \qquad \dots \quad (2.17)$$

Remark 1 : The real difficulty will sometimes be connected with the unknown value γ and there may be many cases in which γ may not be constant throughout all strata. Both from the theoretical and the practical points of view there is need for discussing the cases when (i) γ may change from stratum to stratum and/or (ii) we have no complete information about those values of γ. These problems may however be different from those we have considered just now and in fact they belong to a realm of sequential designs in which our aims of estimation are concerned not only with improvement of allocation but also with requirements to obtain better information about these unknown parameters involving higher moments.

3. USE OF PREVIOUS KNOWLEDGE IN ALLOCATION OF RESOURCES TO A PRELIMINARY SURVEY

In this section we shall discuss the case where we have some previous knowledge which we can utilize without spending any part of the amount available for the survey. We shall consider two types of situations. First, we can assume our population to remain unchanged at least with a respect to within-stratum variances. Secondly, our previous information may be concerned with a certain restriction of the possible domain in which the parameter point $\sigma=(\sigma_1, \sigma_2, \ldots, \sigma_k)$ should lie. We shall consider these cases in Theorems 3.2 and 3.3 . In both cases we can observe how far our allocation to a preliminary survey may be suitably diminished by making use of previous information although the methods are quite different in the two cases.

Theorem 3.2 : *In addition to the Assumptions in Theorem 3.1, let us assume that we have another independent system of unbiased estimates of within-stratum standard deviations* $\{\hat{\sigma}_{0i}\}$ *obtained from a sample of size* m_0 *in each stratum independently and for which Assumption* (2°) *in* §2 *is also valid. Let us put* $km_0 = pC/\xi$.

Let us define the pooled estimates

$$\hat{\sigma}'_i = (m_0\hat{\sigma}_{0i}+n_0\hat{\sigma}_i)/(m_0+n_0) \qquad \ldots \quad (3.01)$$

for $i = 1, 2, \ldots, k$. *The sizes of samples* $\{n_i\}$, *are given by*

$$n_i = \frac{(1-a)CN_i\hat{\sigma}'_i}{\xi(N_1\hat{\sigma}'_1+N_2\hat{\sigma}'_2+\ldots+N_k\hat{\sigma}'_k)} \qquad \ldots \quad (3.02)$$

for $i = 1, 2, \ldots, k$, *in virtue of* (2.01).

Then we have only one relative minimum solution $\hat{a}_p$ *upto the first order approximation to the problem enunciated in* §2 *which will be defined as follows.*

(1) *If* $g_1 \geqslant p(p+g_1)$, *then we have*

$$\hat{a}_p = \{g_1(1+p+g_1)\}^{\frac{1}{2}} - (p+g_1). \qquad \ldots \quad (3.03)$$

(2) *If* $g_1 < p(p+g_1)$, *then we have*

$$\hat{a}_p = 0. \qquad \ldots \quad (3.04)$$

The proof can be obtained simply by seeking the value of a which will minimize

$$\frac{1}{1-a}\left[1+\frac{g}{a+p}\right] \qquad \ldots \quad (3.05)$$

in view of the method adopted in Theorem 3.1.

Remark : In the enunciation of Theorem 3.1 it is important to notice that the following minimax relation holds true in the abbreviated sense adopted in this Theorem,

$$\min_{0<a<1} \sup_{\Omega} W_r(\sigma; a, C) = \sup_{\Omega} \min_{0<a<1} W_r(\sigma; a, C), \qquad \ldots \quad (3.06)$$

whose proof is quite immediate.

Theorem 3.3 : *In addition to the general Assumptions in Theorem 3.1 let us assume further that there is an a priori knowledge according to which our parameter space of σ should be restricted within a certain subspace Ω_1 of Ω which yields us*

$$\max_{\sigma\epsilon\Omega_1} \frac{\sum\sum_{(i\neq j)} N_i\sigma_i N_j\sigma_j}{\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2} = \zeta. \quad \dots \quad (3.07)$$

Then our relative minimum solution in the sense of §2 is now given by

$$\hat{a}_1 = (g_1+g_1^2)^{\frac{1}{2}}-g_1 \quad \dots \quad (3.08)$$

where we have defined

$$g_1 = g_1\frac{k}{k-1}\zeta = \frac{\gamma\bar{c}k}{C}\zeta. \quad \dots \quad (3.09)$$

The proof is immediate from the relation (2.11) in view of Theorem 3.1.

4. Allocation of resources to a sequence of surveys consisting of a preliminary and two successive aimed surveys

Under the general assumptions in §2, let us now proceed to discuss the problem of the optimum allocation of a given total cost C

$$C = a_0C+a_1C+a_2C \quad \dots \quad (4.01)$$

with $a_0+a_1+a_2 = 1, a_i > 0$, in the following process:

(0) Draw a stratified random sample O_0 consisting of k mutually independent random samples of size n_0 from each stratum, where $kn_0 = a_0C/\bar{c}$. Let $\hat{\sigma}_{0i}$ be the unbiased estimates of the i-th within-stratum standard deviation obtained from O_0.

(1) Let us define each of the sizes of samples to be drawn from each i-th stratum independently as follows:-

$$n_{1i} = \frac{a_1C}{\bar{c}}\cdot\frac{N_i\hat{\sigma}_{0i}}{\sum_{j=1}^{k} N_j\hat{\sigma}_{0j}} \quad (4.02)$$

for $i = 1, 2, \dots, k$, and let us denote by O_1 this stratified random sample.

Let $\bar{x}_{1i}$ and $\hat{\sigma}_{1i}$ be unbiased estimates of the i-th stratum mean and the i-th stratum standard deviation obtained from O_1 respectively for $i = 1, 2, \dots, k$.

(2) Let us now define the size of the third sample to be drawn from the i-th stratum independently by

$$n_{2i} = \frac{a_2 C}{\bar{c}} \frac{N_i \phi_2(\hat{\sigma}_{0i}, \hat{\sigma}_{1i})}{\sum_{j=1}^{k} N_j \phi_2(\hat{\sigma}_{0j}, \hat{\sigma}_{1j})} \quad \ldots \quad (4.03)$$

for $i = 1, 2, \ldots k$, where ϕ_2 are certain functions defined a priori, and let us denote this stratified random sample by O_2. Let $\bar{x}_{2i}$ and $\hat{\sigma}_{2i}$ be unbiased estimates of the i-th stratum mean and the i-th stratum standard deviation obtained from O_2 for $i = 1, 2, \ldots, k$.

Here we shall consider the following cases.

Case A. This procedure S_A is characterised by the following conditions:

(i) Our ultimate concern is the last stage estimation of the general mean which is defined by

$$\bar{x}_2 = \frac{\sum_{j=1}^{k} N_j \bar{x}_{2j}}{\sum_{j=1}^{k} N_j} . \quad \ldots \quad (4.04)$$

(ii) The sample sizes $\{n_{2i}\}$ are defined simply by

$$n_{2i} = \frac{a_2 C}{\bar{c}} \cdot \frac{N_i \hat{\sigma}_{1i}}{\sum_{j=1}^{k} N_j \hat{\sigma}_{1j}} \quad (i = 1, 2, \ldots, k), \quad \ldots \quad (4.05)$$

that is to say, we have specially in (4.03)

$$\phi_2(\hat{\sigma}_{0i}, \hat{\sigma}_{1i}) = \hat{\sigma}_{1i} \quad \ldots \quad (4.06)$$

for $i = 1, 2, \ldots, k$.

Case B. This procedure S_B is defined by the following conditions:

(i) We are concerned with the estimation of the general means over the two stages, $\bar{x}_1$ and $\bar{x}_2$.

(ii) The same as condition (ii) of Case A.

Case C. This procedure S_C is characterised by the following conditions:

(i) The same as condition (i) of Case A.

(ii) The sample sizes $\{n_{2i}\}$ are now defined by

$$n_{2i} = \frac{a_2 C}{\bar{c}} \cdot \frac{N_i(n_0 \hat{\sigma}_{0i} + n_{1i} \hat{\sigma}_{1i})/(n_0 + n_{1i})}{\sum_{j=1}^{k} N_j(n_0 \hat{\sigma}_{0j} + n_{1j} \hat{\sigma}_{1j})/(n_0 + n_{1j})} \quad (i = 1, 2, \ldots, k), \quad \ldots \quad (4.07)$$

that is to say,

$$\phi_2(\hat{\sigma}_{0i}, \hat{\sigma}_{1i}) = (n_0+n_{1i})^{-1}(n_0\hat{\sigma}_{0i}+n_{1i}\hat{\sigma}_{1i}). \qquad \dots \quad (4.08)$$

Case D. This procedure S_D is defined by the following conditions :

(i) The same as condition (i) of case B.

(ii) The same as condition (ii) of case C.

The detailed discussion of Cases C and D will not be given in this paper, although it may be worthwhile to point out these interesting procedures.

5. Preliminary considerations : solutions of the zero and the first orders

To solve the four problems enunciated in §4 and to deal with the more general cases of sequential designs, certain notions of solution should be introduced here in order that some manageable answers could be given in practical situations. These solutions are not exact solutions but from the operational points of view these approximate solutions seem to be sufficient in almost all practical situations so far as large-scale surveys are concerned. The object of this section is to discuss the Cases A and B as examples and to observe how and why the notion of solution of the first order may be so useful.

(1) Case A. In this case we have

$$V_1(\bar{x}_2) = \frac{\bar{c}}{a_2 C}\sum_{i=1}^{k} N_i\sigma_i\frac{\sigma_i}{\hat{\sigma}_{1i}}\sum_{j=1}^{k} N_j\,\sigma_j\,\frac{\hat{\sigma}_{1j}}{\sigma_j}$$

$$\equiv W(\sigma; \hat{\sigma}_0, \hat{\sigma}_1, a_0, a_1, C) \equiv W(\sigma; \hat{\sigma}, a, C) \text{ say,} \qquad \dots \quad (5.01)$$

where we have used vector notations

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_k), \qquad \dots \quad (5.02)$$

$$\hat{\sigma}_\nu = (\hat{\sigma}_{\nu 1}, \hat{\sigma}_{\nu 2}, \dots, \hat{\sigma}_{\nu k}) \qquad (\nu = 0, 1), \qquad \dots \quad (5.03)$$

$$a = (a_0, a_1). \qquad \dots \quad (5.04)$$

Our sampling procedure will define the density function of the joint distribution of $(\hat{\sigma}_0, \hat{\sigma}_1)$ such that

$$\phi(\hat{\sigma}_0; \sigma, a_0, C)\phi(\hat{\sigma}_1; \sigma, a_1/\hat{\sigma}_0, a_0, C)$$
$$\equiv \phi(\hat{\sigma}_0, \hat{\sigma}_1; \sigma, a_0, a_1, C) \equiv \phi(\hat{\sigma}; \sigma, a, C), \text{ say.} \qquad \dots \quad (5.05)$$

Consequently the expected value of (5.01) will be given by

$$W(\sigma; a, C) \equiv \int W(\sigma; \hat{\sigma}, a, C)\phi(\hat{\sigma}; \sigma, a, C)d\hat{\sigma} \qquad \dots \quad (5.06)$$

for each assigned σ and a which will lead us to the following definition of the relative loss function useful in discussing our present problem:

$$W_r(\sigma; a, C) \equiv W(\sigma; a, C)/\left\{\frac{\bar{c}}{C}\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2\right\}. \qquad \dots \quad (5.07)$$

Let us introduce the following assumption which seems to us to be indispensable if we aim to make an accurate discussion of sequential procedures.

Assumption (2^{**}) : We have

$$E\left\{\frac{\sigma_i^h}{\hat{\sigma}_{\nu i}^h}\right\} = 1+\frac{\gamma_h}{n_{\nu i}}+\frac{\delta_h}{n_{\nu i}^2} +O\left(\frac{1}{n_{\nu i}^3}\right) \qquad \dots \quad (5.08)$$

for each $\nu = 0, 1$, each $h = 1, 2$, and $i = 1, 2, \dots, k$, where γ_h and δ_h are assumed to be positive constants independent of all $\{n_{\nu i}\}$, ν and i, and $O(n_{\nu i}^{-3})$ stands for functions uniformly bounded irrespective of ν and i when they are divided by $n_{\nu i}^3$ respectively.

In what follows let us denote by E_ν the conditional expectation of a stochastic variable under the condition that the values of the stochastic variables $\{\hat{\sigma}_{pi}\}$, $(p = 0, 1, \dots, \nu;\ i = 1, 2, \dots, k)$ are assumed to be assigned and by E the absolute expectation of a stochastic variable. For the sake of convenience let us write also for $\nu = 0, \pm 1, \pm 2, \dots$

$$A_\nu = \sum_{i=1}^{k} (N_i\sigma_i)^\nu. \qquad \dots \quad (5.09)$$

Lemma 3.1 : *Under the Assumptions* (1^0) *and* (2^0) *we have*

$$E_0\{W(\sigma; \hat{\sigma}, a, C)\}$$

$$= \frac{\bar{c}}{a_2 C}\left\{ A_2+\sum\sum_{i\neq j} N_i\sigma_i N_j\sigma_j E_0\left(\frac{\hat{\sigma}_{1i}}{\sigma_i}\right) E_0\left(\frac{\sigma_j}{\hat{\sigma}_{1j}}\right)\right\}$$

$$= \frac{\bar{c}}{a_2 C}\left\{A_1^2+\sum_{j=1}^{k} N_j\sigma_j(A_1-N_j\sigma_j)(\gamma_1/n_{1j}+\delta_1/n_{1j}^2+O(n_{1j}^{-3}))\right\}. \qquad \dots \quad (5.10)$$

$$E\left\{n_{1j}^{-1}\right\} = \frac{\bar{c}}{a_1 C} E\left\{\left(N_j\hat{\sigma}_{0j}\right)^{-1} \sum_{i=1}^{k} N_i\hat{\sigma}_{0i}\right\}$$

$$= \frac{\bar{c}}{a_1 C}(N_j\sigma_j)^{-1}\left\{A_1+(A_1-N_j\sigma_j)(\gamma_1 n_0^{-1}+\delta_1 n_0^{-2}+O(n_0^{-3}))\right\}. \qquad \dots \quad (5.11)$$

$$E\left\{n_{1j}^{-2}\right\} = \left(\frac{\bar{c}}{a_1 C}\right)^2 E\left\{\left(N_j\hat{\sigma}_{0j}\right)^{-2}\left(\sum_{i=1}^{k} N_i\hat{\sigma}_{0i}\right)^2\right\}$$

$$= \left(\frac{\bar{c}}{a_1 C}\right)^2\left[1+\frac{A_2-N_j^2\sigma_j^2}{N_j^2\sigma_j^2}\left(1+\frac{\gamma_2}{n_0}+\frac{\delta_2}{n_0^2} +O\left(\frac{1}{n_0^3}\right)\right)+\right.$$

$$+\frac{A_1-N_j\sigma_j}{N_j\sigma_j}\left(1+\frac{\gamma_1}{n_0}+\frac{\delta_1}{n_0^2}+O\left(\frac{1}{n_0^3}\right)\right)+$$

$$\left.+\frac{\sum_{p\neq q}^{(j)} N_p\sigma_p N_q\sigma_q}{N_j^2\sigma_j^2}\left(1+\frac{\gamma_2}{n_0}+\frac{\delta_2}{n_0^2}+ O\left(\frac{1}{n_0^3}\right)\right)\right]. \qquad \dots \quad (5.12)$$

where the last summation runs through all p and q such as $p, q = 1, 2, \ldots, k$ under the condition that $p \neq q$, $p \neq j$ and $q \neq j$.

Furthermore we have

$$W_r(\sigma; a, C) \equiv E\{W(\sigma; \hat{\sigma}, a, C)\}A_1^{-2}\bar{c}^{-1}C$$

$$= \frac{1}{a_2}\left\{1+\frac{(k-1)\gamma_1\bar{c}}{a_1C}\right\}+$$

$$+\frac{1}{a_2}\left\{\frac{\gamma_1\bar{c}}{a_1C}\cdot\frac{\gamma_1\bar{c}}{a_0C}\left(k-2+\frac{A_2}{A_1^2}\right)+\delta_1\left(\frac{\bar{c}}{a_1C}\right)^2\sum_{j=1}^{k}\frac{A_1-N_j\sigma_j}{N_j\sigma_j}\right\}+$$

$$+\frac{1}{a_2}\left\{\frac{\gamma_1\bar{c}}{a_1C}\cdot O\left(\left(\frac{\bar{c}}{a_0C}\right)^2\right)\left(k-2+\frac{A_2}{A_1^2}\right)+\right.$$

$$+\delta_1\left(\frac{\bar{c}}{a_1C}\right)^2\cdot\frac{\bar{c}}{a_0C}\left(\gamma+O\left(\frac{\bar{c}}{a_0C}\right)\right)\sum_{j=1}^{k}\frac{A_1-N_j\sigma_j}{N_j\sigma_j}+$$

$$\left.+O\left(\left(\frac{\bar{c}}{a_1C}\right)^3\right)\left(k-2+\frac{A_2}{A_1^2}\right)\right\} \qquad \ldots \quad (5.13)$$

where $\gamma = \max(\gamma_1, \gamma_2)$.

In order to proceed in a way similar to that of §2, there are two difficulties to be overcome. In the first place, we have to assume certain conditions as to the values $\{N_j\sigma_j\}$ according to which we should be able to obtain the finite value of the supremum of the relative weights $W_r(\sigma; a, C)$. Let us introduce

Assumption (3°): Our parameter point $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_k)$ is restricted within a certain domain Ω_1 in Ω such that

$$\sup_{\sigma\epsilon\Omega_1}\ (k-2+A_2A_1^{-2}) = B_1, \qquad \ldots \quad (5.14)$$

$$\sup_{\sigma\epsilon\Omega_1}\sum_{j=1}^{k}\frac{A_1-N_j\sigma_j}{N_j\sigma_j} = B_2 \quad \text{say}, \qquad \ldots \quad (5.15)$$

exist and $0 < B_1, B_2 < \infty$.

Lemma 3.2 : *Under the Assumption* (1°), (2*°) *and* (3°) *we have*

$$\sup_{\sigma\epsilon\Omega_1}\ W_r(\sigma; a, C) = \frac{1}{a_2}\left[1+\frac{(k-1)\gamma_1\bar{c}}{a_1C}\right] \qquad \ldots \quad (5.16)$$

Secondly we shall give approximate minimax solution upto a certain degree of approximation. There are various types of approximation in dealing with the right-hand side of (5.16). The choice among these approximations should depend upon the relative magnitude of the compounds. We may adopt the following

Definition 3.2 : The set of non-negative numbers (a_0, a_1, a_2) such that $a_0+a_1+a_2=1$ is said to be the relative minimax solutions of (i) the zero-order, (ii) the abbreviated first-order, and (iii) the first-order respectively according as it minimizes the following algebraic forms respectively:

$$w_0(a) = \frac{1}{a_2}\left[\ 1+\frac{(k-1)\gamma_1\bar{c}}{a_1C}\right], \qquad \ldots \quad (5.17)$$

$$w_1(a) = \frac{1}{a_2}\left[\ 1+\frac{(k-1)\gamma_1\bar{c}}{a_1C}+B_1\frac{\gamma_1\bar{c}}{a_1C}\cdot\frac{\gamma_1\bar{c}}{a_0C}\right], \qquad \ldots \quad (5.18)$$

$$w_2(a) = \frac{1}{a_2}\left[\ 1+\frac{(k-1)\gamma_1\bar{c}}{a_1C}+B_1\frac{\gamma_1\bar{c}}{a_1C}\cdot\frac{\gamma_1\bar{c}}{a_0C}+B_2\delta_1\left(\frac{\bar{c}}{a_1C}\right)^2\right], \qquad \ldots \quad (5.19)$$

under the condition $a_0+a_1+a_2 = 1,\ a_i \geqslant 0\ (i = 0, 1, 2)$.

It is to be noted that (i) $w_0(a)$ is too much simplified to be useful in Case A, but it will be found to be useful in other cases; (ii) the first step should usually be to apply $w_1(a)$; (iii) the minimum of $A_2A_1^{-2}$ in the whole parameter space Ω being k^{-1}, there is no need of restricting our domain within a narrower one Ω_1 so far as (5.18) may be used.

(2) Case B. In this case we have to take into consideration each estimation of $\bar{x}_1$ and $\bar{x}_2$. According to the general considerations developed in Part I, let us now introduce one consolidated variance

$$\sum_{i=1}^{2} \lambda_i V_{i-1}\{\bar{x}_i\}, \qquad \ldots \quad (5.20)$$

whose expectation and relative efficiency will be denoted by $W_\lambda(\sigma;\ a,\ c)$ and $W_{\lambda,r}(\sigma;\ a,\ C)$ respectively. Thus we shall put

$$W_{\lambda,r}(\sigma; a,C) \equiv W_\lambda(\sigma; a,C)\left\{\frac{\bar{c}}{C}\left(\sum_{i=1}^{k} N_i\sigma_i\right)^2\right\}^{-1}$$

$$= \frac{C}{\bar{c}A_1^2}\left[\ \lambda_1 E\{V_0(\bar{x}_1)\}+\lambda_2 E(V_1\{\bar{x}_2\})\right]. \qquad \ldots \quad (5.21)$$

In combination of (2.11) and (5.12) we shall have, putting $g=\gamma\bar{c}C^{-1}$,

$$W_{\lambda,r}(\sigma; a, C) = \frac{\lambda_2}{a_2}\left[\ 1+\frac{(k-1)g}{a_1}+\left(k-2+\frac{A_2}{A_1^2}\right)\frac{g^2}{a_0a_1}\right]+$$

$$+\frac{\lambda_1}{a_1}\left[\ 1+\left(1-\frac{A_2}{A_1^2}\right)\frac{(k-1)g}{a_0}\right]. \qquad \ldots \quad (5.22)$$

upto the approximation of the abbreviated first order.

Here we have to notice

Lemma 3.3. *The supremum of the value* (5.22) *when σ runs through Ω can be given as follows :*

(1) *In the domain defined by $\lambda_2 g > (k-1)\lambda_1 a_2$, it attains its maximum at the points where $A_2 = A_1^2$ which give us*

$$\sup_{\sigma\epsilon\Omega} W_{\lambda,r}(\sigma; a, C) = \frac{\lambda_2}{a_2}\left[1+\frac{(k-1)g}{a_1}+\frac{(k-1)g^2}{a_0 a_1}\right] + \frac{\lambda_1}{a_1}. \qquad \ldots \quad (5.23)$$

(2) *In the domain defined by $\lambda_2 g = (k-1)\lambda_1 a_2$ the value* (5.22) *is equal to a constant given by*

$$W_{\lambda,r}(\sigma; a, C) = \sup_{\sigma\epsilon\Omega} W_{\lambda,r}(\sigma; a, C)$$

$$= \frac{\lambda_2}{a_2}\left[1+\frac{(k-1)g}{a_1}+\frac{(k-2)g^2}{a_0 a_1}\right]+\frac{\lambda_1}{a_1}\left[1+\frac{(k-1)g}{a_0}\right]. \qquad \ldots \quad (5.24)$$

(3) *In the domain $\lambda_2 g < (k-1)\lambda_1 a_2$, it attains its maximum at the points where $kA_2 = A_1^2$ and which give us*

$$\sup_{\sigma\epsilon\Omega} W_{\lambda,r}(\sigma; a, C) = \frac{\lambda_2}{a_2}\left[1+\frac{(k-1)g}{a_1}+\frac{(k-1)^2 g^2}{k a_0 a_2}\right]+$$

$$+\frac{\lambda_1}{a_1}\left[1+\frac{(k-1)g}{a_0}\cdot\frac{k-1}{k}\right]. \qquad \ldots \quad (5.25)$$

Here it is to be noted that (1°) when $k = 1$ all values of the supremum of $W_{\lambda,r}(\sigma; a, C)$ will reduce to the simplest form $\lambda_1 a_1^{-1}+\lambda_2 a_2^{-1}$ as might be expected and that (2°) among the three cases (1), (2) and (3) above the case (3) will be the one which is of special interest from the practical point of view. In what follows we shall first derive some fundamental relations in a sequence of sample surveys rather from the formal points of view and then turn to the adoption of the zero-order approximations in certain cases. From the theoretical point of view it is necessary to discuss the relative minimax solutions more accurately in view of the extensive use of the relative loss functions. Nevertheless it should also be noticed that such detailed discussion will require more accurate knowledge concerning the distribution function of within-variance estimates. In actual sample surveys where fairly large samples are to be drawn from each stratum, our zero-order approximate solutions besides being simple in form, may serve the purpose of sufficiently approximate minimax solutions.

6. Preliminary considerations in the allocation of resources to a sequence of surveys

Under the general assumptions in §2, let us consider a sequence of surveys consisting of one preliminary and h aimed surveys among which our total cost C will be divided, thus:

$$C = a_0 C + a_1 C + \ldots + a_{h-1} C + a_h C \qquad \ldots \quad (6.01)$$

where $\qquad a_0 + a_1 + \ldots + a_h = 1, a_i > 0 \; (i = 0, 1, 2, \ldots, h).$

Our problem is to determine a set of these fractions $\{a_l\}$ $(l = 0, 1, \ldots, h)$ under the assumption of some sequential process of statistical inference. There may be various approaches to this problem. In this section, we propose to indicate the situations where our approximate approach as given in (1^0) below may be useful. We shall mention some formulae of sequential procedure which will suggest under what conditions our proposed approximate solutions may be useful.

Our procedure may be generally defined as follows :—

(0^0) Draw a stratified random sample O_0 which consists of k mutually independent random samples of size n_0 from each stratum where $kn_0 = a_0 C\bar{c}^{-1}$. Let $\hat{\sigma}_{0i}$ be the unbiased estimate of the i-th stratum standard deviation obtained from O_0.

(1^0) Let us define the sizes of samples to be drawn from the i-th stratum

$$n_{1i} = \frac{a_1 C}{\bar{c}} \quad \frac{N_i \hat{\sigma}_{0i}}{\sum\limits_{j=1}^{k} N_j \hat{\sigma}_{0j}} \qquad \ldots \quad (6.02)$$

for $i = 1, 2, \ldots, k$, and let us denote by O_1 this stratified random sample. Let $\bar{x}_{1i}$ and $\hat{\sigma}_{1i}$ be unbiased estimates of the i-th stratum mean and the i-th stratum standard deviation respectively, obtained from O_1.

(2^0) For $2 \leqslant l \leqslant h$, let us define the sizes of samples to be drawn from the i-th stratum

$$n_{li} = \frac{a_l C}{\bar{c}} \quad \frac{N_i \phi_l(\hat{\sigma}_{0i}, \hat{\sigma}_{1i}, \ldots, \hat{\sigma}_{l-1,i})}{\sum\limits_{j=1}^{k} N_j \phi_l(\hat{\sigma}_{0j}, \hat{\sigma}_{1j}, \ldots, \hat{\sigma}_{l-1,j})} \qquad \ldots \quad (6.03)$$

for $i = 1, 2, \ldots, k$, where ϕ_l are certain functions defined *a priori* and let us denote this stratified random sample by O_l. Let $\bar{x}_{li}$ and $\hat{\sigma}_{li}$ be unbiased estimates of the i-th stratum mean and the i-th stratum standard deviation respectively obtained from O_l.

In such situations the optimum allocation must necessarily depend both upon our real objects of estimation and upon the mathematical models on which our estimates will be based. Successive means and/or means depending on regression relations should be adopted under certain assumptions of the time change of the population means, and the successive pooling of estimates of standard deviations may depend upon some assumptions in which the time change of the within-variances are so slight as to make our poolings effective. These poolings of data were discussed by Bancroft (1944), Paull (1950) and the present author (1950a).

Here we shall consider the following cases:

Case A. This procedure S_A is characterised by the following two conditions:

(i) Our ultimate concern is with the last stage estimation of the grand mean defined by

$$\bar{x}_h = \frac{N_1 \bar{x}_{h1} + N_2 \bar{x}_{h2} + \ldots + N_k \bar{x}_{hk}}{N_1 + N_2 + \ldots + N_k}. \qquad \ldots \quad (6.04)$$

(ii) The sample sizes $\{n_{lj}\}$ are defined by

$$n_{lj} = \frac{a_l C}{\bar{c}} \cdot \frac{N_j \hat{\sigma}_{l-1,j}}{\sum_{i=1}^{k} N_i \hat{\sigma}_{l-1,i}} \quad \dots \quad (6.05)$$

that is to say $$\phi_l(\hat{\sigma}_{0j}, \hat{\sigma}_{1j}, \dots, \hat{\sigma}_{l-1,j}) = \hat{\sigma}_{l-1,j} \quad \dots \quad (6.06)$$

for $j = 1, 2, \dots, k$.

Case B. This procedure S_B is defined by the following two conditions :

(i) We are concerned with the estimation of the grand means over the h stages, namely, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{h-1}$ and $\bar{x}_h$.

(ii) The same as condition (ii) of Case A.

Case A. In this case we shall have the last stage unbiased estimation of the population mean with the variance

$$V_{h-1}(\bar{x}_h) = \frac{\bar{c}}{a_h C} \sum_{i=1}^{k} N_i \sigma_i \frac{\sigma_i}{\hat{\sigma}_{(h-1)i}} \sum_{j=1}^{k} N_j \sigma_j \frac{\hat{\sigma}_{(h-1)j}}{\sigma_j} \quad \dots \quad (6.07)$$

which we may write as

$$V_{h-1}(\bar{x}_h) \equiv W(\sigma; \hat{\sigma}_0, \hat{\sigma}_1, \dots, \hat{\sigma}_{(h-1)}, a_0, a_1, \dots, a_{h-1}, C), \quad \dots \quad (6.08)$$

where we have used vector notations such that

$$\begin{cases} \sigma = (\sigma_1, \sigma_2, \dots, \sigma_k), \\ \hat{\sigma}_l = (\hat{\sigma}_{l1}, \hat{\sigma}_{l2}, \dots, \hat{\sigma}_{lk}), \qquad (l = 0, 1, 2, \dots, h-1). \end{cases} \quad \dots \quad (6.09)$$

Our sampling procedure will define, under the assumption of the time invariance of each within-variance, the density function of the joint distribution of $(\hat{\sigma}_0, \hat{\sigma}_1, \dots, \hat{\sigma}_{h-1})$ such that

$$\prod_{l=0}^{h-1} \phi(\hat{\sigma}_l; \sigma, a_l / \hat{\sigma}_0, \dots, \hat{\sigma}_{l-1}, a_0, \dots, a_{l-1}, C) \equiv \phi(\hat{\sigma}_0, \hat{\sigma}_1, \dots, \hat{\sigma}_{(h-1)}; \sigma, a_0, \dots, a_{h-1}, C) \quad \dots \quad (6.10)$$

which may be written simply $\phi(\hat{\sigma}; \sigma, a, C)$.

Consequently the expected value of (6.10) will be given by

$$W(\sigma; a, C) \equiv \int W(\sigma; \hat{\sigma}, a, C) \phi(\hat{\sigma}; \sigma, a, C) d\hat{\sigma}, \quad \dots \quad (6.11)$$

which leads us to the following definition of the relative loss function

$$W_r(\sigma; a, C) \equiv W(\sigma; a, C) \left\{ \frac{c_0}{C} \left(\sum_{i=1}^{k} N_i \sigma_i \right)^2 \right\}^{-1}. \quad \dots \quad (6.12)$$

Hereafter we shall also apply Assumptions (1°) and (2°) in the following generalised sense

$$E\left\{\frac{\sigma_i}{\hat{\sigma}_{li}}\right\} = 1+\frac{\gamma}{n_{li}}+O\left(\frac{1}{n_{li}^2}\right) \qquad \dots \quad (6.13)$$

for every l-th stage and every i-th stratum where $l = 0, 1, 2, \dots, h-1$; $i = 1, 2, \dots, k$.

It is also necessary to make use of the conditional expectation when the values of the stochastic variables $\{\hat{\sigma}_{pi}\}$ $(p = 0, 1, \dots, l-1;\ i = 1, 2, \dots, k)$ are assumed to be assigned.

Our present purpose is to give the following assumption which leads us to simplified approximate formulae. Although the assumption will impose certain restriction upon the choice $a = (a_0, a_1, \dots, a_h)$ and the concurrence parameter point σ there seems to be no essential loss of generality so far as sample surveys of reasonably large scale are concerned.

Assumption (3°*) : In the allocation procedure defined in (6.03) let us assume that the allocation point $a = (a_0, a_1, \dots, a_h)$ and our concurrence parameter point $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_k)$ should be such that there are a domain of the points a and a domain D_a of the points a and a domain Ω^* of the points σ in which we have

$$\inf_{\substack{a\epsilon D_a \\ \sigma\epsilon\Omega^*}} \min_{1\leqslant j\leqslant k} \{n_{lj}\} = \mu_l > 0. \qquad \dots \quad (6.14)$$

There is now no difficulty in observing the following

Lemma 3.4 : *Under our generalised Assumptions* (1°), (2°) *and* (3°*) *in this section, we have for* $l = 1, 2, \dots, h$, and $j = 1, 2, \dots, k$ *the relations*

$$E_{l-1}\{n_{lj}^{-1}\} = \frac{\bar{c}}{a_l C}\left\{\frac{A_1}{N_j\sigma_j}+\frac{A_1-N_j\sigma_j}{N_j\sigma_j}\cdot\frac{\gamma_1}{n_{l-1j}}\left(1+O\left(\frac{1}{\mu_{l-1}}\right)\right)\right\} \qquad \dots \quad (6.15)$$

and hence for $l = 1, 2, \dots, h$

$$E_{l-1}\{V_{h-1}(\bar{z}_h)\} \Big/ \left(\frac{\bar{c}A_1^2}{a_h C}\right)$$

$$= 1+\sum_{s=l+1}^{h-1} b_s \prod_{\nu=s}^{h-1}\left\{\left(1+O\left(\frac{1}{\mu_\nu}\right)\right)\frac{g}{a_\nu}\right\}$$

$$+\prod_{\nu=l+1}^{h-1}\left\{\left(1+O\left(\frac{1}{\mu_\nu}\right)\right)\frac{g}{a_\nu}\right\}\sum_{j=1}^{k}\frac{(A_1-N_j\sigma_j)^{h-l}}{A_1(N_j\sigma_j)^{h-l-2}n_{lj}}\left(1+O\left(\frac{1}{\mu_l}\right)\right) \quad \dots \quad (6.16)$$

where

$$g \equiv \gamma_1\bar{c}/C, \qquad \dots \quad (6.17)$$

$$b_s \equiv A_1^{-1}\sum_{j=1}^{k}\frac{(A_1-N_j\sigma_j)^{h-s}}{(N_j\sigma_j)^{h-s-1}} \qquad \dots \quad (6.18)$$

for $s = h-1, h-2, \dots, 1, 0.$

Lemma 3.5 : *Under the Assumptions in Lemma 3.4, we have*

$$W_j(\sigma; a, C) \equiv E\{V_{k-1}\{x_k\}\}(A_1^*\bar{c}C^{-1})^{-1} = \frac{1}{a_k}\Big\{1+\sum_{s}^{k-1} b_s \prod_{r=s}^{k-1}\Big\{(1+O(\mu_r^{-1}))\frac{g}{a_r}\Big\}\Big\}. \quad \dots \quad (6.19)$$

Let us now consider the equations $\{b_s\}$.

Assumption (4°) : For $s = 0, 1, \dots, k-1 < \infty$ we have

$$\sup_{\sigma \in \Omega^*} \{b_s\} = B_s < \infty. \quad \dots \quad (6.20)$$

Lemma 3.6 : *Under the Assumption* (4°) *in addition to the general Assumptions in Lemma* 3.5 *we have*

$$\sup_{\sigma \in \Omega^*} W_j(\sigma; a, C) = \frac{1}{a_k}\Big\{1+\sum_{s=0}^{k-1} B_s \prod_{r=s}^{k-1}\Big\{(1+O(\mu_r^{-1}))\frac{g}{a_r}\Big\}\Big\}. \quad \dots \quad (6.21)$$

Case B. In this case we are concerned with

$$W_{\lambda,j}(\sigma; a, C) \equiv W_\lambda(\sigma; a, C)\{A_1^*\bar{c}C^{-1}\}^{-1} = \frac{C}{\bar{c}A_1^*}\sum_{i=1}^{k} \lambda_i E\{V_{i-1}\{x_i\}\}. \quad \dots \quad (6.22)$$

In view of the considerations developed in Theorem 3.2. Lemma 3.6 will give us directly

$$\sup_{\sigma \in \Omega^*} W_{\lambda,j}(\sigma; a, C) = \sum_{i=1}^{k} \frac{\lambda_i}{a_i}\Big\{1+\sum_{s=0}^{i-1} B_s \prod_{r=s}^{i-1}\Big\{(1+O(\mu_r^{-1}))\frac{g}{a_r}\Big\}\Big\} \quad \dots \quad (6.23)$$

provided that the assumptions in Lemma 3.6 are assumed here also, and (6.23) will suggest the approximate approach which makes use of certain fomulae derived from it.

We shall introduce

Definition 3.3 : An allocation point $a = (a_0, a_1, \dots, a_k), a_i \geqslant 0, a_0+\dots+a_k = 1$, is said to be the relative minimax solution of designing a sequence of sample surveys of the type S_B up to the zero-order if the point a gives the minimum value of the following algebraic form

$$\omega_\lambda(a) \equiv \sum_{i=1}^{k}\frac{\lambda_i}{a_i}\Big(1+B_0\frac{g}{a_{i-1}}\Big). \quad \dots \quad (6.24)$$

It is also noticed that without any restriction to the parameter point $B_0 = k-1$.

7. Comparison of relative efficiencies of various procedures upto the zero-order approximation

Now let us consider the following procedures each of which is a special case defined in §6 and in which we are concerned with the estimation of the grand means over the h stages.

Case I. A preliminary survey enunciated in §2 is used to allocate the sample sizes to each stratum not only for the estimation of $\bar{x}_1$ but also for estimating all the following means $\bar{x}_i (i = 2, 3, ..., h)$. That is to say we define the procedure S_I by

$$n_{lj} = \frac{a_l C}{\bar{c}} \cdot \frac{N_j \hat{\sigma}_{0j}}{\sum_{i=1}^{k} N_i \hat{\sigma}_{0i}} \qquad \text{... (7.01)}$$

for $l = 1, 2, ..., h$; $i = 1, 2, ..., k$.

Case II. This procedure S_{II} is the Case B discussed in §6.

Case III. This procedure S_{III} appeals to pooled estimation of within-variances, that is to say, we define

$$n_{lj} = \frac{a_l C}{\bar{c}} \cdot \frac{N_j(n_0\hat{\sigma}_{0j} + n_1\hat{\sigma}_{1j} + ... + n_{l-1}\hat{\sigma}_{l-1j})}{\sum_{i=1}^{k} N_i(n_0\hat{\sigma}_{0i} + n_1\hat{\sigma}_{1i} + ... + n_{l-1}\hat{\sigma}_{l-1i})} \qquad \text{... (7.02)}$$

for $l = 1, 2, ..., h$; $j = 1, 2, ..., k$.

Our preliminary considerations given in §6 give us under the Assumptions (1°), (2°) and (3°*) that the relative minimax solutions upto the zero-order will be concerned with minimizing the following algebraic forms respectively:

$$\omega_{\lambda, I}(a) \equiv \sum_{l=1}^{h} \frac{\lambda_l}{a_l}\left(1 + \frac{g_1}{a_0}\right), \qquad \text{... (7.03)}$$

$$\omega_{\lambda, II}(a) \equiv \sum_{l=1}^{h} \frac{\lambda_l}{a_l}\left(1 + \frac{g_1}{a_{l-1}}\right), \qquad \text{... (7.04)}$$

$$\omega_{\lambda, III}(a) \equiv \sum_{l=1}^{h} \frac{\lambda_l}{a_l}\left(1 + \frac{g_1}{a_{l-1} + a_{l-2} + ... + a_0}\right), \qquad \text{... (7.05)}$$

where

$$g_1 = (k-1)\gamma c_0 c^{-1}. \qquad \text{... (7.06)}$$

In order that these algebraic forms may be useful as reasonably good approximations to each of their true relative minimum solution, there are certain restrictions as to the domains to which $a = (a_0, a_1, ..., a_h)$ and $\sigma = (\sigma_1, \sigma_2, ..., \sigma_k)$ should belong as we have pointed out in (7.04) in particular. The constants should be compared after obtaining approximate solutions to (7.03), (7.04) and (7.05) in order that we may be convinced that these will serve as approximate relative minimum solutions. Inspite of their simpler expressions the exact solutions which give the minimum values of (7.04) and (7.05) respectively are not easy to give in any compact mathematical formula. We shall therefore adopt some broad views in which effective differences between these three procedures may be briefly observed.

Case I. The problem of minimizing (7.03) under the restriction $a_0+a_1+\ldots+a_k = 1$, $a_i \geqslant 0$ $(i = 0, 1, \ldots, k)$ can be exactly solved by the usual method of Lagrange's undetermined multipliers. Thus we get

$$\frac{\lambda_1^{\frac{1}{2}}}{a_1} = \frac{\lambda_2^{\frac{1}{2}}}{a_2} = \ldots = \frac{\lambda_k^{\frac{1}{2}}}{a_k} = \frac{\lambda_1^{\frac{1}{2}}+\ldots+\lambda_k^{\frac{1}{2}}}{1-a_0}, \qquad \ldots \quad (7.07)$$

which gives $$\min_a \ \omega_I(a) = \frac{H^2}{1-a_0}\Big(1+\frac{g_1}{a_0}\Big) = H^2(1+g_1)\Big\{1+\Big(\frac{g_1}{1+g_1}\Big)^{\frac{1}{2}}\Big\}, \qquad \ldots \quad (7.08)$$

where $$H = \sum_{i=1}^{k} \lambda_i^{\frac{1}{2}}. \qquad \ldots \quad (7.09)$$

Case II. The problem of minimizing (7.04) under the restriction $a_0+a_1+\ldots+a_k = 1$, $a_i \geqslant 0$ may be solved approximately as follows. Instead of applying the usual procedure we shall make use of a certain point $\hat{a}$ such that

$$\frac{\lambda_1^{\frac{1}{2}}}{\hat{a}_1}\Big(1+\frac{g_1}{\hat{a}_0}\Big) = \frac{\lambda_2^{\frac{1}{2}}}{\hat{a}_2}\Big(1+\frac{g_1}{\hat{a}_1}\Big) = \ldots = \frac{\lambda_k^{\frac{1}{2}}}{\hat{a}_k}\Big(1+\frac{g_1}{\hat{a}_{k-1}}\Big), \qquad \ldots \quad (7.10)$$

These equations have some resemblance to (7.07) and may be considered to give an approximate solution.

Indeed (7.10) gives us

$$\omega_{II}(\hat{a}) = \frac{H^2}{1-\hat{a}_0}\Big\{1+\frac{g_1}{H}\sum_{i=1}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\hat{a}_{i-1}}\Big\} \qquad \ldots \quad (7.11)$$

for which we have $$\min_a \ \omega_{II}(a) \leqslant \omega_{II}(\hat{a}). \qquad \ldots \quad (7.12)$$

Case III. The problem of minimizing (7.05) under the restriction $a_0+a_1+\ldots+a_k = 1$, $a_i \geqslant 0$ may be solved approximately as in Case II. Thus we shall be concerned with point $\bar{a}$ for which

$$\frac{\lambda_1^{\frac{1}{2}}}{\bar{a}_1}\Big(1+\frac{g_1}{\bar{a}_0}\Big) = \frac{\lambda_2^{\frac{1}{2}}}{\bar{a}_2}\Big(1+\frac{g_1}{\bar{a}_0+\bar{a}_1}\Big) = \ldots = \frac{\lambda_k^{\frac{1}{2}}}{\bar{a}_k}\Big(1+\frac{g_1}{\sum_{i=0}^{k-1}\bar{a}_i}\Big) \qquad \ldots \quad (7.13)$$

which yields us $$\omega_{III}(\bar{a}) = \frac{H^2}{1-\bar{a}_0}\Big(1+\frac{g_1}{H}\sum_{i=1}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\bar{a}_0+\ldots+\bar{a}_{i-1}}\Big) \quad \ldots \quad (7.14)$$

and which is subject to $$\min_{a}\ \omega_{III}(a) < \omega_{III}(\bar{a}). \quad \ldots \quad (7.15)$$

In order to compare the relative magnitudes among (7.08), (7.11) and (7.14) the following observations may be of some use. In the first place let us compare (7.11) with (7.08). For this purpose let us rewrite the former as

$$\omega_{II}(\hat{a}) = \frac{H^2}{1-\hat{a}_0}\Big\{1+\frac{g_1}{\hat{a}_0}\Big(\frac{\lambda_1^{\frac{1}{2}}}{H}+\sum_{i=2}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\lambda_{i-1}^{\frac{1}{2}}}\cdot\frac{\lambda_{i-1}^{\frac{1}{2}}}{H}\frac{\hat{a}_0}{\hat{a}_{i-1}}\Big)\Big\}$$

$$= \frac{H^2}{1-\hat{a}_0}\Big\{1+\frac{g_1}{\hat{a}_0}\Big(\frac{\lambda_1^{\frac{1}{2}}}{H}+\frac{\hat{a}_0}{1-\hat{a}_0}\sum_{i=2}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\lambda_{i-1}^{\frac{1}{2}}}\times\Big(1+\frac{g_1}{a_{i-2}}\Big)\Big(1+\frac{g_1}{H}\sum_{i=1}^{k}\frac{\lambda_i}{a_{i-1}}\Big)\Big\} \quad (7.16)$$

which broadly speaking may be approximated by a simpler expression

$$\omega_{II}(\hat{a}) = \frac{H^2}{1-\hat{a}_0}\Big\{1+\frac{g_1}{\hat{a}_0}\Big(\frac{\lambda_1^{\frac{1}{2}}}{H}+\frac{\hat{a}_0}{1-\hat{a}_0}\sum_{i=2}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\lambda_{i-1}^{\frac{1}{2}}}\Big)\Big\} \quad \ldots \quad (7.17)$$

Comparing with the formulae (7.08) we shall find that

$$\frac{\omega_{II}(\hat{a})}{\min_{a}\omega_I(a)} \simeq \Big(1+\frac{g}{\hat{a}_0}\Big(\frac{\lambda_1^{\frac{1}{2}}}{H}+\frac{\hat{a}_0}{1-\hat{a}_0}\sum_{i=2}^{k}\frac{\lambda_i^{\frac{1}{2}}}{\lambda_{i-1}^{\frac{1}{2}}}\Big)\Big)\Big(1+\frac{g}{a_0}\Big)^{-1}, \quad \ldots \quad (7.18)$$

where a_0 is the value defined in Case I.

Secondly in order to compare (7.11) and (7.14), the following relation will be of some use

$$\frac{\omega_{III}(a)}{\omega_{II}(a)} < \Big(1+\frac{g}{\hat{a}_0}\sum_{i=1}^{k}\frac{\lambda_i}{\hat{a}_0+\ldots+\hat{a}_{i-1}}\Big)\Big(1+\frac{g}{\hat{a}_0}\sum_{i=1}^{k}\frac{\lambda_i}{\hat{a}_{i-1}}\Big)^{-1} \quad \ldots \quad (7.19)$$

where the right-hand side may be approximated under certain conditions such as

$$1-\frac{g}{\hat{a}_0}\sum_{i=1}^{k}\frac{\lambda_i(\hat{a}_0+\ldots+\hat{a}_{i-2})}{\hat{a}_{i-1}(\hat{a}_0+\ldots+\hat{a}_{i-1})}. \quad \ldots \quad (7.20)$$

8. Detailed comparison of various procedures in view of the abbreviated first order approximation

In this section Cases I and II will be compared more accurately by using the abbreviated first order approximation introduced in §5. On the contrary we shall restrict ourselves to the case when $h = 2$.

For Case I, we shall make use of the following relation

$$\omega_I(a) = \frac{\lambda_2}{a_2}\left(1 + \frac{g_1}{a_0}\right) + \frac{\lambda_1}{a_1}\left(1 + \frac{g_1}{a_0}\right), \qquad \ldots \quad (8.01)$$

which is a special case of that treated in §7. On the other hand for Case II, we shall use the one introduced in Lemma 3.3(3) which will now be written as

$$\omega_{II}^*(a) = \frac{\lambda_2}{a_2}\left\{1 + \frac{g_1}{a_1} + \frac{g_1^2}{ka_0a_1}\right\} + \frac{\lambda_1}{a_1}\left\{1 + \frac{g_1}{a_0}\,\frac{k-1}{k}\right\} \qquad \ldots \quad (8.02)$$

in accordance with the abbreviated first order approximation.

Case I has already been discussed in §7, and let the point $a = (a_0, a_1, a_2)$ be the one where $\omega_I(a)$ will take its minimum value. As we have already seen in §7, this point will be defined as

$$a_0 = (g_1 + g_1^2)^{\frac{1}{2}} - g_1, \qquad \ldots \quad (8.03)$$

$$\frac{a_1}{\lambda_1^{\frac{1}{2}}} = \frac{a_2}{\lambda_2^{\frac{1}{2}}} = \frac{1-a_0}{\lambda_1^{\frac{1}{2}} + \lambda_2^{\frac{1}{2}}}. \qquad \ldots \quad (8.04)$$

On the other hand the exact solution of the point $\hat{a} = (\hat{a}_0, \hat{a}_1, \hat{a}_2)$ which will minimize (8.02) is still difficult to be obtained in any explicit form. One approach may be as follows. First let us define an approximate value $\hat{a}_0^*$ of $\hat{a}_0$ as the value a_0 defined in (8.04) and then let us define $(\hat{a}_1^*, \hat{a}_2^*)$, $\hat{a}_1^* + \hat{a}_2^* = 1 - a_0$ so as to minimize (8.02) under this restriction.

For the latter part, we shall make use of the following

Lemma 3.7 : *Let A, B, C and Θ be a set of any assigned positive numbers and let $a_i (i = 1, 2)$ be positive numbers such that $a_1 + a_2 = \Theta$. Then we have the following assertions:*

(a) *The allocation $(\hat{a}_1^*, \hat{a}_2^*)$ which will minimize the function*

$$V = \frac{A}{a_1} + \frac{B}{a_2} + \frac{C}{a_1a_2} \qquad \ldots \quad (8.05)$$

is given by

$$\hat{a}_1^* = \frac{\Theta(A\Theta + C)^{\frac{1}{2}}}{(A\Theta + C)^{\frac{1}{2}} + (B\Theta + C)^{\frac{1}{2}}} \qquad \ldots \quad (8.06)$$

$$\hat{a}_2^* = \frac{\Theta(B\Theta + C)^{\frac{1}{2}}}{(A\Theta + C)^{\frac{1}{2}} + (B\Theta + C)^{\frac{1}{2}}}, \qquad \ldots \quad (8.07)$$

which yield us the following minimum value of V

$$V_{\min} = \theta^{-2}\{(A\theta+C)^{\frac{1}{2}}+(B\theta+C)^{\frac{1}{2}}\}^2. \qquad \text{... (8.08)}$$

(b) *The allocation* $a = (a_1, a_2)$ *defined by the relation*

$$a_1 = \theta A^{\frac{1}{2}}(A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-1}, \qquad \text{... (8.09)}$$

$$a_2 = \theta B^{\frac{1}{2}}(A^{\frac{1}{2}}+B^{\frac{1}{2}})^{-1}, \qquad \text{... (8.10)}$$

gives us

$$V_a = \theta^{-2}(A^{\frac{1}{2}}+B^{\frac{1}{2}})(\theta+CA^{-\frac{1}{2}}B^{-\frac{1}{2}}). \qquad \text{... (8.11)}$$

(c) *The difference between the two values* (8.08) *and* (8.11) *is approximately given by*

$$V_a - V_{\min} = \frac{(A-B)^2C^2}{4A^{3/2}B^{3/2}\theta^3} \qquad \text{... (8.12)}$$

provided that C is so small compared with $A\theta$ *and* $B\theta$ *that* $(1+C(A\theta)^{-1})^{\frac{1}{2}}$ *and* $(1+C(B\theta)^{-1})^{\frac{1}{2}}$ *may be sufficiently approximated by the Taylor series upto the second term.*

The proof is immediate. To apply to our present case it suffices to put $\theta = 1-a_0$ and to define

$$A = \lambda_1\left(1+\frac{g_1}{1-\theta}\,\frac{k-1}{k}\right), \qquad \text{... (8.13)}$$

$$B = \lambda_2, \qquad \text{... (8.14)}$$

$$C = \lambda_2 g_1\left(1+\frac{g_1}{1-\theta}\cdot\frac{1}{k}\right). \qquad \text{... (8.15)}$$

Theorem 3.4 : *Under the general assumptions in this section, we have the following assertions :*

(1) *The relative difference between the procedures* S_I *and* S_{II} *under the allocation* $a = (a_0, a_1, a_2)$ *which minimizes* $w_I(a)$ *is given by*

$$\{\omega_I(a)-\omega_{II}(a)\}/\omega_I(a) = \left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\left\{1+\left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\right\}\times$$

$$\times W\left\{\frac{1}{1+W} - \{(g_1+g_1^2)^{\frac{1}{2}}-g_1\}\,\frac{2+W}{1+W}\right\}+$$

$$+\frac{1}{k}\left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\left\{1+\left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\right\}\times$$

$$\times\left\{\frac{1}{1+W} - Wg_1\left(1+\left(\frac{g_1}{1+g_1}\right)^{\frac{1}{2}}\right)\right\} \qquad (8.16)$$

where

$$W = \lambda_2^{\frac{1}{2}}/\lambda_1^{\frac{1}{2}}.$$

(2) *The difference between* $\omega_{II}(a)$ *and* $\omega_{II}(\hat{a}^*)$ *where* $a = (a_0, a_1, a_2)$ *and* $\hat{a}^* = (\hat{a}^*_0, \hat{a}^*_1, \hat{a}^*_2)$ *is approximately given by*

$$\omega_{II}(a) - \omega_{II}(\hat{a}^*) = \frac{\lambda_1^{\frac{1}{2}}\lambda_2^{\frac{1}{2}}\left(W^2 - \left(1 + \frac{g_1(k-1)}{(1-\theta)k}\right)\right)^2\left(1 + \frac{g_1}{(1-\theta)k}\right)^2}{4\theta^3\left(1 + \frac{g_1}{1-\theta}\cdot\frac{k-1}{k}\right)^{3/2}} \quad \dots \quad (8.17)$$

provided that the following two conditions are satisfied :

$$g_1(1-\theta)k^{-1} < 1, \quad \dots \quad (8.18)$$

$$\lambda_2 g_1 < \lambda_1 \theta. \quad \dots \quad (8.19)$$

Broadly speaking (8.17) will be approximated by

$$g_1^{\frac{1}{2}}(W + \tfrac{1}{k})(1+W)^{-1} \quad \dots \quad (8.20)$$

and (8.18) by

$$(\lambda_1^{\frac{1}{2}} + \lambda_2^{\frac{1}{2}})W(W-1)^2 g^2 (4\theta^3)^{-1}. \quad \dots \quad (8.21)$$

References

Bancroft, T. A. (1944): On biases in estimation due to the use of preliminary tests of significance. *Ann. Math. Stat.*, 15, 190–204.

Birnbaum, Z. W. and Sirken, M. G. (1950): On the total error due to non-interview and to random sampling. *Int. Opinion and Attitude Res.*, 4, 179–191.

Cochran, W. G. (1939): The use of analysis of variance in enumeration by sampling. *J. Amer. Stat. Ass.*, 34, 492–510.

——— (1953): *Sampling Techniques*, John Wiley & Sons, New York.

Cox, D. R. (1952): Estimation by double sampling. *Biometrika*, 39, 217–227.

Deming, W. E. (1944): On errors in surveys. *Amer. Sociological Review*, 9, 359–69.

——— (1950): *Some Theory of Sampling*, John Wiley and Sons, New York.

Ghosh, B. (1947): Bias introduced by changing the system of stratification. *Bull. Cal. Stat. Ass.*, 1, 43–45.

Hansen, H. M. and Hurwitz, W. N. (1943): On the theory of sampling from finite populations. *Ann. Math. Stat.*, 14, 333–362.

——— (1946): The problem of non-response in sample surveys. *J. Amer. Stat. Ass.*, 41, 517–529.

Hansen, H. M., Hurwitz, W. N., Marks, E. S. and Mauldin, W. P. (1951): Response errors in surveys. *J. Amer. Stat. Ass.*, 46, 147–190.

Hubback, J. A. (1927): Sampling for rice yield in Bihar and Orissa. *Bull. Agr. Res. Inst. Pusa*, No. 166, 1–23.

Jessen, R. J. (1942): Statistical investigation of a sample survey for obtaining farm factor. *Iowa Agric. Exp. Stat. Res. Bull.*, 304.

Jessen, R. J. and Houseman, E. E. (1944): Statistical investigation of farm sample survey taken in Iowa, Florida and California. *Iowa. Agric. Exp. Sta. Res. Bull.*, 329.

Kinashi, K. (1953): Forest investories by sampling methods. *Bull. Kyushu Univ. Forests*, No. 23, 1–153.

KING, A. J. AND JESSEN, R. J. (1945) : The master sample of agriculture. *J. Amer. Stat. Ass.*, 40, 38–56.

KITAGAWA, T. (1950a) : Successive process of statistical inferences (1). *Mem. Fact. Sci. Kyushu Univ.*, Ser. A, 5, 139–180.

——— (1950b) : Estimation-formulae used in the sampling survey of the Fishing catches in Fukuoka Prefecture : Appendix to Sampling survey of fishery catch in Fukuoka Prefecture. Statistics and Research Division, Agricultural Improvement, Bureau, Ministry of Agriculture and Forestry, 1–39.

——— (1951a) : Successive process of statistical inferences (2). *Mem. Frac. Sci. Kyushu Univ.*, Ser. A, 6, 55–95.

——— (1951b) : Random integration. *Bull. Math. Stat.*, 4.

——— (1952a) : Successive process of statistical inferences (3). *Mem. Frac. Sci. Kyushu Univ.*, Ser. A, 7, 131–135.

——— (1952b) : Successive process of statistical inferences (4). *Bull. Math. Stat.*, 5, 35–50.

——— (1953a) : Successive process of statistical inferences (5). *Mem. Frac. Sci. Kyusyu Univ.*, Ser. A, 7, 95–120.

——— (1953b) : Successive process of statistical inferences (6). *Mem. Fact. Sci. Kyusyu Univ.*, Ser. A, 8, 1–29.

——— (1953c) : Some stochastic considerations concerning empirical functions. *Bull. Math. Stat.*, 5, 19–33.

KITAGAWA, T. AND FUJITA, T. (1951) : Sampling surveys of living costs of labourers in Mieke coal mine, (in Japanese). Toyo Keizai Shimpo, Tokyo.

MAHALANOBIS, P. C. (1940) : A sample survey of the acreage under jute in Bengal. *Sankhyā*, 4, 511–530.

——— (1944) : On large-scale sample surveys. *Phil. Trans. Roy. Soc.*, B, 231, 329–451.

——— (1946) : Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Stat. Soc.*, 109, 325–370.

——— (1952) : Some aspects of the design of sample surveys. *Sankhyā*, 12, 1–7.

NEYMAN, J. (1925) : Contributions to the theory of small sample drawn from a finite population. *Biometrika*, 27, 472–478.

——— (1934) : On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection. *J. Roy. Stat. Soc.*, 97, 558–606.

——— (1938) : Contribution to the theory of sampling human populations. *J. Amer. Stat. Ass.*, 33, 101–116.

PAULL, A. E. (1950) : On a preliminary test for pooling mean squares in the analysis of variance. *Ann. Math. Stat.*, 21, 539–556.

ROBBINS, H. (1952) : Some aspects of the sequential designs of experiments. *Bull. Amer. Math. Soc.*, 58, 527–535.

ROBBINS, H. AND MONRO, S. (1951) : A stochastic approximation method. *Ann. Math. Stat.*, 22 400–407.

STEPHAN, F. F. (1945) : The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate. *Ann. Math. Stat.*, 16, 50–61.

SUKHATME, P. V. (1935) : Contribution to the theory of the representative method. *J. Roy. Stat. Soc.*, *Supplement*, 2, 253–268.

WALD, A. (1950) : *Statistical Decision Functions*, John Wiley and Sons, New York.

YATES, F. (1939) : *Sampling Methods for Censuses and Surveys*, Charles Griffin & Co., London.

Paper received : August, 1953.