OPTIMUM STRATIFIED SAMPLING USING PRIOR INFORMATION

Kallappa M. Koti, Southern Methodist University, Dallas, TX 75275

Benjamin S. Duran and Truman O. Lewis, Texas Tech University

Key words and phrases: Superpopulation model; conditional expectation; squared-error loss; maximum ξp-risk; data based prior; inverse gamma distribution, preposterior variance.

1 INTRODUCTION

Consider a population $U = (u_{ij})$, $j = 1, 2, \cdots, N_i$; $i = 1, 2, \cdots, k$, of distinguishable elements which are classified into k strata, the size of the ith stratum being N_i, a known integer and $N = N_1 + N_2 + \cdots + N_k$. Let $\underline{\pi} = (\pi_1, \pi_2, \cdots, \pi_k)$, where $\pi_i = N_i/N$. With each element u_{ij} of U, is associated a real unknown value $y_{ij} = y(u_{ij})$. The N-dimensional vector $\underline{y} = (y_{ij})$, $j = 1, 2, \cdots, N_i$; $i = 1, 2, \cdots, k$, is called the parameter vector of a Euclidean N-space R_N. We associate with the parameter-space R_N the couple $(\mathcal{B}^N, \mathcal{C})$ where $\mathcal{B}^N$ is the Borel σ-field on R_N, and $\mathcal{C}$ is a family of prior probability measures on $(R_N, \mathcal{B}^N)$. The class $\mathcal{C}$ of probability distributions ξ on R_N, $\mathcal{C} = \{\xi\}$, is called a superpopulation model. The population vector y is unknown and a complete survey is not feasible. Hence the objective is to infer about the unknown population mean $\bar{y} = (1/N)\sum_i \sum_j y_{ij}$, on the basis of a sample.

Any subset s of U is called a sample. Let $\mathcal{S} = \{s : s \subset U\}$. A sampling design is a function $p : \mathcal{S} \to [0, 1]$ such that $s \subset \mathcal{S}$, $p(s) \geq 0$ and $\Sigma_{\mathcal{S}} p(s) = 1$. A stratified sample is a vector $\underline{n} = (n_1, n_2, \cdots n_k)$, where $n_i \geq 0$, $i = 1, 2, \cdots, k$, is the number of observations drawn independently from the ith stratum. It is assumed that sampling is without replacement. Let $(\mu_1, \mu_2, \cdots, \mu_k)$ be the vector of y-means of k strata so that

$$\bar{y} = \sum_{i=1}^{k} \pi_i \mu_i. \tag{1.1}$$

Let $\bar{y}_{n_i}$ be the arithmetic mean of a random sample of size n_i taken from the ith stratum, $i = 1, 2, \cdots, k$. In standard texts such as Cochran (1977), the stratified sample mean

$$\bar{y}_{st} = \sum_{i=1}^{k} \pi_i \bar{y}_{n_i} \tag{1.2}$$

is proved to be p-unbiased with p-variance

$$V(\bar{y}_{st}) = \sum_{i=1}^{k} \pi_i^2 S_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right), \tag{1.3}$$

where

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (y_{ij} - \mu_i)^2. \tag{1.4}$$

The allocation of resources in a stratified sampling design aimed at estimating the population mean is usually carried out, keeping in mind either one of the following two alternatives: (i) to achieve maximum precision for a given total cost of the survey, or (ii) to achieve a given precision at a minimum cost. The well known Neyman optimum allocation is one, based on this type of approach. From the Neyman optimum allocation (c.f., Cochran (1977)), it is apparent that the statistician planning the stratified sample survey needs to have some prior information on the behavior of the character under consideration, other than the size of each stratum. The available literature on optimum stratified sampling using prior information consists of the papers by Aggarwal (1959), Ericson (1965), Draper and Guttman (1968), Zacks (1970), Rao and Ghangurde (1972), and many others. An excellent review and criticism of many of these is found in Solomon and Zacks (1970). The objective of this paper is to present noninformative as well as two-phase minimax type stratified sampling designs.

2 NONINFORMATIVE DESIGNS

Let c_i, $i = 1, 2, \cdots, k$, be the per unit sampling cost in the ith stratum. Consider the augmented bounded squared error loss

$$L = L(\bar{y}, \bar{y}_{st}|\underline{n}) = \beta(\bar{y}_{st} - \bar{y})^2 + \sum_{i=1}^{k} c_i n_i, \tag{2.1}$$

where $\beta(> 0)$ is a weighting constant. We know that the p-expectation of L, given $\underline{n} = (n_1, n_2, \cdots, n_k)$, is

$$E_p(L\,|\underline{n}) = \beta \sum_{i=1}^{k} \pi_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2 + \sum_{i=1}^{k} c_i n_i, \tag{2.2}$$

where S_i^2 is defined in (1.4). Let all N_i's be fairly large so that $(N_i - 1)^{-1} \simeq N_i^{-1}$. At this point we assume for each $j = 1, 2, \cdots, N_i$, that

$$E_\xi\,(y_{ij} - \mu_i)^2 = \Theta_i, \quad i = 1, 2, \cdots, k.$$

Or one may consider stratum-i to be a sample from a superstratum whose variance with respect to y is Θ_i. Thus we have the ξp-risk R for a given vector as follows.

$$\begin{aligned} R(\underline{n}) &= E_\xi E_p(L|\underline{n}) \\ &= \beta \sum_{i=1}^{k} \pi_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \Theta_i + \sum_{i=1}^{k} c_i n_i. \end{aligned} \tag{2.3}$$

The concept of maximum ξp-risk of this paper, is based on the assumption that $\Theta_1, \Theta_2, \cdots, \Theta_k$ are random variables with known distribution functions $F_1, F_2, \cdots, F_k$, respectively. It is also assumed that $E(\Theta_i) < \infty$, $i = 1, 2, \cdots, k$. We use $\underline{\theta}^t$ to denote the vector $(\theta_1, \theta_2, \cdots, \theta_k)$. Let

$$\Delta_i = \left\{ \underline{\theta} \in R^k : \quad \theta_i \geq \max_{1 \leq j \leq k} (\theta_j, j \neq i) \right\}, \; i = 1, 2, \cdots, k. \tag{2.4}$$

We assume that $p(\Delta_i) > 0$ for each $i = 1, 2, \cdots, k$. We define the maximum ξp-risk $R_1(\underline{n})$ as follows.

Definition 2.1 *Let $\underline{n} = (n_1, n_2, \cdots, n_k)$ be a given vector.*

$$R_1(\underline{n}) - \sum_{i=1}^{k} c_i n_i =$$

$$\begin{cases} \pi_1^2 \left(\frac{1}{n_i} - \frac{1}{N-i}\right) \beta \Theta_1 + \sum_{j=2}^{k} \pi_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j}\right) \sup(\Theta_j) \\ \text{if } \underline{\Theta} \in \Delta_1, \\ \cdots \qquad \cdots \\ \pi_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \beta \Theta_i + \sum_{j \neq i}^{k} \pi_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j}\right) \sup(\Theta_j) \\ \text{if } \underline{\Theta} \in \Delta_k, \\ \cdots \qquad \cdots \\ \pi_k^2 \left(\frac{1}{n_k} - \frac{1}{N_k}\right) \beta \Theta_k + \sum_{j=1}^{k-1} \pi_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j}\right) \sup(\Theta_j) \\ \text{if } \underline{\Theta} \in \Delta_k. \end{cases}$$

Assume that $\Theta_1, \Theta_2, \cdots, \Theta_k$ are nonnegative random variables having continuous support over $(0, \infty)$.

Theorem 2.1 *Let $r_1(\underline{n}) = E[R_1(\underline{n})]$.*

$$r_1(\underline{n}) = \sum_{i=1}^{k} \left\{ \pi_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \beta \tau_i + c_i n_i \right\},$$

where

$$\tau_i = \sum_{j=1}^{k} \int_0^\infty x f_j(x) \left(\prod_{\ell \neq j}^{k} F_\ell(x) \right) dx.$$

Proof: See the proof of Theorem 3.1.

Remark 1 *Note that*

$$R_1(\underline{n}) \geq R(\underline{n}) \quad a.s..$$

Therefore, by monotonocity property of mathematical expectation, we have

$$r_1(\underline{n}) = E[R_1(\underline{n})] \geq E[R(\underline{n})].$$

Thus the adjective 'maximum' when referred to R_1 makes sense.

Remark 2 *If $F_1, F_2, \cdots, F_k$ have continuous support over $(0, \infty)$, all τ_i's are identical. Derivation of τ_i's for a case of $k = 2$, when F_1, F_2 are discrete, is discussed in Koti (1988). Next subsection deals with a case of $k = 2$, when F_1, F_2 are uniform cumulative distribution functions.*

The idea is to obtain $\underline{n}^*$ that minimizes r_1. The resulting solution is the desired stratified sample minimax allocation.

Aggarwal (1959) considered the Bayes risk

$$r_a = \sum_{i=1}^{k} \left[\pi_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \beta \sigma_i^2 + c_i n_i \right],$$

where σ_i^2 ($i = 1, 2, \cdots, k$) was assumed to be a known upper bound for $E_\xi(y_{ij} - \mu_i)^2$. He showed that

$$n_{ia}^* = \text{ integer nearest to } \left(\frac{\pi_i^2 {\sigma_i}^2}{c_i} \beta + \frac{1}{4} \right)^{\frac{1}{2}} \text{ and } \leq N_i$$

minimizes r_a.

Theorem 2.2 n_i^* = *integer nearest to* $\left(\frac{\pi_i^2 \tau_i}{c_i} \beta + \frac{1}{4} \right)^{\frac{1}{2}}$ *and* $\leq N_i$.

Proof: Follows from Aggarwal (1959).

2.1 An Example

Consider the case of two ($k = 2$) strata. Assume that strata variances Θ_0 and Θ_1 have prior uniform distributions, respectively, over (a_0, b_0) and (a_1, b_1). The constants a_0, a_1, b_0, b_1 are positive and

$$a_0 < a_1 < b_0 < b_1.$$

Then

$$\tau_0 = \int_{a_1}^{b_0} \frac{x}{b_0 - a_0} \left(\int_{a_1}^{x} \frac{du}{b_1 - a_1} \right) dx + \int_{a_1}^{b_0} \frac{x}{b_1 - a_1} \left(\int_{a_0}^{x} \frac{du}{b_0 - a_0} \right) dx.$$

On integrating and simplifying we obtain

$$(b_0 - a_0)(b_1 - a_1)\tau_0 = \frac{2}{3} \left(b_0^3 - a_1^3 \right) - \frac{a_0 + a_1}{2} \left(b_0^2 - a_1^2 \right).$$

Next by definition,

$$\begin{aligned} \tau_1 &= \int_{a_1}^{b_0} \frac{x}{b_0 - a_0} \left(\int_{a_1}^{x} \frac{du}{b_1 - a_1} \right) dx + \\ & \int_{a_1}^{b_1} \frac{x}{b_1 - a_1} \left(\int_{a_0}^{x} \frac{du}{b_0 - a_0} \right) dx \\ &= \int_{a_1}^{b_0} \frac{x}{b_0 - a_0} \left(\int_{a_1}^{x} \frac{x}{b_1 - a_1} \right) dx + \\ & \int_{a_1}^{b_0} \frac{x}{b_1 - a_1} \left(\frac{du}{b_0 - a_0} \right) dx \\ &+ \int_{b_0}^{b_1} \frac{x}{b_1 - a_1} dx. \end{aligned}$$

$$\text{i.e.,} \quad \tau_1 = \tau_0 + \frac{1}{2} \left(b_1^2 - b_0^2 \right) / (b_1 - a_1).$$

Consider the following constants.

$$c_0 = c_1 = 1.0$$

$$\pi_0 = \pi_1 = 0.5$$

$$a_0 = 10, a_1 = 20, b_0 = 110, b_1 = 125.$$

We find $\tau_0 = 67.2857$ and $\tau_1 = 84.0714$. Finally applying Theorem 2.2, we get

$$n_0^* = 130 \text{ and } n_1^* = 145.$$

3 TWO-PHASE STRATIFIED SAMPLING DESIGNS

We adopt the framework of Draper and Guttman (1968). Suppose we have k strata from which arise observations x_{ij} (or y_{ij}) $(i = 1, 2, \cdots, k)$. Assume that known proportion π_i of the total population lies in the ith stratum, where $\pi_1 + \pi_2 + \cdot + \pi_k = 1$. The principal objective of the sample survey is to estimate the overall mean $\mu = \Sigma\pi_i\mu_i$ as precisely as possible. The letters x and y will refer to the first and second phases, respectively. Let x_{ij} (or y_{ij}) $\sim N(\mu_i, \theta_i)$. We assume that the prior information available before the first phase sampling can be represented by independent, locally uniform prior distributions on μ_i and θ_i so that

$$P(\mu_i)d\mu_i \;\alpha\; d\,\mu_i \quad \text{and} \quad p(\theta_i)d\theta_i \;\alpha\; \frac{d\theta_i}{\theta_i} \tag{3.1}$$

Suppose that a sample $\{x_{ij}\}(i = 1, 2, \cdots, k)$ of $m_i(> 3)$ observations is taken in the first phase of sampling from the ith stratum. These observations are used to enhance the prior information concerning θ_i's. Let n_i denote the number of observations taken from the ith stratum in the second phase of sampling. The n_i's will be chosen by minimizing a maximum Bayes risk defined below.

3.1 Maximum Posterior Risk: Definition and Expectation

Again we borrow some notation and results from Draper and Guttman (1968). Let

$$m_i\bar{x}_i = \sum_{j=1}^{m_i} x_{ij} \quad \text{and} \quad (m_i - 1)s_i^2 = \sum_{j=1}^{m_i}(x_{ij} - \bar{x}_i)^2.$$

Also let

$$n_i\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} \quad \text{and} \quad (n_i - 1)w_i^2 = \sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2.$$

In order to obtain an allocation for the n_i, we must make a decision before the second-phase observations are available. It is appropriate that this decision be based on the variances of $\Sigma\pi_i\mu_i$ and the sampling cost for the second phase. Again c_i $(i = 1, 2, \cdots, k)$ denotes the cost of taking one observation from the ith stratum.

Draper and Guttman (1968) showed that

$$(m_i + n_i - 3)(m_i + n_i)V(\mu_i) = \frac{m_i n_i}{m_i + n_i}(\bar{x}_i - \bar{y}_i)^2 + (m_i - 1)s_i^2 + (n_i - 1)w_i^2. \tag{3.2}$$

Note that the conditional expectation of $V(\mu_i)$ given $\Theta_i = \theta_i$ $(i = 1, 2, \cdots k)$ is

$$E(V_i|\theta_i) = \frac{(m_i - 1)\frac{m_i - 1}{m_i - 3}s_i^2 + (n_i - 1)\theta_i}{(m_i + n_i - 3)(m_i + n_i)} \tag{3.3}$$

Let

$$\nu_i = (m_i - 1)\frac{m_i - 2}{m_i - 3}s_i^2 \tag{3.4}$$

and

$$\alpha_i = \frac{\pi_i^2}{(m_i + n_i - 3)(m_i + n_i)}. \tag{3.5}$$

We define $R_2(\underline{n})$, the maximum posterior risk as follows.

Definition 3.1 *Let $\underline{n} = (n_1, n_2, \cdots, n_k)$ be a given vector.*

$$R_2(\underline{n}) - \sum_{i=1}^{k}\alpha_i\nu_i - \beta\sum_{i=1}^{k}c_i n_i =$$

$$\begin{cases} (n_1 - 1)\alpha_1\Theta_1 + \sum_{j=2}^{k}(n_j - 1)\alpha_j \sup(\Theta_j) & \text{if } \underline{\Theta} \in \Delta_1, \\ \cdots & \cdots \quad \cdots \\ (n_i - 1)\alpha_i\Theta_i + \sum_{j\neq i}^{k}(n_j - 1)\alpha_j \sup(\Theta_j) & \text{if } \underline{\Theta} \in \Delta_i, \\ \cdots & \cdots \quad \cdots \\ (n_k - 1)\alpha_k\Theta_k + \sum_{j=1}^{k-1}(n_j - 1)\alpha_j \sup(\Theta_j) & \text{if } \underline{\Theta} \in \Delta_k, \end{cases}$$

where Δ_i is defined in (2.4).

Let f_i denote the first-phase sample posterior density of Θ_i and F_i be the corresponding c.d.f.. Recall that under the above setting, f_i is an inverse gamma density with parameters $\frac{m_i - 1}{2}$ and $\left[\frac{m_i - 1}{2}s_i^2\right]^{-1}$. See Berger (1985). These f_i's form the prior information for the second-phase sample allocation. Rao and Ghangurde (1972) called them (f_i's) as 'data based' prior densities. By f is inverse gamma density with parameters a and b we mean

$$f(x) = (\Gamma(a)b^a)^{-1}\left(\frac{1}{x}\right)^{a+1} e^{\frac{1}{bx}}, \quad x > 0, a > 0, b > 0. \tag{3.6}$$

Theorem 3.1 *Let $r_2(\underline{n}) = E[R_2(\underline{n})]$.*

$$r_2(\underline{n}) = \sum_{i=1}^{k}\left[(n_i - 1)\alpha_i\tau + \alpha_i\nu_i + \beta c_i n_i\right], \tag{3.7}$$

where

$$\tau = \sum_{j=1}^{k}\int_0^\infty x f_j(x)\left[\prod_{\ell\neq j}^{k} F_\ell(x)\right]dx.$$

Proof: Since Θ's have absolutely continuous support, $R_2(\underline{n})$ of Definition 3.1 can be written as

$$R_2(\underline{n}) = \sum_{i=1}^{k}\alpha_i\nu_i + \beta\sum_{i=1}^{u}c_i n_i + \sum_{i=1}^{k}(n_i - 1)\alpha_i \max(\Theta_1, \Theta_2, \cdots, \Theta_k). \tag{3.8}$$

We need to find the expected value of this R_2. The first two terms on the right hand side of (3.8) are constants. To find the expectation of $\Theta_{(k)} = \max(\Theta_1, \colon \Theta_k)$, we note that the Θ's are statistically independent and therefore the density of $\Theta_{(k)}$ is

$$\sum_{j=1}^{k} f_j(y)\left(\prod_{\ell\neq j}^{k} F_\ell(y)\right).$$

Taking expectation of $\Theta_{(k)}$ and writing the integral of a sum as a sum of the integrals we have

$$\tau = E(\Theta_{(k)}) = \sum_{j=1}^{k}\int_0^\infty y f_j(y)\left(\prod_{\ell\neq j}^{k} F_\ell(y)\right) dy.$$

Thus the proof is complete.

3.2 The Minimax Solution

The idea is to obtain n_i's by minimizing $r_2(\underline{n})$ given in Theorem 3.1. Note that for a given vector $\underline{n}$, the risk of Definition 3.1 as a function of the random vector $\underline{\Theta}$ satisfies the inequality:

$$R_2(\underline{\Theta}|\underline{n}) \geq \sum_{i=1}^{k}\left[\pi_i^2 E(V(\mu_i)|\Theta_i) + c_i n_i\right] \quad a.s..$$

Thus it is appropriate to call R_2 as a maximum Beyes risk. Now by monotonocity property of mathematical expectation, we have

$$r_2(\underline{n}) \geq E[V(\mu)] + \sum_{i=1}^{u} c_i n_i.$$

Therefore, the resulting solution is called a minimax solution.

Theorem 3.2 *For $i = 1, 2, \cdots, k$, let*

$$a_{i0} = \beta c_i\left(m_i^2 - 3m_i\right)\left(m_i^2 - m_i - 2\right) + \pi_i^2(2-2m_i)\nu_i - \pi_i^2(m_i - m_i^2)\tau$$

$$a_{i1} = \beta c_i\left(4m_i^3 - 12m_i^2 + 2m_i + 6\right) + \pi_i^2(3-4m_i)\tau - 2\nu_i\pi_i^2$$

$$a_{i2} = \beta c_i\left(6m_i^2 - 12m_i + 1\right) - \pi_i^2\tau$$

$$a_{i3} = \beta c_i(4m_i - 4) \text{ and } a_{i4} = \beta c_i.$$

The optimal allocation n_i^* is the smallest positive integer for which

$$P_i(n_i) = a_{i4}n_i^4 + a_{i3}n_i^3 + a_{i2}n_i^2 + a_{i1}n_i + a_{i0} > 0.$$

Proof: Since the ith term on the right hand side of (3.7) depends on n_i alone, it is sufficient to minimize

$$h_i(n_i) = (n_i - 1)\alpha_i\tau + \alpha_i\nu_i + \beta c_i n_i.$$

We see that

$$h_i(n_i+1) - h_i(n_i) = \frac{P_i(n_i)}{(m_i+n_i-3)(m_i+n_i-2)(m_i+n_i)\cdot(m_i+n_i+1)}.$$

To minimize $h_i(n_i)$, we choose the smallest positive integral value for which the difference $h_i(n_i+1) - h_i(n_i)$ is positive. The proof is complete.

3.3 An Example

As mentioned earlier, Draper and Guttman (1968) minimized $E_y\{V(\mu)\}$ and obtained

$$n_{Di}^* = \frac{c}{c_i}q_i - m_i, \quad i = 1, 2, \cdots, k, \tag{3.9}$$

where

$$q_i = \pi_i\left(\frac{m_i-1}{m_i-3}\right)^{\frac{1}{2}}\sqrt{c_i}s_i \Big/ \sum_{j=1}^{k}\pi_j\left(\frac{m_i-1}{m_i-3}\right)^{\frac{1}{2}}\sqrt{c_j}s_j.$$

Note that $\underline{n}_D^*$ denotes the Draper-Guttman allocation while $\underline{n}^*$ denotes the minimax allocation proposed in the Subsection 3.2. Draper and Guttman (1968) also showed the variance of the preposterior estimator $\sum\pi_i\bar{y}_i$ of μ, is

$$V(\Sigma\pi_i\bar{y}_i|\underline{n}) = \sum_{i=1}^{k}\pi_i^2\left(1-\frac{1}{m_i}\right)\left(1+\frac{m_i}{n_i}\right)\frac{s_i^2}{m_i-3}. \tag{3.10}$$

In this subsection, we consider an example to compare the preposterior variance (3.10) under $\underline{n}_D^*$ and $\underline{n}^*$. Let V_1 and V_2 denote the preposterior variance, respectively, under $\underline{n}_D^*$ and $\underline{n}^*$.

Consider the following constants:

$$k = 2: \qquad m_1 = m_2 = 21$$
$$\pi_1 = \pi_2 = 0.5, \quad s_1^2 = 5.0, \quad s_2^2 = 6.5$$
$$c_1 = c_2 = 0.02$$

With a FORTRAN subroutine, we got $\tau = 9.25$. We calculate the minimax allocation $\underline{n}^*$ for various values of β. For each β, we computed the Draper-Guttman allocation $\underline{n}_D^*$. The total cost C used is the total cost incurred by the minimax allocation. Following table summarizes all results.

β	(n_1^*, n_2^*)	C	(n_{D1}^*, n_{D2}^*)	$V_2 - V_1$
0.01	(186, 190)	8.36	(174, 202)	0.000099
0.02	(123, 127)	5.84	(115, 135)	0.000114
0.03	(94, 99)	4.70	(89, 104)	0.000094
0.04	(77, 82)	4.02	(73, 86)	0.000071
0.05	(66, 70)	3.56	(62, 74)	0.000073
0.06	(57, 62)	3.22	(54, 65)	-0.000019
0.07	(50, 55)	2.94	(48, 57)	-0.000011
0.08	(45, 50)	2.74	(43, 42)	-0.000093
0.09	(40, 46)	2.56	(39, 47)	-0.00011
0.10	(36, 42)	2.40	(35, 43)	-0.000205
0.11	(33, 39)	2.28	(32, 40)	-0.000316

References

[1] Aggarwal, O.P. (1959). "Bayes and Minimax Procedures in Sampling from Finite and Infinite Populations." *The Annals of Mathematical Statistics*, Vol. 30, pp. 206-218.

[2] Berger, James O. (1985). *Statistical Decision Theory and Bayes Analysis*. Springer-Verlag, New York.

[3] Cochran, W.G. (1977). *Sampling Techniques*. 3rd edition, John Wiley & Sons, New York.

[4] Draper, N.R., and I. Guttman. (1968). "Some Bayesian Stratified Two-Phase Sampling Results." *Biometrika*, 55, pp. 131-139.

[5] Ericson, W.A. (1965). "Optimum Stratified Sampling Using Prior Information." *Journal of the American Statistical Association*, Vol. 60, pp. 750-71.

[6] Koti, K.M. (1988). *Optimum Stratified Sampling Using Prior Information*. Ph.D. dissertation submitted to Texas Tech University.

[7] Rao, J.N.K., and P.D. Ghangurde. (1972). "Baysian Optimization in Sampling Finite Populations." *Journal of the American Statistical Association*, Vol. 67, pp. 439-443.

[8] Solomon, H., and S. Zacks. (1970). "Optimal Design of Sampling from Finite Populations: A Critical Review and Indication of New Research Areas." *Journal of the American Statistical Association*, Vol. 65, pp. 653-677.

[9] Zacks, S. (1970). "Bayesian Design of Single and Double Stratified Sampling for Estimating Proportion in Finite Population." *Technometrics*, Vol. 12, pp. 119-130.