The Journal of National Chengchi University Vol. 29, 1974

MINIMUM NORM QUADRATIC UNBIASED ESTIMATION IN STRATIFIED RANDOM SAMPLING

by

Ing-Tzer Wey

I. Introduction

Let a population of N units be divided up into L strata, the h-th stratum containing N_h units with stratum total Y_h, stratum mean $\overline{Y}_h$ and stratum variance S^2_{yh} for the character y under study. The population total

$$Y=\sum_{h=1}^{L} Y_h$$

is to be estimated from a stratified random sample. If the strata sample sizes n_h are all greater than one, the well-known standard estimation theory applies. However, in many practical situations it may be desirable to draw only one unit from each stratum. Denoting the y-value associated with the single unit drawn from the h-th stratum by y_h, an unbiased estimator of Y is given by

$$\hat{Y}=\sum_{h=1}^{L} N_h y_h \qquad (1)$$

with a variance of

$$Var(\hat{Y})=\sum_{h=1}^{L} N_h(N_h-1)\, S^2_{yh} \qquad (2)$$

The problem arises in estimating the stratum variance S^2_{yh} from a single observation per stratum. The techniques of collapsed strata have been suggested for solving this problem (Cochran [1] and Hansen et al [2]). For considering a new solution to this problem, we assume in this study that we can associate with the strata one or more concomitant variables which are correlated with the strata means $\overline{Y}_h$ in the sense of providing a good regression fit. When this is the case, we shall consider the principle components of concomitant variables and use the method of minimum norm quadratic unbiased estimation (MINQUE) for estimating all the different strata variances. The method of MINQUE is suggested by Rao [6, 7] for the estimation of heteroscedastic variances in linear models.

Consider a linear model

$$\mathbf{y}=\mathbf{X}\beta+\mathbf{e}$$

where $\mathbf{y}$ is a vector of n observations, $\mathbf{X}$ is an nxp known matrix of rank p ($\leqq$n), β is a vector of p unknown parameters and $\mathbf{e}$ is an error vector such that

$$E(\mathbf{e}) = \mathbf{0}$$

$$E(\mathbf{e}\mathbf{e}')=\begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}$$

Let $\sum_{i=1}^{n} c_i \sigma_i^2$ be a linear function of the variances to be estimated. The quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y}$ in the observations $\mathbf{y}$ is said to be a MINQUE of $\sum_{i=1}^{n} c_i\sigma_i^2$ if the matrix $\mathbf{A}=(a_{ij})$ is chosen such that $\|\mathbf{A}\|$, the Euclidean norm of $\mathbf{A}$, which is the same as the square root of trace $\mathbf{A}^2$, is a minimum, subject to the conditions

(1) $\mathbf{A}\mathbf{X}=\mathbf{0}$

(2) $\sum_{i=1}^{n} a_i\sigma_i^2 \equiv \sum_{i=1}^{n} c_i\sigma_i^2$

Denote by $\mathbf{M}=(m_{ij})$ the projection matrix $\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, by $\mathbf{v}$ the vector of squares of the residuals $\mathbf{My}$, by σ^2 the vector of variances $\sigma_1^2, \cdots\cdots,$ σ_n^2 and define $\mathbf{F}=(m_{ij}^2)$. The followinga two lemmas are given by Rao [6] .

<u>Lemma 1.</u> Iet $\sigma_1^2, \cdots\cdots \sigma_n^2$ be all different. Then the MINQUE of $\sigma_1^2, \cdots\cdots, \sigma_n^2$ are solution of $\mathbf{F}\ \sigma^2=\mathbf{v}$ provided $\mathbf{F}$ is non-singular. The MINQUE of the linear function $\mathbf{c}'\sigma^2$ is $\mathbf{c}'\mathbf{F}^{-1}\mathbf{v}$.

<u>Lemma 2.</u> Let all σ_i^2 be different. Then the MINQUE has minimum average variance in the class of quadratic unbiased estimators for any symmetric <u>a priori</u> distribution of $\sigma_1^2, \cdots\cdots, \sigma_n^2$ over which the average is taken.

Rao [5] showed that no uniformly minimum variance quadratic estimator exists when σ_i^2 are different and unknown. Even in the case when all σ_i^2 are equal, such an estimator exists only under restrictive conditions such as $\beta_2=3$, where β_2 is the Pearsonian coefficient of kurtosis.

We shall also extend the principle of MINQUE to regression estimation of the population mean $\overline{Y}$ in the stratified random sampling where the sample of size n_h (>1) is drawn from the h-th stratum ($h = 1, \cdots\cdots, L$).

II. MINQUE in Stratified Random Sampling

We assume that we can associate with the strata a set of p ($\geqq 1$) concomitant variables x_{ih} ($i = 1, \cdots\cdots, p$; $h = 1, \cdots\cdots, L$) which are correlated to the strata means $\overline{Y}_h$ in the sense of providing a good regression fit. The concomitant variables x_{ih} must be fixed values, i. e.,values which do not depend on the particular sample drawn. Let the relation between the strata means $\overline{Y}_h$ and the concomitant variables x_{ih} be of the form

$$\mu = X\beta + d \qquad (3)$$

where
$$\mu = \begin{pmatrix} \overline{Y}_1 \\ \vdots \\ \overline{Y}_L \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} \cdots x_{p1} \\ \vdots & \vdots \\ 1 & x_{1L} \cdots x_{pL} \end{pmatrix}$$

β is the (p + 1)x1 vector of parameters and **d** is the Lx1 vector of residuals.

In many practical applications, the following situations may occur:

(1) The underlying assumption of independence among the concomitant variables in the linear model is not satisfied by the presence of multicollinearity.

(2) The number of concomitant variables available in the estimation is greater than that of dependent variables.

(3) The number of degrees of freedom for the residual sum of squares becomes small as the number of concomitant variables increases.

The problems arisen from the above mentioned situations can be solved by the theory of principal components. Since $X'X$ is a symmetric positive definite matrix, it possesses p positive characteristic roots $\lambda_1, \cdots\cdots, \lambda_p$ which satisfy

$$|X'X - \lambda I| = 0$$

We choose the r ($<$ p) largest characteristic roots $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ such that

$$\sum_{i=1}^{r} \lambda_i / tr(X'X) \approx 1$$

where $tr(X'X)$ is the trace of $X'X$. Further, let $\mathbf{a}_i$ be the px1 characteristic vector associated with λ_i such that

$$X'X\mathbf{a}_i = \lambda_i \mathbf{a}_i$$
$$\mathbf{a}_i'\mathbf{a}_i = 1, \quad \text{for } i = 1, \cdots\cdots, r \qquad (4)$$
and
$$\mathbf{a}_i'\mathbf{a}_j = 0 \quad \text{for } i \neq j$$

Denote the i-th principle component of $(1, x_{1h}, \cdots\cdots, x_{ph})$ by $z_{i.}$, that is,

$$z_{i.} = Xa_i \qquad i = 1, \cdots\cdots r \tag{5}$$

Then we have

$$z_{i.}'z_{i.} = \lambda_i \qquad \text{for } i = 1, \cdots\cdots, r \tag{6}$$
$$z_{i.}'z_{j.} = 0 \qquad \text{for } i \neq j$$

Now we can rewrite equation (3) as follows:

$$\mu = Z\theta + d \tag{7}$$

where $Z = (z_{1.}, \cdots\cdots, z_{r.})$ is the Lxr matrix of r principal components and θ is the rxl vector of unknown parameters. By the method of least squares, we have the fitted hyperplane of μ, denoted by μ_f, as follows:

$$\mu_f = Z(Z'Z)^{-1}Z'\mu = Z\Lambda^{-1}Z'\mu \tag{8}$$

where
$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_r \end{pmatrix}$$

The deviation from the fitted hyperplane is given by

$$d = \mu - \mu_f = (I - Z\Lambda^{-1}Z')\mu \equiv M\mu \tag{9}$$

Let y be the Lxl vector whose elements are y_h, the y-value associated with the single unit drawn from the h-th stratum, and let $e = y - \mu$, then

$$E(e) = 0$$

$$E(ee') = \begin{pmatrix} S^2_{y1} & & 0 \\ & \ddots & \\ 0 & & S^2_{yL} \end{pmatrix}$$

Consider the following identity:

$$y = \mu_f + (\mu - \mu_f) + (z - \mu) = Z\theta + M\mu + e \tag{10}$$

Since $Z'M = 0$, an unbiased estimator of θ is given by

$$\hat{\theta} = \Lambda^{-1}Z'y \tag{11}$$

Define the Lxl vector of residuals by

$$v = y - Z\theta = (I - Z\Lambda^{-1}Z')y = My \tag{12}$$

Then
$$E(v) = M\mu \tag{13}$$

Let v^2 denote the Lxl vector whose elements are squares of elements of v, and define the LxL symmetric matrix F by

$$F = (m^2_{ij}) \tag{14}$$

where m_{ij} are elements of $M = I - Z\Lambda^{-1}Z'$, i.e., the i-th diagonal element m_{ii} is given by

$$m_{ii} = 1 - z._i{}'\Lambda^{-1}z._i$$

and (i , j)-th off-diagonal element m_{ij} is given by

$$m_{ij} = z._i{}' \Lambda^{-1} z._j$$

where $z._i{}' = (z_{1i}, z_{2i}, \cdots\cdots, z_{ri})$ is the i-th row of $\mathbf{Z}$. Let $\mathbf{S}^2$ denote the Lxl vector whose elements are the L strata variances S^2_{yh}. Then by by Rao's lemma 1 the MINQUE of $\mathbf{S}^2$ is given by

$$\hat{\mathbf{S}}^2 = \mathbf{F}^{-1}\, \mathbf{v}^2 \qquad (15)$$

It is obvious that the expected value of $\mathbf{v}^2$ is

$$E(\mathbf{v}^2) = \mathbf{F}\mathbf{S}^2 + \mathbf{d}^2 \qquad (16)$$

where $\mathbf{d}^2$ is the Lxl vector whose elements are squares of elements of $\mathbf{d} = \mathbf{M}\,\mu$. Thus, the bias in $\mathbf{S}^2$ is given by

$$\mathbf{b} = \mathbf{F}^{-1}\mathbf{d}^2 \qquad (17)$$

Let $\mathbf{t}$ denote the Lxl vector with elements $N_h(N_h - 1)$, then the MINQUE of $\mathrm{Var}(\hat{Y})$ is given by

$$\mathrm{var}(\hat{Y}) = \mathbf{t}'\hat{\mathbf{S}}^2 = \mathbf{t}'\mathbf{F}^{-1}\,\mathbf{v}^2 \qquad (18)$$

and the bias in $\mathrm{var}(\hat{Y})$ is given by

$$\mathrm{Bias}\,[\mathrm{var}\,(\hat{Y})] = \mathbf{t}'\mathbf{F}^{-1}\,\mathbf{d}^2 \qquad (19)$$

We summarize the foregoing in the following theorem:

Theorem 1. In stratified random sampling with a single unit drawn from each stratum, let $\mathbf{y}$ be the Lxl vector of observations and S^2 the Lxl vector of strata variances S^2_{yh}. Suppose that $\mathbf{X}$ is the Lx(p + 1)matrix of concomitant variables which are correlated with the strata means $\overline{Y}_h$. Let $\mathbf{Z}$ be the Lxr ($r < p + 1$) matrix whose columns are principal components of $\mathbf{X}$, and Λ the rxr diagonal matrix whose diagonal elements are the r largest charadteristic roots of $\mathbf{X}'\mathbf{X}$. Then the minimum norm quadratic estimator of S^2 is given by

$$\hat{\mathbf{S}}^2 = \mathbf{F}^{-1}\, \mathbf{v}^2$$

where $\mathbf{F}$ is the LxL symmetric matrix whose elements are squares of the elements of $\mathbf{M} = \mathbf{I} - \mathbf{Z}\Lambda^{-1}\mathbf{Z}'$ and $\mathbf{v}^2$ is the Lxl vector whose elements are squares of elements of $\mathbf{v} = \mathbf{M}\mathbf{y}$.

Corollary 1. The bias in $\hat{S}^2$ is given by $\mathbf{b} = \mathbf{F}^{-1}\,\mathbf{d}^2$, where $\mathbf{d}^2$ is the Lxl vector whose elements are squares of elements of $\mathbf{d} = \mathbf{M}\mu$, and μ is the Lxl vector of strata means $\overline{Y}_h$.

Corollary 2. In stratified random sampling with a single unit drawn from each stratum, let $\hat{Y}$ denote an unbiased estimator of the population total Y,

i. e.,

$$\hat{Y}=\sum_{h=1}^{L} N_h y_h$$

Then the minimum norm quadratic estimator of the variance of $\hat{Y}$ is given by

$$\hat{\text{var}}(\hat{Y})=\mathbf{t}'\mathbf{F}^{-1}\mathbf{v}^2$$

and the bias in $\hat{\text{var}}(\hat{Y})$ is given by

$$\text{Bias}\ [\hat{\text{var}}(\hat{Y})]\ =\mathbf{t}'\mathbf{F}^{-1}\mathbf{d}^2$$

where $\mathbf{t}$ is the Lxl vector with elements $N_h(N_h-1)$.

Hartley, et al [3] considered an estimator of the variance of the estimator of the population total in stratified random sampling with one unit per stratum by using the matrix $\mathbf{X}$ of order $L\times(p+1)$ in the estimation procedure. It is expected that the goodness of regression fit to μ is better by using $\mathbf{X}$ than by the first r principle components of X, i. e.,$\mathbf{d}^2$ in equation (17) and (19) is larger than $\mathbf{d}^2_*$ whose elements are squares of elements of $\mathbf{d}_*=(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mu$. But a better fit does not necessarily lead to a smaller bias since $\mathbf{b}$ and Bias $[\text{var}(\hat{Y})]$ also depend on $\mathbf{F}^{-1}$.

III. Regression Estimation with the Method of MINQUE

In the stratified random sampling, let us consider the normal case where the strata sample size n_h are all greater than one. Let y_{hi} denote the i-th observation in the h-th stratum($i=1,\cdots\cdots,n_h$; $h=1,\cdots\cdots,L$). For simplicity, we shall assume that we can associate with the strata only one concomitant variable x_h ($h=1,\cdots\cdots,L$) which is correlated with y_{hi} in the sense of providing agood regression fit. For instance, the strata are regions and the concomitant variable is rainfall in a region.

Let the relation between y_{hi} and x_h be of the following form:

$$y_{hi}=\alpha+\beta x_h+e_{hi} \tag{20}$$

where

$$E(e_{hi}\mid x_h)=0 \quad \text{for all } i$$
$$E(e_{hi}^2\mid x_h)=\sigma_h^2 \quad \text{for all } i$$
$$E(e_{hi}e_{hj}\mid x_h)=0 \quad \text{for } i\neq j.$$

Then the best linear unbiased estimators of α and β are given dy

$$\hat{\alpha}_w=\bar{y}_w-\hat{\beta}_w\bar{x}_w \tag{21}$$

$$\hat{\beta}_w=\sum_{h=1}^{L} w_h(x_h-\bar{x}_w)\,\bar{y}_w \Big/ \sum_{h=1}^{L} w_h(x_h-\bar{x}_w)^2 \tag{22}$$

where $\bar{y}_w = \sum_{h=1}^{L} w_h \bar{y}_h / w, \quad \bar{x}_w = \sum_{h=1}^{L} w_h x_h / w$

$$w_h = n_h / \sigma_h^2 \tag{23}$$

$$w = \sum_{h=1}^{L} w_h \tag{24}$$

Thus, the regression estimator of the population mean $\bar{Y}$ is given by

$$\bar{y}_r = \bar{y}_w + \hat{\beta}_w (\bar{X} - \bar{x}_w) \tag{25}$$

where $\bar{X} = \sum_{h=1}^{L} N_h x_h / N$

Suppose that the strata finite population correction terms $(N_h - n_h)/N_h$ are all negligible, then variance of $\bar{y}_r$ is given by

$$Var(\bar{y}_r) = \frac{1}{w} + \frac{(\bar{X} - \bar{x}_w)^2}{\sum_{h=1}^{L} w_h (x_h - \bar{x}_w)^2} \tag{26}$$

An examination of the expression for $\bar{y}_r$ in (25) shows that a knowledge of true weights $w_h = n_h / \sigma_h^2$ is required for calculating $\bar{y}_r$. Since the true weights w_h are rarely known, we have to estimate the strata residual variances σ_h^2. For obtaining an estimator of σ_h^2, Jacquez, et al [4] used only the observations y_{hi} ($i = 1, \cdots\cdots, n_h$), i.e.,

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \tag{27}$$

In this study we shall use the method of MINQUE to estimate σ_h^2. Let us assume for a moment that the linear model in (20) has homoscedastic variances. then the two parameters α and β in the model are estimated by

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$

$$\hat{\beta} = \sum_{h=1}^{L} n_h (x_h - \bar{x}) \bar{y}_h \Big/ \sum_{h=1}^{L} n_h (x_h - \bar{x})^2$$

where $\bar{y} = \sum_{h=1}^{L} n_h \bar{y}_h / n, \quad \bar{x} = \sum_{h=1}^{L} n_h x_h / n, \quad n = \sum_{h=1}^{L} n_h.$

Denote the h-th stratum residual sum of squares by

$$v_h^2 = \sum_{i=1}^{n_h} [y_{hi} - \bar{y} - \hat{\beta} (x_h - \bar{x})]^2 \tag{28}$$

If we use the matrix notation to express the linear model in (20), then the matrix X is of order nx2 as follows:

$$X' = \begin{pmatrix} 1 \cdots 1 & 1 \cdots 1 \cdots 1 \cdots 1 \\ \underbrace{x_1 \cdots x_1}_{n_1} & \underbrace{x_2 \cdots x_2}_{n_2} \cdots \underbrace{x_L \cdots x_L}_{n_L} \end{pmatrix}$$

The inverse of $X'X$ is given by

$$(X'X)^{-1} = \frac{1}{T_x} \begin{pmatrix} \sum_{h=1}^{L} n_h x_h^2 & -\sum_{h=1}^{L} n_h x_h \\ -\sum_{h=1}^{L} n_h x_h & n \end{pmatrix}$$

where $T_x = n \sum_{h=1}^{L} n_h x_h^2 - (\sum_{h=1}^{L} n_h x_h)^2$

Let $a_i = (n-1) \sum_{h=1}^{L} n_h x_h^2 - (\sum_{h=1}^{L} n_h x_h)^2 + 2x_i (\sum_{h=1}^{L} n_h x_h) - n x_i^2$

$$b_i = -\sum_{h=1}^{L} n_h x_h^2 + 2x_i (\sum_{h=1}^{L} n_h x_h) - n x_i^2$$

$$c_{ij} = -\sum_{h=1}^{L} n_h x_h^2 + (x_i + x_j) \sum_{h=1}^{L} n_h x_h - n x_i x_j$$

Further let

$$\underset{n_i x n_i}{A_{ii}} = \begin{pmatrix} a_i & b_i \cdots b_i \\ b_i & a_i \cdots b_i \\ \vdots & \vdots \quad \vdots \\ b_i & b_i \cdots a_i \end{pmatrix} \qquad i = 1, \cdots, L$$

$$\underset{n_i x n_j}{C_{ij}} = \begin{pmatrix} c_{ij} & c_{ij} \cdots c_{ij} \\ c_{ij} & c_{ij} \cdots c_{ij} \\ \vdots & \vdots \quad \vdots \\ c_{ij} & c_{ij} \cdots c_{ij} \end{pmatrix} \qquad i \neq j$$

Then the projection matrix $I - X(X'X)^{-1}X'$ can be written as follows:

$$I - X(X'X)^{-1}X' = \frac{1}{T_x} \begin{pmatrix} A_{11} & C_{12} \cdots C_{1L} \\ C_{21} & A_{22} \cdots C_{2L} \\ \vdots & \vdots \quad \vdots \\ C_{L1} & C_{L2} \cdots A_{LL} \end{pmatrix}$$

Then from Rao's lemma 1 we have the following equations:

$$n_1 [(\frac{a_1}{T_x})^2 + (n_1 - 1)(\frac{b_1}{T_x})^2] \hat{\sigma}_1^2 + n_1 n_2 (\frac{c_{12}}{T_x})^2 \hat{\sigma}_2^2 + \cdots\cdots + n_1 n_L (\frac{c_{1L}}{T_x})^2 \hat{\sigma}_L^2 = v_1^2$$

$$n_2 n_1 \left(\frac{c_{21}}{T_x}\right)^2 \hat{\sigma}_1^2 + n_2 \left[\left(\frac{a_2}{T_x}\right)^2 + (n_2 - 1)\left(\frac{b_2}{T_x}\right)^2 \right] \hat{\sigma}_2^2 + \cdots\cdots + n_2 n_L \left(\frac{c_{2L}}{T_x}\right)^2 \hat{\sigma}_L^2 = v_2^2$$

$$\vdots \qquad\qquad \vdots$$

$$n_L n_1 \left(\frac{c_{L1}}{T_x}\right)^2 \hat{\sigma}_1^2 + n_L n_2 \left(\frac{c_{L2}}{T_x}\right)^2 \hat{\sigma}_2^2 + \cdots + n_L \left[\left(\frac{a_L}{T_x}\right)^2 + (n_L - 1)\left(\frac{b_L}{T_x}\right)^2 \right] \hat{\sigma}_L^2 = v_L^2$$

The coefficients of $\hat{\sigma}_h^2$ can be simplified as follows: The i-th diagonal element of the coefficient matrix is

$$n_i \left[\left(\frac{a_i}{T_x}\right)^2 + (n_i - 1)\left(\frac{b_i}{T_x}\right)^2 \right] = n_i^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right]^2 + n_i \left[1 - 2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \right]$$

and the (i , j)th off-diagonal element is

$$n_i n_j \left(\frac{c_{ij}}{T_x}\right)^2 = n_i n_j \left[\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}} \right]^2$$

where $\quad S_{xx} = \sum_{h=1}^{L} n_h (x_h - \bar{x})^2$

Iet $\quad k_{ij} = \dfrac{1}{n} + \dfrac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}$

$$m_{ii} = n_i^2 k_{ii}^2 + n_i (1 - 2k_{ii})$$

$$m_{ij} = n_i n_j k_{ij}^2$$

Then the equations for obtaining $\hat{\sigma}_1^2, \cdots, \hat{\sigma}_L^2$ are given by

$$\begin{aligned} m_{11}\hat{\sigma}_1^2 + m_{12}\hat{\sigma}_2^2 + \cdots\cdots + m_{1L}\hat{\sigma}_L^2 &= v_1^2 \\ m_{21}\hat{\sigma}_1^2 + m_{22}\hat{\sigma}_2^2 + \cdots\cdots + m_{2L}\hat{\sigma}_L^2 &= v_2^2 \\ \vdots \qquad\qquad & \vdots \\ m_{L1}\hat{\sigma}_1^2 + m_{12}\hat{\sigma}_2^2 + \cdots\cdots + m_{LL}\hat{\sigma}_L^2 &= v_L^2 \end{aligned} \tag{29}$$

We summarize the foregoing in the following theorem:

Theorem 2. In a stratified random sample, suppose that the relation between the observations y_{hi}(h = 1 , ……, L; f = 1 , ……, n_h)and a concomitant variable x_h is of the following linear form

$$y_{hi} = \alpha + \beta x_h + e_{hi}$$

where $\quad E(e_{hi} \mid x_h) = 0 \qquad$ for all i

$\qquad\quad E(e_{hi}^2 \mid x_h) = \sigma_h^2 \qquad$ for all i

$$E(e_{hi}e_{hj} \mid x_h) = 0 \quad \text{for} \quad i \neq j.$$

Then the MINQUE of $\sigma_1^2, \cdots\cdots, \sigma_L^2$ are obtained from equation (29).
The regression estimator of the population mean $\overline{Y}$ obtained by using the MINQUE of $\sigma_1^2, \cdots\cdots, \sigma_L^2$ as weights is given by

$$\bar{y}_r = \bar{y} + \hat{\beta}_w (\overline{X} - \bar{x}_w)$$

where

$$\bar{y}_w = \sum_{h=1}^{L} \hat{w}_h \bar{y}_j / w, \qquad \bar{x}_w = \sum_{h=1}^{L} \hat{w}_h x_h / \hat{w}$$

$$\hat{\beta}_w = \sum_{h=1}^{L} \hat{w}_h (x_h - \bar{x}_w) y_h \Big/ \sum_{h=1}^{L} \hat{w}_h (x_h - \bar{x}_w)^2$$

$$\hat{w}_h = n_h / \hat{\sigma}_h^2, \quad \hat{w} = \sum_{h=1}^{L} \hat{w}_h.$$

If the residual variances $\sigma_1^2, \cdots\cdots, \sigma_L^2$ are known, then the variance of $\bar{y}_r$ is given by

$$\mathrm{Var}(\bar{y}_r) = \frac{1}{w} + \frac{(\overline{X} - \bar{x}_w)^2}{\sum_{h=1}^{L} w_h (x_h - \bar{x}_w)^2}$$

Since the MINQUE of σ_h^2 is obtaind from utilizing the information supplied by all the observations (y_{hi}, x_h), $h = 1, \cdots\cdots, L$; $i = 1, \cdots\cdots, n_h$ and not merely based on the observations $(y_{h1}, \cdots\cdots, y_{hn_h})$ corresponding to x_h, the regression estimator using the MINQUE of σ_h^2 may be more efficient than those using the estimator of σ_h^2 based on the observations $(y_{h1}, \cdots\cdots y_{hn_h})$ only. HoAever, a detailed investigation is necessary.

IV. Summary

In stratified random sampling with one unit per stratum, the strata variances are estimated by the method of minimum norm quadratic unbiased estimation (MINQUE) and using the first r ($<$p) principal components of a set of p concomitant variables which are correlated with the strata means. If the underlying assumption of independence among the concomitant variables in the linear model is not satisfied by the presence of multicollinearity or if the number of concomitant variables available in the estimation is greater than that of dependent variables, the present method is a good device for solving the problems arisen in such situations. The estimated variance obtained from the method of MINQUE may be more efficient than the method of collapsed strata.

If the relation between the observations y_{hi} ($h = 1, \cdots\cdots, L$; $i = 1, \cdots\cdots, n_h$) and a concomitant variable x_h in a stratified random sample is of the linear form $y_{hi} = \alpha + \beta x_h + e_{hi}$ with heteroscedastic variances σ_h^2, then the population mean $\overline{Y}$ is estimated by weighted least squares estimator $\bar{y}_r = \bar{y}_w + \hat{\beta}_w(\overline{X} - \bar{x}_w)$, where estimated variances $\hat{\sigma}_h^2$ in the weights $\hat{w}_h = n_h / \hat{\sigma}_h^2$ are obtained from the method of MINQUE. Since the MINQUE of σ_h^2 is obtained from utilizing the information supplied by all the observations (y_{hi}, x_h) in the whole sample and not merely based on the observations $(y_{h1}, \cdots\cdots, y_{hn_h})$ corresponding to x_h, the regression estimator using the MINQUE of σ_h^2 in the weights may be more efficient than those using the estimator of σ_h^2 based on the observations $(y_{h1}, \cdots\cdots, y_{hn_h})$ only.

References

1. Cochran, W. G. Sampling Techniques, Second Edition. John Wiley, 1963.
2. Hansen, M. H., Hurwitz, W. N., and Madow, W. G. Sample Survey Methods and Theory, Vol. II, John Wiley, 1953.
3. Hartley, H. O., Rao, J. N. K., and Kiefer, G. Variance estimation with one unit per stratum. Journal of American Statistical Association, 64 (1969), 841-851.
4. Jacquez, J. A., Mathur, F. J., and Crawford, C. R. Linear regression with non-constant, unknown error variances: Sampling experiments with least squares, weighted least squares and maximum likelihood estimators. Biometrics, 24(1968), 607-627.
5. Rao, C. R. Some theorems on minimum variance estimation, Sankhya, 12 (1952), 27-42.
6. Rao, C. R. Estimation of hoteroscedastic variances in linear models. Journal of American Statistical Association, 65(1970), 161-172.
7. Rao, C. R. Estimating variance and covariance components in linear models. Journal of American Statistical Association, 67(1972), 112-115.