RATIO METHOD OF ESTIMATION IN MULTI-PHASE SAMPLING WITH SEVERAL AUXILIARY VARIABLES

By J. N. Goswami and B. V. Sukhatme

Institute of Agricultural Research Statistics, New Delhi

1. Introduction and Summary

In sample surveys, it happens sometimes that several auxiliary variables are available for measurement which are highly correlated with the variable under study. It is therefore of interest to investigate the effect of using all or some of them in building up an estimate of the population mean or total of the character under study. Olkin (1958) has discussed the extension of ratio method of estimation to several auxiliary variables for uni-stage designs. Sukhatme and Koshal (1959) have discussed the ratio method of estimation in the case of a single auxiliary variable with unknown population mean for multi-stage designs. The object of this paper is to extend these results to several auxiliary variables with unknown means for a three-stage design. The treatment is perfectly general and the results can be extended to designs with any number of stages.

2. Notation and the Procedure of Multi-Phase Sampling

Let

p = number of auxiliary variables available for measurement.

N = total number of primary units in the population.

M_i = total number of secondary units in the i-th primary unit ($i = 1, 2, \ldots, N$).

B_{ij} = total number of tertiary units in the (i, j)-th secondary unit ($j = 1, 2, \ldots, M_i$).

y_{ijk} = the value of the variable under study for the k-th tertiary unit of the j-th secondary unit in the i-th primary unit ($k = 1, 2, \ldots, B_{ij}$).

$x_{\alpha ijk}$ = the value of α-th auxiliary variable for the k-th tertiary unit of the j-th secondary unit in the i-th primary unit ($\alpha = 1, 2, \ldots, p$).

Further, let

$$\bar{B}_{i.} = \frac{1}{M_i}\sum_{j=1}^{M_i} B_{ij}, \quad v_{ij} = \frac{B_{ij}}{\bar{B}_{i.}} \quad Q_i = M_i\bar{B}_{i.}, \quad \bar{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i$$

$$u_i = \frac{Q_i}{\bar{Q}}, \quad y_{ij.} = \frac{1}{B_{ij}}\sum_{k=1}^{B_{ij}} y_{ijk}, \quad \bar{y}_{i..} = \frac{1}{M_i}\sum_{j=1}^{M_i} v_{ij}\,\bar{y}_{ij.} \tag{2.1}$$

and

$$\bar{y}_{...} = \frac{1}{N}\sum_{i=1}^{N} u_i\bar{y}_{i..}$$

The quantities corresponding to each of the auxiliary variables are also defined in a similar fashion.

The problem considered here is to estimate the population mean $\bar{y}_{...}$ using the information on the p auxiliary variables when the population means of these auxiliary variables are not known. In such a case, the procedure of multi-phase sampling is usually adopted as given in the following scheme.

We shall consider a scheme of sampling where at each stage, sampling units are selected with equal probabilities and without replacement. The results can however be extended to other schemes of sampling also. To estimate the population means of the auxiliary variables, we take a random sample of n' primary units out of N. In the i-th selected primary unit, we select a random sample of m_i' secondary units out of M_i. In the (i, j)-th selected secondary unit, we select a random sample of b_{ij} tertiary units out of B_{ij} and observe the auxiliary variables on the selected units. Then an unbiased estimate of $\bar{x}_{\alpha\,...}$ is given by

$$\bar{x}_{\alpha n'} = \frac{1}{n'}\sum^{n'}\frac{u_i}{m_i'}\sum^{m_i'} v_{ij}\,\bar{x}'_{\alpha ij} \tag{2.2}$$

where

$$\bar{x}'_{\alpha ij} = \frac{1}{b_{ij}'}\sum^{b'_{ij}} x_{\alpha ijk} \qquad (\alpha = 1, 2, \cdots, p).$$

and the summation is taken over all the units in the sample.

A sub-sample of n units is now selected at random with equal probabilities and without replacement from the selected n' primary units. If the i-th primary unit is in the sub-sample, a simple random sample of m_i secondary units is drawn out of m_i'. If the (i, j)-th secondary unit is in the sub-sample, a simple random sample of b_{ij} tertiary units is drawn from b_{ij}'. This sub-sample is used to observe the main variable under study. Then the estimate proposed for the population mean $\bar{y}_{...}$ using information on all the auxiliary variables is

$$\bar{y}_R = \sum_{\alpha=1}^{p} w_\alpha r_\alpha \bar{x}_{\alpha n'} \tag{2.3}$$

where

$$r_\alpha = \frac{\bar{y}_n}{\bar{x}_{\alpha n}}$$

$$\bar{y}_n = \frac{1}{n} \sum^{n} \frac{u_i}{m_i} \sum^{m_i} v_{ij} \bar{y}_{ij} \tag{2.4}$$

$$\bar{y}_{ij} = \frac{1}{b_{ij}} \sum^{b_{ij}} y_{ijk}.$$

$\bar{x}_{\alpha n}$ based on (n, m_i, b_{ij}) are defined as in (2.2) and $w = (w_1, w_2, \ldots, w_p)$ are the weights such that $\sum_{\alpha=1}^{p} w_\alpha = 1$.

Then it is clear that

$$\text{Cov.}\,(\bar{x}_{\alpha n}, \bar{y}_n) = \left(\frac{1}{n} - \frac{1}{N}\right) S_{byx_\alpha} + \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) S_{iyx_\alpha}$$
$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{u_i^2}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) S_{ijyx_\alpha} \tag{2.5}$$

where

$$S_{byx_\alpha} = \frac{1}{N-1} \sum_{i=1}^{N} (u_i \bar{y}_{i..} - \bar{y}_{..})(u_i \bar{x}_{\alpha i..} - \bar{x}_{\alpha ...})$$

$$S_{iyx_\alpha} = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (v_{ij} \bar{y}_{ij.} - \bar{y}_{i..})(v_{ij} \bar{x}_{\alpha ij.} - \bar{x}_{\alpha i..}) \tag{2.6}$$

$$S_{ijyx_a} = \frac{1}{B_{ij} - 1} \sum_{j=1}^{B_{ij}} (y_{ijk} - \bar{y}_{ij.}) (x_{aijk} - \bar{x}_{aij.})$$

$$(a = 1, \quad 2, \ldots, p).$$

In what follows, any estimate with a dash will indicate an estimate based on the larger sample while one without it will indicate an estimate based on the smaller sub-sample, the estimator being the same in both the cases unless specified otherwise.

3. EXPECTATION AND VARIANCE OF $\bar{y}_R$

It can be shown that when x's are positive and sample size so large that terms of order higher than $1/n$ can be neglected:

$$E (r_a \bar{x}_{an'})$$
$$= \bar{y}_{...} \left[1 + \frac{V(\bar{x}_{an}) - V(\bar{x}_{an'})}{\bar{x}^2_{a...}} - \frac{\text{Cov.}(\bar{y}_n, \bar{x}_{an}) - \text{Cov.}(\bar{y}_{n'}, \bar{x}_{an'})}{\bar{y}_{...}\bar{x}_{a...}} \right] \qquad (3.1)$$

Hence the relative bias in the estimate $\bar{y}_R$ to the first degree of approximation is given by:

$$\sum_{a=1}^{p} w_a \left[\frac{v(\bar{x}_{an}) - v(\bar{x}_{an'})}{\bar{x}^2_{...}} - \frac{\text{Cov.}(\bar{y}_n, \bar{x}_{an}) - \text{Cov.}(\bar{y}_{n'}, \bar{x}_{an'})}{\bar{y}_{...} \bar{x}_{a...}} \right] \qquad (3.2)$$

Define

$$S_{byx_a} = \rho_{bya} S_{by} S_{bx_a}, \quad \frac{S^2_{bx_a}}{\bar{x}^2_{a...}} = C_{ba}^2$$

$$S_{iyx_a} = \rho_{iy} \, S_{iy} \, S_{ix_a}, \quad \frac{S^2_{ix_a}}{\bar{x}^2_{a...}} = C_{ia}^2 \qquad (3 \cdot 3)$$

$$S_{ijyx_a} = \rho_{ijy} \, S_{ijy} \, S_{ijx_a} \quad \frac{S^2_{ijx_a}}{\bar{x}^2_{a...}} = C^2_{ija}$$

$$S_{bx_a x_\beta} = \rho_{b\beta a} S_{x_a} S_{x_\beta}.$$

Using (3.3), the expression for the relative bias reduces to

$$\sum_{a=1}^{P} w_a \left[\left(\frac{1}{n} - \frac{1}{N} \right) (C_{ba}^2 - \rho_{bya} C_{by} C_{ba}) \right.$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} u_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right)(C_{ia}^2 - \rho_{iya}\, C_{iy}\, C_{ia})$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{u_i^2}{m_i M_i} \sum_{j=1}^{M_i} v_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)$$

$$\times \left. (C^2_{ija} - \rho_{ijya}\, C_{ijy}\, C_{ija}) \right]$$

— (Similar terms with n, m_i and b_{ij} replaced by n', m_i' and b_{ij}' respectively). (3.4)

In particular, assume that

(*A*) $m_i' = m'$, $m_i = m$, $b_{ij}' = b'$, $b_{ij} = b$.

(*A'*) $M_i = M$, $B_{ij} = B$.

(*B*) Finite correction factors can be ignored.

Let

$$\bar{S}_{wz}^{\,2} = \frac{1}{N} \sum_{i=1}^{N} S_{iz}^{\,2}, \quad \bar{\bar{S}}_{wz}^{\,2} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} S^2_{ijz} \tag{3.5}$$

Then it can be seen that the bias vanishes exactly if

$$\frac{\bar{y}_{...}}{\bar{x}_{a...}} = \frac{S_{byx_a}}{S^2_{bx_a}} = \frac{\bar{S}_{wyx_a}}{\bar{S}^2_{wx_a}} = \frac{\bar{\bar{S}}_{wyx_a}}{\bar{\bar{S}}^2_{wx_a}} \qquad (a = 1, 2, \ldots, p). \tag{3.6}$$

If further, we assume that:

(*C*) $n' = N$, $m_i' = m_i = M_i$, $b_{ij} = b_{ij}' = B_{ij}$

i.e., there is no sub-sampling and there is no double-sampling, the relative bias is given by

$$\sum_{a=1}^{p} w_a\, (C^2_{bx_a} - \rho_{byx_a}\, C_{by}\, C_{bx_a}) \tag{3.7}$$

which is the expression given by Olkin (1958).

It can be seen that to the first order of approximation

$$\text{Cov.}\,(r_\alpha \bar{x}_{\alpha n'},\ \ r_\beta \bar{x}_{\beta n'})$$

$$= \bar{y}^2_{...}\left[\frac{V(\bar{y}_n)}{\bar{y}^2_{...}} - \frac{\text{Cov.}(\bar{y}_n, \bar{x}_{\alpha n})}{\bar{y}_{...}\,\bar{x}_{\alpha ...}} - \frac{\text{Cov.}\,(\bar{y}_n, \bar{x}_{\beta n})}{\bar{y}_{...}\,\bar{x}_{\beta ...}}\right.$$

$$\left. + \frac{\text{Cov.}\,(\bar{x}_{\alpha n}, \bar{x}_{\beta n})}{\bar{x}_{\alpha ...}\,\bar{x}_{\beta ...}}\right] - \bar{y}^2_{...}\left[\frac{V(\bar{y}_{n'})}{\bar{y}^2_{...}} - \frac{\text{Cov.}\,(\bar{y}_{n'}, \bar{x}_{\alpha n'})}{\bar{y}_{...}\,\bar{x}_{\alpha ...}}\right.$$

$$\left. - \frac{\text{Cov.}\,(\bar{y}_{n'}, \bar{x}_{\beta n'})}{\bar{y}_{..}\,\bar{x}_{\beta ...}} + \frac{\text{Cov.}\,(\bar{x}_{\alpha n'}, \bar{x}_{\beta n'})}{\bar{x}_{\alpha ...}\,\bar{x}_{\beta ...}}\right] + V(\bar{y}_{n'}) \quad (3.8)$$

If we write the matrix $D_n = (d_{\alpha\beta})$, where the (α, β)-th element $d_{\alpha\beta}$ is given by

$$d_{\alpha\beta} = \frac{V(\bar{y}_n)}{\bar{y}^2_{...}} - \frac{\text{Cov.}\,(\bar{y}_n, \bar{x}_{\alpha n})}{\bar{y}_{...}\,\bar{x}_{\alpha ...}} - \frac{\text{Cov.}\,(\bar{y}_n, \bar{x}_{\beta n})}{\bar{y}_{...}\,\bar{x}_{\beta ...}} + \frac{\text{Cov.}\,(\bar{x}_{\alpha n}, \bar{x}_{\beta n})}{\bar{x}_{\alpha ...}\,\bar{x}_{\beta ...}}$$

and $D_{n'} = (d'_{\alpha\beta})$ with n, m_i, b_{ij} in $d_{\alpha\beta}$ replaced by n', m_i', b_{ij}', the variance of $\bar{y}_R$ can simply be written as

$$V(\bar{y}_R) = \sum_{\alpha=1}^{p} w_\alpha^2\, V(r_\alpha \bar{x}_{\alpha n'}) + \sum_{\alpha \neq B}^{p} w_\alpha w_\beta\, \text{Cov.}\,(r_\alpha \bar{x}_{\alpha n'}, r_\beta\, \bar{x}_{\beta n'})$$

$$= \bar{y}^2_{...}\ \ WDW' + V(\bar{y}_{n'}) \quad (3.9)$$

where $D = D_n - D_{n'}$.

Using (3.8), the variance of $\bar{y}_R$ can be written as

$$V(\bar{y}_R) = \bar{y}_{...} \sum_{\alpha,\beta}^{p} w_\alpha w_\beta \left[\left(\frac{1}{n} - \frac{1}{n'}\right) C_b^{\,2} + \frac{1}{nN} \sum_{i=1}^{N} \left(\frac{1}{m_i} - \frac{1}{M_i}\right) u_i^2 C_i^2\right.$$

$$- \frac{1}{n'N} \sum_{i=1}^{N} \left(\frac{1}{m_i'} - \frac{1}{M_i}\right) u_i^2 C_i^2$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{u_i^2}{m_i M_i} \sum_{j=1}^{M_i} \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) v_{ij}^2\, C_{ij}^2$$

$$\left. - \frac{1}{n'N} \sum_{i=1}^{N} \frac{u_i^2}{m_i' M_i} \sum_{j=1}^{M_i} \left(\frac{1}{b_{ij}'} - \frac{1}{B_{ij}}\right) v_{ij}^2\ \ C_{ij}^2\right]$$

$$+ \bar{y}^2_{\cdots}\left[\left(\frac{1}{n'} - \frac{1}{N}\right) C_{by}^2 + \frac{1}{n'N}\sum_{i=1}^{N}\left(\frac{1}{m_i'} - \frac{1}{M_i}\right)u_i^2 C_{iy}^2 + \frac{1}{n'N}\sum_{i=1}^{N}\frac{u_i^2}{m_i' M_i}\sum_{j=1}^{M_i}\left(\frac{1}{b_{ij}'} - \frac{1}{B_{ij}}\right) v_{ij}^2 C^2_{ijy}\right] \quad (3.10)$$

Where

$$\begin{aligned} C_b^2 &= C_{by}^2 - \rho_{bya}C_{by}C_{ba} - \rho_{by\beta}C_{by}C_{b\beta} + \rho_{ba\beta}C_{ba}C_{b\beta} \\ C_i^2 &= C_{iy}^2 - \rho_{iya}C_{iy}C_{ia} - \rho_{iy\beta}C_{iy}C_{i\beta} + \rho_{ia\beta}C_{ia}C_{i\beta} \\ C_{ij}^2 &= C^2_{ijy} - \rho_{ijya}C_{ijy}C_{ija} - \rho_{ijy\beta}C_{ijy}C_{ij\beta} + \rho_{ija\beta}C_{ija}C_{ij\beta} \end{aligned} \quad (3.11)$$

Under the assumptions (A), (A'), (B) and (3.5), the expression for $V(\bar{y}_R)$ simplifies to:

$$V(\bar{y}_R) = \sum_{\alpha,\beta}^{p} w_\alpha w_\beta \left[\left(\frac{1}{n} - \frac{1}{n'}\right) S_b^2 + \left(\frac{1}{nm} - \frac{1}{n'm'}\right)\bar{S}_w^2 + \left(\frac{1}{nmb} - \frac{1}{n'm'b'}\right)\bar{\bar{S}}_w^2\right] + \frac{S_{by}^2}{n'} + \frac{\bar{S}_{wy}^2}{n'm'} + \frac{\bar{\bar{S}}_{wy}^2}{n'm'b'} \quad (3.12)$$

Where

$$\begin{aligned} S_b^2 &= S_{by}^2 - R_\alpha S_{byx_\alpha} - R_\beta S_{byx_\beta} + R_\alpha R_\beta S_{bx\alpha} S_{bx_\beta} \\ \bar{S}_w^2 &= \bar{S}_{wy}^2 - R_\alpha \bar{S}_{wyx_\alpha} - R_\beta \bar{S}_{wyx_\beta} + R_\alpha R_\beta \bar{S}_{wx_\alpha}\bar{S}_{wx_\beta} \\ \bar{\bar{S}}_w^2 &= \bar{\bar{S}}_{wy}^2 - R_\alpha \bar{\bar{S}}_{wyx_\alpha} - R_\beta \bar{\bar{S}}_{wyx_\beta} + R_\alpha R_\beta \bar{\bar{S}}_{wx_\alpha}\bar{\bar{S}}_{wx_\beta} \\ R_\alpha &= \frac{\bar{y}_{\dots}}{\bar{x}_{\alpha\dots}} \quad \text{and} \quad R_\beta = \frac{\bar{y}_{\dots}}{\bar{x}_{\beta\dots}} \end{aligned} \quad (3.13)$$

4. Optimum Weight Function

Of all the weight vectors with the condition $\sum_{1}^{p} w_\alpha = 1$, we choose that vector for which $V(\bar{y}_R)$ is minimum. From the expression for the variance of $\bar{y}_R$ given in (3.9), it is clear that it will suffice to minimize WDW'.

Let

$$\phi = WDW' - 2\lambda(we' - 1)$$

Where λ is the lagrangian multiplier and $e = (1, 1, \ldots, 1)$. Differentiating ϕ with respect to W and equating the result to zero vector for a minimum, we obtain:

$$WD - \lambda e = 0.$$

Assuming D^{-1} to exist, we obtain

$$W = \lambda e D^{-1},$$

so that

$$\lambda = \frac{1}{eD^{-1}e'}$$

and therefore the optimum vector W is given by

$$W = \frac{eD^{-1}}{eD^{-1}e'}, \tag{4.1}$$

and the optimum variance is given by

$$V(\bar{y}_R)_{\text{opt.}} = \frac{\bar{y}^2_{\ldots}}{eD^{-1}e'} + V(\bar{y}_{n'}) \tag{4.2}$$

Under the assumption (C), it is clear that $V(\bar{y}_{n'}) = 0$ and the matrix D reduces to G/n where the (α, β)-th element of the matrix G is given by

$$g_{\alpha\beta} = \frac{N-n}{N}(C_y^2 - \rho_{yx_\alpha} - C_y C_{x_\alpha} - \rho_{yx_\beta} C_y C_{x_\beta} + \rho_{x_\alpha x_\beta} C_{x_\alpha} C_{x_\beta}) \tag{4.3}$$

and the optimum variance of $\bar{y}_R$ reduces to

$$V(\bar{y}_R)_{\text{opt.}} = \frac{\bar{y}_{\ldots}}{neG^{-1}e'}, \tag{4.4}$$

which is the result given by Olkin (1958).

Using the method of Olkin (1958), it can again be shown that in the case of multiphase sampling, the estimate based on q variables is more efficient than the one based on p auxiliary variables whenever q is greater than p.

5. ESTIMATION OF THE VARIANCE OF $\bar{y}_R$

If the bias is negligible, R_α may be estimated by $\hat{R}_\alpha = \bar{y}_n/\bar{x}_{\alpha n}$, so that a consistent estimate of $\bar{y}^2_{...}\ d_{\alpha\beta}$ is given by

$$\text{Est. } V(\bar{y}_n) - \hat{R}_\alpha \text{ Est. Cov. } (\bar{y}_n, \bar{x}_{\alpha n}) - \hat{R}_\beta \text{ Est. Cov. } (\bar{y}_n, \bar{x}_{\beta n}) + \hat{R}_\alpha \hat{R}_\beta \text{ Est. Cov. } (\bar{x}_{\alpha n}, \bar{x}_{\beta n}) \tag{5.1}$$

with a similar expression for the estimate of $\bar{y}^2_{...}\ d'_{\alpha\beta}$, where Est. $V(\bar{y}_n)$, Est. Cov. $(\bar{y}_n, \bar{x}_{\alpha n})$, Est. Cov. $(\bar{y}_n, \bar{x}_{\beta n})$ and Est. Cov. $(\bar{x}_{\alpha n}, \bar{x}_{\beta n})$ denote unbiased estimates of $V(\bar{y}_n)$, Cov. $(\bar{y}_n, \bar{x}_{\alpha n})$, Cov. $(\bar{y}_n, \bar{x}_{\beta n})$ and Cov. $(\bar{x}_{\alpha n}, \bar{x}_{\beta n})$ respectively. Ignoring the bias, we have approximately,

$$\text{Est. } D^{-1} = (\text{Est. } D)^{-1} = \hat{D}^{-1} \tag{5.2}$$

Also the matrix D is estimated by

$$\hat{D} = (\text{Est. } \bar{y}^2_{...}\, d_{\alpha\beta} - \text{Est. } \bar{y}^2_{...}\, d'_{\alpha\beta}) \tag{5.3}$$

Hence, a consistent estimate of the variance of $\bar{y}_R$ is given by

$$\text{Est. } V(\bar{y}_R) = \frac{1}{e\hat{D}^{-1}e'} + \text{Est. } V(\bar{y}_{n'}) \tag{5.4}$$

6. EFFICIENCY OF MULTI-PHASE SAMPLING

Let

$$\frac{\bar{S}_{wy}{}^2}{\bar{y}^2_{...}} = \bar{C}_{wy}{}^2, \quad \frac{\bar{\bar{S}}_{wy}{}^2}{\bar{y}^2_{...}} = \bar{\bar{C}}_{wy}{}^2,$$

$$\frac{\bar{S}_{wx}{}^2}{\bar{x}_{\alpha...}} = \bar{C}_{w\alpha}{}^2, \quad \frac{\bar{\bar{S}}^2_{wx\alpha}}{\bar{x}^2_{\alpha...}} = \bar{\bar{C}}_w{}^2,$$

$$\bar{S}_{wx_\alpha x_\beta} = \bar{\rho}_{w\alpha\beta}\, \bar{S}_{wx_\alpha} \bar{S}_{wx_\beta}, \quad \bar{\bar{S}}_{wx_\alpha x_\beta} = \bar{\bar{\rho}}_{wx_\alpha x_\beta} \bar{\bar{S}}_{wx_\alpha} \bar{\bar{S}}_{wx_\beta}$$

$$\bar{\bar{S}}_{wy} x_\alpha = \bar{\rho}_{wy\alpha} \bar{S}_{wy} \bar{S}_{wx_\alpha}, \quad \bar{\bar{S}}_{wyx_\alpha} = \bar{\bar{\rho}}_{wy\alpha} \bar{\bar{S}}_{wy} \bar{\bar{S}}_{wx_\alpha}.$$

Further, for the sake of simplicity, let

$$C_{b\alpha} = C_b, \ \bar{C}_{w\alpha} = \bar{C}_w, \ \bar{\bar{C}}_{w\alpha} = \bar{\bar{C}}_w$$

$$\rho_{by\alpha} = \rho_{by}, \ \bar{\rho}_{wy\alpha} = \bar{\rho}_{wy}, \ \bar{\bar{\rho}}_{wy\alpha} = \bar{\bar{\rho}}_{wy}.$$

$$\rho_{b\alpha\beta} = \rho_b, \ \bar{\rho}_{w\alpha\beta} = \bar{\rho}_w, \ \bar{\bar{\rho}}_{w\alpha\beta} = \bar{\bar{\rho}}_w.$$

Then assuming (*A*), (*A'*), (*B*) and (3.5), it can be seen that the optimum weight vector is given by

$$W = \left(\frac{1}{p}, \frac{1}{p}, \ldots, \frac{1}{p}\right)$$

and the expression for the variance $V(\bar{y}_R)$ reduces to

$$V(\bar{y}_R) = \bar{y}^2_{\ldots}\left(\frac{C_{by}^2}{n} + \frac{\bar{C}_{wy}^2}{nm} + \frac{\bar{\bar{C}}_{wy}^2}{nmb}\right)$$

$$+ \frac{\bar{y}^2_{\ldots}}{p}\left[\left(\frac{1}{n} - \frac{1}{n'}\right) C_b^2 (1 - \rho_b)\right.$$

$$+ \left(\frac{1}{nm} - \frac{1}{n'm'}\right) \bar{C}_w^2 (1 - \bar{\rho}_w)$$

$$+ \left(\frac{1}{nmb} - \frac{1}{n'm'b'}\right) \bar{\bar{C}}_w^2 (1 - \bar{\bar{\rho}}_w)$$

$$+ p\left(\frac{1}{n} - \frac{1}{n'}\right)(\rho_b C_b^2 - 2\rho_{by} C_b C_{by})$$

$$+ p\left(\frac{1}{nm} - \frac{1}{n'm'}\right)(\bar{\rho}_w \bar{C}_w^2 - 2\bar{\rho}_{wy} \bar{C}_w \bar{C}_{wy})$$

$$\left. + p\left(\frac{1}{nmb} - \frac{1}{n'm'b'}\right)(\bar{\bar{\rho}}_w \bar{\bar{C}}_w^2 - 2\bar{\bar{\rho}}_{wy} \bar{\bar{C}}_w \bar{\bar{C}}_{wy})\right] \quad (6.1)$$

Under the same assumptions, the variance of $\bar{y}_n$ reduces to

$$V(\bar{y}_n) = \bar{y}^2_{\ldots}\left(\frac{C_{by}^2}{n} + \frac{\bar{C}_{wy}^2}{nm} + \frac{\bar{\bar{C}}_{wy}^2}{nmb}\right) \quad (6.2)$$

It follows that if the following inequalities hold, namely

$$C_b^2 (1 - \rho_b) + p(\rho_b C_b^2 - 2\rho_{by} C_b C_{by}) < 0$$

$$\bar{C}_w^2 (1 - \bar{\rho}_w) + p(\bar{\rho}_w \bar{C}_w^2 - 2\bar{\rho}_{wy} \bar{C}_w \bar{C}_{wy}) < 0$$

$$\bar{\bar{C}}_w^2 (1 - \bar{\bar{\rho}}_w) + p(\bar{\bar{\rho}}_w \bar{\bar{C}}_w^2 - 2\bar{\bar{\rho}}_{wy} \bar{\bar{C}}_w \bar{\bar{C}}_{wy}) < 0$$

the estimate $\bar{y}_R$ based on multi-phase sampling will be more efficient than the estimate $\bar{y}_n$. On simplification, these conditions reduce to:

$$\frac{p\rho_{by}}{1 + (p-1)\rho_b} > \frac{1}{2}\frac{C_b}{C_y}$$

$$\frac{p\bar{\rho}_{wy}}{1 + (p-1)\bar{\rho}_w} > \frac{1}{2}\frac{\bar{C}_w}{\bar{C}_{wy}} \quad (6.3)$$

$$\frac{p\bar{\bar{\rho}}_{wy}}{1+(p-1)\bar{\bar{\rho}}_{w}} > \frac{1}{2}\frac{\bar{\bar{C}}_{w}}{\bar{\bar{C}}_{wy}}.$$

7. Illustration with Special Reference to Survey on Guava in U.P.

In this section, we shall develop appropriate procedures based on the use of two auxiliary variables for estimating the population total with special reference to the sampling design adopted by the Institute of Agricultural Research Statistics, I.C.A.R., in the pilot sample survey on guava crop, carried out in Allahabad District of Uttar Pradesh during the year 1961–62.

The design adopted for the survey was stratified multi-stage random sampling with tehsils constituting the strata, villages, the primary units of sampling, orchards under guava the secondary units and trees the ultimate units of sampling. In each stratum, a certain number of villages were selected with equal probability and without replacement. In each selected village, all the guava orchards were completely enumerated to determine the number of bearing trees and the area under guava orchards. A sub-sample of these villages was selected for the purpose of yield study. Within each of these selected villages, a specified number of guava orchards were selected at random. Within each selected orchard, a certain number of bearing trees were selected at random to record the yield of the crop.

For the purpose of illustration, we will derive result for one stratum only. Let

N = number of villages in the tehsil,

M_i = number of guava orchards in the i-th village ($i = 1, 2, \ldots, N$).

B_{ij} = number of bearing guava trees in the j-th orchard of the i-th village ($j = 1, 2, \ldots, M_i$).

A_{ij} = area of the j-th orchard in the i-th village.

y_{ijk} = yield of the k-th tree in the j-th orchard of the i-th village ($k = 1, 2, \ldots, B_{ij}$).

n' = number of villages selected for complete enumeration.

n = number of villages sub-sampled for yield study.

m_i = number of orchards selected in the i-th village contained in n.

b_{ji} = number of trees selected in the (i, j)-th orchard contained in n.

$$y_{...} = \sum_{i=1}^{N} y_{i..} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} y_{ij.} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{k=1}^{B_{ij}} y_{ijk.}$$

$$A_{.} = \sum_{i=1}^{N} A_{i.} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} A_{ij}$$

$$B_{..} = \sum_{i=1}^{N} B_{i.} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} B_{ij}$$

$$\bar{y}_B = \frac{y_{...}}{B_{..}}, \quad \bar{y}_A = \frac{y_{...}}{A_{..}}.$$

Then an unbiased estimate of the total number of bearing trees is given by

$$B_{n'} = \frac{N}{n} \sum_{i}^{n'} B_{i}.$$

Similarly, an unbiased estimate of total area under guava orchards is given by:

$$A_{n'} = \frac{N}{n} \sum_{i}^{n'} A_{i}$$

Defining

$$y_n = \frac{N}{n} \sum_{i}^{n} \frac{M_i}{m_i} \sum_{j}^{M} B_{ij} \bar{y}_{ij}$$

$$B_n = \frac{N}{n} \sum_{i}^{n} \frac{M_i}{m_i} \sum_{j}^{m_i} B_{ij}$$

$$A_n = \frac{N}{n} \sum_{i}^{n} \frac{M_i}{m_i} \sum_{j}^{m_i} A_{ij}$$

it can be shown that y_n, B_n and A_n are unbiased estimates of $y_{...}$, $B_{..}$ and $A_{..}$ respectively.

For the purpose of illustration, we shall consider the following four estimates:

(i) y_n

(ii) $T_1 = \frac{y_n}{B_n} \cdot B_{n'}$

(iii) $T_2 = \frac{y_n}{A_n} \cdot A_{n'}$

(iv) $T = wT_1 + (1 - w) T_2$

where w is so determined that $V(T)$ is minimum.

Then, it can be shown that

$$V(T_1) = \left(\frac{1}{n} - \frac{1}{n'}\right)(S_{by}^2 + \bar{y}_B^2 S_{bB}^2 - 2\bar{y}_B S_{bBy}) + \left(\frac{1}{n'} - \frac{1}{N}\right) S_{by}^2$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right)(S_{iy}^2 + \bar{y}_B^2 S_{iB}^2 - 2\bar{y}_B S_{iBy})$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{M_i}{m_i} \sum_{j=1}^{M_i} B_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) S^2_{ijy}$$

with a similar expression for $V(T_2)$ and

$$\text{Cov.}(T_1, T_2) = \left(\frac{1}{n} - \frac{1}{n'}\right)(S_{by}^2 + \bar{y}_B \bar{y}_A S_{bBA} - \bar{y}_B S_{bBy} - \bar{y}_A S_{bAy})$$

$$+ \left(\frac{1}{n'} - \frac{1}{N}\right) S_{by}^2$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right)$$

$$\times (S_{iy}^2 + \bar{y}_B \bar{y}_A S_{iAB} - \bar{y}_B S_{iBy} - \bar{y}_A S_{iAy})$$

$$+ \frac{1}{nN} \sum_{i=1}^{N} \frac{M_i}{m_i} \sum_{j=1}^{M_i} B_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) S^2_{ijy}$$

$$S_{by}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_{i..} - \frac{y_{...}}{N} \right)^2$$

$$S_{iy}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} \left(y_{ij.} - \frac{y_{i..}}{M_i} \right)^2$$

$$S^2_{ijy} = \frac{1}{B_{ij} - 1} \sum_{k=1}^{B_{ij}} \left(y_{ijk} - \frac{y_{ij.}}{B_{ij}} \right)^2$$

$$S_{bB}^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(B_{i.} - \frac{B_{..}}{N} \right)^2$$

$$S_{iB}^2 = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} \left(B_{ij} - \frac{B_{i.}}{M_i} \right)^2$$

$$S_{bBy} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_{i..} - \frac{y_{...}}{N} \right) \left(B_{i.} - \frac{B_{..}}{N} \right)$$

$$S_{iBy} = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} \left(y_{ij.} - \frac{y_{i..}}{M_i} \right) \left(B_{ij} - \frac{B_{i.}}{M_i} \right)$$

$$S_{bBA} = \frac{1}{N-1} \sum_{i=1}^{N} \left(B_{i.} - \frac{B_{..}}{N} \right) \left(A_{i.} - \frac{A_{..}}{N} \right)$$

$$S_{iAB} = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} \left(B_{ij} - \frac{B_{i.}}{M_i} \right) \left(A_{ij} - \frac{A_{i.}}{M_i} \right)$$

and

$$S_{bA}^2,\ S_{iA}^2,\ S_{bAy},\ S_{iAy}$$

are defined in a similar manner.

From the theory developed earlier, it is clear that the optimum weight w is given by

$$W = \frac{V(T_2) - \text{Cov.}\ (T_1, T_2)}{V(T_1) - 2\,\text{Cov.}\ (T_1, T_2) + V(T_2)}.$$

To estimate the variances, consider the quantities

$$s^2_{ijy} = \frac{1}{b_i - 1} \sum^{b_{ij}} (y_{ijk} - \bar{y}_{ij})^2$$

$$s_{iy}^2 = \frac{1}{m_i - 1} \sum^{m_i} \left(B_{ij}\,\bar{y}_{ij} - \frac{1}{m_i} \sum^{m_i} B_{ij}\bar{y}_{ij} \right)^2$$

$$s_{by}^2 = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i} \sum^{m_i} B_{ij}\,\bar{y}_{ij} - \frac{1}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} B_{ij}\bar{y}_{ij} \right)^2$$

$$s_{bB}^2 = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i} \sum^{m_i} B_{ij} - \frac{1}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} B_{ij} \right)^2$$

$$s_{iB}^2 = \frac{1}{m_i - 1} \sum^{m_i} \left(B_{ij} - \frac{1}{m_i} \sum^{m_i} B_{ij} \right)^2$$

$$s_{bBy} = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i} \sum^{m_i} B_{ij} - \frac{1}{n} \sum^{n_i} \frac{M_i}{m_i} \sum B_{ij} \right) \times \left(\frac{M_i}{m_i} \sum^{m_i} B_{ij}\,\bar{y}_{ij} - \frac{1}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} B_{ij}\,\bar{y}_{ij} \right)$$

$$s_{iBy} = \frac{1}{m_i - 1} \sum^{m_i} \left(B_{ij} - \frac{1}{m_i} \sum^{m_i} B_{ij} \right) \times \left(B_{ij}\,\bar{y}_{ij} - \frac{1}{m_i} \sum^{m_i} B_{ij}\ \bar{y}_{ij} \right)$$

$$s_{bAB} = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i} \sum^{m_i} B_{ij} - \frac{1}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} B_{ij} \right) \times \left(\frac{M_i}{m_i} \sum^{m_i} A_{ij} - \frac{1}{n} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} A_{ij} \right)$$

$$s_{iAB} = \frac{1}{m_i - 1} \sum^{m_i} \left(B_{ij} - \frac{1}{m_i} \sum^{m_i} B_{ij} \right) \times \left(A_{ij} - \frac{1}{m_i} \sum^{m_i} A_{ij} \right)$$

with similar expressions for s_{bA}^2 s_{iA}^2, s_{bAy} and s_{iAy}.

It can then be shown that consistent estimates of $V(T_1)$, $V(T_2)$ and Cov. (T_1, T_2) are given by

$$\frac{1}{N^2} \text{ Est. } V(T_1)$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) s_{by}^2 + \frac{1}{Nn} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{iy}^2$$

$$+ \frac{1}{Nn} \sum^{n} \frac{M_i}{m_i} \sum^{m_i} B_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) s^2_{ijy} + \hat{\bar{y}}_B^2 \left(\frac{1}{n} - \frac{1}{n'}\right) s_{bB}^2$$

$$+ \frac{\hat{\bar{y}}_B^2}{nn'} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{iB}^2 - 2\hat{\bar{y}}_B \left(\frac{1}{n} - \frac{1}{n'}\right) s_{bBy}$$

$$- \frac{2\hat{\bar{y}}_B}{nn'} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{iBy}$$

with a similar expression for Est. $V(T_2)$, and

$$\frac{1}{N^2} \text{ Est. Cov. } (T_1, T_2)$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) s_{by}^2 + \frac{1}{Nn} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_t}\right) s_{iy}^2$$

$$+ \frac{1}{Nn} \sum^{n} \frac{M_i}{m_i} \sum^{mi} B_{ij}^2 \left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right) s^2_{ijy}$$

$$+ \hat{\bar{y}}_B \hat{\bar{y}}_A \left(\frac{1}{n} - \frac{1}{n'}\right) s_{bAB} + \frac{\hat{\bar{y}}_B \hat{\bar{y}}_A}{nn'} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i}\right) s_{iAB}$$

$$- \hat{\bar{y}}_B \left(\frac{1}{n} - \frac{1}{n'}\right) s_{bBy} - \hat{\bar{y}}_A \left(\frac{1}{n} - \frac{1}{n'}\right) s_{bAy}$$

$$- \frac{\hat{y}_B}{nn'} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_{iBy}$$

$$- \frac{\hat{y}_A}{nn'} \sum^{n} M_i^2 \left(\frac{1}{m_i} - \frac{1}{M_i} \right) s_{iAy}.$$

Where

$$\hat{y}_B = \frac{y_n}{B_n} \text{ and } \hat{y}_A = \frac{y_n}{A_n}.$$

For the purpose of illustration, we shall consider the data for chail tehsil only. Here $N = 153$, $n' = 88$, $n = 27$;

$\hat{y}_B = 0 \cdot 46$ md. per bearing tree.

$\hat{y}_A = 49 \cdot 71$ md. per acre.

Est. $V(T_1) = 189700528 \cdot 1$

Est. $V(T_2) = 238179970 \cdot 0$

Est. Cov. $(T_1, T_2) = 198656922 \cdot 0$

Est. $V(y_n) = 430975066 \cdot 8$

Est. $w = 1 \cdot 29$.

Est. $V(T) = 187076041 \cdot 8$.

The table below gives the relative efficiencies for the different estimates:

Estimate	..	y_n	T_1	T_2	T
% Relative efficiency	..	100	227	181	230

These results clearly show that the use of one or more auxiliary variables has proved efficient, the gain in efficiency ranging from 81% to 130%. The weighted estimate T based on both the auxiliary variables, is the most efficient of all the estimates.

8. Application to the Case of Varying Probabilities with Special Reference to the Survey on Mango in U.P.

In the theory developed so far, it has been assumed that at each stage, the sampling units are selected with equal probability and without replacement. It is however a common practice to find situations, especially in large-scale surveys, where the primary units are selected with varying probabilities and with replacement. The extension of the theory to such cases does not present any special difficulties and is straightforward. We shall merely give an indication with special reference to the sampling design adopted by the Institute of Agricultural Research Statistics, in the pilot sample survey on mango carried out in Varanasi District of U.P. during the year 1960–61. The design adopted for the survey was the same as in the case of survey on guava described in the previous section, except that the villages were selected with probability proportional to area under fresh fruits and with replacement.

Let p_i ($i = 1, 2, \ldots, N$) be the selection probabilities corresponding to N villages in a stratum. It is then clear that

$$Y_n' = \frac{1}{n} \sum^{n} \frac{M_i}{p_i m_i} \sum^{m_i} B_{ij} \bar{y}_{ij}$$

$$B_n' = \frac{1}{n} \sum^{n} \frac{M_i}{p_i m_i} \sum^{m_i} B_{ij}$$

$$A_n' = \frac{1}{n} \sum^{n} \frac{M_i}{p_i m_i} \sum^{m_i} A_{ij}$$

are unbiased estimates of $y_{...}$, $B_{..}$ and $A_{..}$ respectively. As in the previous section, we shall only consider the following three estimates

$$T_1' = \frac{y_n'}{B_n'} \cdot B'_{n'}$$

$$T_2' = \frac{y_n'}{A_n'} \cdot A'_{n'}$$

and

$$T = wT_1' + (1 - w)\, T_2',$$

where

$$B'_{n'} = \frac{1}{n'} \sum^{n'} \frac{B_{i.}}{p_i} \quad \text{and} \quad A'_{n'} = \frac{1}{n'} \sum^{n'} \frac{A_{i.}}{p_i}$$

are unbiased estimates of $B_{..}$ and $A_{..}$ respectively.

Then it can be shown that

$$V(y_n') = \frac{\sigma_{by}^2}{n} + \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\left(\frac{1}{m_i} - \frac{1}{M_i}\right)S_{iy}^2$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\frac{M_i}{m_i p_i}\sum_{j=1}^{M_i} B_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S^2_{ijy}$$

$$V(T_1') = \left(\frac{1}{n} - \frac{1}{n'}\right)(\sigma_{by}^2 + \bar{y}_B^2\,\sigma_{bB}^2 - 2\bar{y}_B\sigma_{bBy})$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\left(\frac{1}{m_i} - \frac{1}{M_i}\right)(S_{iy}^2 + \bar{y}_B^2 S_{iB}^2 - 2\bar{y}_B S_{iBy})$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\frac{M_i}{m_i p_i}\sum_{j=1}^{M_i} B_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S^2_{ijy} + \frac{\sigma_{by}^2}{n'}$$

with a similar expression for $V(T_2')$, and

Cov. (T_1', T_2')

$$= \left(\frac{1}{n} - \frac{1}{n'}\right)(\sigma_{by}^2 - \bar{y}_B\,\sigma_{bBy} - \bar{y}_A\sigma_{bAy} + \bar{y}_B\bar{y}_A\,\sigma_{bAB})$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\left(\frac{1}{m_i} - \frac{1}{M_i}\right)$$

$$\times\, (S_{iy}^2 - \bar{y}_B S_{iBy} - \bar{y}_A\, S_{iAy} + \bar{y}_B\,\bar{y}_A\, S_{iAB})$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\frac{M_i}{m_i p_i}\sum_{j=1}^{M_i} B_{ij}^2\left(\frac{1}{b_{ij}} - \frac{1}{B_{ij}}\right)S^2_{ijy} + \frac{\sigma_{by}^2}{n'}$$

where

$$\sigma_{by}^2 = \sum_{i=1}^{N} p_i\left(\frac{y_{i..}}{p_i} - \bar{y}_{...}\right)^2$$

$$\sigma_{bB}^2 = \sum_{i=1}^{N} p_i\left(\frac{B_{i.}}{p_i} - B_{..}\right)^2$$

$$\sigma_{bBy} = \sum_{i=1}^{N} p_i \left(\frac{y_{i..}}{p_i} - y_{...} \right) \left(\frac{B_{i.}}{p_i} B_{.} \right)$$

$$\sigma_{bAB} = \sum_{i=1}^{N} p_i \left(\frac{B_{i.}}{p_i} - B_{..} \right) \left(\frac{A_i}{p_i} - A_{..} \right)$$

with similar expressions for $\sigma_{bA}{}^2$, σ_{bAy}. The other quantities $S_{iy}{}^2$, $S^2{}_{ijy}$, $S_{iB}{}^2$, S_{iBy}, S_{iAy}, etc., are defined as in the previous section.

For the purpose of estimation, define

$$s_{by}{}'^2 = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i p_i} \sum^{m_i} B_{ij} \bar{y}_{ij} - y_n' \right)^2$$

$$s_{bB}{}'^2 = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i p_i} \sum^{m_i} B_{ij} - B_n' \right)^2$$

$$s'_{bBy} = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i p_i} \sum^{m_i} B_{ij} \dot{y}_{ij} - y_n' \right)$$

$$\times \left(\frac{M_i}{m_i p_i} \sum^{m_i} B_{ij} - \dot{B}_n' \right)$$

$$s'_{bAB} = \frac{1}{n-1} \sum^{n} \left(\frac{M_i}{m_i p_i} \sum^{m_i} B_{ij} - B_n' \right)$$

$$\times \left(\frac{M_i}{m_i p_i} \sum^{m_i} A_{ij} - A_n' \right)$$

with similar expressions for $S_{bA}{}'^2$, S'_{bAy}. The quantities $s_{iy}{}^2$, s_{iBy}, s_{iAy}, $s^2{}_{ijy}$, $s_{iA}{}^2$, $s_{iB}{}^2$ and s_{iAB} are defined as in the previous section.

Then it can be shown that consistent estimates of $V(T_1')$, $V(T_2')$ and Cov. (T_1', T_2') are given by:

Est. $V(T_1')$

$$= \frac{s_{by}'^2}{n'} + \left(\frac{1}{n} - \frac{1}{n'}\right)(s_{by}'^2 + \hat{y}_B^2 s_{bB}'^2 - 2\hat{y}_B s'_{bBy})$$

$$+ \frac{1}{nn'} \sum^n \frac{M_i^2}{p_i^2}\left(\frac{1}{m_i} - \frac{1}{M_i}\right)(\hat{y}_B^2 s_{iB}^2 - 2\hat{y}_B s_{iBy})$$

with a similar expression for Est. $V(T_2')$, and

Est. Cov. (T_1', T_2')

$$= \frac{s_{by}'^2}{n'} + \left(\frac{1}{n} - \frac{1}{n'}\right)(s_{by}'^2 + \hat{y}_A\hat{y}_B s'_{bAB} - \hat{y}_B s'_{bBy} - \hat{y}_A s'_{bAy})$$

$$+ \frac{1}{nn'} \sum^n \frac{M_i^2}{p_i}\left(\frac{1}{m_i} - \frac{1}{M_i}\right)$$

$$\times (\hat{y}_A\hat{y}_B s'_{iAB} - \hat{y}_B s_{iBy}) - \hat{y}_A s_{iAy}$$

with

$$\hat{y}_A = \frac{y_n'}{A_n'} \quad \text{and} \quad \hat{y}_B = \frac{y_n'}{B_n'}.$$

For the purpose of illustration, we shall consider the data from Varanasi tehsil only. Here $n' = 66$, $n = 27$. Results concerning the relative efficiency of different estimates based on this data are given below:

Estimate	.. y_n'	T_1'	T_2'	T'
% Relative efficiency	.. 100	139	92	142

It will be seen that the use of area under mango as an auxiliary variable has not proved efficient. There is a positive gain in the efficiencies of the estimates T_1', T', the gain in efficiency ranging from 39% to 42%. As before we observe that the weighted estimate T is the most efficient of all the estimates considered here.

9. References

1. Olkin, Ingram .. "Multivariate ratio estimation for finite populations," *Biometrika*, 1958, **45**, 154–65.
2. Sukhatme, B. V. and Koshal, R. S. "A contribution to double sampling," *Jour. Ind. Soc. Agri. Stat.*, 1959, **11**, 128–44.

CERTAIN DISTRIBUTION-FREE TESTS OF REGRESSION*

By R. S. Kurup ‡

University of Kerala, Trivandrum

1. The Problem

Suppose we are given n pairs of observations (x_i, y_i), $i = 1, 2 \cdots n$ from a continuous bivariate distribution and we are required to fit a relation of the form $Y = f(x, \theta)$ where 'θ' denotes a set of parameters whose values may be found by any method of estimation. To test the significance of regression, the null hypothesis is $H_0 : \theta = 0$. Classical workers tested regression by assuming that the errors are normally and independently distributed and this forms the basis of the x^2-test. In this paper the problem is tackled without any such assumptions.

For this problem, Brown and Mood (1950),[1] (1951)[2] suggested a statistic,

$$A = \frac{8}{n}\left\{\left(r_1 - \frac{n}{4}\right)^2 + \left(r_2 - \frac{n}{4}\right)^2\right\}$$

where r_1 and r_2 are the number of positive ϵ's below and above the median of the x's, ϵ being the discrepancy between the observed 'y' and the value of 'y' under the null hypothesis. For moderately large 'n', this is distributed as a 'x^2' with 2 degrees of freedom. This statistic considers the 4 possible arrangements of signs, as shown below:

$n/2$	$n/2$
+ + + + ··· +	− − − − ··· −
+ + + + ··· +	+ + + + ··· +
− − − − ··· −	− − − − ··· −
− − − − ··· −	+ + + + ··· +

* Paper prepared in 1955 under a Government of India Senior Scholarship.

‡ The author is at present employed in the Bereau of Economics and Statistics, Trivandrum.