Estimating Population Mean in Sample Surveys

Dr. Vyas Dubey¹, Dr. Minal Uprety² and Ujjwal Dubey³
¹Professor and Head, School of Studies in Statistics, Pt. Ravishankar Shukla University, Raipur, C.G., INDIA
²Assistant Professor, Department of Quantitative Techniques/Systems, Prestige Institute of Management and Research, Indore, M.P., INDIA
³Computer Science Engineer, Department of Computer Science and Engineering, Shri Shankaracharya Group of Institutions, Bhilai, C.G., INDIA

³Corresponding Author: ujjwalubey1212@gmail.com

ABSTRACT
The paper deals with a generalized estimator of population mean which includes several estimators as its particular cases. Under certain conditions, the proposed estimator is more efficient than existing estimators. Results are supported by numerical illustration.

Keywords— Bias, Mean Squared Error (MSE), Relative Efficiency, Simple Random Sampling without Replacement (SRSWOR), Probability Proportional to Size (PPS) Sampling

I. INTRODUCTION
Let y and x study and auxiliary variables, taking values (Yᵢ, Xᵢ) on ith unit of a finite population U= (1,2,…,N). Further, let (Ŷ, X̂) be unbiased estimators of population means (Ŷ, X̂) of (y, x) respectively. If y and x are highly positively correlated and X̂ is known; then for estimating population mean Ŷ ratio estimator Ŷᵣ = ŶX̂/X̂ is used in practice while for negatively correlated variables product estimator Ŷ_p = ŶX̂/X̂ is suggested [Murthy (1964)]. Many attempts have been made in literature to improve such estimators. Srivastava (1967, 1971) proposed exponential type estimators of Ŷ. Reddy (1974) discussed a transformed estimator of Ŷ after making transformation on auxiliary variable x. Chakraborty (1968), Vos (1980), Adhvaryu and Gupta (1983), Chaubey, Singh and Dwivedi (1984) and others proposed several weighted estimators of Ŷ, Ŷᵣ and Ŷ_p. But all these estimators are equally efficient as linear regression estimator

Ŷ₀ = Ŷ + β(X̂ - X̂) (1.1)

where ̂β is sample estimate of β = Cov(Ŷ, X̂)/V(Ŷ̂) [Sarndal et al. (1992)].

Das and Tripathi (1980) and Das (1988) considered

e₀ = α₁Ŷ + α₂(X̂ - X̂) , α₁ + α₂ ≠ 1 (1.2)

e₁ = WŶ₀ (1.3)

for improving Ŷ₀. The minimum MSE of both estimators e₀ and e₁ is same, which is almost equal to MSE of Ŷ₀ for large samples [see also Rao (1991)]. Dubey (2003) re-studied the problem and suggested

e₂ = ψ₁Ŷ + ψ₂(X̂ - X̂) + (1 - ψ₁)X̂ (1.4)

which is considerably more efficient than all the above estimators if
The condition \( e_3 = \tilde{Y} + \varphi_1 (\tilde{X} - \tilde{\hat{X}}) + \varphi_2 \{\mu_2(x) - \hat{\mu}_2(x)\} \)
is more precise than \( \tilde{Y}_\lambda \) whenever a quadratic type relationship between \( y \) and \( x \) exists. In section 2, we propose an estimator of \( \bar{Y} \) which is further more precise than all the above estimators.

\[
\hat{\tilde{Y}}_g = \lambda_1 \tilde{Y} + \lambda_2 (\tilde{X} - \tilde{\hat{X}}) + \lambda_3 \{\mu_2(x) - \hat{\mu}_2(x)\} + (1 - \lambda_1) \bar{X}
\]

where \( \lambda_i \), \( i = 1,2,3 \) are suitably chosen constants.

\[
V_{abc} = E[(\tilde{Y} - \bar{Y})^a(\tilde{X} - \bar{X})^b(\hat{\mu}_2(x) - \mu_2(x))^c]; (a,b,c)=0,1,2
\]

\[
\phi = \frac{\bar{Y}}{\bar{X}}, \quad C^2(\hat{Y}) = \frac{V_{200}}{\bar{Y}^2}, \quad \beta_1g(x) = \frac{V_{011}^2}{V_{020}^3}, \quad \beta_2g(x) = \frac{V_{002}}{V_{020}^2}, \quad \gamma_{12g}(y,x) = \frac{V_{110}}{V_{020}\sqrt{V_{200}}}
\]

\[
\rho_g = \frac{V_{110}}{\sqrt{V_{020} V_{200}}},
\]

\[
\xi_g^2(y,x) = \frac{\{\hat{\gamma}_{12g}(y,x) - \rho_g \sqrt{\beta_1g(x)}\}^2}{\beta_2g(x) - \beta_1g(x)}.
\]

The proposed estimator \( \hat{\tilde{Y}}_g \) has bias are

\[
B(\hat{\tilde{Y}}_g) = (\lambda_1 - 1)(\bar{Y} - \bar{X})
\]

and MSE

\[
M(\hat{\tilde{Y}}_g) = \lambda_1^2 V_{200} + \lambda_2^2 V_{020} + \lambda_3^2 V_{002} - 2\lambda_1 \lambda_2 V_{110} - 2\lambda_1 \lambda_3 V_{101} + 2\lambda_2 \lambda_3 V_{011} + (\lambda_1 - 1)^2(\bar{Y} - \bar{X})^2
\]
The estimator $\hat{Y}_g$ has minimum MSE, if values of $\lambda_i$, $i = 1,2,3$ are taken as

$$\lambda_{01} = \frac{(\bar{Y} - \bar{X})^2}{(\bar{Y} - \bar{X})^2 + V_{200} (1 - \rho_g^2 - \xi_g^2(y,x))},$$

(2.4)

$$\lambda_{02} = -\lambda_{01}B_{1g}$$

(2.5)

$$\lambda_{03} = -\lambda_{01}B_{2g}$$

(2.6)

where

$$B_{1g} = \frac{V_{002}V_{110} - V_{101}V_{011}}{V_{020}V_{002} - V_{011}^2}$$

(2.7)

$$B_{2g} = \frac{V_{020}V_{101} - V_{110}V_{011}}{V_{020}V_{002} - V_{011}^2}$$

(2.8)

Therefore, minimum MSE of $\hat{Y}_g$ is given by

$$M_0(\hat{Y}_g) = \frac{V(\hat{Y})(1 - \rho_g^2 - \xi_g^2(y,x))(1 - \phi)^2}{(1 - \phi)^2 + V(\hat{Y})(1 - \rho_g^2 - \xi_g^2(y,x))}$$

(2.9)

We note that for this situation (2.2) reduces to

$$B_0(\hat{Y}_g) = \frac{V(\hat{Y})(1 - \rho_g^2 - \xi_g^2(y,x))(\bar{Y} - \bar{X})}{(\bar{Y} - \bar{X})^2 + V(\hat{Y})(1 - \rho_g^2 - \xi_g^2(y,x))}$$

(2.10)

which is of order $n^{-1}$.

**III. EFFICIENCY COMPARISIONS**

For comparing efficiency of proposed estimator, we consider minimum MSE attained by the above estimators as under

$$V(\hat{Y}_\lambda) = V(\hat{Y}) (1 - \rho_g^2)$$

(3.1)

$$M_0(e_1) = \frac{V(\hat{Y})(1 - \rho_g^2)}{1 + C^2(\hat{Y})(1 - \rho_g^2)}$$

(3.2)
\[ M_0(e_2) = \frac{V(\hat{Y}) (1-\rho_g^2)(1-\phi)^2}{(1-\phi)^2 + C^2(\hat{Y})(1-\rho_g^2)} \] 

(3.3)

\[ M_0(e_3) = V(\hat{Y})(1-\rho_g^2 - \xi_g^2(y,x)) \] 

(3.4)

Again, one may consider Searls (1964) type estimator

\[ e_4 = \lambda_1 \hat{Y} + \lambda_2 (\hat{X} - \hat{\hat{X}}) + \lambda_3 \{ \mu_2(x) - \mu_2(x) \} \] 

(3.5)

with minimum MSE

\[ M_0(e_4) = \frac{V(\hat{Y})(1-\rho_g^2 - \xi_g^2(y,x))}{1+C^2(\hat{Y})(1-\rho_g^2 - \xi_g^2(y,x))} \] 

(3.6)

Now, it can be seen that

\[ V(\hat{Y}_g) - M_0(\hat{Y}_g) = V_{200} \left[ \frac{(1-\phi)^2 \xi_g^2(y,x) + \bar{Y}^2 M_0(e_3)}{(1-\phi)^2 + \bar{Y}^2 M_0(e_3)} \right] > 0 \] 

(3.7)

\[ M_0(e_2) - M_0(\hat{Y}_g) = \frac{V_{200} (1-\phi)^4 \eta_g^2(y,x)}{[ (1-\phi)^2 + \bar{Y}^2 V(\hat{Y}_{rg}) ] [ (1-\phi)^2 + \bar{Y}^2 M_0(e_3) ]} > 0 \] 

(3.8)

\[ M_0(e_3) - M_0(\hat{Y}_g) = \frac{M_0^2(e_3)}{(1-\phi)^2 + M_0(e_3)} > 0 \] 

(3.9)

Again, if condition (1.5) is satisfied

\[ M_0(e_1) - M_0(\hat{Y}_g) = \frac{G^2 \bar{Y}^4 (1-\phi)^2 V_{200} + \phi(2-\phi) \bar{Y}^2 M(\hat{Y}_{rg}) M_0(e_3)}{[ (1-\phi)^2 + M_0(e_3) ] [ \bar{Y}^2 + M(\hat{Y}_{rg}) ]} > 0 \] 

(3.10)

\[ M_0(e_4) - M_0(\hat{Y}_g) = \frac{M_0^2(e_3) \phi(\phi-2)}{[ \bar{Y}^2 + M_0(e_3) ] [ (1-\phi)^2 + M_0(e_3) ]} > 0 \] 

(3.11)

Again comparing (2.7), (3.1) and (3.4), we have

\[ M_0(\hat{Y}_g) < M_0(e_3) < M(\hat{Y}_g) \] 

(3.12)
while comparison of \((2.7)\) with \((3.3)\) reveals

\[
M_0(\hat{Y}_g) < M_0(e_2)
\]  

(3.13)

If the condition \((1.5)\) holds, we have from \((2.7), (3.2)\) and \((3.6)\) that

\[
M_0(\hat{Y}_g) < M_0(e_4) < M_0(e_1)
\]  

(3.14)

Combining \((3.12), (3.13)\) and \((3.14)\); it is concluded that the proposed estimator is better than all existing estimators under condition \((1.5)\).

### IV. SAMPLE ESTIMATES OF \(\lambda_{0i}, i=1,2,3\)

The value of \(\lambda_{0i} \) in \((2.4)\) may also be expressed as

\[
\lambda_{0i} = 1 - \frac{M_0(e_3)}{E(e_3(\text{opt}) - \bar{X})^2}
\]  

(4.1)

where,

\[
e_{3(\text{opt})}^{\text{opt}} = \hat{Y} + B_{1g} (\bar{X} - \hat{X}) + B_{2g} \{\hat{\mu}_2(x) - \hat{\mu}'_2(x)\}
\]  

(4.2)

Substituting optimum values of \(\lambda_i\) in \((2.1)\), we find that \(\hat{Y}_g\) reduces to

\[
\hat{Y}_{g(\text{opt})} = \lambda_{0i} e_{3(\text{opt})}^{\text{opt}} + (1 - \lambda_{0i}) \bar{X}
\]  

(4.3)

Let \(\hat{B}_{1g}, \hat{B}_{2g}\) and \(\hat{M}_0(e_3)\) be sample estimates of \(B_{1g}, B_{2g}\) and \(M_0(e_3)\) which may easily be obtained by replacing \(V_{abc}\) by its unbiased estimator \(\hat{V}_{abc}\) in \((2.7), (2.8)\) and \((3.4)\) respectively. Again, let

\[
e_3^{\ast} = \hat{Y} + B_{1g} (\bar{X} - \hat{X}) + B_{2g} \{\hat{\mu}_2(x) - \hat{\mu}'_2(x)\}
\]  

(4.4)

The estimate of \(\lambda_{0i}\) is given by

\[
\hat{\lambda}_{0i} = 1 - \frac{\hat{M}_0(e_1)}{(e_3^{\ast} - \bar{X})^2}
\]  

(4.5)

Thus using \((4.3)\), the estimator of \(\bar{Y}\) is given by

\[
\hat{Y}_{g^{\ast}} = \hat{\lambda}_{0i} e_3^{\ast} + (1 - \hat{\lambda}_{0i}) \bar{X}
\]  

(4.6)

Writing \(\hat{\lambda}_{0i} = \lambda_{0i} + \lambda_{01}^{\ast}\)

(4.7)
with \( \lambda_{01} \leq O(n^{-3/2}) \).

we have
\[
\hat{Y}_g = \hat{Y}_{g(\text{opt})} + \lambda_{01}(e^* - \bar{X})
\]

(4.8)

The MSE of \( \hat{Y}_g \) is
\[
M(\hat{Y}_g) = M(\hat{Y}_{g(\text{opt})}) + O(n^{-2})
\]

(4.9)

Therefore, \( \hat{Y}_g \) is equally efficient as \( \hat{Y}_{g(\text{opt})} \) up to first order of approximation.

The estimators \( \hat{Y}_e, e_1, e_2 \) in SRSWOR were discussed by Hansen, Hurwitz and Madow (1946), Bedi and Hajela (1984), Dubey and Singh (2001) respectively. The quadratic type estimator \( e_3 \) under SRSWOR was proposed by Dubey and Sharma (2003) where they have shown their estimator to be more efficient than Srivastava and Jhajj (1981) classes of estimators which utilizes population mean and variance of auxiliary variable \( x \).

Let \( \theta = n^{-1}(1 - f) \), \( f = n/N \), the dispersion terms are as follows:

\[
\sum_{i=U} (Y_i - \bar{Y})^a (X_i - \bar{X})^b (X_i^2 - \mu_2(x))^c.
\]

(5a.1)

For simplicity of presentation, we write
\[
S_y^2 = S_{200}(y, x, x^2), \quad S_x^2 = S_{020}(y, x, x^2)
\]

\[
N\mu_{ab0}(y, x, x^2) = (N-1)S_{ab0}(y, x, x^2),
\]

\[
\mu_{200}(y, x, x^2) = \mu_2(y), \quad \mu_{0a0}(y, x, x^2) = \mu_a(x), a=2,3,4
\]

\[
\beta_1(x) = \frac{\mu_3(x)}{\mu_2(x)}, \quad \beta_2(x) = \frac{\mu_4(x)}{\mu_2(x)}, \quad \gamma_{12}(y, x) = \frac{\mu_{12}(y, x)}{\mu_2(x) \sqrt{\mu_2(y)}}.
\]
\[ \xi^2(y, x) = \frac{\left[ y_{12}^2(1, x) - \rho \sqrt{\beta_1(x)} \right]^2}{\beta_2(x) - \beta_1(x)}, \rho = \text{correlation coefficient between } y \text{ and } x \]

The minimum MSE of regression estimator \( \hat{Y}_\lambda \),

\( e_1, e_2, e_3, e_4 \) and proposed estimator \( \hat{Y}_g \) in SRSWOR

are correspondingly given by

\[ V(\bar{y}_{lr}) = \theta S_\gamma^2 (1 - \rho^2) \]
\[ M_0(e_1(srs)) = \frac{\theta S_\gamma^2 (1 - \rho^2)}{1 + \theta C_\gamma^2 (1 - \rho^2)} \]
\[ M_0(e_2(srs)) = \frac{\theta S_\gamma^2 (1 - \rho^2)(1 - \phi^2)}{(1 - \phi)^2 + \theta C_\gamma^2 (1 - \rho^2)} \]
\[ M_0(e_3(srs)) = \theta S_\gamma^2 (1 - \rho^2 - \xi^2(y, x)) \]
\[ M_0(e_4(srs)) = \frac{\theta S_\gamma^2 (1 - \rho^2 - \xi^2(y, x))}{1 + \theta C_\gamma^2 (1 - \rho^2 - \xi^2(y, x))} \]
\[ M_0(\hat{Y}_g(srs)) = \frac{\theta S_\gamma^2 (1 - \phi^2)(1 - \rho^2 - \xi^2(y, x))}{(1 - \phi)^2 + \theta S_\gamma^2 (1 - \rho^2 - \xi^2(y, x))} \]

An unbiased estimate of \( S_{abc}(y, x, x^2) \)

\[ s_{abc}(y, x, x^2) = (n - 1)^{-1} \sum_{i \in S} (y_i - \overline{y})^a (x_i - \overline{x})^b (x_i^2 - m_2(x))^c \]

Therefore sample estimates of \( B_1 \) and \( B_2 \) are respectively given by

\[ b_1 = \frac{s_{yx}s_u^2 - s_{yu}s_{xu}}{s_x^2s_u^2 - s_{xu}^2} \]
\[ b_2 = \frac{s_{xu}s_x^2 - s_{yx}s_{yu}}{s_x^2s_u^2 - s_{yu}^2} \]
Again, following Cochran (1977, pp. 195) where estimate of $S_y^2 (1 - \rho_y^2)$ is given as

$$V_{lr} = (n - 2)^{-1} \sum_{i \in s} \{y_i - \bar{y} - b(x_i - \bar{x})\}^2,$$

we consider estimate of $S_y^2 (1 - \rho^2 - G^2)$ as

$$\hat{V}(y, x) = (n - 2)^{-1} \sum_{i \in s} \{y_i - \bar{y} - b_1(x_i - \bar{x}) - b_2(x_i^2 - m_2(x))\}^2$$

Therefore, estimating $\lambda_{01R}$ by

$$\hat{\lambda}_{01R} = 1 - \frac{\hat{V}(y, x)}{(\bar{y}_k - \bar{X})^2}$$  \hspace{1cm} (5a.8)

where $\bar{y}_k = \bar{y} + b_1(\bar{X} - \bar{x}) + b_2(\mu_2(x) - m_2(x))$  \hspace{1cm} (5a.9)

Thus optimum estimator of $\bar{Y}$ in SRSWOR as

$$\bar{y}_{opt} = \hat{\lambda}_{01R} \bar{y}_k + (1 - \hat{\lambda}_{01R}) \bar{X}$$  \hspace{1cm} (5a.10)

which has MSE (5a.7) up to first order of approximation.

5a.2. Numerical Illustration

Let us consider the population from Tripathi et al. (2002) which consists summarized data of 142 cities of India with population (number of persons) 1,00,000 and above. Let $x$ be census population in the year 1961 (in 00's) and $y$ be census population in the year 1971 (in 00's). Values of the required population parameters are given below:

$$\bar{Y} = 4015.2183, \quad \bar{X} = 2900.3872, \quad S_y = 8564.546, \quad S_x = 6417.636,$$

$$\rho = 0.9948, \quad \beta_2(x) = 48.157, \quad \beta_1(x) = 35.2524, \quad \gamma_1 = 6.1772.$$  

Table-1 shows relative efficiencies of various estimators with respect to conventional estimator $\bar{Y}$, defined by

$$\left[ \frac{V(\bar{Y})}{M(.)} \right] \times 100.$$
Table 1

| Estimator     | Relative efficiency n=20 | Relative efficiency n=30 | Relative efficiency n=50 |
|---------------|--------------------------|--------------------------|--------------------------|
| $\bar{y}_{lr}$| 9640.45                  | 9640.45                  | 9640.45                  |
| $e_{1(srs)}$  | 9659.61                  | 9652.41                  | 9646.35                  |
| $e_{2(srs)}$  | 9893.98                  | 9795.62                  | 9715.41                  |
| $e_{3(srs)}$  | 12068.78                 | 12048.32                 | 12032.56                 |
| $e_{4(srs)}$  | 12087.94                 | 12060.15                 | 12038.36                 |
| $\hat{Y}_g(srs)$ | 21043.73                | 20886.32                 | 20767.73                 |

Table-1 shows that usual regression estimator $\bar{y}_{lr}$ and Searls (1964) type estimator $e_1$ are almost equally efficient. Similarly the quadratic type estimator $e_3$ and its Searls type estimator $e_4$ are almost equally efficient. The modified estimator $e_2$ has its superiority over $\bar{y}_{lr}$ and $e_1$. The suggested estimator $\hat{Y}_g$ is most efficient than all the existing estimators.

5b.1. Probability Proportional to Size Sampling with Replacement (PPSWR)

Let $p_i$ be the probability of selecting $i$th unit from the population then in PPSWR sampling, the values of $\hat{Y}$, $\hat{X}$ and $\hat{\mu}^2(x)$ are correspondingly given by

$$\hat{Y} = \sum_{j=1}^{n} \frac{y_j}{N_p j}, \quad \hat{X} = \sum_{j=1}^{n} \frac{x_j}{N_p j} \quad \text{and} \quad m'(x)_{pps} = \sum_{j=1}^{n} \frac{x_j^2}{N_p j}$$

Thus regression estimator $\hat{Y}_\lambda$, modified estimators $e_1, e_2, e_3, e_4$ and proposed estimator $\hat{Y}_g$ in PPSWR sampling are respectively as follows:

$$\bar{y}_{lr}(pps) = \bar{y}_{pps} + \beta_{pps}(\bar{X} - \bar{x}_{pps}) \quad (5b.1)$$

$$e_1(pps) = \lambda_{0p} \bar{y}_{lr}(pps) \quad (5b.2)$$

$$e_2(pps) = \lambda_{1p} \bar{y}_{pps} + \lambda_{2p} (\bar{X} - \bar{x}_{pps}) + (1 - \lambda_{1p}) \bar{X} \quad (5b.3)$$

$$e_3(pps) = \bar{y}_{pps} + \lambda_{3p} (\bar{X} - \bar{x}_{pps}) + \lambda_{4p} (\mu'_2(x) - m'_2(x)_{pps}) \quad (5b.4)$$

This Work is under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
\( e_4(\text{pps}) = \lambda_5 p e_3(\text{pps}) \tag{5b.5} \)

\( \bar{y}_g(\text{pps}) = \lambda_6 p \bar{y}_{\text{pps}} + \lambda_7 p (\bar{X} - \bar{x}_{\text{pps}}) + \lambda_8 p (\mu_2(x) - m_2(x))_{\text{pps}} + (1 - \lambda_6 p) \bar{X} \tag{5b.6} \)

where \( \text{Cov}(\bar{y}_{\text{pps}}, \bar{x}_{\text{pps}})/\sqrt{\bar{x}_{\text{pps}}} \) and \( \lambda_{ip} ; i=1,2,\ldots,8 \) are constants. The estimator \( \bar{y}_{\text{lr}}(\text{pps}) \) was suggested by Tripathi (1969), \( e_1(\text{pps}) \) and \( e_2(\text{pps}) \) were discussed by Dubey (2003) while \( e_3(\text{pps}) \) has been illustrated by Dubey (2006). In PPSWR sampling, dispersion term

\[
V_{abc} = n^{-1} \sigma_{abc}(\text{pps})(y, x, x^2), \quad (a, b, c)=0, 1, 2 \tag{5a.1}
\]

where

\[
\sigma_{abc}(\text{pps})(y, x, x^2) = \sum_{i \in U} \left( \frac{Y_i - \bar{Y}}{Np_i} \right)^a \left( \frac{X_i - \bar{X}}{Np_i} \right)^b \left( \frac{Y_i^2 - \mu_2(x)}{Np_i} \right)^c.
\]

For simplicity of presentation, we write

\[
\begin{align*}
\sigma_{yp}^2 &= \sigma_{200}(\text{pps})(y, x, x^2), \\
\sigma_{xp}^2 &= \sigma_{020}(\text{pps})(y, x, x^2), \\
\sigma_{up}^2 &= \sigma_{002}(\text{pps})(y, x, x^2), \\
\sigma_{yp} &= \sigma_{110}(\text{pps})(y, x, x^2), \\
\sigma_{xp} &= \sigma_{011}(\text{pps})(y, x, x^2).
\end{align*}
\]

Define

\[
\begin{align*}
\rho_p &= \frac{\sigma_{yp}}{\sigma_{yp}\sigma_{xp}}, & \beta_{1p}(x) &= \frac{\sigma_{xp}}{\sigma_{xp}^3}, & \beta_{2p}(x) &= \frac{\sigma_{up}}{\sigma_{xp}^4}, \\
\gamma_{12g}(y, x) &= \frac{\sigma_{up}}{\sigma_{xp}^2\sigma_{yp}}, & \varepsilon_p^2 &= \frac{(\gamma_{12p}(y, x) - \rho_p \sqrt{\beta_{1p}(x)})^2}{\beta_{2p}(x) - \beta_{1p}(x)}
\end{align*}
\]

Therefore, minimum MSE of estimators \( \hat{y}_{\text{rg}}, e_1, e_2, e_3, e_4 \) and proposed estimator are respectively as follows

\[
\begin{align*}
V(\hat{y}_{\text{lr}}(\text{pps})) &= \frac{\sigma_{yp}(1 - \rho_p^2)}{n} \tag{5b.7} \\
M_0(e_1(\text{pps})) &= \frac{\sigma_{yp}(1 - \rho_p^2)}{n + C_{yp}(1 - \rho_p^2)} \tag{5b.8} \\
M_0(e_2(\text{pps})) &= \frac{\sigma_{yp}(1 - \rho_p^2)(1 - \phi)^2}{n(1 - \phi)^2 + C_{yp}(1 - \rho_p^2)} \tag{5b.9}
\end{align*}
\]
\[ M_0(e_3(pps)) = \frac{\sigma_{yp}^2(1-\rho_p^2-\xi_p^2(y,x))}{n} \quad (5b.10) \]

\[ M_0(e_4(pps)) = \frac{\sigma_{yp}^2(1-\rho_p^2-\xi_p^2(y,x))}{n + C_{yp}^2(1-\rho_p^2-\xi_p^2(y,x))} \quad (5b.11) \]

\[ M_0(\bar{y}_g(pps)) = \frac{\sigma_{yp}^2(1-\phi)^2(1-\rho_p^2-\xi_p^2(y,x))}{(1-\phi)^2 + C_{yp}^2(1-\rho_p^2-\xi_p^2(y,x))}, \quad (5b.12) \]

Considering

\[ s_{abc} = (n-1)^{-1} \sum_{i \in S} (\frac{y_i}{Np_i} - \bar{y}_{pps})^a (\frac{x_i}{Np_i} - \bar{x}_{pps})^b (\frac{m_i}{Np_i} - m^i_{2(x)})^c \]

as unbiased estimate of \( \sigma_{abc} \), the value of \( \tilde{\lambda}_01 \) in ppswr sampling may easily be found.

5b.2. Numerical Example

Consider the data from Gupta and Rao (1997), which relates to the population of 16 districts of West Bengal. Let \( z, x, \) and \( y \) be population of the districts in 1951, 1961 and 1971 respectively. Let \( z \) be the variable which measures size of the units. For this data we have

\[ \bar{Y} = 2777.51, \quad \bar{X} = 2182.88, \quad \sigma_{yp}^2 = 1.554998 \times 10^{-11}, \quad \sigma_{xp}^2 = 39744.23, \]

\[ \sigma_{up}^2 = 1.6616 \times 10^{-13}, \quad \sigma_{yx} = 71153288.10, \quad \sigma_{yp} = 5.6460 \times 10^{11}, \quad \sigma_{xp} = 179054687 \]

\[ \rho_p = 0.90509, \quad \gamma_{12p}(y, x) = 36.0248 ; \quad \beta_{lp}(x) = 510.680 ; \quad \beta_{2p}(x) = 10519.09. \]

Efficiency of proposed estimator with respect to \( \bar{y}_{pps} \), defined by \[ \{ V(\bar{y}_{pps}) / M(.) \} \times 100 \] is given in Table 2.

| Estimator | R.Eff. |
|-----------|--------|
| \( \bar{y}_{lr}(pps) \) | n=5  | n=8  | n=10 |
| \( e_1(pps) \) | 553.06 | 553.06 | 553.06 |
| \( e_2(pps) \) | 553.09 | 553.08 | 553.07 |
| \( e_3(pps) \) | 561.86 | 558.56 | 557.46 |
| \( e_4(pps) \) | 683.64 | 683.64 | 683.64 |
| \( \bar{y}_g(pps) \) | 683.65 | 683.64 | 683.63 |
| \( \bar{y}_{lr}(pps) \) | 692.43 | 689.13 | 688.02 |

This Work is under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
Table 2 shows that for both populations, the proposed estimator $\bar{y}_g(pps)$ is considerably more efficient than all the existing estimators.

REFERENCES

[1] Adhvaryu, D. & Gupta, P. C. (1983). On some alternative sampling strategies using auxiliary information. Metrika, 30, 217-221.
[2] Bedi, P. K. & Hajela, D. (1984). An estimator for mean utilizing known coefficient of variation and auxiliary variable. Jour. Stat. Res., 18, 29-30.
[3] Chakrabarty, R. P. (1968). Contribution to the theory of ratio type estimator. Ph.D Thesis, Texas A and M University College Station, U.S.A.
[4] Chaubey, Y. P., Singh M., & Dwivedi, T. D. (1984). A note on an optimality property of the regression estimator. Biometrika Journal, 26(4), 465-467.
[5] Cochran, W. G. (1977). Sampling techniques. (3rd ed.). New York: John Wiley and Sons.
[6] Das, A. K. (1988). Contribution to the theory of sampling strategies based on auxiliary information. Ph.D. thesis submitted to B. C. K.V University College Station, U.S.A.
[7] Das, A. K. & Tripathi, T. P. (1980). Sampling strategies for population mean when the coefficient of variation of an auxiliary character is known. Sankhya, 42(1), 76-86.
[8] Dubey, V. & Singh, S. K. (2001). An improved regression estimator for estimating population mean. Journal Ind. Soc. Agri. Stat., 52(2), 179-183.
[9] Dubey, V. (2006). A modified regression estimator of population mean using auxiliary information in general sampling scheme. Transitions in Statistics, 7(6), 1265-1275.
[10] Dubey, V. (2003). On estimating population mean using auxiliary information. Proceeding the International Conference on Recent Statistical Techniques in Life Testing. Reliability, Sampling Theory and Quality Control, B.H.U., Varanasi, 183-193.
[11] Dubey, V. & Sharma, H. K. (2003). A new class of estimators of population mean using auxiliary information. Proceeding the Mathematical Society, BHU, Varanasi, 19, 47-55.
[12] Gupta, B. K. & Rao, T. J. (1997). Stratified PPS sampling and allocation of sample size, Jour. Ind. Soc. Agr. Statistics, 50(2), 199-208.
[13] Hansen, M.H., Hurwitz, W.N., & Madow, W.G. (1953). Sample survey methods and theory. (2nd ed.). New York: John Wiley and Sons.
[14] Murthy, M.N. (1964). Product method of estimation. Sankhya, 26(A), 294-307.
[15] Rao, T.J. (1991). On certain methods of improving ratio and regression estimators. Commun. Satist. Theory Math., 20(10), 3325-3340.
[16] Reddy, V.N. (1974). On a transformed ratio method of estimation. Sankhya, 36(C), 59-70.
[17] Sarndal, C. E., Swensson, B., & Wretman, J. H. (1992). Model Assisted Survey sampling. New York: Springer-Verlag.
[18] Searls, D.T. (1964). The utilization of known coefficient of variation in the estimation procedure. Jour. Amer. Stat. Assoc., 59, 1225-1226.
[19] Srivastava, S.K. (1967). An estimator using auxiliary information in sample surveys. Cal. Stat. Assoc. Bull., 16, 121-132.
[20] Srivastava, S.K. (1971). A generalized estimator for the population mean of a finite population using multi-auxiliary inform. Jour. Amer. Stat. Assoc., 66, 404-407.
[21] Srivastava, S.K. & Jhajj, H.S. (1981). A class of estimators of the population mean in survey sampling using auxiliary information. Biometrika, 68(1), 341-343.
[22] Tripathi, T. P. (1969). A regression-type estimator in sampling with PPS and with replacement. Aust. Jour. Stat., 11, 140-148.
[23] Tripathi, T. P., Singh, H. P., & Upadhayaya, L. N. (2002). A general method of estimation and its application to the estimation of coefficient of variation. Statistics in Transition, 5(6), 1081-1102.
[24] Vos, J. W. E. (1980). Mixing of direct ratio and product method of estimation. Stat. Neerl., 34, 209-218.