Ratio Estimators in Stratified Random Sampling

C. Kadilar* and H. Cingi

Department of Statistics, Hacettepe University, Beytepe, Ankara, Turkey

Abstract

This paper considers some ratio-type estimators and their properties are studied in stratified random sampling. The results are supported by an application with original data.

Key Words: Ratio-type estimators; Stratified random sampling; Mean square errors.

1. Introduction

A ratio estimate of the population mean $\bar{Y}$ can be made in two ways. One is to make a separate ratio estimate of the total of each stratum and add these totals. An alternative estimate is derived from a single combined ratio. From the sample data, we compute sample mean of the variates in stratified random sampling method as

$$\bar{y}_{st} = \sum_{h=1}^{k} \omega_h \bar{y}_h \, ; \qquad \bar{x}_{st} = \sum_{h=1}^{k} \omega_h \bar{x}_h$$

where k is the number of stratum, $\omega_h = \dfrac{N_h}{N}$ is stratum weight, N is the number of units in population, N_h is the number of units in stratum h, $\bar{y}_h$ is the sample mean of variate of interest in stratum h and $\bar{x}_h$ is the sample mean of auxiliary variate in stratum h. The combined ratio estimate is

$$\bar{y}_{RC} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} = \hat{R}_c \bar{X} \tag{1.1}$$

where $\bar{X}$ is the population mean of auxiliary variate (Cochran, 1977). The variance of combined ratio estimate is

$$V(\bar{y}_{RC}) = \sum_{h=1}^{k} \omega_h^2 \gamma_h \left(S_{yh}^2 - 2R S_{yxh} + R^2 S_{xh}^2 \right) \tag{1.2}$$

where $\gamma_h = \dfrac{1 - (n_h/N_h)}{n_h}$, $R = \dfrac{\bar{Y}}{\bar{X}}$ is the population ratio, n_h is the number of units

---

* Corresponding author: kadilar@hacettepe.edu.tr

in sample stratum h, S^2_{yh} is the population variance of variate of interest in stratum h, S^2_{xh} is the population variance of auxiliary variate in stratum h and S_{yxh} is the population covariance between auxiliary variate and variate of interest in stratum h.

2. Ratio Estimators and their Mean Square Errors

When first degree approximation is used in obtaining the mean square error (MSE) of a ratio estimate, it is known that MSE is equal to the variance, so MSE of combined ratio estimate can be written as follows:

$$\text{MSE}(\bar{y}_{\text{RC}}) = V(\bar{y}_{\text{RC}}) = \sum_{h=1}^{k} \omega_h^2 \gamma_h (S^2_{yh} - 2RS_{yxh} + R^2 S^2_{xh}). \tag{2.1}$$

We can derive the bias of combined ratio estimator as

$$B(\bar{y}_{\text{RC}}) = \frac{1}{\bar{X}} \sum_{h=1}^{k} \omega_h^2 \gamma_h (R S^2_{xh} - S_{yxh}). \tag{2.2}$$

2.1 *Sisodia-Dwivedi estimator*

When the population coefficient of variation C_x is known, SISODIA and DWIVEDI (1981) suggested a modified ratio estimator for $\bar{Y}$ as

$$\bar{y}_{\text{SD}} = \bar{y} \frac{\bar{X} + C_x}{\bar{x} + C_x}. \tag{2.3}$$

In stratified random sampling, we propose this estimator as

$$\bar{y}_{st\text{SD}} = \bar{y}_{st} \frac{\sum\limits_{h=1}^{k} \omega_h (\bar{X}_h + C_{xh})}{\sum\limits_{h=1}^{k} \omega_h (\bar{x}_h + C_{xh})}. \tag{2.4}$$

We can define $x_{\text{SD}} = \sum\limits_{h=1}^{k} \omega_h (\bar{x}_h + C_{xh})$ and $X_{\text{SD}} = \sum\limits_{h=1}^{k} \omega_h (\bar{X}_h + C_{xh})$. Then equation (2.4) will be

$$\bar{y}_{st\text{SD}} = \frac{\bar{y}_{st}}{x_{\text{SD}}} X_{\text{SD}} = \hat{R}_{\text{SD}} X_{\text{SD}}$$

where $\hat{R}_{\text{SD}} = \dfrac{\bar{y}_{st}}{x_{\text{SD}}}$. It should be noted that the difference between combined ratio estimator and Sisodia-Dwivedi estimator is only $\hat{R}_{\text{SD}}$. Thus, MSE and bias of this estimator can be given in the same way like equation (2.1) and equation (2.2),

respectively as

$$\text{MSE}\,(\bar{y}_{stSD}) = \sum_{h=1}^{k} \omega_h^2 \gamma_h (S_{yh}^2 - 2R_{SD}S_{yxh} + R_{SD}^2 S_{xh}^2) \tag{2.5}$$

$$B(\bar{y}_{SD}) = \frac{1}{X_{SD}} \left(\sum_{h=1}^{k} \omega_h^2 \gamma_h (R_{SD}S_{xh}^2 - S_{yxh}) \right) \tag{2.6}$$

where $R_{SD} = \dfrac{\bar{Y}_{st}}{X_{SD}} = \dfrac{\sum_{h=1}^{k} \omega_h \bar{Y}_h}{\sum_{h=1}^{k} \omega_h (\bar{X}_h + C_{xh})}$.

2.2 *Singh-Kakran estimator*

Motivated by SISODIA and DWIVEDI (1981), Singh-Kakran ratio-type estimator for $\bar{Y}$ is developed as

$$\bar{y}_{SK} = \bar{y} \frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)} \tag{2.7}$$

where $\beta_2(x)$ is the population coefficient of kurtosis of auxiliary variate x (UPADHYAYA and SINGH, 1999).

In stratified random sampling, we suggest this estimator as

$$\bar{y}_{stSK} = \bar{y}_{st} \frac{\sum_{h=1}^{k} \omega_h (\bar{X}_h + \beta_{2h}(x))}{\sum_{h=1}^{k} \omega_h (\bar{x}_h + \beta_{2h}(x))} . \tag{2.8}$$

Again, we can define $x_{SK} = \sum_{h=1}^{k} \omega_h (\bar{x}_h + \beta_{2h}(x))$ and $X_{SK} = \sum_{h=1}^{k} \omega_h (\bar{X}_h + \beta_{2h}(x))$. Then equation (2.8) will be

$$\bar{y}_{stSK} = \frac{\bar{y}_{st}}{x_{SK}} X_{SK} = \hat{R}_{SK} X_{SK} .$$

The only difference of MSE and bias of this estimator from MSE and bias of Sisodia-Dwivedi estimator is to replace R_{SD} by $R_{SK} = \dfrac{\bar{Y}_{st}}{\sum_{h=1}^{k} \omega_h (\bar{X}_h + \beta_{2h}(x))}$ in equation (2.5) and X_{SD} by X_{SK} in equation (2.6) as

$$\text{MSE}\,(\bar{y}_{stSK}) = \sum_{h=1}^{k} \omega_h^2 \gamma_h (S_{yh}^2 - 2R_{SK}S_{yxh} + R_{SK}^2 S_{xh}^2) \tag{2.9}$$

$$B(\bar{y}_{SK}) = \frac{1}{X_{SK}} \left(\sum_{h=1}^{k} \omega_h^2 \gamma_h (R_{SK}S_{xh}^2 - S_{yxh}) \right)$$

2.3. *Upadhyaya-Singh estimator*

UPADHYAYA and SINGH (1999) considered both coefficients of variation and kurtosis in their ratio-type estimator as

$$\bar{y}_{US1} = \bar{y} \frac{\bar{X}\beta_2(x) + C_x}{\bar{x}\beta_2(x) + C_x} \,. \tag{2.10}$$

We modify this estimator for stratified random sampling as

$$\bar{y}_{stUS1} = \bar{y}_{st} \frac{\displaystyle\sum_{h=1}^{k} \omega_h(\bar{X}_h\beta_{2h}(x) + C_{xh})}{\displaystyle\sum_{h=1}^{k} \omega_h(\bar{x}_h\beta_{2h}(x) + C_{xh})} \tag{2.11}$$

where we can define

$$x_{US1} = \sum_{h=1}^{k} \omega_h(\bar{x}_h\beta_{2h}(x) + C_{xh}) \quad \text{and} \quad X_{US1} = \sum_{h=1}^{k} \omega_h(\bar{X}_h\beta_{2h}(x) + C_{xh}) \,.$$

By this way, we can rewrite equation (2.11) as

$$\bar{y}_{stUS1} = \frac{\bar{y}_{st}}{x_{US1}} X_{US1} = \hat{R}_{US1}X_{US1}$$

MSE and bias of this estimator, will be

$$\text{MSE}(\bar{y}_{stUS1}) = \sum_{h=1}^{k} \omega_h^2\gamma_h(S_{yh}^2 - 2R_{US1}S_{yxh} + R_{US1}^2 S_{xh}^2) \tag{2.12}$$

$$B(\bar{y}_{stUS1}) = \frac{1}{X_{US1}} \left(\sum_{h=1}^{k} \omega_h^2\gamma_h(R_{US1}S_{xh}^2 - S_{yxh}) \right) \tag{2.13}$$

respectively, where $R_{US1} = \dfrac{\bar{Y}_{st}}{\displaystyle\sum_{h=1}^{k} \omega_h(\bar{X}_h\beta_{2h}(x) + C_{xh})} \,.$

UPADHYAYA and SINGH (1999) proposed another estimate by changing the place of coefficient of kurtosis and coefficient of variation as

$$\bar{y}_{US2} = \bar{y} \frac{\bar{X}C_x + \beta_2(x)}{\bar{x}C_x + \beta_2(x)} \tag{2.14}$$

In stratified random sampling this estimator will definitely be

$$\bar{y}_{stUS2} = \bar{y}_{st} \frac{\displaystyle\sum_{h=1}^{k} \omega_h(\bar{X}_hC_{xh} + \beta_{2h}(x))}{\displaystyle\sum_{h=1}^{k} \omega_h(\bar{x}_hC_{xh} + \beta_{2h}(x))} \,. \tag{2.15}$$

Let $x_{US2} = \sum_{h=1}^{k} \omega_h (\bar{x}_h C_{xh} + \beta_{2h}(x))$ and $X_{US2} = \sum_{h=1}^{k} \omega_h (\bar{X}_h C_{xh} + \beta_{2h}(x))$. In this case,

$$\bar{y}_{stUS2} = \frac{\bar{y}_{st}}{x_{US2}} X_{US2} = \hat{R}_{US2} X_{US2} .$$

The only difference between this estimator and Upadhyaya-Singh first estimator for MSE and bias terms is to replace R_{US1} by $R_{US2} = \dfrac{\bar{Y}_{st}}{\sum_{h=1}^{k} \omega_h (\bar{X}_h C_{xh} + \beta_{2h}(x))}$ in equation (2.12) and X_{US1} by X_{US2} in equation (2.13), respectively.

3. Efficiency Comparisons

We compare the combined ratio estimator with Singh-Kakran estimator. We will have the conditions as follows:

$$\text{MSE}(\bar{y}_{stSK}) < \text{MSE}(\bar{y}_{RC})$$

$$\sum_{h=1}^{k} \omega_h^2 \gamma_h (S_{yh}^2 - 2R_{SK}S_{yxh} + R_{SK}^2 S_{xh}^2) < \sum_{h=1}^{k} \omega_h^2 \gamma_h (S_{yh}^2 - 2RS_{yxh} + R^2 S_{xh}^2) \quad (3.1)$$

$$\sum_{h=1}^{k} \omega_h^2 \gamma_h (-2R_{SK}S_{yxh} + R_{SK}^2 S_{xh}^2) < \sum_{h=1}^{k} \omega_h^2 \gamma_h (-2RS_{yxh} + R^2 S_{xh}^2)$$

Let

$$A = \sum_{h=1}^{k} \omega_h^2 \gamma_h S_{yxh} \quad \text{and} \quad B = \sum_{h=1}^{k} \omega_h^2 \gamma_h S_{xh}^2$$

Thus, (3.1) becomes

$$-2R_{SK}A + R_{SK}^2 B < -2RA + R^2 B$$
$$-2R_{SK}A + 2RA + R_{SK}^2 B - R^2 B < 0$$
$$-2A(R_{SK} - R) + B(R_{SK}^2 - R^2) < 0$$
$$-2A(R_{SK} - R) + B(R_{SK} - R)(R_{SK} + R) < 0$$

where there are two conditions as follows:
   (i) when $(R_{SK} - R)(R_{SK} + R) > 0$,

$$\frac{-2A}{R_{SK} + R} + B < 0$$
$$B < \frac{2A}{R_{SK} + R}$$

(ii) when $(R_{SK} - R)(R_{SK} + R) < 0$,

$$\frac{-2A}{R_{SK} + R} + B > 0$$

$$B > \frac{2A}{R_{SK} + R} \ .$$

When either of these conditions is satisfied, Singh-Kakran estimator will be more efficient than combined ratio estimator. These conditions are also valid while comparing MSE of combined ratio estimator with MSE of Sisodia-Dwivedi estimator and MSE of combined ratio estimator with MSE of Upadhyaya-Singh estimators. However, it is noted that R_{SK} in the condition (i) or (ii) should be replaced by R_{SD}, R_{US1} or R_{US2} according to the estimator, which is compared.

4. Application

We have applied the ratio estimators on the data of apple production amount (as interest of variate) and number of apple trees (as auxiliary variate) in 854 villages of Turkey in 1999 (Source: Institute of Statistics, Republic of Turkey). First, we have stratified the data by regions of Turkey and from each stratum (region); we have randomly selected the samples (villages). By using Neyman allocation (COCHRAN, 1977),

$$n_h = n \frac{N_h S_h}{\sum\limits_{h=1}^{k} N_h S_h} \tag{4.1}$$

we have computed sample size in stratum h. Here we take sample size as $n = 140$ (CINGI, 1994). From the results of n_h, we decide to join two regions so we take six strata (as 1: Marmara, 2: Agean, 3: Mediterranean, 4: Central Anatolia, 5: Black Sea, 6: East and Southeast Anatolia) for this data. Then by using this stratified random sampling, MSE of ratio estimators are computed as described in section 2 and these estimators are compared between each other with respect to their MSE values.

In Table 1, we observe the statistics about the population, strata and sample size. Note that the correlation between the variates is 92%. In Table 2, the values of MSE are given. From Table 2, it can be concluded that the combined ratio estimator has the minimum MSE and therefore it is the best estimator for the data.

In the same way, when we analyze MSE of Singh-Kakran estimator and combined ratio estimator, we see

$$(R_{SK} - R)(R_{SK} + R) < 0$$

Table 1

Data Statistics

$N = 854$	$N_1 = 106$	$N_2 = 106$	$N_3 = 94$	$N_4 = 171$	$N_5 = 204$	$N_6 = 173$	$A = 5207110$
$n = 140$	$n_1 = 9$	$n_2 = 17$	$n_3 = 38$	$n_4 = 67$	$n_5 = 7$	$n_6 = 2$	$B = 57208120$
$\bar{X} = 37600$	$\bar{X}_1 = 24375$	$\bar{X}_2 = 27421$	$\bar{X}_3 = 72409$	$\bar{X}_4 = 74365$	$\bar{X}_5 = 26441$	$\bar{X}_6 = 9844$	$R = 0.07793$
$\bar{Y} = 2930$	$\bar{Y}_1 = 1536$	$\bar{Y}_2 = 2212$	$\bar{Y}_3 = 9384$	$\bar{Y}_4 = 5588$	$\bar{Y}_5 = 967$	$\bar{Y}_6 = 404$	$R_{SK} = 0.07784$
$\beta_x = 312.07$	$\beta_{x1} = 25.71$	$\beta_{x2} = 34.57$	$\beta_{x3} = 26.14$	$\beta_{x4} = 97.60$	$\beta_{x5} = 27.47$	$\beta_{x6} = 28.10$	$R_{SD} = 0.07792$
$\beta_y = 195.84$	$C_{x1} = 2.02$	$C_{x2} = 2.10$	$C_{x3} = 2.22$	$C_{x4} = 3.84$	$C_{x5} = 1.72$	$C_{x6} = 1.91$	$R_{US1} = 0.0014$
$C_x = 3.85$	$C_{y1} = 4.18$	$C_{y2} = 5.22$	$C_{y3} = 3.19$	$C_{y4} = 5.13$	$C_{y5} = 2.47$	$C_{y6} = 2.34$	$R_{US2} = 0.0285$
$C_y = 5.84$	$S_{x1} = 49189$	$S_{x2} = 57461$	$S_{x3} = 160757$	$S_{x4} = 285603$	$S_{x5} = 45403$	$S_{x6} = 18794$	$X_{SK} = 37602$
$S_x = 144794$	$S_{y1} = 6425$	$S_{y2} = 11552$	$S_{y3} = 29907$	$S_{y4} = 28643$	$S_{y5} = 2390$	$S_{y6} = 946$	$X_{SD} = 37642$
$S_y = 17106$	$\varrho_1 = 0.82$	$\varrho_2 = 0.86$	$\varrho_3 = 0.90$	$\varrho_4 = 0.99$	$\varrho_5 = 0.71$	$\varrho_6 = 0.89$	$X_{US1} = 2086644$
$\varrho = 0.92$	$\gamma_1 = 0.102$	$\gamma_2 = 0.049$	$\gamma_3 = 0.016$	$\gamma_4 = 0.009$	$\gamma_5 = 0.138$	$\gamma_6 = 0.006$	$X_{US2} = 102815$
$\varkappa = 0.975$	$\omega_1^2 = 0.015$	$\omega_2^2 = 0.015$	$\omega_3^2 = 0.012$	$\omega_4^2 = 0.04$	$\omega_5^2 = 0.057$	$\omega_6^2 = 0.041$	

since $R = 0.07793$ and $R_{SK} = 0.07784$. Then the condition

$$B = 57208120 > \frac{2A}{R_{SK} + R} = 66856382$$

is not satisfied. Thus, combined ratio estimator is more efficient than Singh-Kak-ran estimator. Because of the same reason, combined ratio estimator is also more efficient than Sisodia-Dwivedi estimator and Upadhyaya-Singh estimators. For example, when we compare first estimator of Upadhyaya-Singh with combined ratio estimator, again we find

$$(R_{US1} - R)(R_{US1} + R) < 0$$

so we investigate the condition

$$B = 57208120 > \frac{2A}{R_{US1} + R} = 131272434$$

and we see that this condition is not satisfied. Therefore, combined ratio estimator is more efficient than ratio-type estimator proposed by Upadhyaya and Singh (1999).

Table 2

MSE Values of Ratio Estimators

Estimators	MSE Values
Combined Ratio	215710.432
Sisodia-Dwivedi	215717.639
Singh-Kakran	215841.602
Upadhyaya-Singh 2	429527.205
Upadhyaya-Singh 1	665347.704

5. Conclusion

We have examined ratio-type estimators in stratified random sampling and obtained their MSE equations. By these equations, MSE of estimators have been compared in theory and by this comparison, the conditions which the estimators have smaller MSE with respect to each other, have been found. These theoretical conditions are also satisfied by the results of an application with original data. In this application, it is concluded that the traditional combined ratio estimator is more efficient than the ratio estimators developed in recent years. This conclusion shows that, in the forthcoming studies, new ratio-type estimators should be proposed not only in simple random sampling but also in stratified random sampling.

References

COCHRAN, W. G., 1977: *Sampling Techniques*, John Wiley and Sons, New-York.
CINGI, H., 1994: *Sampling Theory*, Hacettepe University Press.
SINGH, H. P. and KAKRAN, M. S., 1993: A Modified Ratio Estimator Using Known Coefficient of Kurtosis of an Auxiliary Character. (unpublished)
SISODIA, B. V. S. and DWIVEDI, V. K., 1981: A Modified Ratio Estimator Using Coefficient of Variation of Auxiliary Variable. *Journal of Indian Society Agricultural Statistics* **33**, 13–18.
UPADHYAYA, L. N. and SINGH, H. P., 1999: Use of Transformed Auxiliary Variable in Estimating the Finite Population Mean. *Biometrical Journal* **41**, 5, 627–636.