Available online at http://scik.org
J. Math. Comput. Sci. 7 (2017), No. 2, 364-374
ISSN: 1927-5307

REVIEW OF SOME RATIO ESTIMATORS IN STRATIFIED RANDOM SAMPLING

TOLGA ZAMAN* AND SERHAT YILMAZ

Ondokuz Mayıs University, Department of Statistics, 55139 Atakum SAMSUN TURKEY

Abstract: In this paper, some regression and mean ratio estimators, which suggested by using ancillary variable information in stratified random sampling in literature, and variances belonging to these estimators are investigated. Priority of estimators is explored by making paired comparisons of these estimators. As a result of the comparison, It is determined that combined regression estimator is more efficient than estimators developed by Kadılar-Çıngı. All these results are supported with application of an original real data.

Keywords: stratified random sampling; regression estimates; mean ratio estimates; variance.

2010 AMS Subject Classification: 62P10.

1. Introduction

Let a N sized population be divided into l stratum. Estimating of sum of populations, Y and X, in the stratified random sampling, which n_h units were chosen from N_h units in every stratum using simple random sampling, and using them in rational estimate is called compound rational estimate. In this situation, sample means of variables are as follows:

---

*Corresponding author

E-mail address: tolga.zaman@omu.edu.tr

Received November 24, 2016; Published March 1, 2017

$$\bar{y}_{st} = \sum_{h=1}^{l} W_h \bar{y}_h \tag{1}$$

$$\bar{x}_{st} = \sum_{h=1}^{l} W_h \bar{x}_h \tag{2}$$

Population means of variables are as follows:

$$Y_h = \sum_{i=1}^{N_h} y_{hi} = N_h \bar{Y}_h \tag{3}$$

$$Y = \sum_{h=1}^{l} Y_h \tag{4}$$

In a stratified population, real value of population mean defines as $\bar{Y}_{st}$ and estimate of this from sample defines as $\bar{y}_{st}$ and shown as the following: Real value is

$$\overline{Y} = \bar{Y}_{st} = \frac{\sum_{h=1}^{l} Y_h}{N} = \frac{\sum_{h=1}^{l} \sum_{i=1}^{n_h} y_{hi}}{N} = \frac{\sum_{h=1}^{l} N_h \bar{Y}_h}{N} \tag{5}$$

estimate value is obtained through writing $\bar{y}_h$ in place of $\bar{Y}_h$ as seen below.

$$\bar{y}_{st} = \frac{\sum_{h=1}^{l} N_h \bar{y}_h}{N} \tag{6}$$

If $\rho(x, y) > 0$, population mean belonging to variable y is

$$\hat{\bar{Y}}_{rc} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \overline{X} \tag{7}$$

Ratio of variables is

$$R = \frac{\overline{Y}}{\overline{X}} \tag{8}$$

And estimator of $R\ '$ is

$$\hat{R}_{st} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \tag{9}$$

If equation (9) replaces equation (7),

$$\hat{\bar{Y}}_{rc} = \hat{R}_{st}\bar{X} \tag{10}$$

is obtained [1]. To calculate estimator or random variable $\hat{\bar{Y}}_r$, firstly, let us look into variance of estimate $\hat{R}_{st}$.

$$\hat{R}_{st} - R = \frac{\bar{y}_{st}}{\bar{x}_{st}} - R = \frac{\bar{y}_{st} - R\bar{x}_{st}}{\bar{x}_{st}} \tag{11}$$

Here, if $\bar{x}_{st} = \bar{X}$ is taken, and then both sides squared and their estimate value is calculated, it becomes as follows:

$$E(\hat{R}_{st} - R)^2 \cong \frac{E(\bar{y}_{st} - R\bar{x}_{st})^2}{\bar{X}^2} \tag{12}$$

Now, let us define a new

$$u_{hi} = y_{hi} - Rx_{hi} \tag{13}$$

variable. Its stratified sample mean and population mean is

$$\bar{u}_{st} = \bar{y}_{st} - R\bar{x}_{st}; \quad \bar{U}_{st} = \bar{Y}_{st} - R\bar{X}_{st} = 0 \tag{14}$$

$$Var(\bar{u}_{tb}) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)S_{uh}^2}{n_h} \tag{15}$$

$$S_{uh}^2 = \sum_{i=1}^{N_h} \frac{(u_{hi} - \bar{U}_h)^2}{N_h - 1} \tag{16}$$

$$S_{uh}^2 = \sum_{i=1}^{N_h} \frac{[(y_{hi} - Rx_{hi}) - (\bar{Y}_h - R\bar{X}_h)]^2}{N_h - 1} \tag{17}$$

$$S_{uh}^2 = \sum_{i=1}^{N_h} \frac{[(y_{hi} - \bar{Y}_h) - R(x_{hi} - \bar{X}_h)]^2}{N_h - 1} \tag{18}$$

$$S_{uh}^2 = S_{yh}^2 - 2RCov(x_h, y_h) + R^2 S_{xh}^2 \tag{19}.$$

If equation (18) is put in its position in equation (14), it is calculated as

$$Var(\bar{u}_{tb}) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)(S_{yh}^2 - 2RCov(x_h, y_h) + R^2 S_{xh}^2)}{n_h} \tag{20}$$

It should be pointed out that obtained equation (20) is equal to numerator of equation (12). In this regard, it is

$$Var(\hat{R}_{st}) \cong \frac{1}{\bar{X}^2}\sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2-2RCov(x_h,y_h)+R^2S_{xh}^2)}{n_h} \tag{21}$$

$$Var\left(\hat{\bar{Y}}_{rc}\right) \cong \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2-2RCov(x_h,y_h)+R^2S_{xh}^2)}{n_h} \tag{22}$$

Hence, compound rational estimate is calculated as above mentioned [1].

2. Ratio Estimators

There are various rational estimators which are suggested in literature. Here, we examined the compound rational estimator, which is one of these estimators, Kadılar-Çıngı 1 ratio estimator, which is obtained through adaptation of the estimator, which is suggested by Sisodia-Dwidedi (1981) in simple random sampling, to the stratified random sampling by Kadılar-Çıngı (2003), Kadılar-Çıngı 2 ratio estimator, which is obtained through adaptation of the estimator suggested by Singh and Kahran (1993) to the stratified random sampling by Kadılar-Çıngı (2003), Kadılar-Çıngı 3 and 4 ratio estimators and variances, which are obtained through adaptation of estimators suggested by Upadhyaya-Singh (1999) to the stratified random sampling by Kadılar-Çıngı (2003). Later, while considering variance belonging to compound regression estimate, compound rational estimate compared Kadılar-Çıngı estimates and some results are obtained. Under these obtained circumstances, situations of these estimators compared to each other is discussed. Examined rational estimates and mean squared errors,

Separate rational estimation is given as follows

$$\hat{\bar{Y}}_{rs} = \sum_{h=1}^{l} W_h\,\hat{R}_h\bar{X} \tag{23}$$

Here, $\hat{R}_h = \frac{\bar{y}_h}{\bar{x}_h}$.

$$Var\left(\hat{\bar{Y}}_{rs}\right) = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2-2R_hCov(x_h,y_h)+{R_h}^2S_{xh}^2)}{n_h} \tag{24}$$

Variance belonging to separate rational estimation is calculated as aforementioned [2].

Combined regression estimate is calculated as follows

$$\hat{\bar{Y}}_{cre} = \bar{y}_{st} + b_c(\bar{X} - \bar{X}_{st}) \tag{25}$$

Here, if $\vartheta = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)S_{yxh}}{n_h}$; $\varphi = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)S_{xh}^2}{n_h}$, then it is calculated as $b_c = \frac{\vartheta}{\varphi}$. As for variance of combined regression estimate

$$Var\left(\hat{\bar{Y}}_{cre}\right) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)S_{yh}^2}{n_h}\{1 - \delta_{cer}^2\} \tag{26}$$

Here, if $\xi = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)S_{yh}^2}{n_h}$, then $\delta_{cer} = \frac{\vartheta}{\sqrt{\varphi}*\sqrt{\xi}}$.

Separate regression estimate is

$$\hat{\bar{Y}}_{sre} = \sum_{h=1}^{l} W_h[\bar{y}_h + b_h(\bar{X}_h - \bar{x}_h)] \tag{27}$$

Here $b_h = \frac{s_{xyh}}{s_{xh}^2}$. As for variance of separate regression estimate,

$$Var\left(\hat{\bar{Y}}_{rs}\right) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)(S_{yh}^2 - 2\beta_h Cov(x_h, y_h) + {\beta_h}^2 S_{xh}^2)}{n_h} \tag{28}$$

Here, $\beta_h = \frac{S_{xyh}}{S_{xh}^2}$ [5].

In stratified random sampling, estimators and variances developed by Kadılar-Çıngı (2003)[3], Kadılar-Çıngı 1 estimate, which is developed based on estimate of Sisodia-Dwidedi (1981)[7], is calculated as follows.

$$\hat{\bar{Y}}_{stSD} = \bar{y}_{st} \frac{\sum_{h=1}^{l} W_h(\bar{X}_h + C_{xh})}{\sum_{h=1}^{l} W_h(\bar{x}_h + C_{xh})} \tag{29}$$

As for calculation of the variance in this estimate,

$$Var\left(\hat{\bar{Y}}_{stSD}\right) = \sum_{h=1}^{l} \frac{W_h^2(1-f_h)(S_{yh}^2 - 2R_{SD}Cov(x_h, y_h) + {R_{SD}}^2 S_{xh}^2)}{n_h} \tag{30}$$

Here $R_{SD} = \frac{\sum_{h=1}^{l} W_h \bar{Y}_h}{\sum_{h=1}^{l} W_h(X_h + C_{xh})}$. C_{xh} is defined as coefficient of variance belonging to stratum $h.$ of helper variable and is calculated as $C_{xh} = \frac{S_{xh}}{\bar{X}_h}$.

Kadılar-Çıngı 2 estimate, which is developed based on estimate of Sink-Kakran (1993)[6], is calculated as follows,

$$\hat{\bar{Y}}_{stSK} = \bar{y}_{st}\frac{\sum_{h=1}^{l} W_h\left(\bar{X}_h + \beta_{2h}(x)\right)}{\sum_{h=1}^{l} W_h\left(\bar{x}_h + \beta_{2h}(x)\right)} \tag{31}$$

$\beta_{2h}(x)$ indicates the coefficient of kurtosis belonging to stratum $h.$ of helper variable. As for calculation of the variance in this estimate,

$$Var\left(\hat{\bar{Y}}_{stSK}\right) = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2 - 2R_{SK}Cov(x_h,y_h) + {R_{SK}}^2 S_{xh}^2)}{n_h} \tag{32}$$

$R_{SK} = \frac{\sum_{h=1}^{l} W_h\bar{Y}_h}{\sum_{h=1}^{l} W_h(\bar{X}_h+\beta_{2h}(x))}$

Kadılar-Çıngı 3 estimate, which is developed based on first estimates of Upadhyaya-Singh (1999)[8], is calculated as follows,

$$\hat{\bar{Y}}_{stUS1} = \bar{y}_{st}\frac{\sum_{h=1}^{l} W_h(\bar{X}_h\beta_{2h}(x) + C_{xh})}{\sum_{h=1}^{l} W_h(\bar{x}_h\beta_{2h}(x) + C_{xh})} \tag{33}$$

Variance of this estimate is calculated as follows,

$$Var\left(\hat{\bar{Y}}_{stUS1}\right) = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2 - 2R_{US1}Cov(x_h,y_h) + {R_{US1}}^2 S_{xh}^2)}{n_h} \tag{34}$$

$R_{US1} = \frac{\sum_{h=1}^{l} W_h\bar{Y}_h}{\sum_{h=1}^{l} W_h(\bar{X}_h\beta_{2h}(x)+C_{xh})}$

Kadılar-Çıngı 4 estimate, which is developed based on second estimates of Upadhyaya-Singh (1999), is calculated as follows,

$$\hat{\bar{Y}}_{stUS2} = \bar{y}_{st}\frac{\sum_{h=1}^{l} W_h\left(\bar{X}_h C_{xh} + \beta_{2h}(x)\right)}{\sum_{h=1}^{l} W_h\left(\bar{x}_h C_{xh} + \beta_{2h}(x)\right)} \tag{35}$$

Variance of this estimate is calculated as follows,

$$Var\left(\hat{\bar{Y}}_{stUS2}\right) = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2 - 2R_{US2}Cov(x_h,y_h) + {R_{US2}}^2 S_{xh}^2)}{n_h} \tag{36}$$

$R_{US2} = \frac{\sum_{h=1}^{l} W_h\bar{Y}_h}{\sum_{h=1}^{l} W_h\left(\bar{X}_h C_{xh}+\beta_{2h}(x)\right)}$ [3,4]

3. Efficiency Comparisons

Here, we compared combined regression and combined rational estimate, and Kadılar-Çıngı estimates. As results of comparison, following conditions are obtained.

$$Var\left(\widehat{\overline{Y}}_{cre}\right) < Var\left(\widehat{\overline{Y}}_{rc}\right)$$

$$\sum_{h=1}^{l}\frac{W_h^2(1-f_h)S_{yh}^2}{n_h}\{1-\delta_{cer}^2\} < \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(S_{yh}^2-2RCov(x_h,y_h)+R^2S_{xh}^2)}{n_h}$$

$$-\sum_{h=1}^{l}\frac{W_h^2(1-f_h)S_{yh}^2\delta_{cer}^2}{n_h} < \sum_{h=1}^{l}\frac{W_h^2(1-f_h)(-2RCov(x_h,y_h)+R^2S_{xh}^2)}{n_h} \tag{37}$$

Here, in equation (37), if $\xi = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)S_{yh}^2}{n_h}$; $\vartheta = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)S_{yxh}}{n_h}$; $\varphi = \sum_{h=1}^{l}\frac{W_h^2(1-f_h)S_{xh}^2}{n_h}$ and $\delta_{cer} = \frac{\vartheta}{\sqrt{\varphi}*\sqrt{\xi}}$, then,

$$-\xi*\left(\frac{\vartheta}{\sqrt{\xi}*\sqrt{\varphi}}\right)^2 < -2R\vartheta + R^2\varphi$$

$$-\vartheta^2 + 2\vartheta\varphi R - R^2\varphi^2 < 0$$

$$\vartheta^2 - 2\vartheta\varphi R + R^2\varphi^2 > 0$$

$$(\vartheta - \varphi R)^2 > 0 \tag{38}$$

In the condition of equation (38), combined regression estimate is more efficient than combined rational estimate. Moreover, when R_{SD}, R_{SK}, R_{US} , R_{US} is put in place of R, combined regression estimate is more efficient than above estimators developed by Kadılar-Çıngı. Also, because condition (38) is always bigger than zero, it can be said that combined regression estimate is more efficient than both combined rational estimate and Kadılar-Çıngı estimates.

4. Real Data Application

Crop output differs based on plantation. Because plantations in countries from different continents and output in this field will vary, the world is divided into stratums.

Summary data information between 1960-2014 belonging to 1050 crop outputs (y) and crop plantation (x), which is of countries randomly chosen from taken stratums, is as below. Data set is taken from world data bank [9]. Firstly, each continent taken from stratified data are defined as a stratum and countries are chosen randomly from these continents. $n = 408$ sized sample is chosen with $\%\ 99$ confidence level and $d = 0.05$ margin of error from 1050 sized population. Stratums are taken as Europe, America, Africa and Asia.

Allocation of sample size to stratums is calculated through using Neyman Allocation method given in equation below assuming transportation cost to the units does not differ from stratum to stratum [2],

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{l} N_h S_h}, \qquad h = 1,2,3,4 \tag{39}$$

Table1. Population Information According to Stratums belonging to Crop Output (y) between 1960 - 2014 and Crop Plantation (x) Variables

$N_1 = 265$	$N_2 = 262$	$N_3 = 260$	$N_4 = 263$	$N = 1050$
$\bar{Y}_1$=2623.725	$\bar{Y}_2 = 1921.344$	$\bar{Y}_3 = 1161.951$	$\bar{Y}_4 = 2199.289$	$\bar{Y} = 2232.571$
$\bar{X}_1 = 2858318$	$\bar{X}_2 = 4389540$	$\bar{X}_3 = 784537.5$	$\bar{X}_4 = 5931675$	$\bar{X} = 3496690$
$R_1 = 0.001268$	$R_2 = 0.000438$	$R_3 = 0.001481$	$R_4 = 0.000371$	$R = 0.000638$
$S_{y1} = 1344.653$	$S_{y2} = 877.5063$	$S_{y3} = 318.4846$	$S_{y4} = 1093.45$	$\vartheta = 2550941.342$
$S_{x1} = 3302283$	$S_{x2} = 4847253$	$S_{x3} = 675797.5$	$S_{x4} = 5201781$	$\varphi = 1117.743$
$C_{y1} = 0.371069$	$C_{y2} = 0.456715$	$C_{y3} = 0.274095$	$C_{y4} = 0.497183$	$\xi = 19475285263$
$C_{x1} = 1.155324$	$C_{x2} = 1.104273$	$C_{x3} = 0.861396$	$C_{x1} = 0.87695$	$R_{SD} = 0.000638$
$\beta_{21}(y) = 4.264$	$\beta_{22}(y) = 3.258$	$\beta_{23}(y) = 2.636$	$\beta_{24}(y) = 2.515$	$R_{SK} = 0.000638$
$\beta_{21}(x) = 3.162$	$\beta_{22}(x) = 1.34$	$\beta_{23}(x) = 2.598$	$\beta_{24}(x) = 1.752$	$R_{US1} = 0.000326$
$\rho_1 = 0.633$	$\rho_2 = 0.522$	$\rho_3 = 0.521$	$\rho_4 = 0.629$	$R_{US2} = 0.000635$
$\beta_1 = 0.000258$	β_2 =9.46005E-05	$\beta_3 = 0.000246$	$\beta_4 = 0.000132$	$\delta_{cer} = 0.546$

$n = 408$ sized sample is chosen from 1050 sized population. Allocation of chosen sample size to stratums is made according to Neyman allocation. Obtained data belonging to sample is shown on Table 2.

Table2. Sample Information According to Stratums belonging to Crop Output (y) between 1960 - 2014 and Crop Plantation (x) Variables

$n_1 = 151$	$n_2 = 98$	$n_3 = 35$	$n_4 = 122$
$\bar{y}_1 = 3663.241$	$\bar{y}_2 = 1945.909$	$\bar{y}_3 = 1163.802$	$\bar{y}_4 = 2222.121$
$\bar{x}_1 = 2670549$	$\bar{x}_2 = 5425402$	$\bar{x}_3 = 852112.2$	$\bar{x}_4 = 6317300$
$s_{y1} = 1339.715$	$s_{y1} = 845.8587$	$s_{y3} = 362.8792$	$s_{y4} = 1058.419$
$s_{x1} = 3141607$	$s_{x2} = 5032321$	$s_{x3} = 664442$	$s_{x4} = 5184948$

If we summarize ratio estimators examined in the study and variances belonging to these in Table 3,

Table3. Means of Ratio Estimations and Numeric Values of Their Variances

Estimators	Mean Values	Variances
$\hat{\bar{Y}}_{rc}$	2063.417	5752.496
$\hat{\bar{Y}}_{rs}$	2170.321	5335.044
$\hat{\bar{Y}}_{cre}$	2214.883	783.6117
$\hat{\bar{Y}}_{sre}$	2229.356	730.287
$\hat{\bar{Y}}_{stSD}$	2063.417	5799.546
$\hat{\bar{Y}}_{stSK}$	2063.417	5799.541
$\hat{\bar{Y}}_{stUS1}$	2127.856	1512.498
$\hat{\bar{Y}}_{stUS2}$	2061.348	5740.400

When Table 3 is examined, variance values of regression estimators are seen smaller than variances of examined mean ratio estimators.

Conclusions

We examined some regression and ratio estimators, which are suggested in stratified random sampling, and variances of these estimators. Theoretical comparisons of combined regression and combined rational estimate, and estimators developed by Kadılar-Çıngı and condition in equation (38) is obtained. The condition in (38) is occurred as a result of theoretical comparison of combined rational estimate and combined regression estimate. In this equation, when $R_{SD}, R_{SK}, R_{US1}, R_{US2}$ is put in place of R, estimations, which are developed by Kadılar-Çıngı, and theoretical comparisons of combined regression estimate can be seen separately. This equation is always bigger than zero. Therefore, it is seen that combined regression estimate is more efficient than mean rational estimates, which are developed by Kadılar-Çıngı. All these results are supported by application of an original data set.

Conflict of Interests

The authors declare that there is no conflict of interests.

REFERENCES

[1] Cingi, H. Sampling Theory, Hacettepe University Press, Ankara, (2009).

[2] Cochran, W. G. Sampling techniques. 1977. New York: John Wiley and Sons. (1977).

[3] Kadilar, C., & Cingi, H. Ratio estimators in stratified random sampling. Biometrical journal. 45(2) (2003), 218-225.

[4] Kadilar, C., & Cingi, H. A new ratio estimator in stratified random sampling. Communications in Statistics—Theory and Methods, 34(3) (2005), 597-602.

[5] Koyuncu, N.,. Tabakalı Rasgele Örneklemede Yardımcı Değişken Bilgisi Kullanılarak Kitle Ortalaması ve Varyansının Tahmin Edilmesi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi, Ankara, (2007).

[6] Singh, H. P., & Kakran, M. S. A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. unpublished paper (1993).

[7] Sisodia, B. V. S., & Dwivedi, V. K. Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics (1981).

[8] Upadhyaya, L. N., & Singh, H. P. Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal. 41(5) (1999), 627-636.

[9] URL-1. http://data.worldbank.org/