J. Ind. Soc. Agril. Statist.
58(2), 2004 : 223-230

An Improved Estimator of Population Mean Using Power Transformation

Housila, P. Singh, Rajesh Tailor, Ritesh Tailor and M.S. Kakran[1]
Vikram University, Ujjain - 456010, M.P
(Received: May, 2003)

SUMMARY

This article presents a modified ratio estimator using prior value of coefficient of kurtosis of an auxiliary character x, with the intention to improve the efficiency of ratio estimator. The first order large sample approximations to the bias and the mean square error of the proposed estimator are obtained and compared with sample mean estimator and usual ratio estimator. A generalized version of the suggested estimator is also presented. Numerical illustrations are listed to compare the performance of different estimators.

Key words : Coefficient of kurtosis, Ratio estimator, Auxiliary character, Bias, Mean square error.

1. Introduction

The use of prior value of coefficient of kurtosis in estimating the population variance of study character y was first made by Singh *et al.* (1973). Later, used by Sen (1978), Upadhyaya and Singh (1984), Singh (1984) and Searls and Intrapanich (1990) in the estimation of population mean of study character. The knowledge of coefficient of kurtosis of the character under study is seldom available. However the coefficient of kurtosis of an auxiliary character can easily be obtained. In this paper we have made the use of known value of coefficient of kurtosis of an auxiliary character in proposing modified ratio estimator for the population.

Let y and x be the real valued functions defined on a finite population $U = \{U_1, U_2, ..., U_N\}$ and $\overline{Y}$ and $\overline{X}$ be the population means of the study character y and auxiliary character x respectively. Consider a simple random sample of size n drawn without replacement from population U. Quite often we have surveys in which some auxiliary character x is relatively less expensive

[1] College of Agriculture, Gwalior, M.P.

(with regard to time and money) to observe than the study character y. In order to have a survey estimate of the population mean $\overline{Y}$ of the study character y, assuming the knowledge of the population mean $\overline{X}$ of the auxiliary character x, we mention below a well known ratio estimator

$$\hat{\overline{Y}}_R = \overline{y}\left(\frac{\overline{X}}{\overline{x}}\right) \tag{1.1}$$

where $\overline{y}$ and $\overline{x}$ are the unweighted sample mean of the characters y and x respectively.

The bias and mean square error (MSE) of $\hat{\overline{Y}}_R$ to the first order large sample approximations are given by

$$B(\hat{\overline{Y}}_R) = \theta\overline{Y}C_x^2(1-K) \tag{1.2}$$

and

$$MSE(\hat{\overline{Y}}_R) = \theta\overline{Y}^2\left[C_y^2 + C_x^2(1-2K)\right] \tag{1.3}$$

where $\theta = \frac{(N-n)}{(Nn)}$, $K = \rho\left(\frac{C_y}{C_x}\right)$, C_y and C_x are coefficient of variation of y and x respectively and ρ is correlation coefficient between y and x.

Assume that information on all the units of auxiliary variable x is available and thus the value of the coefficient of kurtosis $\beta_2(x)$ is known. Now using the transformation $u_i = x_i + \beta_2(x), (i = 1, 2, \ldots, N)$ we suggest the following modified ratio estimator for the population mean $\overline{Y}$ as

$$\hat{\overline{Y}}_M = \overline{y}\left(\frac{\overline{X}+\beta_2(x)}{\overline{x}+\beta_2(x)}\right) \tag{1.4}$$

To obtain the bias and MSE of $\hat{\overline{Y}}_M$ we put $\overline{y} = \overline{Y}(1+e_0)$, $\overline{x} = \overline{X}(1+e_1)$ so that $E(e_0) = E(e_1) = 0$ and $V(e_0) = \theta C_y^2$, $V(e_1) = \theta.C_x^2$ and $Cov(e_0, e_1) = \theta\rho C_y C_x$. We can reasonably assume that the sample size n is large to make $|e_0|$ and $|e_1| < 1$. Further to validate first degree of approximation we are going to obtain, we assume that the sample size is large enough to get $|e_0|$ and $|e_1|$ as small so that the terms involving e_0 and/or e_1 in a degree greater than two will be negligible, an assumption which is generally not unrealistic. Now, we have

$$\hat{\overline{Y}}_M = \overline{Y}(1+e_0)(1+\lambda e_1)^{-1}$$

where $\lambda = \dfrac{\overline{X}}{\overline{X} + \beta_2(x)}$

Suppose $|\lambda e_1| < 1$ so that $(1 + \lambda e_1)^{-1}$ is expandable. Therefore to the first degree of approximation, the bias and MSE of $\hat{\overline{Y}}_M$ are respectively given by

$$B(\hat{\overline{Y}}_M) = \theta\overline{Y}\lambda C_x^2(\lambda - K) \tag{1.5}$$

and

$$MSE(\hat{\overline{Y}}_M) = \theta\overline{Y}^2\left[C_y^2 + \lambda C_x^2(\lambda - 2K)\right] \tag{1.6}$$

2. *Theoretical Comparisons*

We shall now compare $\hat{\overline{Y}}_M$ with simple mean estimator $\overline{y}$ and ratio estimator $\hat{\overline{Y}}_R$.

The variance of sample mean estimator $\overline{y}$, is given by

$$V(\overline{y}) = \theta\overline{Y}^2 C_y^2 \tag{2.1}$$

The estimator $\hat{\overline{Y}}_M$ will dominate over sample mean estimator $\overline{y}$ if

$$\rho > \frac{1}{2}\left(\frac{\overline{X}C_x}{\left(\overline{X} + \beta_2(x)\right)C_y}\right) \tag{2.2}$$

From (1.3) and (1.6) we observe that $\hat{\overline{Y}}_M$ will be more precise than the ratio estimator $\hat{\overline{Y}}_R$ if

$$\rho < \frac{1}{2}\frac{C_x}{C_y}\left(\frac{2\overline{X} + \beta_2(x)}{\overline{X} + \beta_2(x)}\right) \tag{2.3}$$

Now combining (2.2) and (2.3) we find that the proposed estimator $\hat{\overline{Y}}_M$ is more efficient than simple mean estimator $\overline{y}$ and traditional ratio estimator $\hat{\overline{Y}}_R$ if the following inequality

$$\frac{1}{2}\frac{C_x}{C_y}\left(\frac{\overline{X}}{(\overline{X} + \beta_2(x))}\right) < \rho < \frac{1}{2}\frac{C_x}{C_y}\left(\frac{2\overline{X} + \beta_2(x)}{\overline{X} + \beta_2(x)}\right) \tag{2.4}$$

holds.

Further, if follows from (1.2) and (1.5) that the absolute relative bias of $\hat{\overline{Y}}_M$ is less than that of $\hat{\overline{Y}}_R$ if $|\lambda - K| < |1 - K|$

i.e. if $\rho < \frac{1}{2}\frac{C_x}{C_y}\left(\frac{2\overline{X}+\beta_2(x)}{\overline{X}+\beta_2(x)}\right)$ (2.5)

Hence, it follows from (2.3) and (2.5) that the proposed estimator $\hat{\overline{Y}}_M$ is more efficient as well as less biased than commonly used ratio estimator $\hat{\overline{Y}}_R$ if the inequality (2.5) holds good.

3. A Generalized Version of $\hat{\overline{Y}}_M$

By applying the power transformation on $\hat{\overline{Y}}_M$ in (1.4), the generalized estimator is

$$\hat{\overline{Y}}_{M\alpha} = \overline{y}\left(\frac{\overline{X}+\beta_2(x)}{\overline{x}+\beta_2(x)}\right)^{\alpha} \tag{3.1}$$

where α is a suitably chosen scalar.

The bias and MSE of the estimator $\hat{\overline{Y}}_{M\alpha}$ to the first degree of approximation are respectively given by

$$B(\hat{\overline{Y}}_{M\alpha}) = \theta\alpha\left(\frac{\overline{Y}}{2}\right)\lambda C_x^2\{\lambda(\alpha+1)-2K\} \tag{3.2}$$

and

$$MSE(\hat{\overline{Y}}_{M\alpha}) = \theta\overline{Y}^2\left[C_y^2+\alpha\lambda C_x^2(\alpha\lambda-2K)\right] \tag{3.3}$$

The MSE of $\hat{\overline{Y}}_{M\alpha}$ at (3.3) is minimized for

$$\alpha = \frac{K}{\lambda} \tag{3.4}$$

Thus the resulting bias and minimum MSE of $\hat{\overline{Y}}_{M\alpha}$ are respectively given by

$$B_0(\hat{\overline{Y}}_{M\alpha}) = \theta\left(\frac{\overline{Y}}{2}\right)KC_x^2\{\lambda-K\} \tag{3.5}$$

and

$$\min.MSE(\hat{\overline{y}}_{M\alpha}) = \theta\overline{Y}^2C_y^2(1-\rho^2) \tag{3.6}$$

The min. $MSE(\hat{\overline{Y}}_{M\alpha})$ at (3.6) is same as that of the approximate variance of the usual linear regression estimator.

From (2.1) and (3.3) it follows that MSE $(\hat{\bar{Y}}_{M\alpha}) < V(\bar{y})$ if

$$\left.\begin{array}{ll}\text{either} & 0 < \alpha < \dfrac{2K}{\lambda} \\ \text{or} & \dfrac{2K}{\lambda} < \alpha < 0\end{array}\right\} \tag{3.7}$$

Expressions (1.3) and (3.3) show that $\hat{\bar{Y}}_{M\alpha}$ is more efficient than usual ratio estimator $\bar{y}_R$ if

$$\left.\begin{array}{ll}\text{either} & \dfrac{1}{\lambda} < \alpha < \dfrac{(2K-1)}{\lambda} \\ \text{or} & \dfrac{(2K-1)}{\lambda} < \alpha < \dfrac{1}{\lambda}\end{array}\right\} \tag{3.8}$$

Comparing the MSE expressions (1.6) and (3.3), it is observed that $\hat{\bar{Y}}_{M\alpha}$ is more efficient than $\hat{\bar{Y}}_M$ if

$$\left.\begin{array}{ll}\text{either} & 1 < \alpha < \dfrac{(2K-\lambda)}{\lambda} \\ \text{or} & \dfrac{(2K-\lambda)}{\lambda} < \alpha < 1\end{array}\right\} \tag{3.9}$$

4. Numerical Illustration

The following three natural populations are considered to illustrate the relative behaviour of proposed estimators.

Population - I [Sources: Das (1988)]

It consists of 142 cities of India with population (number of persons) 100,000 and above; the characters x and y being

x : Census population in the year 1961

y : Census population in the year 1971

$\bar{Y} = 4015.2183$, $\bar{X} = 2900.3872$, $C_y = 2.1118$

$C_x = 2.1971$, $\rho = 0.9948$, $\beta_2(x) = 48.1567$

Population - II [Source : Das 1988)]

The population consists of 278 village towns/wards under Gajole police station of Malada district of West Bangal, India, (in fact only those villages of towns/wards have been considered which are shown as inhabited and common to both census 1961 and census 1971 list). The variates considered are

x : The number of agricultural labourers for 1961
y : The number of agricultural labourers for 1971

$\bar{Y} = 39.0680$, $\bar{X} = 25.1110$ $C_y = 1.4451$

$C_x = 1.6198$, $\rho = 0.7213$, $\beta_2(x) = 38.8898$

Population - III [Source : Cochran (1977)]

The variates are defined as follows
y : Number of persons per block
x : Number of rooms per block

$\bar{Y} = 101.1$, $\bar{X} = 58.80$, $C_y = 0.14450$

$C_x = 0.1281$, $\rho = 0.6500$, $\beta_2(x) = 2.2387$

The percent relative efficiencies (PREs) of different estimators with respect to $\bar{y}$ and ranges have been computed and presented in Tables 4.1, 4.2, and 4.3.

Table 4.1. Percent relative efficiency of $\hat{\bar{Y}}_{M\alpha}$ with respect to $\bar{y}$ for different values of α

			POPULATION I		
α	0.00	0.25	0.50	0.75	$\alpha_{opt} = .972054$
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	100.00	179.71	450	1612.50	9640.45
α	1.00	1.25	1.50	1.75	2.00
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	8935.82	1095.47	330.80	155.23	89.52
			POPULATION II		
α	0.00	0.25	0.50	0.75	1.00
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	100.00	117.17	136.77	157.99	178.90
α	1.25	1.50	$\alpha_{opt} = 1.64$	1.75	2.00
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	196.40	206.82	208.45	207.44	198.12
α	2.25	2.50	2.75	3.00	3.25
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	181.27	119.42	101.94	119.42	101.94
α	3.50				
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	87.05				
			POPULATION III		
α	0.00	0.25	0.50	0.75	$\alpha_{opt.} = .761463$
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	100.00	130.19	159.41	173.13	173.16
α	1.00	1.25	1.50	1.75	
PRE($\hat{\bar{Y}}_{M\alpha}, \bar{y}$)	161.56	133.08	102.50	77.48	

Table 4.2. Percent relative efficiency of $\bar{y}, \bar{y}_R$ and $\hat{\bar{Y}}_M$ with respect to $\hat{\bar{y}}$

Estimator	PREs of (.) with respect to $\bar{y}$		
	POPULATION		
	I	II	III
$\bar{y}$	100.00	100.00	100.00
$\bar{y}_R$	8031.10	156.40	157.87
$\hat{\bar{Y}}_M$	8935.82	178.90	161.56

Table 4.3. Range of α for $\hat{\bar{Y}}_{M\alpha}$ to be more efficient than the estimators $\bar{y}, \bar{y}_R$ and $\hat{\bar{Y}}_M$

Estimator	Range of α		
	POPULATION		
	I	II	III
$\bar{y}$	(0.00, 1.9441)	(0.00, 3.2802)	(0.00, 1.5223)
$\bar{y}_R$	(0.9275, 1.0166)	(0.7315, 2.5487)	(0.4842, 1.0381)
$\hat{\bar{Y}}_M$	(0.9441, 1.000)	(1.00, 2.2802)	(0.5223, 1.00)

Table 4.1 and 4.2 exhibit that there is suitable gain in efficiency by using the suggested estimators $\hat{\bar{Y}}_M$ and $\hat{\bar{Y}}_{M\alpha}$ over usual unbiased estimator $\bar{y}$ and ratio estimator $\bar{y}_R$. The estimator $\hat{\bar{Y}}_{M\alpha}$ attained its maximum efficiency at optimum value of $\alpha = \alpha_{opt}$. There is enough scope of choosing the scalar α to obtain better estimator from $\hat{\bar{Y}}_{M\alpha}$.

Table 4.3 gives the range of α for $\hat{\bar{Y}}_{M\alpha}$ to be better than the estimators $\bar{y}, \bar{y}_R$ and $\hat{\bar{Y}}_M$. It is observed from Table 4.3 that the common ranges of α for which the estimator $\hat{\bar{Y}}_{M\alpha}$ is better than all the three estimators $\bar{y}, \bar{y}_R$ and $\hat{\bar{Y}}_M$ are (0.9441, 1.0000), (1.000, 2.2802) and (0.5223, 1.0000) respectively for I, II and III populations. This shows that even if the scalar α deviates from its exact optimum value (α_{opt}) the proposed estimator $\bar{Y}_{M\alpha}$ will yield better estimators than usual estimators $\bar{y}, \bar{y}_R$ and $\hat{\bar{Y}}_M$.

ACKNOWLEDGEMENT

Authors are thankful to the referee for his valuable suggestions regarding the improvement of the paper.

REFERENCES

Cochran, W.G. (1977). *Sampling Techniques.* Third U.S. Edition. Wiley Eastern Limited, 325.

Das, A.K. (1988). Contribution to theory of sampling strategies based on auxiliary information. Ph.D. Thesis submitted to BCKV, Mohanpur, Nadia, West Bangal, India.

Sen, A.R. (1978). Estimation of the population mean when the coefficient of variation is known. *Comm. Stat.–Theory Methods,* **A7**, 657-672.

Singh, J., Pandey, B.N. and Hirano, K. (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Ann. Inst. Statist. Math.,* **25**, 51-55.

Searls, D.T. and Intrapanich, R. (1990). A note on an estimator for variance that utilizes the kurtosis. *Amer. Stat.*, **44(4)**, 295-296.

Singh, H.P. (1984). Use of auxiliary information in estimation procedure for some population parameters. Ph.D. Thesis submitted to Indian School of Mines, Dhanbad, Jharkhand.

Upadhyaya, L.N. and Singh, H.P. (1984). On the estimation of the population mean with known coefficient of variation. *Biom. J.*, **26(8)**, 915-922.