Jour. Ind. Soc. Ag. Statistics
46(2), 1994 : 183–189

A Class of Almost Unbiased Regression — Type Estimators For Finite Population Mean Applying Jack-Knife Technique

N.P. Katyar, Housila P. Singh* and D.K. Gangwar
J. N. Agriculture University, Jabalpur
(Received : July, 1988)

Summary

The paper defines a class of almost unbiased regression - type estimatiors for finite population mean using jack-knife technique suggested by Quenouille [3]. The variance expression of the proposed class of estimators is obtained. In particular, classes of almost unbiased ratio-type and exactly unbiased product-type estimators are generated.

Key words : Almost unbiased regression-type estimators, Jack-Knife technique, ratio and product-type estimatiors.

Introduction

The linear regression estimate is designed to increase precision by the use of an auxiliary variate x that is correlated with the study variate y. When the relation between y and x is examined, it may be found that although the relation is approximately linear, the line does not go through the origin. This suggests an estimate based on the linear regression of y on x rather than on the ratio of the two variables.

Let the variates y, x take values (y_i, x_i) on the ith unit (i = 1, 2, . . ., N) of a finite population. A common situation in surveys is that the population mean $\overline{X}$ of an auxilary characteristic x is known and we are to estimate $\overline{Y}$, the population mean of the study character y. For illustration, consider a simple random sample of size n (< N) units without replacement from the population. Let $(\overline{y}, \overline{x})$ be unbiased estimators of $(\overline{Y}, \overline{X})$ based on n observations. Then the usual linear regression estimator of $\overline{Y}$ is

$$\overline{y}_{lr} = [\overline{y} + \hat{\beta}(\overline{X} - \overline{x})] \tag{1.1}$$

where $\hat{\beta} = \hat{S}_{xy}/\hat{S}_x^2$ is the sample regression coefficient of y on x,

$$\hat{S}_{xy} = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \text{ and } \hat{S}_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

---

* Vikram University, Ujjain, (M.P.).

The bias and mean square error of $\bar{y}_{er}$ to the first degree of approximation are respectively given by

$$B(\bar{y}_{lr}) = -\frac{(N-n)}{(N-2)} \cdot \frac{\beta}{n} \left(\frac{\mu_{21}}{\mu_{11}} - \frac{\mu_{30}}{\mu_{20}} \right) \tag{1.2}$$

(See Sukhatme etal [10], page 239, equation (15))

and

$$MSE(\bar{y}_{lr}) = \frac{(N-n)}{nN} S_y^2 (1-\rho^2) \tag{1.3}$$

where $\beta = S_{xy}/S_x^2$ is the population regression coefficient of y on x,

$$S_{xy} = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y}), \quad S_x^2 = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \bar{X})^2,$$

$$S_y^2 = (N-1)^{-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 ,$$

$\rho = S_{xy}/(S_x S_y)$ is the population correlation coeficient between x and y and $\mu_{rs} = N^{-1} \sum_{i=1}^{n} (x_i - \bar{X})^r (y_i - \bar{Y})^s$, $r, s = 1, 2, 3, 4$; (r, s) being non-negative integers.

It is obvious from (1.2) that $\bar{y}_{lr}$ is biased. It is therefore, desirable to reduce or completely eliminate it. Few authors, including Mickey [2] and Williams [11] have developed estimators that are unbiased, but have not yet been extensively tried. Rao [4] found Mickey's estimator usually inferior to the standard regression and ratio estimators in the natural populations. In this paper an attempt has been made to construct a class of almost unbiased regression type estimators using jack-knife technique envisaged by Quenouille [3], which is further developed by Schucany, Gray and Owen [5].

2. *The Class of Estimators*

A simple random sample of size n = km drawn without replacement, is split at random into K subsamples each of size m. Then the jack knifed regression estimator is defined by

$$\hat{Y}_{lrJ} = \frac{1}{K} \sum_{j=1}^{K} \bar{y}'_{(lr)_j} \tag{2.1}$$

where $\bar{y}'_{(lr)j}$ [$= \bar{y}'_j + \hat{\beta}'(\bar{X} - \bar{x}'_j)$] is the standard regression estimate, computed from the sample with group j omitted, (j = 1, 2, . . ., K);

$\bar{y}'_j = (n\bar{y} - m\bar{y}_j)/(n-m)$, $\bar{x}'_j = (n\bar{x} - m\bar{x}'_j)/(n-m)$, $n\bar{y} = \sum_{i=1}^{n} y_i$, $n\bar{x} = \sum_{i=1}^{n} x_i$,

$\hat{\beta}'_j = \{(n-1)\hat{S}_{xy} - (m-1)\hat{S}_{xyj}\}/\{(n-1)\hat{S}^2_x - (m-1)\hat{S}^2_{xj}\}$, $(\bar{y}_j, \bar{x}_j)$ being the j^th^ sub sample means, $\hat{S}_{xyj} = \sum_{i=1}^{m}(x_{ji} - \bar{x}_j)(y_{ji} - \bar{y}_j)/(m-1)$ and $\hat{S}^2_{xj} = \sum_{i=1}^{m}(x_{ji} - \bar{x}_j)^2/(m-1)$.

The bias of $\hat{Y}_{lrj}$ to the first degree of approximation, is given by

$$B(\hat{Y}_{lrj}) = -\frac{(N-n+m)\beta}{(n-m)(N-2)} \cdot \left(\frac{\mu_{21}}{\mu_{11}} - \frac{\mu_{30}}{\mu_{20}}\right) \tag{2.2}$$

The mean square error of $\hat{Y}_{lrj}$ to the first degree of approximation, is obtained as follows :

Define

$$\bar{y}'_j = \bar{Y}(1 + e'_{oj}), \bar{x}'_j = \bar{X}(1 + e'_{1j}) \text{ and } \hat{\beta}'_j = \beta(1 + e'_{2j})$$

such that

$$E(e'_{oj}) = E(e'_{1j}) = 0 \text{ and}$$

$$E(\hat{\beta}'_j) = \beta + O(n^{-1})$$

$$\Rightarrow \quad E(e'_{2j}) = O(n^{-1})$$

Expressing $\hat{Y}_{lrj}$ in terms of e's we have

$$\hat{Y}_{lrj} = \frac{1}{K}\sum_{j=1}^{K} [\bar{Y}(1 + e'_{oj}) + \beta(1 + e'_{2j})\{\bar{X} - \bar{X}(1 + e'_{1j})\}]$$

$$= \bar{Y} + \frac{\bar{Y}}{K}\sum_{j=1}^{K} [e'_{oj} - \frac{\beta\bar{X}}{\bar{Y}}(e'_{1j} + e'_{1j}e'_{2j})]$$

or

$$(\hat{Y}_{lrJ} - \bar{Y}) = \frac{\bar{Y}}{K}\sum_{j=1}^{K}\left[e'_{oj} - \left(\rho\frac{C_y}{C_x}\right)(e'_{1j} + e'_{1j}e'_{2j})\right] \tag{2.3}$$

where $C_y = S_y/\overline{Y}$ and $C_x = S_x/\overline{X}$.

Squaring both sides of (2.3) and ignoring terms involving e's having power greater than two, we have

$$(\hat{\overline{Y}}_{lrj} - \overline{Y})^2 = \frac{\overline{Y}^2}{K^2}\left[\sum_{j=1}^{K}\{e'^2_{0j} - \left(\rho\frac{C_y}{C_x}\right)^2 e'^2_{1j} - 2\left(\rho\frac{C_y}{C_x}\right)e'_{0j}e'_{1j}\}\right.$$

$$\left. + \sum_{j \neq l = 1}^{K}\{e'_{0j}e'_{0l} - \left(\rho\frac{C_y}{C_x}\right)(e'_{0j}e'_{1l} + e'_{1j}e'_{0l}) + \left(\rho\frac{C_y}{C_x}\right)^2 e'_{1j}e'_{1l}\}\right] \quad (2.4)$$

The following results can easily be derived :

$$E(e'^2_{0j}) = \lambda C_y^2,\ E(e'^2_{1j}) = \lambda C_x^2,\ E(e'_{0j}e'_{1j}) = \lambda\rho C_y C_x,$$

$$E(e'_{0j}e'_{0l}) = \lambda^* C_y^2,\ E(e'_{0j}e'_{1l}) = E(e'_{1j}e'_{0l}) = \lambda^*\rho C_y C_x,\ E(e'_{1j}e'_{1l}) = \lambda^* C_x^2,$$

where

$$\lambda = \frac{(N-n+m)}{N(n-m)} \text{ and } \lambda^* = \frac{1}{(K-1)^2}\left[(k^2 - 2K)\left(\frac{1}{n} - \frac{1}{N}\right) - \frac{1}{N}\right].$$

Taking expectation of both sides of (2.4), using the above results and after simplification, we get the MSE of $\overline{Y}_{lrj}$ to the first degree of approximation as

$$\text{MSE}(\hat{\overline{Y}}_{lrj}) = \frac{(N-n)}{nN} S_y^2(1-\rho^2) = \text{MSE}(\overline{y}_{lr}) \quad (2.5)$$

Now, define a class of regression - type estimators for $\overline{Y}$ as

$$\hat{\overline{Y}}_\theta = (\theta_1\overline{y} + \theta_2\overline{y}_{lr} + \theta_3\hat{\overline{Y}}_{lrj}) \quad (2.6)$$

where θ_1, θ_2 and θ_3 are suitably chosen constants such that $\sum_{i=1}^{3}\theta_i = 1$.

It follows from (1.2) and (2.2) that an estimator in the class (2.6) would be almost unbiased if and only if

$$\delta\theta_2 + \theta_3 = 0 \quad (2.7)$$

where

$$\delta = \frac{(N-n)(n-m)}{(N-n+m)n} \quad (2.8)$$

If we set $\theta_1 = (1-\alpha-\varphi), \theta_2 = \alpha$ and $\theta_3 = \varphi$ in (2.6) then unbiasedness condition takes the form

$$\varphi = -\delta\,\alpha \tag{2.9}$$

Thus, get the general class of almost unbiased regression type estimators

$$\hat{\bar{Y}}_g = \left[\{1-\alpha(1-\delta)\}\bar{y} + \alpha\,\bar{y}_{lr} - \alpha\,\delta\frac{1}{K}\sum_{j=1}^{K}\bar{y}'_{(lr)j}\right] \tag{2.10}$$

of $\bar{Y}$, where α is a suitably chosen constant.

Remark 2.1 : One can observe that $\alpha = 0$ gives the usual unbiased estimator $\bar{y}$ while $\alpha = (1-\delta)^{-1}$ yields the estimator

$$\hat{\bar{Y}}_1 = \left[\frac{(N-n+m)}{N}K\,\bar{y}_{lr} - \frac{(N-n)}{N}\cdot\frac{(K-1)}{K}\sum_{j=1}^{K}\bar{y}'_{(lr)j}\right] \tag{2.11}$$

When N is very large or the population is infinite the estimator $\hat{\bar{Y}}_1$ reduces to

$$\hat{\bar{Y}}_1 = \left[K\,\bar{y}_{lr} - \frac{(K-1)}{K}\sum_{j=1}^{K}\bar{y}'_{(lr)j}\right] \tag{2.12}$$

which is reported in Cochran ([1], equation (7.62), page 203).

Remark 2.2 : For (i) $\hat{\beta} = (\bar{y}/\bar{x})$ and $\hat{\beta}'_j = (\bar{y}'_j/\bar{x}'_j)$; and (ii) $\hat{\beta} = -(\bar{y}/\bar{x})$ and $\hat{\beta}'_j = -\left(\frac{\bar{y}'_j}{\bar{X}}\right)$; the class of estimators $\hat{\bar{Y}}_g$ in (2.10) reduces to the class of almost unbiased ratio type and class of exactly unbiased product-type estimators, resprectively, as

$$\hat{\bar{Y}}_{rC} = \left[\{1-\alpha(1-\delta)\}\,\bar{y} + \alpha\,\bar{y}\,(\bar{X}/\bar{x}) - \alpha\,\delta\frac{1}{k}\sum_{j=1}^{K}\bar{y}'_j\,(\bar{X}/\bar{x}'_j)\right] \tag{2.13}$$

and

$$\hat{\bar{Y}}_{pC} = \left[\{1-\alpha(1-\delta)\}\,\bar{y} + \alpha\,\bar{y}\,(\bar{x}/\bar{X}) - \alpha\,\delta\frac{1}{K}\sum_{j=1}^{K}\bar{y}'_j\,(\bar{x}'_j/\bar{X})\right] \tag{2.14}$$

which are due to Singh [13]. In addition to Singh [6, 7, 8, 9] type estimators many other almost unbiased estimators of $\bar{Y}$ can be had by substituting the proper choices of α in $\hat{\bar{Y}}_g$.

3. *Optimum Estimator in the Class* $\hat{Y}_g$

The variance of $\hat{Y}_g$, to the first degree of approximation is given by

$$V(\hat{Y}_g) = \frac{(N-n)}{nN} S_y^2 [1 - \alpha(1-\delta)\{2 - \alpha(1-\delta)\rho^2]\tag{3.1}$$

which is minimized for

$$\alpha = (1-\delta)^{-1} = \alpha_{opt} \text{ (say)}\tag{3.2}$$

Hence the resulting variance to the first degree of approximation, is

$$\text{Min. } V(\hat{Y}_g) = V(\bar{y}_{lr}) = \frac{(N-n)}{Nn}(1-\rho^2) S_y^2\tag{3.3}$$

Substituting $\alpha_{opt} = (1-\delta)^{-1}$ for α in (2.10) one can get optimum estimator of the class (2.10) as given by

$$\hat{Y}_1 = \left[\frac{(N-n+m)}{N} K\,\bar{y}_{lr} - \frac{(N-n)}{N} \cdot \frac{(K-1)}{K} \sum_{j=1}^{K} \bar{y}'_{(lr)j}\right]\tag{3.4}$$

with the approximate variance equals to that of usual biased linear regression estimator $\bar{y}_{lr}$ given by (1.3).

ACKNOWLEDGEMENT

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

REFERRENCES

[1] Cochran, W.G., 1977. Sampling techniques, third edition, fourth Wiley Eastern Reprint, October 1985.

[2] Mickey, M.R., 1959. Some finite population unbiased ratio and regression estimators. *J. Amer. Statist. Assoc.*, **54**, 594–612.

[3] Quenouille, M.H., 1956. Notes on bias in estimation. *Biometrika*, **43**, 353–360.

[4] Rao, J.N.K., 1969. Ratio and regression estimators. New Developments in survey sampling, N.L. Johnson and H. Smith, Jr. (Eds.), John Wiley and Sons, New York, 213–234.

[5] Schucany, W.R., Gray, H.L. and Owen, D.B., 1971. On bias reduction in estimation, *J. Amer. Statist. Assoc.* **66**, 524–533.

[6] Singh, H.P., 1984. Use of auxiliary information in estimation procedure for some population parameters. Unpublisehed Ph.D. thesis, submitted to Indian School of Mines, Dhanbad, India.

[7] Singh, H.P., 1986. A note on unbiased product type estimators. *Alig. J. Statist.* **6**, 45–52.

[8] Singh, H.P., 1987. On applying jack-knife technique to ratio estimator. *Guj. Stat. Rev.* **14**, (1), 9–18.

[9] Singh, H.P., 1987. Class of almost unbiased ratio and product type estimators for finite population mean applying Quenouille's method. *J. Ind. Soc. Agri. Stat.* **39** (3), 280–288.

[10] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C., 1984. Sampling theory of surveys with applications. Third Revised Edition, Iowa State University Press, Ames, Iowa (USA) and Indian Society of Agricultural Statistics, New Delhi, India.

[11] Williams, W.H., 1963. The precision of some unbiased regression estimators. *Biometrika*, **17**, 267–274.