Rev.R.Acad.Cienc.Exact.Fís.Nat. (Esp)
Vol. 93, N.º 2, pp 217-220, 1999
Matemáticas

A CLASS OF PPS ESTIMATORS OF POPULATION VARIANCE USING AUXILIARY INFORMATION

(auxiliary information/bias/finite population/mean squared error/PPS estimators)

HOUSILA P. SINGH AND M. RUIZ ESPEJO

School of Studies in Statistics, Vikram University, Ujjain 456010, M.P., India; and UNED, Apartado 19207, 28080 Madrid, Spain

Presentado por Pedro Jiménez Guerra el 6 de octubre de 1998. Aceptado el 10 de Marzo de 1999.

ABSTRACT

A class of estimators for the variance of a finite population is defined using information on two auxiliary variables, one of which is used at the sample selection stage and other for improving the estimator at the estimation stage. Asymptotic expressions for bias and mean squared error are obtained. Asymptotically optimum estimator (AOE) in the class is investigated with its approximate mean squared error. Further, a more general class of estimators is also proposed and its properties are discussed. For illustration, an empirical study is provided.

RESUMEN

Se define una clase de estimadores de la varianza de una población finita, usando información auxiliar de dos variables auxiliares, una de las cuales se usa en la etapa de la selección muestral y otra para mejorar el estimador en la etapa de estimación. Se obtienen expresiones asintóticas para el sesgo y el error cuadrático medio. Se investiga el estimador óptimo asintótico (EOA) en la clase, con su error cuadrático medio aproximado. Además, se propone también una clase de estimadores más general y se discuten sus propiedades. Se proporciona un estudio empírico como ilustración.

MATERIAL AND METHODS

Suppose that information on two auxiliary variables related with the study variable y is available. Let a sample of size n be drawn with PPS (probability proportional to size) with replacement. Let P_i denote the probability of selection (based on one of the two auxiliary variables) of ith unit $i = 1, 2, ..., N$. Let y_i and x_i denote the values of the variable under study y and the auxiliary variable x for the ith unit of the population and $(\bar{Y}, \bar{X})$ denote their population means respectively. We write

$$u_i = y_i/(NP_i),\ v_i = x_i/(NP_i),\ \bar{u}_n = \sum_{i=1}^{n} u_i/n,\ \bar{v}_n = \sum_{i=1}^{n} v_i/n,$$

$$s_u^2 = \sum_{i=1}^{n}(u_i - \bar{u}_n)^2/(n-1) \text{ and } s_v^2 = \sum_{i=1}^{n}(v_i - \bar{v}_n)^2/(n-1).$$

Further, we define

$$\sigma_u^2 = \sum_{i=1}^{N} P_i(u_i - \bar{Y})^2,\ \sigma_v^2 = \sum_{i=1}^{N} P_i(v_i - \bar{X})^2,$$

$$c_u = \sigma_u/\bar{Y},\ c_v = \sigma_v/\bar{X},$$

$$\rho_{uv} = \mu_{11}/(\sigma_u\sigma_v),\ \phi = \mu_{21}/(\sigma_u^2\sigma_v),$$

$$\beta_2(x) = \mu_{04}/\sigma_v^4,\ \beta_2(y) = \mu_{40}/\sigma_u^4,$$

$$\gamma_1^2 = \beta_1(x),\ \gamma_1 = \mu_{03}/\sigma_v^3,\ \theta = \mu_{22}/(\sigma_u^2\sigma_v^2),$$

$$\mu_{rs} = \sum_{i=1}^{N} P_i(u_i - \bar{Y})^r(v_i - \bar{X})^s;\ r \geq 0,\ s \geq 0;$$

$$\varepsilon = (s_u^2/\sigma_u^2) - 1,\ \delta = (\bar{v}_n/\bar{X}) - 1,\ \eta = (s_v^2/\sigma_v^2) - 1$$

such that $E(\varepsilon) = E(\delta) = E(\eta) = 0$ and to the first degree of approximation,

$$E(\varepsilon^2) = \beta_2^*(y)/n,\ E(\eta^2) = \beta_2^*(x)/n,\ E(\delta^2) = c_v^2/n,$$

$$E(\varepsilon\delta) = \phi c_v/n,\ E(\delta\eta) = \gamma_1 c_v/n \text{ and } E(\varepsilon\eta) = \theta^*/n,$$

where

$$\beta_2^*(y) = (\beta_2(y) - 1),\ \beta_2^*(x) = (\beta_2(x) - 1) \text{ and } \theta^* = (\theta - 1).$$

In this paper, we define a class of estimators of the population variance σ_u^2 of the study character y on the lines of Srivastava and Jhajj (1980) when sampling is done by the method of probability proportional to a suitable size variable which is different from the auxiliary variable used at the estimation stage. Asymptotic expressions for the bias and mean squared error of the proposed class of estimators are derived and the minimum value of mean squared error is also obtained. Comparison has been made with the conventional estimator of the variance. A more general class of estimators is also defined with its properties.

RESULTS AND DISCUSSION

The estimator and its properties

Let $w=\bar{v}_n/\bar{X}$ and $z=s_v^2/\sigma_v^2$. Whatever be the sample chosen, let (w, z) assume values in a bounded closed convex subset, D of the two dimensional real space containing the point (1, 1). Let $h(w, z)$ be a function of w and z such that $h(1, 1) = 1$ and such that it satisfies the following conditions:

1. In D, the function $h(w, z)$ is continuous and bounded.

2. The first and second partial derivatives of $h(w, z)$ exist and are continuous and bounded.

We define the class of estimators of σ_u^2 as

$$S_h = s_u^2 h(w,z). \tag{1}$$

Expanding $h(w, z)$ about the point (1, 1) in a second order Taylor's series and noting that the second order partial derivatives of $h(w, z)$ are bounded, we have

$$E(S_h)=\sigma_u^2+o(n^{-1})$$

and so the bias is of the order of n^{-1}.

Using the results of the section «Material and methods», we get the mean squared error (MSE) of S_h up to terms of order n^{-1},

$$\begin{aligned} MSE(S_h)&=E(S_h-\sigma_u^2)^2= \\ &=(\sigma_u^4/n)\left[\beta_2^*(y)+c_v^2h_1^2(1,1)+\beta_2^*(x)h_2^2(1,1)+\right. \\ &\left.+2\gamma_1 c_v h_1(1,1)h_2(1,1)+2\phi c_v h_1(1,1)+2\theta^* h_2(1,1)\right], \end{aligned} \tag{2}$$

where $h_1(1, 1)$ and $h_2(1, 1)$ denote the first partial derivatives of $h(w, z)$ with respect to w and z respectively.

The MSE of S_h is minimized for

$$\left.\begin{aligned} h_1(1,1)&=\frac{(\theta^*\gamma_1-\beta_2^*(x)\phi)}{(\beta_2(x)-\beta_1(x)-1)c_v} \\ h_2(1,1)&=\frac{(\gamma_1\phi-\theta^*)}{(\beta_2(x)-\beta_1(x)-1)} \end{aligned}\right\}. \tag{3}$$

Hence the resulting minimum MSE of S_h is given by

$$\min.MSE(S_h)=(\sigma_u^4/n)\left[\beta_2^*(y)-\phi^2-\frac{(\gamma_1\phi-\theta^*)^2}{(\beta_2(x)-\beta_1(x)-1)}\right]. \tag{4}$$

Thus we state the following theorem.

Theorem 1. Up to terms of order n^{-1},

$$MSE(S_h)\geq(\sigma_u^4/n)\left[\beta_2^*(y)-\phi^2-\frac{(\gamma_1\phi-\theta^*)^2}{(\beta_2(x)-\beta_1(x)-1)}\right] \tag{5}$$

with equality holding if

$$\left.\begin{aligned} h_1(1,1) &= (\theta^*\gamma_1-\beta_2^*(x)\phi)/\{c_v(\beta_2(x)-\beta_1(x)-1)\} \\ h_2(1,1) &= (\gamma_1\phi-\theta^*)/(\beta_2(x)-\beta_1(x)-1) \end{aligned}\right\}. \tag{6}$$

Any parametric function $h(w, z)$ such that $h(1, 1) = 1$ and satisfying the conditions 1 and 2 can generate an estimator of the class. The class of such estimators is very large. The following functions for illustration give five estimators of the class (1),

$$h^{(1)}(w,z)=w^\alpha z^\beta,\ h^{(2)}(w,z)=\frac{\{1+\alpha(w-1)\}}{\{1-\beta(z-1)\}},$$

$$h^{(3)}(w,z) = [1+\alpha(w-1)+\beta(z-1)],$$

$$h^{(4)}(w,z) = [1-\alpha(w-1)-\beta(z-1)]^{-1},$$

$$h^{(5)}(w,z) = \exp\{\alpha(w-1)+\beta(z-1)\},$$

where α and β are suitable chosen constants.

Let the five estimators generated by $h^{(i)}(w, z)$ be denoted by

$$S_{h_i}=s_u^2 h^{(i)}(w,z);\ (i=1,2,3,4,5).$$

It is easily seen that the optimum values of constants α and β in all the five estimators are same and are given by the right hand side of (3) (or (6)) and with these optimum values, the asymptotic minimum MSE is given by (4). For more illustrations the papers of Srivastava and Jhajj (1980), Singh (1986) and Upadhyaya and Singh (1985) may be seen.

It is well known that

$$V(s_u^2)=(\sigma_u^2/n)\beta_2^*(y). \tag{7}$$

From (4) and (7) we have

$$V(s_u^2) - \min.MSE(S_h) = (\sigma_u^2 / n)\left[\phi^2 + \frac{(\gamma_1\phi - \theta^*)^2}{(\beta_2(x) - \beta_1(x) - 1)}\right]$$

which is always positive. Thus the proposed class of estimators S_h is more efficient than conventional unbiased estimator s_u^2.

The bias of S_h

To obtain the bias of S_h in (1), we assume that the third partial derivative of h (w, z) also exist and are continuous and bounded. Then expanding h (w, z) about (1, 1) in third order Taylor's series, the bias of S_h is obtained as

$$B(S_h) = \left(\frac{\sigma_u^2}{2n}\right)\left[2\phi c_v h_1(1,1) + 2\theta^* h_2(1,1) + \right.$$
$$\left. + \frac{1}{2}\left\{c_v^2 h_{11}(1,1) + 2\gamma_1 c_v h_{12}(1,1) + \beta_2^*(x) h_{22}(1,1)\right\}\right], \quad (8)$$

where h_{11} (1, 1), h_{12} (1, 1) and h_{22} (1, 1) denote the second order partial derivatives of the function h (w, z) about (1, 1). Thus we observe that the bias of the estimator S_h also depends upon the second orden partial derivatives of the function h (w, z) at the point (1, 1) and hence will be different for different optimum estimators of the class (1). From (8), the asymptotic bias of any estimator belonging to the class (1) can easily be obtained.

A wider class of estimators

In this section we define a class of estimators of σ_u^2 wider as well as more efficient than S_h defined in (1). It is to be noted that the class of estimators S_h fails to include the difference type estimator

$$S_{g(1)} = \left[s_u^2 + \alpha(w-1) + \beta(z-1)\right]$$

and the estimator

$$S_{g(2)} = \alpha^* s_u^2$$

of Singh, Pandey and Hirano (1973) and Searls and Intarapanich (1990) -type estimator. These estimators led authors to propose the following class of estimators for σ_u^2 as

$$S_g = g(s_u^2,\ w, z), \quad (9)$$

where $g(s_u^2, w, z)$ is a function of s_u^2, w and z satisfies the following conditions:

(*i*) (s_u^2, w, z) assume the values in a bounded closed convex subset P, of the three dimensional real space containing the point $(\sigma_u^2, 1, 1)$;

(*ii*) $g(s_u^2, u, v)$ is a continuous and bounded in P;

(*iii*) $g(\sigma_u^2, 1, 1) = \sigma_u^2 g_1(\sigma_u^2, 1, 1)$, (10)

where $g_1(\sigma_u^2, 1, 1)$ denotes the first partial derivatives with respect to s_u^2 at $(s_u^2, w, z) = (\sigma_u^2, 1, 1,)$;

(*iv*) the first and second order partial derivatives of $g(s_u^2, w, z)$ exist and are continuous and bounded in P.

Expanding $g(s_u^2, w, z)$ about the point $(\sigma_u^2, 1, 1)$ in a second order Taylor's series, we have that

$$E(S_g) = \sigma_u^2 g_1(\sigma_u^2, 1, 1) + o(n^{-1}) \quad (11)$$

and so the bias of S_g is of order o $(1/n)$.

The MSE of S_g upto terms of order o (n^{-1}) is

$$MSE(S_g) = E(S_g - \sigma_u^2)^2 =$$
$$= \left[\sigma_u^4\{1 - g_1(\sigma_u^2,1,1)\}^2 + (1/n)\{\beta_2^*(y)\sigma_u^4 g_1^2(\sigma_u^2,1,1) + \right.$$
$$+ c_v^2 g_2^2(\sigma_u^2,1,1) + \beta_2^*(x) g_3^2(\sigma_u^2,1,1) + 2\phi c_v \sigma_u^2 g_1(\sigma_u^2,1,1) g_2(\sigma_u^2,1,1) +$$
$$+ 2\theta^* \sigma_u^2 g_1(\sigma_u^2,1,1) g_3(\sigma_u^2,1,1)$$
$$\left. + 2\gamma_1 c_v g_2(\sigma_u^2,1,1) g_3(\sigma_u^2,1,1)\}\right] \quad (12)$$

where $g_2(\sigma_u^2, 1, 1)$ and $g_3(\sigma_u^2, 1, 1)$ denote the first order partial derivatives of w and z respectively about the point $(s_u^2, w, z) = (\sigma_u^2, 1, 1)$.

The MSE (S_g) at (12) is minimized for

$$\left.\begin{aligned} g_1(\sigma_u^2,1,1) &= d_1/d, \\ g_2(\sigma_u^2,1,1) &= -\left\{(R\bar{X}c_u)^2 / c_v\right\}\frac{d_2}{d}, \\ g_3(\sigma_u^2,1,1) &= (R\bar{X}c_u)^2 d_3/d. \end{aligned}\right\} \quad (13)$$

where $R = \bar{Y}/\bar{X}$, $d_1 = (\beta_2(x) - \beta_1(x) - 1)$,

$d_2 = (\phi\beta_2^*(x) - \gamma_1\theta^*)$, $d_3 = (\phi\gamma_1 - \theta^*)$

and $d = \left[d_1\{1 + (1/n)(\beta_2^*(y) - \phi^2)\} - (1/n)(\theta^* - \phi\gamma_1)\right]$.

Substituting optimum values of $g_i(\sigma_u^2, 1, 1)$, (i = 1, 2, 3) from (13) into (12) we obtain the minimum MSE of S_g to terms of order o (n^{-1}),

$$\min.MSE\ (S_g) = \left(\frac{\sigma_u^4}{n}\right)\left[d_1(\beta_2^*(y) - \phi^2) - (\theta^* - \phi\gamma_1)\right]/d. \quad (14)$$

From (4) and (14) we have

$$\min.MSE\,(S_h) - \min.MSE\,(S_g) = \left(\frac{\sigma_u^2}{n}\right)\frac{\left[d_1\left(\beta_2^*(y)-\phi^2\right)-\left(\theta^*-\phi\gamma_1\right)\right]^2}{dd_1} \tag{15}$$

which is always positive. Hence the class of estimators S_g defined in (9) is more wide as well as more efficient than the class of estimators S_h. If we set $g_1(\sigma_u^2,1,1)=1$ then the minimum *MSE* of the class S_g equals to the minimum *MSE* of S_h given in (4).

Empirical illustration

The data under consideration is given in Cochran (1977, Ch. 2) dealing with data on y = weekly expenditure on food, x_1 = the number of persons in a family, and x_2 = the weekly family income. Here x_1 is used at the selection stage and x_2 at the estimation stage. The following values were computed and comparison among different estimators is made for a PPS sample of size n drawn with replacement.

$$\bar{Y}=27.49,\ \sigma_u^2=111.83,\ \beta_2(y)=2.05,\ \beta_2(x_2)=3.894,$$
$$\bar{X}_2=72.49,\ \sigma_v^2=1163.84,\ \phi=0.2607,\ \beta_1(x_2)=1.663,$$
$$\gamma_1=\sqrt{\beta_1(x_2)}=1.2894 \text{ and } \theta=1.5037$$

Using these values the relative minimum *MSE* of various estimators are computed and assembled in Table I, where $\hat{c}_v^2 = s_v^2/\bar{v}_n^2$ and (α^*, α, β) are constants.

Further, to see the pattern of relative efficiencies of these estimators with respect to s_u^2 are computed for n = 10 and tabled in Table II.

Table II exhibits that the performances of the estimator e_3 is best among all the estimators followed by $e_{3(3)}$.

No.	Estimator	min.*MSE* $(\cdot)$ / (σ_u^4/n)
1.	$e_0 = s_u^2$	1.0465
2.	$e_1 = s_u^2(\bar{v}_n/\bar{X}_2)^\alpha (s_v^2/\sigma_v^2)^\beta$	0.9557
3.	$e_{1(1)} = s_u^2(\bar{v}_n/\bar{X}_2)^\alpha$	0.9785
4.	$e_{1(2)} = s_u^2(s_v^2/\sigma_v^2)$	0.9588
5.	$e_{1(3)} = s_u^2(\hat{c}_v^2/c_v^2)$	0.9972
6.	$e_2 = s_u^2+\alpha(w-1)+\beta(z-1)$	0.9557
7.	$e_3 = \alpha^* s_u^2+\alpha(w-1)+\beta(z-1)$	0.9557 / (1 + 0.9557/n)
8.	$e_{3(1)} = \alpha^* s_u^2$	1.0465 / (1 + 1.0465/n)
9.	$e_{3(2)} = \alpha^* s_u^2+\alpha(w-1)$	0.9785 / (1 + 0.9785/n)
10.	$e_{3(3)} = \alpha^* s_u^2+\beta(z-1)$	0.9588 / (1 + 0.9588/n)
11.	$e_4 = s_u^2(\sigma_v^2/s_v^2)$	1.9366
12.	$e_5 = s_u^2(\bar{X}_2/\bar{v}_n)$	1.0226
13.	$e_6 = s_u^2-(w-1)-(z-1)$	1.0391

Table I. The relative minimum *MSE* of different estimators of σ_u^2.

Estimator	Relative efficiency (%)
e_0	100
e_1	109.50
$e_{1(1)}$	107.00
$e_{1(2)}$	109.15
$e_{1(3)}$	105.00
e_2	109.7
e_3	120.00
$e_{3(1)}$	110.45
$e_{3(2)}$	117.41
$e_{3(3)}$	119.61
e_4	54.038
e_5	102.337
e_6	100.712

Table II. Showing the per cent relative efficiencies of various estimators w.r.t. s_u^2.

REFERENCES

1. Cochran, W.G. (1977) *Sampling Techniques* (Third edition). John Wiley, New York.
2. Searls, D.T. & Intarapanich, P. (1990) A note on an estimator for the variance that utilizes the kurtosis. *Amer. Statist.* **44**, 295-296.
3. Singh, H.P. (1986) A generalized class of estimators of ratio, product and mean using supplementary information on an auxiliary character in PPSWR sampling scheme. *Gujarat Statist. Rev.* **13** (2), 1-30.
4. Singh, J., Pandey, B.N. & Hirano, K. (1973) On the utilization of a known coefficient of kurtosis in the estimation procedures for variance. *Ann. Inst. Statist. Math.*, **25**, 51-55.
5. Srivastava, S.K. & Jhajj, H.S. (1980) A class of estimators using auxiliary information for estimating finite population variance. *Sankhyā Ser. C* **42**, 87-96.
6. Upadhyaya, L.N. & Singh, H.P. (1985) A class of estimators using auxiliary information for estimating ratio of two finite population means. *Gujarat Statist. Rev.* **12** (2), 7-16.