COVARIANCE OF SAMPLE MOMENTS FOR SIMPLE RANDOM SAMPLING WITH REPLACEMENT OF A FINITE POPULATION

By

M. Ruiz Espejo
Facultad de Ciencias Económicas, UNED, Madrid, Spain

and

Housila P. Singh
School of Studies in Statistics, Vikram University, Ujjain, India

[Received 7 September 1998. Read 24 May 1999. Published 30 December 1999.]

Abstract

In this paper we derive the exact covariance of some sample moments for 'simple random sampling with replacement' (SRSWR) of a finite population. An application of the results is given for estimating the finite population mean with linear regression estimators.

1. Introduction

When the finite population is available in data base form it is simplest to use the 'simple random sampling with replacement' (SRSWR) design for programming the sample selection (see [6]).

Ruiz Espejo [5] dealt with the problem of the determination of the covariance of sample moments under 'simple random sampling without replacement' (SRSWOR) for a finite population. For concrete concepts of sampling designs the reader is referred to [4].

This paper considers the problem of obtaining the covariance of (i) the sample non-central moments of any orders, for two variables x and y, and (ii) the sample non-central moment of any order for x, with the sample variance for y (sample central moment of order 2 for y).

Applications of these basic investigations are provided in §3 for generalised regression punctual estimators for the finite population mean.

2. Results

Theorem 2.1. *For simple random sample of size n drawn with replacement from a finite population of size N, we have:*

(2.1.i)
$$Cov\left(a_{i0}, a_{0j}\right) = \frac{\alpha_{ij} - \alpha_{i0}\alpha_{0j}}{n},$$

(2.1.ii)
$$Cov\left(a_{i0}, m_{02}\right) = \frac{n-1}{n^2}(\alpha_{i2} - \alpha_{i0}\alpha_{02} - 2\alpha_{01}\alpha_{i1} + 2\alpha_{i0}\alpha_{01}^2) + \frac{2(n-1)(n-2)}{n^2N^2}\alpha_{i2},$$

where $a_{ij} = (1/n)\sum_{k=1}^{n} x_k^i y_k^j$ *is the sample joint non-central moment of order i and j;* $m_{02} = (1/n)\sum_{k=1}^{n}(y_k - a_{01})^2$ *is the sample marginal variance of the second variable y; and* $\alpha_{ij} = (1/N)\sum_{k=1}^{N} x_k^i y_k^j$ *is the population joint non-central moment of order i and j.*

The first result (2.1.i) is also valid for any 'population' with such finite parameters, since for any 'population' we can build a succession of finite populations convergent in law to such 'population'. The second result (2.1.ii) is also valid for an infinite population, when $N \to \infty$.

Corollary 2.2. *For simple random sampling of size n from a finite population of size N,*

$$Cov\,(a_{0i}, m_{02}) = \frac{n-1}{n^2}\left(\alpha_{0,i+2} - \alpha_{0i}\alpha_{02} - 2\alpha_{01}\alpha_{0,i+1} + 2\alpha_{0i}\alpha_{01}^2\right) + \frac{2\,(n-1)\,(n-2)}{n^2 N^2}\alpha_{0,i+2}.$$

The result is a particular case of (2.1.ii) when the first (x) and the second (y) variables are the same ($x_k = y_k : k = 1, 2, ..., N$).

Corollary 2.3. *For simple random sampling of size n from a finite population of size N,*

(2.3.i)

$$Cov\,(a_{10}, a_{01}) = \frac{\mu_{11}}{n},$$

(2.3.ii)

$$Cov\,(a_{10}, m_{02}) = \frac{n-1}{n^2}\mu_{12} + \frac{2(n-1)(n-2)}{n^2 N^2}\alpha_{12},$$

where $\mu_{ij} = (1/N)\sum_{k=1}^{N}(x_k - a_{10})^i (y_k - a_{01})^j$ *is the population joint central moment of order i and j.*

This is a consequence of Theorem 2.1 when $i = j = 1$.

Corollary 2.4. *Denoting* $s_y^2 = nm_{02}/\,(n-1)$*, for simple random sampling of size n from a finite population of size N, we have*

$$Cov\,\left(a_{10}, s_y^2\right) = \frac{\mu_{12}}{n} + \frac{2(n-2)}{nN^2}\alpha_{12}.$$

This follows from (2.3.ii).

Corollary 2.5. *For simple random sampling of size n,*

$$Cov\left(a_{0i}, a_{0j}\right) = \frac{\alpha_{0,i+j} - \alpha_{0i}\alpha_{0j}}{n}.$$

This follows from (2.1.i) when the first and second variables are the same.

Corollary 2.6. *For simple random sampling of size n,*

$$Cov\left(a_{10}, a_{02}\right) = \frac{\alpha_{12} - \alpha_{10}\alpha_{02}}{n}.$$

This follows from (2.1.i).

Corollary 2.7. *For simple random sampling of size n from a finite population of size N,*

(2.7.i)

$$V\left(a_{10}\right) = \frac{\mu_{20}}{n},$$

(2.7.ii)

$$Cov\left(a_{10}, m_{20}\right) = \frac{(n-1)\,\mu_{30}}{n^2} + \frac{2\,(n-1)\,(n-2)\,\alpha_{30}}{n^2 N^2}.$$

This follows from Corollary 2.3, when the first and second variables are the same. The result (2.7.ii) was given by Cramér [1, p. 401] for infinite populations (when $N \to \infty$) and then the second term of (2.7.ii) would be $(n-1)\,\mu_{30}/n^2$.

3. Applications

Some regression estimators of α_{10} are:

$$\hat{\alpha}_{10} = a_{10} + b_1\left(\alpha_{01} - a_{01}\right) + b_2\left(\mu_{02} - s_y^2\right), \tag{1}$$

$$\hat{\alpha}'_{10} = a_{10} + b_1\left(\alpha_{01} - a_{01}\right) + b_2\left(\alpha_{02} - a_{02}\right). \tag{2}$$

The treatment of these estimators can be solved with the previous results. For example, for (1), considering b_1 and b_2 two constants, the variance is

$$V\left(\hat{\alpha}_{10}\right) = V\left(a_{10}\right) + b_1^2 V\left(a_{01}\right) + b_2^2 V\left(s_y^2\right) - 2b_1 Cov\left(a_{10}, a_{01}\right)$$

$$-2b_2 Cov\left(a_{10}, s_y^2\right) + 2b_1 b_2 Cov\left(a_{01}, s_y^2\right),$$

which is minimised when

$$b_1 = \frac{Cov\left(a_{10}, a_{01}\right) V\left(s_y^2\right) - Cov\left(a_{01}, s_y^2\right) Cov\left(a_{10}, s_y^2\right)}{V\left(a_{01}\right) V\left(s_y^2\right) - Cov^2\left(a_{01}, s_y^2\right)} \tag{3}$$

and

$$b_2 = \frac{V(a_{01})\,Cov\left(a_{10}, s_y^2\right) - Cov\left(a_{01}, s_y^2\right) Cov\left(a_{10}, a_{01}\right)}{V(a_{01})\,V\left(s_y^2\right) - Cov^2\left(a_{01}, s_y^2\right)}. \tag{4}$$

If the auxiliary variable or second variable y is known for all population units, then b_1 and b_2 of (3) and (4) are known, except the values of $Cov\,(a_{10}, a_{01})$ and $Cov\,(a_{10}, s_y^2)$, which can be estimated, following Olkin [3]. The consistent estimates of $Cov\,(a_{10}, a_{01})$ and $Cov\,(a_{10}, s_y^2)$ are respectively given by

$$\widehat{Cov}\,(a_{10}, a_{01}) = \frac{\hat{\mu}_{11}}{n} = \frac{m_{11}}{n} \tag{5}$$

and

$$\widehat{Cov}\left(a_{10}, s_y^2\right) = \frac{\hat{\mu}_{12}}{n} + \frac{2(n-2)}{nN^2}\hat{\alpha}_{12} = \frac{m_{12}}{n} + \frac{2(n-2)}{nN^2}a_{12}, \tag{6}$$

where $m_{ij} = \left(1/n\right) \sum_{k=1}^{n} (x_k - a_{10})^i\,(y_k - a_{01})^j$.

Alternative unbiased estimators of $Cov\,(a_{10}, a_{01})$ and $Cov\,(a_{10}, s_y^2)$ are respectively given by

$$\left[\widehat{Cov}\,(a_{10}, a_{01})\right]_u = (a_{11} - a_{10}\alpha_{01})\,/n \tag{7}$$

and

$$\left[\widehat{Cov}\left(a_{10}, s_y^2\right)\right]_u = \frac{1}{n}\left[a_{12} - 2a_{11}\alpha_{01} + a_{10}\left(2\alpha_{01}^2 - \alpha_{02}\right)\right] + \frac{2(n-2)}{nN^2}a_{12}. \tag{8}$$

Using (5) and (6) in (3) and (4) we can get the consistent estimates of b_1 and b_2 as

$$\hat{b}_1 = \frac{\left[\widehat{Cov}\,(a_{10}, a_{01})\,V\left(s_y^2\right) - Cov\left(a_{01}, s_y^2\right) \widehat{Cov}\left(a_{10}, s_y^2\right)\right]}{\left\{V(a_{01})\,V\left(s_y^2\right) - \left[Cov\left(a_{01}, s_y^2\right)\right]^2\right\}}, \tag{9}$$

and

$$\hat{b}_2 = \frac{\left[V(a_{01})\,\widehat{Cov}\left(a_{10}, s_y^2\right) - Cov\left(a_{01}, s_y^2\right) \widehat{Cov}\,(a_{10}, a_{01})\right]}{\left\{V(a_{01})\,V\left(s_y^2\right) - \left[Cov\left(a_{01}, s_y^2\right)\right]^2\right\}}. \tag{10}$$

Alternative estimates of b_1 and b_2 can be obtained by using (7) and (8) in (3) and (4) as

$$\hat{b}_1^* = \frac{\left\{\left[\widehat{Cov}\,(a_{10}, a_{01})\right]_u V\left(s_y^2\right) - Cov\left(a_{01}, s_y^2\right) \left[\widehat{Cov}\left(a_{10}, s_y^2\right)\right]_u\right\}}{\left\{V(a_{01})\,V\left(s_y^2\right) - \left[Cov\left(a_{01}, s_y^2\right)\right]^2\right\}}, \tag{11}$$

and

$$\hat{b}_2^* = \frac{\left\{V(a_{01})\left[\widehat{Cov}\left(a_{10}, s_y^2\right)\right]_u - Cov\left(a_{01}, s_y^2\right) \left[\widehat{Cov}\,(a_{10}, a_{01})\right]_u\right\}}{\left\{V(a_{01})\,V\left(s_y^2\right) - \left[Cov\left(a_{01}, s_y^2\right)\right]^2\right\}}. \tag{12}$$

Thus the regression estimator in (1), based on 'estimated optimum', is given by

$$\hat{\alpha}_{10}^{\circ} = a_{10} + \hat{b}_1 (\alpha_{01} - a_{01}) + \hat{b}_2 \left(\mu_{02} - s_y^2\right) ; \tag{13}$$

alternatively,

$$\hat{\alpha}_{10}^{*} = a_{10} + \hat{b}_1^{*} (\alpha_{01} - a_{01}) + \hat{b}_2^{*} \left(\mu_{02} - s_y^2\right) . \tag{14}$$

To obtain an estimator of $V(\hat{\alpha}_{10})$, we write the consistent estimate of $V(a_{10})$ as

$$\hat{V}(a_{10}) = \frac{\hat{\mu}_{20}}{n} = \frac{m_{20}}{n}. \tag{15}$$

Alternatively, the unbiased estimator of $V(a_{10})$ is given by

$$\hat{V}_u(a_{10}) = \frac{m_{20}}{n-1}. \tag{16}$$

Also,

$$V(a_{01}) = \frac{\mu_{02}}{n}, \quad \text{from (2.7.i)}, \tag{17}$$

$$V\left(s_y^2\right) = \frac{\mu_{04}}{n} - \frac{(n-3)\,\mu_{02}^2}{n(n-1)}, \quad \text{from [2, p. 99]}, \tag{18}$$

and

$$Cov\left(a_{01}, s_y^2\right) = \frac{\mu_{03}}{n} + \frac{2(n-2)\,\alpha_{03}}{nN^2}, \quad \text{from (2.7.ii)}. \tag{19}$$

Using (5) (or (7)), (6) (or (8)), (9) (or (11)), (10) (or (12)), (15) (or (16)), (17), (18) and (19) one can easily get the estimate of $V(\hat{\alpha}_{10})$. Thus the problem of making the estimator (1) practicable is solved.

Other generalisations to univariate auxiliary variable regression estimators, and for other population parameters, are straightforward.

4. Proof of Theorem 2.1

(2.1.i) For the population of size N, we have

$$\begin{aligned}
E\left(a_{i0}a_{0j}\right) &= E\left[\left(\frac{1}{n}\sum_{k=1}^{N} x_k^i e_k\right)\left(\frac{1}{n}\sum_{h=1}^{N} y_h^j e_h\right)\right] \\
&= \frac{1}{n^2} E\left(\sum_{k=1}^{N} x_k^i y_k^j e_k^2 + \sum_{k=1}^{N}\sum_{h\neq k}^{N} x_k^i y_h^j e_k e_h\right) \\
&= \frac{1}{n^2}\left[N\alpha_{ij}\frac{n(N-1)+n^2}{N^2} + \left(N^2\alpha_{i0}\alpha_{0j} - N\alpha_{ij}\right)\frac{n(n-1)}{N^2}\right] \\
&= \frac{\alpha_{ij} + (n-1)\,\alpha_{i0}\alpha_{0j}}{n},
\end{aligned}$$

where $(e_1, e_2, ..., e_N)$ follows the multinomial distribution of n trials and equal probability $1/N$ of success. Thus,

$$Cov\left(a_{i0}, a_{0j}\right) = E\left(a_{i0}a_{0j}\right) - E\left(a_{i0}\right)E\left(a_{0j}\right)$$
$$= \frac{\alpha_{ij} + (n-1)\,\alpha_{i0}\alpha_{0j}}{n} - \alpha_{i0}\alpha_{0j} = \frac{\alpha_{ij} - \alpha_{i0}\alpha_{0j}}{n}.$$

This proves (2.1.i).

(2.1.ii) Since

$$Cov\left(a_{i0}, m_{02}\right) = Cov\left(a_{i0}, a_{02}\right) - Cov\left(a_{i0}, a_{01}^2\right),$$

we have from (2.1.i) that

$$Cov\left(a_{i0}, a_{02}\right) = \left(\alpha_{i2} - \alpha_{i0}\alpha_{02}\right)/n.$$

The covariance between a_{i0} and a_{01}^2 can be expressed as

$$Cov\left(a_{i0}, a_{01}^2\right) = E\left(a_{i0}a_{01}^2\right) - E\left(a_{i0}\right)E\left(a_{01}^2\right),$$

where

$$E\left(a_{i0}a_{01}^2\right) = E\left(\frac{1}{n^3}\sum_{k=1}^{N}\sum_{h=1}^{N}\sum_{m=1}^{N} x_k^i y_h y_m e_k e_h e_m\right)$$
$$= \frac{1}{n^3}\left[\sum_{k=1}^{N} x_k^i y_k^2 E\left(e_k^3\right) + 2\sum_{k=1}^{N}\sum_{h\neq k}^{N} x_k^i y_k y_h E\left(e_k^2 e_h\right)\right.$$
$$\left. + \sum_{k=1}^{N}\sum_{h\neq k}^{N} x_k^i y_h^2 E\left(e_k e_h^2\right) + \sum_{k=1}^{N}\sum_{h\neq k}^{N}\sum_{\substack{m\neq k\\ m\neq h}}^{N} x_k^i y_h y_m E\left(e_k e_h e_m\right)\right]$$
$$= \alpha_{i2}\frac{N^2 - 2(n-1)(n-2)}{n^2N^2} + \alpha_{i1}\alpha_{01}\frac{2(n-1)}{n^2}$$
$$+ \alpha_{i0}\alpha_{02}\frac{n-1}{n^2} + \alpha_{i0}\alpha_{01}^2\frac{(n-1)(n-2)}{n^2},$$

$$E\left(a_{i0}\right) = \alpha_{i0},$$

and

$$E\left(a_{01}^2\right) = V\left(a_{10}\right) + \left[E\left(a_{01}\right)\right]^2 = \frac{\alpha_{02} - \alpha_{01}^2}{n} + \alpha_{01}^2.$$

Thus

$$Cov\left(a_{i0}, m_{02}\right) = \frac{n-1}{n^2}\left(\alpha_{i2} - \alpha_{i0}\alpha_{02} - 2\alpha_{01}\alpha_{i1} + 2\alpha_{i0}\alpha_{01}^2\right) + \frac{2(n-1)(n-2)}{n^2N^2}\alpha_{i2},$$

which proves (2.1.ii).

ACKNOWLEDGEMENT

The authors are grateful to the referee for helpful suggestions and comments.

REFERENCES

[1] H. Cramér, *Métodos matemáticos de estadística*, Aguilar, Madrid, 1953. (Transl. *Mathematical methods of statistics*, Princeton University Press, Princeton, 1953).
[2] M.H. Hansen, W.N. Hurwitz and W.G. Madow, *Sample survey methods and theory*, vol. 2, John Wiley, New York, 1953.
[3] I. Olkin, Multivariate ratio estimation for finite populations, *Biometrika* **45** (1958), 154–65.
[4] M. Ruiz-Espejo, *Basic sampling for applied sciences*, Academy of Sciences, Zaragoza, 1996.
[5] M. Ruiz-Espejo, Covariance of sample moments for finite populations, *Proceedings of the Royal Irish Academy* **97**A (1997), 163–7.
[6] M. Ruiz-Espejo, Sampling: disjoint and lapelled groups, *Investigación y Ciencia* **270** (1999), 42–4.