Jour. Ind. Soc. Ag. Statistics
48(2), 1996 : 141–149

Improved Estimator of Finite Population Correlation Coefficient

Sarjinder Singh*, N.S. Mangat¤ and Jai P. Gupta*
Punjab Agricultural University, Ludhiana - 141 004
(Received : September, 1990)

SUMMARY

In the present investigation, an improved estimator, r_s, of the finite population correlation coefficient is suggested. A numerical example is also given.

Key words : Correlation coefficient, Finite population, Simple random sampling, Mean square error.

1. Introduction

The correlation coefficient, ρ, was first introduced by Bravais [1] and an estimate 'r' of this parameter from n pair of observations from a bivariate normal population was proposed by Pearson [9]. Wakimoto [18] has studied the behaviour of 'r' under stratified random sampling for continuous populations. Gupta, Singh and Lal [3,4] have studied the behaviour of 'r' under simple and stratified random sampling respectively for finite populations. Gupta and Singh [5], Kumar [8] and Chand [2] have studied the behaviour of 'r' under PPSWR, two stage and cluster sampling respectively. Rana [12] has also studied the behaviour of 'r' under simple random sampling.

Srivastava and Jhajj [15] have proposed a general class of estimators for estimating finite population correlation coefficient, ρ, using prior information $\bar{X}$ and S_x^2 of the auxiliary variable. The estimators belonging to the class suggested by Srivastava and Jhajj [15] takes the inadmissible value, that is, the value of their estimators may or may not lie in the interval (-1, +1) for a given sample.

In the present paper, we have proposed an admissible estimator, 'r_s', of finite population correlation coefficient, ρ, by following Prasad [10] and Prasad and Singh [11].

* Department of Mathematics and Statistics
¤ Department of Plant Breeding

2. *Notations and Expectations*

Assume that a simple random sample of size n is drawn from the given population of size N. Let the values of the variable y and x be denoted by Y_i and X_i as usual for the ith unit of the population i=1,2,...N and by y_i and x_i for the ith unit in the sample, i=1,2,...,n. Let us denote by

$$\bar{y} = n^{-1} \sum_{i=1}^{n} y_i, \qquad \bar{x} = n^{-1} \sum_{i=1}^{n} x_i, \quad \bar{Y} = N^{-1} \sum_{i=1}^{N} Y_i$$

$$\bar{X} = N^{-1} \sum_{i=1}^{N} X_i$$

$$\hat{U}_{pq} = (n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^p \; (x_i - \bar{x})^q$$

$$U_{pq} = (N-1)^{-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^p \; (X_i - \bar{X})^q$$

$$\rho = U_{11}/(U_{02}\, U_{20})^{1/2}, \qquad r = \hat{U}_{11}/(\hat{U}_{02}\; \hat{U}_{20})^{1/2}$$

For simplicity assume that population size N is quite large as compared to sample size n so that the finite population correction terms are ignored throughout:

Writing

$$\delta = (\hat{U}_{02}/U_{02}) - 1, \; \epsilon = (\hat{U}_{20}/U_{20}) - 1, \; \phi = (\hat{U}_{11}/U_{11}) - 1$$

such that

$$E(\delta) = E(\epsilon) = E(\phi) = 0$$

and

$$E(\delta^2) = n^{-1} \left(\frac{U_{04}}{U_{02}^2} - 1 \right) \qquad E(\epsilon^2) = n^{-1} \left(\frac{U_{40}}{U_{20}^2} - 1 \right)$$

$$E(\phi^2) = n^{-1} \left(\frac{U_{22}}{U_{11}^2} - 1 \right) \qquad E(\delta\, \phi) = n^{-1} \left(\frac{U_{13}}{U_{02}\, U_{11}} - 1 \right)$$

$$E(\in \phi) = n^{-1}\left(\frac{U_{31}}{U_{20}\,U_{11}} - 1\right) \qquad E(\in \delta) = n^{-1}\left(\frac{U_{22}}{U_{20}\,U_{02}} - 1\right)$$

These expected values may easily be obtained by following Sukhatme *et al.* [17]. The variance of 'r' (up to terms of order, n^{-1}) is given by :

$$\text{Var}(r) = n^{-1}\rho^2\left[\frac{U_{22}}{U_{11}^2} + \frac{1}{4}\left(\frac{U_{04}}{U_{02}^2} + \frac{U_{40}}{U_{20}^2}\right) - \frac{1}{2}\left(\frac{U_{13}}{U_{02}\,U_{11}} + \frac{U_{31}}{U_{20}\,U_{11}}\right) + \frac{U_{22}}{2U_{20}\,U_{02}}\right] \quad ...(2.1)$$

3. *Proposed Estimator*

We have proposed an estimator, r_s of finite population correlation coefficient ρ as:

$$r_s = \frac{\hat{U}_{11}^*}{(\hat{U}_{20}\;\hat{U}_{02})^{1/2}} \quad ...(3.1)$$

where, $\hat{U}_{11}^* = K\hat{U}_{11}$ and K is some constant. We shall use the following results to find the mean square error of, r_s defined at (3.1).

Result I : We have

$$V(\hat{U}_{11}^*) = E[\hat{U}_{11}^* - U_{11}]^2 = E[KU_{11}(1+\phi) - U_{11}]^2$$

$$= U_{11}^2\left[(K-1)^2 + n^{-1}K^2\left(\frac{U_{22}}{U_{11}^2} - 1\right)\right] \quad ...(3.2)$$

Variance of $\hat{U}_{11}^*$ is minimized for

$$K = 1\Big/\left[1 + n^{-1}\left(\frac{U_{22}}{U_{11}^2} - 1\right)\right] \quad ...(3.3)$$

Minimum variance of $\hat{U}_{11}^*$ is given by :

$$\text{Min. Var.}(\hat{U}_{11}^*) = \frac{n^{-1}U_{11}^2\left(\frac{U_{22}}{U_{11}^2} - 1\right)}{1 + n^{-1}\left(\frac{U_{22}}{U_{11}^2} - 1\right)} = n^{-1}KU_{11}^2\left(\frac{U_{22}}{U_{11}^2} - 1\right) \quad ...(3.4)$$

Result II :

$$V[(\hat{U}_{20}\hat{U}_{02})^{1/2}] = E[(\hat{U}_{20}\hat{U}_{02})^{1/2} - (U_{20}U_{02})^{1/2}]^2$$

$$= (4n)^{-1} U_{20}U_{02}\left[\frac{U_{04}}{U_{02}^2} + \frac{U_{40}}{U_{20}^2} + \frac{2U_{22}}{U_{20}U_{02}}\right] \quad ...(3.5)$$

Result III :

$$\text{Cov}[\hat{U}_{11}^*, (\hat{U}_{20}\hat{U}_{02})^{1/2}] = K\,\text{Cov}[\hat{U}_{11}, (\hat{U}_{20}\hat{U}_{02})^{1/2}]$$

$$= K[E\{\hat{U}_{11}(\hat{U}_{20}\hat{U}_{02})^{1/2}\} - E(\hat{U}_{11})E(\hat{U}_{20}\hat{U}_{02})^{1/2}]$$

$$= (2n)^{-1} KU_{11}\ (U_{20}U_{02})^{1/2}\left[\frac{U_{13}}{U_{02}U_{11}} + \frac{U_{31}}{U_{20}U_{11}} - 2\right] \quad ...(3.6)$$

4. *Variance of the Proposed Estimator*

The variance of the proposed estimator, r_s, at (3.1) is given by

$$\text{Var}(r_s) = E[r_s - \rho]^2 = E\left[\frac{\hat{U}_{11}^*}{(\hat{U}_{20}\hat{U}_{02})^{1/2}} - \rho\right]^2$$

$$= \frac{E[\hat{U}_{11}^* - \rho(\hat{U}_{20}\hat{U}_{02})^{1/2}]^2}{(\hat{U}_{20}\hat{U}_{02})}$$

$$= E\frac{[\hat{U}_{11}^* - \rho(\hat{U}_{20}\hat{U}_{02})^{1/2}]^2}{U_{20}U_{02}\left[1 + \frac{\hat{U}_{20} - U_{20}}{U_{20}}\right]\left[1 + \frac{\hat{U}_{02} - U_{02}}{U_{02}}\right]}$$

$$= E\frac{[\hat{U}_{11}^* - \rho(\hat{U}_{20}\hat{U}_{02})^{1/2}]^2}{U_{20}U_{02}}\left[1 - \left(\frac{\hat{U}_{20} - U_{20}}{U_{20}}\right) - \left(\frac{\hat{U}_{02} - U_{02}}{U_{02}}\right) + O(\in^2)\right] \quad ...(4.1)$$

Neglecting higher order terms at (4.1), we get

$$\text{Var}(r_s) = E\frac{[\hat{U}_{11}^* - (\hat{U}_{20}\hat{U}_{02})^{1/2}]^2}{U_{20}U_{02}}$$

$$= (U_{20}\,U_{02})^{-1}\ E[(\hat{U}^*_{11} - U_{11}) - \rho\ \{(\hat{U}_{20}\,\hat{U}_{02})^{1/2} - (U_{20}\,U_{02})^{1/2}\}]^2$$

$$= (U_{20}\,U_{02})^{-1}\ [E\,(\hat{U}^*_{11} - U_{11})^2 + \rho^2\,E\ \{(\hat{U}_{20}\,\hat{U}_{02})^{1/2} - (U_{20}\,U_{02})^{1/2}\}^2$$

$$- 2\rho E\,(\hat{U}^*_{11} - U_{11})\ \{(\hat{U}_{20}\,\hat{U}_{02})^{1/2} - (U_{20}\,U_{02})^{1/2}\}]$$

$$= (U_{20}\,U_{02})^{-1}\ [V\,(\hat{U}^*_{11}) + \rho^2\,V\{(\hat{U}_{20}\,\hat{U}_{02})^{1/2}\} - 2\rho\ Cov\ \{\hat{U}^*_{11}, (\hat{U}_{20}\,\hat{U}_{02})^{1/2}\}] \quad ...(4.2)$$

Using the results mentioned at (3.4), (3.5) and (3.6) in the relation (4.2), we get

$$Var\,(r_s) = n^{-1}\,\rho^2\left[K\left(\frac{U_{22}}{U_{11}^2} - \frac{U_{13}}{2U_{02}\,U_{11}} - \frac{U_{31}}{2U_{20}\,U_{11}}\right) + \frac{1}{4}\left(\frac{U_{04}}{U_{02}^2} + \frac{U_{40}}{U_{20}^2}\right) + \frac{U_{22}}{2U_{20}\,U_{02}}\right] \quad ...(4.3)$$

Relation (2.1) and (4.3) shows that, $Var(r_s) \le Var\,(r)$ if

$$U_{22} \ge \frac{U_{11}}{2}\left(\frac{U_{31}}{U_{20}} + \frac{U_{13}}{U_{02}}\right) = L,\ \ (say) \quad ...(4.4)$$

5. *Bias of the Proposed Estimator*

One can easily see that,

$$B(r_s) = (K-1)\,\rho + KB\,(r) \quad ...(5.1)$$

where $B\,(r_s)$ and $B(r)$ denote the bias in the estimators r_s and r respectively. Thus we have to show that

$$|\,B\,(r_s)\,| \le |\,B\,(r)\,| \quad ...(5.2)$$

Now we will take the following cases :

Case-I. Both ρ and $B(r)$ are positive; then (5.2) will hold if

$$\rho + B(r) \ge 0, \text{ because } K \le 1 \text{ (always)} \quad ...(5.3)$$

which is always true.

Case-II. When ρ is positive and B(r) is negative;

Condition (5.2) will hold if

$$\rho \geq |B(r)| \qquad ...(5.4)$$

Case-III. When ρ is negative and B(r) is positive;

Condition (5.2) will hold if

$$B(r) \geq |\rho| \qquad ...(5.5)$$

Case-IV. Both ρ and B(r) are negative;

The condition (5.2) will not hold. Proposed estimator r_s will be more biased than the usual one. These conditions will be quite helpful to the investigator to decide the situation where the proposed estimator can safely be used.

6. *On Optimum Choice of K*

The optimum choice of K given at (3.3) depends on unknown population parameters U_{22} and U_{11} and hence restricts the practical use of the proposed estimator. In this case we will discuss two cases:

Case-I. If the population is bivariate normal, then following Kendall and Stuart [7] we have

$$U_{22} = \rho\,(U_{20}\,U_{02})^{1/2} \text{ and } U_{11} = (1 + 2\rho^2)\; U_{20}\,U_{02}$$

Thus the optimum value of K becomes

$$K = 1/[1 + n^{-1}\,\{(1 + 2\,\rho^2)/\rho^2\}] \qquad ...(6.1)$$

and its estimator is given by

$$\hat{K} = 1/[1 + n^{-1}\,\{\,(1 + 2r^2)/r^2\}] \qquad ...(6.2)$$

Case-II. One can also replace U_{22} and U_{11} by their consistent or unbiased estimates to get an estimator of K given by

$$\hat{K} = 1/[1 + n^{-1}\,(\frac{\hat{U}_{22}}{\hat{U}_{11}^{2}} - 1)\,] \qquad ...(6.3)$$

It is shown by several research workers viz. Srivastava and Jhajj [14], Sampth [13] and Prasad and Singh [11] that if we replace K by its estimator

$\hat{k}$, then the asymptotic mean square error of the proposed estimator remains same. These results hold equally good for the proposed estimator in the present paper.

For the purpose of numerical illustration, we have considered the population described below and observe that condition (4.4) is satisfied.

Table 1.

Population	Horvitz and Thompson [6]	Sukhatme and Sukhatme [16]
Y_i	No. of house-holds on the ith block	No. of banana bunches
X_i	Eye estimated no. of house holds in ith block	No. of banana pits
ρ	0.8662	0.7737
U_{11}	64.6400	18896.3000
U_{13}	10554.3800	759797648.2000
U_{31}	15478.6700	
U_{02}	61.8000	2746278278.0000
U_{20}	90.1200	13433.4000
N	20.0000	20.0000
U_{22}	11261.8300	1446954322.0000
L	11070.8600	1061772488.0000

ACKNOWLEDGEMENTS

The authors are thankful to the referee for his constructive comments to bring the original manuscript in the present form. They are also thankful to Dr. Ravindra Singh, Professor of Statistics, Deptt. of Math. and Stat. for his keen interest during the revision of original manuscript.

REFERENCES

[1] Bravais, A., 1846. Analyse mathematique sur les probabilities des erreurs de situation d'um point; Memories presents par divers savants. *l'A-academie Royal des Sciences de l'Institute de France*, **9**, 255-332.

[2] Chand, P., 1989. Study of the usual correlation coefficient estimator in stratified sub-sampling. M.Sc. Thesis submitted to Punjab Agricultural University, Ludhiana.

[3] Gupta, J.P., Singh, R. and Lal, B., 1978. On the estimation of the finite population correlation coefficient-I. *Sankhya*, **40**, Ser. C, Pt.1, 38-59.

[4] Gupta, J.P., Singh, R. and Lal, B., 1979. On the estimation of the finite population correlation coefficient-II. *Sankhya*, **41**, Ser. C, Pt.1, 1-39.

[5] Gupta, J.P. and Singh, R., 1989. Usual correlation coefficient in PPSWR sampling. *Jour. of Indian Statistical Association*, **27**, 13-16.

[6] Horvitz, D.G. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *Jour. Amer. Stat. Assoc.*, **47**, 260, 663-685.

[7] Kendall, M.G. and Stuart, A., 1952. The advanced theory of statistics: Distribution theory. Vol. 1, Fifth Edition, Butler and Tanner Ltd., Frome and London. pp. 83.

[8] Kumar, P., 1989. Study of the usual correlation coefficient estimator in two-stage sampling. M.Sc. Thesis submitted to the Punjab Agricultural University, Ludhiana.

[9] Pearson, K., 1896. Mathematical contribution to the theory of evolution-III-regression; heridity and panmixia. Phil. Trans. (A) *Royal Soc. London*, **187**, 253-318.

[10] Prasad, B., 1989. Some improved ratio type estimators of population mean and ratio in finite population sample surveys. *Commun. Statist. Theory Meth.*, **18** (1), 379-392.

[11] Prasad, B. and Singh, H.P., 1990. Some improved ratio-type estimators of finite population variance in sample surveys. *Commun. Statist., Theory Meth.*, **19**, 1127-1139.

[12] Rana, R.S., 1989. Concise estimator of bias and variance of the finite population correlation coefficient. *Jour. Indian Soc. Agril. Statist.*, **41**, (1), 69-76.

[13] Sampth, S., 1989. On the optimal choice of unknowns in ratio-type estimators. *J. Indian Soc. Agric. Statist.*, **41**, 166-172.

[14] Srivastava, S.K. and Jhajj, H.S., 1983. A class of estimators of the population mean using multi-auxiliary information. *Calcutta Statistical Assoc. Bull.*, **32**, 47-56.

[15] Srivastava, S.K. and Jhajj, H.S., 1986. On the estimation of finite population correlation coefficient. *Jour. Indian Soc. Agril. Statist.*, **38** **(1)**, 82-91.

[16] Sukhatme, P.V. and Sukhatme, B.V., 1970. Sampling theory of surveys with applications. Asia Publishing House, Second revised edition, 166.

[17] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Ashok, C., 1984. Sampling theory of surveys with applications. Indian Soc. of Agril. Statist. New Delhi, India and Iowa State University Press, Iowa, USA.

[18] Wakimoto, K., 1971. Stratified random sampling (III) : Estimation of the correlation coefficient. *Ann. Inst. Statist. Math.*, **23**, 339-355.