Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 69(3) 2015 257-262

An Estimator of the Correlation Coefficient in Probability Proportional to Size without Replacement Sampling

Jai P. Gupta[1] and P.K. Mahajan[2]

[1]*32/3 Nehru Road, Civil Lines, Meerut, India*

[2]*Dr. Y.S. Parmar University of Horticulture and Forestry, Nauni, Solan*

Received 19 January 2015; Accepted 19 December 2015

SUMMARY

The present paper deals with the problem of estimating the correlation coefficient for finite population in case of Probability Proportional to Size without Replacement Sampling. Asymptotic expressions for the bias, an upper limit of the bias, variance and an estimate of variance of the proposed estimator for finite population correlation ρ have been obtained. An empirical investigation has also been made.

Keywords: Correlation coefficient, Probability proportional to size without replacement sampling, Midzuno's scheme of sampling, mean square error

1. INTRODUCTION

The correlation coefficient ρ was first introduced by Bravais (1846) and an estimate

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1.1}$$

of this parameter from *n* pairs of observations from a bi-variate normal population was advocated by Pearson (1896). This estimator *r* has only been studied for infinite bi-variate populations. It’s variance,

$$V(r) = \frac{\rho^2}{n}\left[\frac{\mu_{22}}{\mu_{11}^2} + \frac{1}{2}\frac{\mu_{22}}{\mu_{20}\cdot\mu_{02}} + \frac{1}{4}\frac{\mu_{40}}{\mu_{20}^2} + \frac{1}{4}\frac{\mu_{04}}{\mu_{02}^2} - \frac{\mu_{31}}{\mu_{11}\cdot\mu_{20}} - \frac{\mu_{13}}{\mu_{11}\cdot\mu_{02}}\right] \tag{1.2}$$

where μ's are the central moments of the joint distribution of the variables x and y was obtained by Karl Pearson and reported by Kendall (1958). When the joint distribution of (*x, y*) is bi-variate normal, the variance of the estimator r, for large n, becomes approximately $\frac{(1-p^2)^2}{n-1}$ (Fisher, 1915 and 1921). In socio-economic surveys, there is an extensive use of sampling methods, where normal or infinite populations are not generally available; thus a study of the correlation coefficient for finite population is thus needed. The estimation procedure of correlation coefficient for finite population in case of some sampling methods is available in literature. For reference see Wakimoto (1971), Gupta *et al.* (1978, 1979, 1993), Gupta and Singh (1989, 1990) and Gupta (2002).

The estimation of the correlation coefficient in probability proportional to size without replacement sampling (PPSWOR) has not received due attention so far. Keeping in view the importance of this sampling scheme in field surveys, the present paper considers the estimation of the population correlation

Corresponding author: P.K. Mahajan
E-mail address: pawan_uhf@yahoo.com

coefficient ρ when the units in the sample are selected with unequal initial probabilities [p_i, $\sum p_i=1$] and the probability of drawing a specified unit of the population at a given draw changes with the draw. The paper proposes an estimator of correlation coefficient (r_1) along with the expressions for bias, variance and estimate of the variance for the estimator. The numerical illustration has also been made using a real life situation.

Let the units in the given finite population be denoted by $U_1, U_2,..,U_N$ and a sample of size n is taken with PPSWOR sampling and the measurements on variables X and Y are recorded. In what follows, we define the following

$$t_i = \begin{cases} 1 & \text{if } U_i \text{ is included in the sample} \\ 0 & \text{otherwise} \end{cases}$$

Obviously, $E(t_i) = \pi_i$ (the prob. that U_i is selected) so that $E(t_i^2) = (t_i^3) = (t_i^4) = \pi_i$

$E(t_i t_j) = \pi_{ij}$ (the prob. that U_i and U_j both occur in the sample);

$E(t_i t_j t_k) = \pi_{ijk}$ (the prob. that U_i, U_j and U_k occur in the sample) and

$E(t_i t_j t_k t_l) = \pi_{ijkl}$ (the prob. that U_i, U_j, U_k and U_1 are included in the sample).

We shall also observe the following notations throughout the paper:

$$\sum_1 \text{for} \sum_{i=1}^{N}; \sum_2 \text{for} \sum_{i\neq j=1}^{N}; \sum_3 \text{for} \sum_{i\neq j\neq k=1}^{N}; \sum_4 \text{for} \sum_{i\neq j\neq k\neq l=1}^{N}; \text{ and } \sum' \text{ for}$$

sample counterparts.

2. THE PROPOSED ESTIMATOR

Let (x_1, y_1), (x_2, y_2), ⋯⋯,(x_n, y_n) be n pairs of observations on the variables X and Y for a sample of size n, drawn with PPSWOR sampling, from a population of size N. The proposed estimator r_1 of ρ is defined as

$$r_1 = \frac{\hat\theta_{11}}{\sqrt{\hat\theta_{21}.\hat\theta_{31}}} \tag{2.1}$$

where $\hat\theta_{11} = \frac{(N-1)}{N}\sum_{i=1}^{N}\frac{X_iY_it_i}{\pi_i} - \frac{1}{N}\sum_{i\neq j=1}^{N}\frac{X_iY_jt_it_j}{\pi_{ij}}$,

$$\hat\theta_{21} = \frac{(N-1)}{N}\sum_{i=1}^{N}\frac{X_i^2t_i}{\pi_i} - \frac{1}{N}\sum_{i\neq j=1}^{N}\frac{X_iX_jt_it_j}{\pi_{ij}} \tag{2.2}$$

and $\hat\theta_{31} = \frac{(N-1)}{N}\sum_{i=1}^{N}\frac{Y_i^2t_i}{\pi_i} - \frac{1}{N}\sum_{i\neq j=1}^{N}\frac{Y_iY_jt_it_j}{\pi_{ij}}$,

let $E(\hat\theta_{11}) = \theta_{11}$, $E(\hat\theta_{21}) = \theta_{21}$ and $E(\hat\theta_{31}) = \theta_{31}$, so that $\hat\theta_{11} = \theta_{11} + \varepsilon_{11}$,

$\hat\theta_{21} = \theta_{21} + \varepsilon_{21}$ and $\hat\theta_{31} = \theta_{31} + \varepsilon_{31}$. Hence $E(\varepsilon_{11}) = 0$, $E(\varepsilon_{21}) = 0$, $E(\varepsilon_{31}) = 0$,

$V(\varepsilon_{11}) = E(\varepsilon_{11}^2)$, $V(\varepsilon_{21}) = E(\varepsilon_{21}^2)$, $V(\varepsilon_{31}) = E(\varepsilon_{31}^2)$, $cov(\varepsilon_{11}, \varepsilon_{21}) = E(\varepsilon_{11}\,\varepsilon_{21})$,

$cov(\varepsilon_{11}, \varepsilon_{31}) = E(\varepsilon_{11}\varepsilon_{31})$ and $cov(\varepsilon_{21}, \varepsilon_{31}) = E(\varepsilon_{21}\,\varepsilon_{31})$.

With these notations, we get the following results for the case of PPSWOR sampling. These results are obtained after straightforward but lengthy algebraic simplifications and thus the proofs for the same have been omitted in this paper.

Lemma 1: For PPSWOR sampling

The expected value of $\hat\theta_{11}$ is given by E ($\hat\theta_{11}$) = θ_{11}= $(N-1)\,S_{xy}$

The expected value of $\hat\theta_{21}$ is given by E ($\hat\theta_{21}$) = θ_{21}= $(N-1)\,S_x^2$

Lemma 2: The variance of ε_{11}, in case of PPSWOR sampling, is given by

$$V(\varepsilon_{11}) = E(\varepsilon_{11}^2) = E\left\{\left(\frac{N-1}{N}\sum_1\frac{X_iY_it_i}{\pi_i} - \frac{1}{N}\sum_2\frac{X_iY_it_it_j}{\pi_{ij}}\right)^2\right\} - \theta_{11}^2$$

$$= \left(1-\frac{1}{N}\right)^2 \sum_1 \frac{X_i^2 Y_i^2}{\pi_i} + \frac{1}{N^2}\sum_2 \frac{X_i^2 Y_i^2}{\pi_{ij}} + \sum_2 X_i X_j Y_i Y_j \left\{\left(1-\frac{1}{N}\right)^2 \frac{\pi_{ij}}{\pi_i \pi_j} + \frac{1}{N^2}.\frac{1}{\pi_{ij}}\right\} - 2\frac{N-1}{N^2}\sum_2 \frac{X_i X_j Y_i^2}{\pi_i} + \frac{1}{N^2}\sum_3 X_i X_j Y_k^2 \frac{\pi_{ijk}}{\pi_{ik}\pi_{jk}} - 2\frac{N-1}{N^2}\sum_2 \frac{X_i^2 Y_i Y_j}{\pi_i} + \frac{1}{N^2}\sum_3 X_i^2 Y_j Y_k \frac{\pi_{ijk}}{\pi_{ik}\pi_{ij}} + \sum_3 X_i X_j Y_i Y_k \pi_{ijk}\left(\frac{2}{N^2}.\frac{1}{\pi_{ik}\pi_{ij}} - \frac{2(N-1)}{N^2}.\frac{1}{\pi_i \pi_{jk}}\right) + \frac{1}{N^2}\sum_4 X_i X_j Y_k Y_l \frac{\pi_{ijkl}}{\pi_{ik}\pi_{jl}} - \theta_{11}^2$$

Corollary 1: In case of PPSWOR sampling $V(\varepsilon_{21})$ can be put as

$$V(\varepsilon_{21}) = \left(\frac{N-1}{N}\right)^2 \sum_1 \frac{X_i^4}{\pi_i} + \sum_2 X_i^2 X_j^2 \left(\left(\frac{N-1}{N}\right)^2 \frac{\pi_{ij}}{\pi_i \pi_j} + \frac{2}{N^2}.\frac{1}{\pi_{ij}}\right) - \frac{4(N-1)}{N^2}\sum_2 \frac{X_i^3 X_j}{\pi_i} + \frac{1}{N^2}\sum_4 X_i X_j X_k X_l \frac{\pi_{ijkl}}{\pi_{ij}\pi_{kl}} + \sum_3 X_i^2 X_j X_k \frac{\pi_{ijk}}{\pi_i \pi_j \pi_k}\left(\frac{4}{N^2}\frac{\pi_i \pi_j \pi_k}{\pi_{ij}\pi_{ik}} - \frac{2(N-1)}{N^2}\frac{\pi_k \pi_j}{\pi_{kj}}\right) - \theta_{21}^2 \qquad (2.3)$$

where $\theta_{21} = (N-1)S_X^2$.

The variance of ε_{31} is obtained from (2.3) by replacing X by Y

Lemma 3: The covariance between ε_{11} and ε_{21} is given by

$$\mathrm{Cov}(\varepsilon_{11}, \varepsilon_{21}) = E(\varepsilon_{11}\varepsilon_{21}) = \left(\frac{N-1}{N}\right)^2 \sum_1 \frac{X_i^3 Y_i}{\pi_i} - \sum_2 \frac{X_i^2 X_j Y_j \pi_{ij}}{\pi_i \pi_j}\left\{\frac{(N-1)^2}{N^2} + \frac{2\pi_i \pi_j}{N^2 \pi_{ij}^2}\right\} + \sum_3 X_i^2 X_j Y_k \frac{\pi_{ijk}}{\pi_i \pi_j \pi_k}\left\{\frac{1-N}{N^2}.\frac{\pi_j \pi_k}{\pi_{jk}} + \frac{2}{N^2}.\frac{\pi_i \pi_j \pi_k}{\pi_{ij}\pi_{ik}}\right\} + \frac{3(1-N)}{N^2}\sum_2 \frac{X_i^2 X_j Y_i}{\pi_i} + \sum_3 X_i X_j X_k Y_k \pi_{ijk}\left(\frac{2}{N^2}.\frac{1}{\pi_{ik}\pi_{jk}} + \frac{1-N}{N^2}.\frac{1}{\pi_k \pi_{ij}}\right) + \frac{1}{N^2}\sum_4 X_i X_j X_k Y_l \frac{\pi_{ijkl}}{\pi_{ij}\pi_{kl}} - \theta_{11}\theta_{21}$$

Corollary 2: The covariance between ε_{11} and ε_{31} is given by

$$\mathrm{Cov}(\varepsilon_{11}, \varepsilon_{31}) = E(\varepsilon_{11}\varepsilon_{31}) = \left(\frac{N-1}{N}\right)^2 \sum_1 \frac{Y_i^3 X_i}{\pi_i} - \frac{1-N}{N^2}\sum_2 \frac{Y_i^3 X_j}{\pi_i} + \sum_2 Y_i^2 Y_j X_j \frac{\pi_{ij}}{\pi_i \pi_j}\left\{\left(\frac{N-1}{N}\right)^2 + \frac{2}{N^2}.\frac{\pi_i \pi_j}{\pi_{ij}^2}\right\} + \sum_3 Y_i^2 Y_j X_k \frac{\pi_{ijk}}{\pi_i \pi_j \pi_k}\left\{\frac{1-N}{N^2}.\frac{\pi_j \pi_k}{\pi_{jk}} + \frac{2}{N^2}.\frac{\pi_i \pi_j \pi_k}{\pi_{ij}\pi_{ik}}\right\} + \frac{3(1-N)}{N^2}\sum_2 \frac{Y_i^2 Y_j X_i}{\pi_i} + \sum_3 Y_i Y_j Y_k X_k \pi_{ijk}\left(\frac{1-N}{N^2}.\frac{1}{\pi_k \pi_{ij}} + \frac{2}{N^2}.\frac{1}{\pi_{ik}\pi_{jk}}\right) + \frac{1}{N^2}\sum_4 Y_i Y_j Y_k X_l \frac{\pi_{ijkl}}{\pi_{ij}\pi_{kl}} - \theta_{11}\theta_{31}$$

Lemma 4: The covariance between ε_{21} and ε_{31} is given by

$$\mathrm{Cov}(\varepsilon_{21}, \varepsilon_{31}) = \left(\frac{N-1}{N}\right)^2 \sum_1 \frac{X_i^2 Y_i^2}{\pi_i} + \left(\frac{N-1}{N}\right)^2 \sum_2 X_i^2 Y_j^2 \frac{\pi_{ij}}{\pi_i \pi_j} - \frac{2(N-1)}{N^2}\sum_2 \frac{Y_i^2 X_i X_j}{\pi_i} - \frac{(N-1)}{N^2}\sum_3 X_i X_j Y_k^2 \frac{\pi_{ijk}}{\pi_k \pi_{ij}} - \frac{2(N-1)}{N^2}\sum_2 \frac{X_i^2 Y_i Y_j}{\pi_i} - \frac{N-1}{N^2}\sum_3 X_i^2 Y_j Y_k \frac{\pi_{ijk}}{\pi_i \pi_{jk}} + \frac{4}{N^2}\sum_3 X_i X_j Y_i Y_k \frac{\pi_{ijk}}{\pi_{ik}\pi_{ij}} + \frac{2}{N^2}\sum_2 \frac{X_i X_j Y_i Y_j}{\pi_{ij}} + \frac{1}{N^2}\sum_4 X_i X_j Y_k Y_l \frac{\pi_{ijkl}}{\pi_{ij}\pi_{kl}} - \theta_{21}\theta_{31}$$

Since $r_1 = \frac{\hat{\theta}_{11}}{\sqrt{\hat{\theta}_{21}.\hat{\theta}_{31}}}$, we have

$$E(r_1) = E\left\{\frac{\theta_{11}}{\sqrt{\theta_{21}\theta_{31}}}.\left(1+\frac{\varepsilon_{11}}{\theta_{11}}\right).\left(1+\frac{\varepsilon_{21}}{\theta_{21}}\right)^{-1/2}.\left(1+\frac{\varepsilon_{31}}{\theta_{31}}\right)^{-1/2}\right\} \qquad (2.4)$$

Assuming that the sample size is sufficiently large so that $\left|\frac{\varepsilon_{21}}{\theta_{21}}\right| < 1$, and $\left|\frac{\varepsilon_{31}}{\theta_{31}}\right| < 1$ and the binomial expansion of $\left(1+\frac{\varepsilon_{21}}{\theta_{21}}\right)^{-1/2}$ and $\left(1+\frac{\varepsilon_{31}}{\theta_{31}}\right)^{-1/2}$ as a convergent series in the powers of ε_{21} and ε_{31} can be made, one get from (2.4).

$$E(r_1) = E\left[\rho\left\{1+\frac{\varepsilon_{11}}{\theta_{11}} - \frac{\varepsilon_{21}}{2\theta_{21}} - \frac{\varepsilon_{31}}{2\theta_{31}} - \frac{\varepsilon_{11}\varepsilon_{21}}{2\theta_{11}\theta_{21}} - \frac{\varepsilon_{11}\varepsilon_{31}}{2\theta_{11}\theta_{31}} + \frac{\varepsilon_{21}\varepsilon_{31}}{4\theta_{21}\theta_{31}} + \frac{3\varepsilon_{21}^2}{8\theta_{21}^2} + \frac{3\varepsilon_{31}^2}{8\theta_{31}^2} + 0(n^{-2})\right\}\right],$$

Neglecting the terms of the order $0(n^{-2})$

$$E(r_1) \cong \rho\left[1+\frac{3E(\varepsilon_{21}^2)}{8\theta_{21}^2} - \frac{E(\varepsilon_{11}\varepsilon_{21})}{2\theta_{11}\theta_{21}} - \frac{E(\varepsilon_{11}\varepsilon_{31})}{2\theta_{11}\theta_{31}} + \frac{E(\varepsilon_{21}\varepsilon_{31})}{4\theta_{21}\theta_{31}} + \frac{3E(\varepsilon_{31}^2)}{8\theta_{31}^2}\right] \qquad (2.5)$$

Thus r_1 is not unbiased for ρ and the bias of r_1 is $B(r_1) = E(r_1) - \rho$.

3. THE VARIANCE OF THE ESTIMATOR R_1

An average measure of the divergence of the different estimates from the true value is given by the expected value of the squared error i.e. mean square error (M.S.E), which in this case is equal to $V(r_1) + \left(B(r_1)\right)^2$. The $E(r_1)$ given in (2.5) has been obtained by neglecting the terms of the order $O(n^{-2})$ while all the terms of the order $O(n^{-1})$ have been taken into account. Thus $\left(B(r_1)\right)^2$ will be of the order $O(n^{-2})$. Then to the order $O(n^{-1})$ both M.S.E. and $V(r_1)$ will be same. Although it is the general practice that for a biased estimator M.S.E. is used, but since in this case both M.S.E.(r_1) and $V(r_1)$ are same, we shall make use of $V(r_1)$, the approximate expression for which can be obtained as below:

$$V(r_1) = E(r_1^2) - \left(E(r_1)\right)^2$$

$$= E\left[\frac{\theta_{11}}{\sqrt{\theta_{21}\theta_{31}}}.\left(1+\frac{\varepsilon_{11}}{\theta_{11}}\right).\left(1+\frac{\varepsilon_{21}}{\theta_{21}}\right)^{-1/2}. \left(1+\frac{\varepsilon_{31}}{\theta_{31}}\right)^{-1/2}\right]^2$$

$$-\rho^2\left[1+\frac{3E(\varepsilon_{21}^2)}{8\theta_{21}^2}-\frac{E(\varepsilon_{11}\varepsilon_{21})}{2\theta_{11}\theta_{21}}-\frac{E(\varepsilon_{11}\varepsilon_{31})}{2\theta_{11}\theta_{31}}+ \frac{E(\varepsilon_{21}\varepsilon_{31})}{4\theta_{21}\theta_{31}}+\frac{3E(\varepsilon_{31}^2)}{8\theta_{31}^2}+0(n^{-2})\right]^2$$

Assuming that the sample size is sufficiently large so that $\left|\frac{\varepsilon_{21}}{\theta_{21}}\right| < 1$, and $\left|\frac{\varepsilon_{31}}{\theta_{31}}\right| < 1$ and the binomial expansion of $\left(1+\frac{\varepsilon_{21}}{\theta_{21}}\right)^{-1/2}$ and $\left(1+\frac{\varepsilon_{31}}{\theta_{31}}\right)^{-1/2}$ can be made as a convergent series in the powers of ε_{21} and ε_{31}, we get

$$V(r_1) \cong \rho^2 E\left[1-\frac{\varepsilon_{21}}{\theta_{21}}+\frac{2\varepsilon_{11}}{\theta_{11}}-\frac{\varepsilon_{31}}{\theta_{31}}+\frac{\varepsilon_{21}^2}{\theta_{21}^2}+ \frac{\varepsilon_{11}^2}{\theta_{11}^2}+\frac{\varepsilon_{31}^2}{\theta_{31}^2}-\frac{2\varepsilon_{11}\varepsilon_{21}}{\theta_{11}\theta_{21}}-\frac{2\varepsilon_{11}\varepsilon_{31}}{\theta_{11}\theta_{31}}+\frac{\varepsilon_{21}\varepsilon_{31}}{\theta_{21}\theta_{31}}+ 0(n^{-2})\right] - \rho^2\left[1+\frac{3E(\varepsilon_{21}^2)}{4\theta_{21}^2}-\frac{E(\varepsilon_{11}\varepsilon_{21})}{\theta_{11}\theta_{21}}- \frac{E(\varepsilon_{11}\varepsilon_{31})}{\theta_{11}\theta_{31}}+\frac{E(\varepsilon_{21}\varepsilon_{31})}{2\theta_{21}\theta_{31}}+\frac{3E(\varepsilon_{21}^2)}{4\theta_{31}^2}+0(n^{-2})\right]$$

Neglecting the terms of $0(n^{-2})$ in above expression, we get

$$V(r_1) \cong \rho^2\left[\frac{E(\varepsilon_{11}^2)}{\theta_{11}^2}+\frac{E(\varepsilon_{21}^2)}{4\theta_{21}^2}+\frac{E(\varepsilon_{31}^2)}{4\theta_{31}^2}- \frac{E(\varepsilon_{11}\varepsilon_{21})}{2\theta_{11}\theta_{21}}-\frac{E(\varepsilon_{11}\varepsilon_{31})}{2\theta_{11}\theta_{31}}+\frac{E(\varepsilon_{21}\varepsilon_{31})}{2\theta_{21}\theta_{31}}\right]$$

Now substituting the expressions for different variances and co-variances of ε's involved in above expression from Lemmas 2-4 and Corollaries 2 and 3 we can have the expression.

4. ESTIMATION OF VARIANCE V (R_1)

As variance of any estimator cannot be obtained unless all the population units are known, therefore we try to get an estimate of this variance from the sample. Also, as far as possible, an unbiased estimate is preferred. When an unbiased estimator is not possible, we go in for the biased estimators. In this case an unbiased estimator of $V(r_1)$ cannot be easily derived because of the implicit terms and cumbersome algebra. In this case an estimate $V(r_1)$ obtained by replacing the various summation terms involved in $V(r_1)$ by their unbiased estimates, is presented in theorem 1 below:

Theorem 1: In case of PPSWOR sampling, the variance $V(r_1)$ is estimated by

$$v(r_1) = r_1^2\left(\frac{A}{\hat{\theta}_{11}^2}+\frac{B}{4\hat{\theta}_{21}^2}+\frac{C}{4\hat{\theta}_{31}^2}-\frac{D}{\hat{\theta}_{11}\hat{\theta}_{21}}- \frac{E}{\hat{\theta}_{11}\hat{\theta}_{31}}+\frac{F}{2\hat{\theta}_{21}\hat{\theta}_{31}}\right),$$

where

$$\mathbf{A}=\left(\frac{N-1}{N}\right)^2\sum_1^{'}\frac{x_i^2y_i^2}{\pi_i^2}+\frac{1}{N^2}\sum_2^{'}\frac{x_i^2y_j^2}{\pi_{ij}^2}+\sum_2^{'}x_ix_jy_iy_j \left\{\left(\frac{N-1}{N}\right)^2.\frac{1}{\pi_i\pi_j}+\frac{1}{N^2}.\frac{1}{\pi_{ij}^2}\right\}-\frac{2(N-1)}{N^2}\sum_2^{'}\frac{x_ix_jy_i^2}{\pi_i\pi_{ij}}+\frac{1}{N^2}\sum_3^{'}\frac{x_ix_jy_k^2}{\pi_{ij}\pi_{jk}} -\frac{2(N-1)}{N^2}\sum_2^{'}\frac{x_i^2y_iy_j}{\pi_i\pi_{ij}}\frac{1}{N^2}\sum_3^{'}\frac{x_i^2y_iy_k}{\pi_{ik}\pi_{ij}}+\sum_3^{'}x_ix_jy_iy_k \left(\frac{2}{N^2}.\frac{1}{\pi_{ik}\pi_{ij}}-\frac{2(N-1)}{N^2}.\frac{1}{\pi_i\pi_{jk}}\right)+\frac{1}{N^2}\sum_4^{'}\frac{x_ix_jy_ky_l}{\pi_{ik}\pi_{jl}}$$

$$\mathbf{B} = \left(\frac{N-1}{N}\right)^2 \sum\nolimits_1' \frac{x_i^4}{\pi_i^2} + \sum\nolimits_2' x_i^2 x_j^2 \left(\left(\frac{N-1}{N}\right)^2 \cdot \frac{1}{\pi_i \pi_j} + \frac{2}{N^2} \cdot \frac{1}{\pi_{ij}^2}\right) - \frac{4(N-1)}{N^2} \sum\nolimits_2' \frac{x_i^3 x_j}{\pi_{ij}\pi_i} + \frac{1}{N^2} \sum\nolimits_4' \frac{x_i x_j x_k x_l}{\pi_{ij}\pi_{kl}} + \sum\nolimits_3' x_i^2 x_j x_k \left[\frac{4}{N^2}\frac{1}{\pi_{ij}\pi_{ik}} - \frac{2(N-1)}{N^2}\frac{1}{\pi_i \pi_{kj}}\right],$$

$$\mathbf{C} = \left(\frac{N-1}{N}\right)^2 \sum\nolimits_1' \frac{y_i^4}{\pi_i^2} + \sum\nolimits_2' y_i^2 y_j^2 \left(\left(\frac{N-1}{N}\right)^2 \cdot \frac{1}{\pi_i \pi_j} + \frac{2}{N^2} \cdot \frac{1}{\pi_{ij}^2}\right) - \frac{4(N-1)}{N^2} \sum\nolimits_2' \frac{y_i^3 y_j}{\pi_{ij}\pi_i} + \frac{1}{N^2} \sum\nolimits_4' \frac{y_i y_j y_k y_l}{\pi_{ij}\pi_{kl}} + \sum\nolimits_3' y_i^2 y_j y_k \left[\frac{4}{N^2}\frac{1}{\pi_{ik}\pi_{ij}} - \frac{2(N-1)}{N^2}\frac{1}{\pi_i \pi_{kj}}\right],$$

$$\mathbf{D} = \left(\frac{N-1}{N}\right)^2 \sum\nolimits_1' \frac{x_i^3 y_i}{\pi_i^2} + \frac{1-N}{N^2} \sum\nolimits_2' \frac{x_i^3 y_j}{\pi_i \pi_{ij}} + \sum\nolimits_2' x_i^2 x_j y_j \left[\frac{(N-1)^2}{N^2}\frac{1}{\pi_i\pi_j} + \frac{2}{N^2}\frac{1}{\pi_{ij}^2}\right] + \frac{3(1-N)}{N^2} \sum\nolimits_2' \frac{x_i^2 x_j y_i}{\pi_i \pi_{ij}} + \sum\nolimits_3' \frac{x_i^2 x_j y_k}{\pi_i \pi_j \pi_k}\left[\frac{1-N}{N^2}\frac{\pi_j \pi_k}{\pi_{jk}} + \frac{2}{N^2}\frac{\pi_i\pi_j\pi_k}{\pi_{ij}\pi_{ik}}\right] + \frac{1}{N^2}\sum\nolimits_4' \frac{x_i x_j x_k y_l}{\pi_{ij}\pi_{kl}} + \sum\nolimits_3' x_i x_j x_k y_k \left[\frac{1-N}{N^2}\frac{1}{\pi_k \pi_{ij}} + \frac{2}{N^2}\frac{1}{\pi_{ik}\pi_{jk}}\right],$$

$$\mathbf{E} = \left(\frac{N-1}{N}\right)^2 \sum\nolimits_1' \frac{y_i^3 x_i}{\pi_i^2} - \frac{N-1}{N^2} \sum\nolimits_2' \frac{y_i^3 x_i}{\pi_i \pi_{ij}} + \sum\nolimits_2' \frac{y_i^2 y_j x_j}{\pi_i \pi_j}\left(\left(\frac{N-1}{N}\right)^2 + \frac{2}{N^2}\frac{\pi_i \pi_j}{\pi_{ij}^2}\right) + \sum\nolimits_3' \frac{y_i^2 y_j x_k}{\pi_i \pi_j \pi_k}\left[\frac{1-N}{N^2}\frac{\pi_j\pi_k}{\pi_{jk}} + \frac{2}{N^2}\frac{\pi_i \pi_j \pi_k}{\pi_{ij}\pi_{ik}}\right] + \frac{3(1-N)}{N^2}\sum\nolimits_2' \frac{y_i^2 y_j x_i}{\pi_i \pi_{ij}} + \frac{1}{N^2}\sum\nolimits_4' \frac{y_i y_j y_k x_l}{\pi_{ij}\pi_{kl}} + \sum\nolimits_3' y_i y_j y_k x_k \left[\frac{1-N}{N^2}\frac{1}{\pi_k \pi_{ij}} + \frac{2}{N^2}\frac{1}{\pi_{ik}\pi_{jk}}\right],$$

$$\mathbf{F} = \left(\frac{N-1}{N}\right)^2 \sum\nolimits_1' \frac{x_i^2 y_i^2}{\pi_i^2} + \left(\frac{N-1}{N}\right)^2 \sum\nolimits_2' \frac{x_i^2 y_j^2}{\pi_i \pi_j} - \frac{2(N-1)}{N^2}\sum\nolimits_2' \frac{x_i x_j y_i^2}{\pi_i \pi_{ij}} - \frac{N-1}{N^2}\sum\nolimits_3' \frac{x_i x_j y_k^2}{\pi_k \pi_{ij}} - \frac{2(N-1)}{N^2}\sum\nolimits_2' \frac{x_i^2 y_i y_j}{\pi_i \pi_{ij}} - \frac{N-1}{N^2}\sum\nolimits_3' \frac{x_i^2 y_j y_k}{\pi_i \pi_{jk}} + \frac{2}{N^2}\sum\nolimits_2' \frac{x_i x_j y_i y_j}{\pi_{ij}^2} + \frac{4}{N^2}\sum\nolimits_3' \frac{x_i x_j y_i y_k}{\pi_{ik}\pi_{ij}} + \frac{1}{N^2}\sum\nolimits_4' \frac{x_i x_j y_k y_l}{\pi_{ij}\pi_{kl}}$$

5. AN EMPIRICAL INVESTIGATION

An empirical investigation has been undertaken on ten natural populations of size N = 12 from the published literature (Appendix–1). From each population a sample of size n = 6 has been considered. For this purpose, Midzuno (1950) sampling scheme is used. The unit at the first draw is selected with unequal probabilities {of a selection, while at all subsequent draws, the units are selected with equal probabilities and without replacement.

Let the units in the population be $U_1, U_2, .., U_N$ and a sample of size n is drawn with Midzuno's scheme of sampling. Consider the random variable ($i = 1, 2, \cdots, N$) such that

$$t_i = \begin{cases} 1 \text{ if } U_i \text{ is included in the sample} \\ 0 \text{ otherwise} \end{cases}$$

then for this scheme of sampling (ref. Sukhatme and Sukhatme (1970)),

$$E(t_i) = \frac{N-n}{N-1}P_i + \frac{n-1}{N-1} = \pi_i,$$

$$E(t_i t_j) = \frac{n-1}{N-1}\left[\frac{N-n}{N-2}(P_i + P_j) + \frac{n-2}{N-2}\right] = \pi_{ij},$$

$$E(t_i t_j t_k) = \frac{(n-1)}{(N-1)}\frac{(n-2)}{(N-2)}\left[\frac{N-n}{N-2}(p_i + p_j + p_k) + \frac{n-3}{N-3}\right] = \pi_{ij}k$$

$$E(t_i t_j t_k t_1) = \frac{(n-1)(n-2)(n-3)}{(N-1)(N-2)(N-3)}\left[\frac{N-n}{N-4}(p_i + p_j + p_k + p_1) + \frac{n-4}{N-4}\right] = \pi_{ijk1}$$

The variances for the estimator r and the proposed estimator r_1 for these populations have been worked out and are presented in Table 1.

Table 1. Variance of the Estimators for PPSWOR Sampling (N = 12, n = 6).

Population No.	ρ	V( )	V(r)
1	0.9825	0.000062	0.00074
2	0.4033	0.04083	0.14020
3	0.4922	0.06336	0.11480
4	0.7359	0.02624	0.04204
5	0.7599	0.01269	0.03571
6	0.4112	0.05400	0.13810
7	0.6074	0.02176	0.07965
8	0.6898	0.01907	0.05495
9	0.5576	0.03514	0.09497
10	0.2641	0.08097	0.17310

The perusal of these variances reveals that variances of the proposed estimator r_1 are smaller in comparison to the estimator being used in practice; indicting thereby the usefulness of the proposed estimator r_1 of in practice for PPSWOR sampling.

APPENDIX–1

Population No. 1: Sampling Theory of Surveys with Applications, (Edition II) Sukhatme, P.V. and Sukhatme B.V. (1970), Asia Publishing House, India. Table No. 4.8, 185.

Population No. 2: Sampling Techniques (Edition II) Cochran, W.G. (1963), John Willey and Sons Ltd., New York, 32.

Population No. 3: New Developments in Survey Sampling Norman L. Johnson and Harry Smith (1969), John Wiley and Sons Ltd., New York, 288–89.

Population No. 4: Sampling Theory and Methods, Murthy, M.N. (1967), Statistical Publishing Society, Calcutta (India), 228.

Population No. 5: Sampling Theory and Methods, Murthy, M.N. (1967), Statistical Publishing Society Calcutta (India), 127-130.

Population No. 6: The design of sample surveys, Des Raj (1972), Mc-Graw Hill Book Company, New York (U.S.A.), 81.

Population No. 7: Impact of changing conditions on grain marketing institutions and the structure of grain markets in erstwhile Punjab-India (1968), A.S. Kahlon, Punjab Agricultural University, Ludhiana, 126.

Population No. 8: Family Budgets of thirty two cultivators in the Punjab 1971–72, Publication No. 210, Economic and Statistical Organization Punjab (India), 40–43.

Population No. 9: Family Budgets of thirty two cultivators in the Punjab 1971–72, Publication No. 210, Economic and Statistical Organization Punjab (India), 60–63.

Population No. 10: Farm Accounts in Punjab 1970–71 Publication No. 128-The Board of Economics Inquiry Punjab, Punjab Government, Chandigarh (India), 88-91.

REFERENCES

Bravais A. (1846). Analyse mathematique sur les probabilities des erreurs de situation d' un point' Memoires prisentis par divers savants', 1'-A-academic Royale des Sciences de lInstitut de France, **9,** 255-322.

Fisher R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507-521.

Fisher R.A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3-32.

Gupta, J.P. (2002). Estimation of the correlation coefficient in probability proportional to size with replacement sampling. *Statistical Papers*, **43**, 525-536.

Gupta, J.P. and Singh, R. (1989). Usual correlation coefficient in PPSWR sampling. *Jour. Ind. Statist. Assoc*. **27,** 13-16.

Gupta, J.P. and Singh, R. (1990). A note on usual correlation coefficient in systematic sampling. *Statistica*, **2,** 255-259.

Gupta, J.P., Singh, R., Kashani, HB (1993). An estimator of the correlation coefficient in probability proportional to size with replacement sampling. *Metron*, **3-4**, 165-178.

Gupta, J.P., Singh, R., Lal, B. (1978). On the estimation of the finite correlation coefficient-I. *Sankhya C*, **40**, 38-59.

Gupta, J.P., Singh, R. Lal, B. (1979). On the estimation of the finite correlation coefficient, II. *Sankhya C*, **41**, 1-39.

Kendall, M.G., Stuart, A. (1958). Advanced Theory of Statistics, Vol. I (second edition), Charles Griffin and Company Ltd., London.

Midzuno, H (1950). An outline of the theory of sampling systems. *Ann. Inst. Statist. Math.* **1**, 149-156.

Pearson, K. (1896) Mathematical contributions to the theory of evolution-III-regression, heredity and panmixia. *Phil. Trans. (A) Royal Soc. London*, **187**, 253-318.

Sukhatme, P.V., Sukhatme, B.V. (1970). Sampling theory of surveys with applications, Asia Publishing House, India.

Wakimoto, K. (1971). Stratified random sampling (III): Estimation of the correlation coefficient. *Ann. Inst. Statist. Math.* **23**, 339-353.