STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–11.
Published online in International Academic Press (www.IAPress.org)

New Transformation For The Selection Probability Under Positive Correlation Coefficient

Naser Odat*

Department of Mathematics, Jadara University, P.O. Box (733), postal code 21111, Irbid-Jordan

Abstract In this paper, we suggested a new transformation for the selection a sample with probability proportional to size measure with replacement under a positive correlation coefficient between the study variable y and the measure of size variable x . The relative efficiency of the proposed estimator has been studied under a super-population model. A numerical investigation into the performance of the estimator has been made.

Keywords Hansen Hurwitz, Probability Proportional to size, Estimator, sampling with replacement

AMS 2010 subject classifications 62D05, 62D99

DOI: 10.19139/soic-2310-5070-2417

1. Introduction

When supplementary size metrics are available, Probability Proportional to Size (PPS) sampling is frequently employed in finite population sampling. Even though traditional estimators like the Simple Random Sampling (SRS) estimator and the Hansen-Hurwitz (HH) estimator are well-established, they might not function as effectively when there is a weak positive correlation between the auxiliary variable (x) and the study variable (y). By stratifying the population prior to selection. Recently presented a transformed PPS estimator $(\widehat{Y}p)$ that is intended to increase efficiency in situations where correlation is low.

The performance of four estimators, $SRS,\ HH,$ and the suggested $\widehat{Y}p$ under a superpopulation model with weak positive correlation is assessed in this study for sample sizes ranging from 20 to 150. Our simulation evaluates robustness, relative performance, and variance efficiency, offering insights into the best estimator choice for practical sampling situations.

When sampling from a finite population, probability proportional to size (PPS) sampling is used, where the probability of choosing a unit is proportional to its size and a size measure is available for each population unit prior to sampling. A modified on PPS sampling estimation studded by [12].

Take into consideration a finite population made up of distinct and identifiable units,

$U = (U_1, U_2, ..., U_N)$. Assume that y_i is the study variable's value on the unit $U_i, i = 1, ..., N$. In actuality, we would like to calculate the population's overall, $Y = \sum y_i$ based on the values of the units drawn in a sample with maximum precision. The simple random sampling with replacement (SRSWR) scheme is the simplest of the probability sampling schemes for selecting a sample, and its unbiased estimator of y is given by:

$$\widehat{T}_{srs} = \frac{N}{n}\sum_{i=1}^{n} y_i$$

*Correspondence to: Naser Odat (Email: nodat@jadara.edu.jo). Department of Mathematics, Jadara University, P.O. Box (733), postal code 21111, Irbid-Jordan.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

With variance is given by:

$$V\left(\widehat{T}_{srs}\right) = \frac{N}{n}[\sum_{i=1}^{N} y_i^2 - \left(\sum y\right)^2 /N]$$

The concept of sampling with probability proportional to size and with replacement (PPSWR) is proposed by [9]. The plan calls for only one unit to be chosen at each of the n draws. The selection probability for every individual chosen from the population is provided $p_i = \frac{x_i}{X}$, where $X = \sum_{i=1}^{N} x_i$

The estimator of the population total Y was given by [6], as

$$\widehat{T}_{HH} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{p_i}$$

with variance

$$v\left(\widehat{T}_{HH}\right) = \frac{1}{n}\left[\sum_{i=1}^{n} \frac{y_i^2}{p_i} - Y^2\right]$$

If the line of regression y on x crosses through the origin, PPS sampling should be more effective than SRS sampling. If not, it is possible to change the auxiliary variable to increase the efficiency of PPS sampling with adjusted sizes. [9] proposed a method for estimating variance that consistently has a smaller variance than the standard in a sample with unequal probability with replacement. [20] proposed a new transformation on the auxiliary variable x, which changed the Midzuno sampling strategy. [7] studded the optimum utilization of auxiliary information.

[1] provide a simple alternative estimator of the population total when the correlation between the study and auxiliary variable is positive; the estimators as follows:

$$\widehat{T} = \sum_{i=1}^{N} \frac{y_i}{p_i^*}\ , \quad p_i^* = \frac{1-\rho}{N} + \rho p_i\ , \quad p_i = \frac{x_i}{\sum x_i}$$

An alternative estimator to estimating a population total when the correlation between the certain variables is poor positive with selection probabilities suggested by [18]. [17] suggested the following estimator of population total:

$$p_i^* = \frac{(1-\rho)(1+\rho)}{N} + \frac{1}{2}\left[\rho(1+\rho)p_i^+ - \rho(1-\rho)p_i^-\right]$$

where $p_i^+ = \frac{x_i}{X}$, $X = \sum_{i=1}^{N} x_i$, $p_i^- = \frac{z_i}{\sum x}$, with $z_i = \frac{X - nx_i}{N-n}$

[2] created a novel transformed estimator of population total when the attributes under study are poorly associated with selected probability after observing that the[10] model works with zero correlation. A straightforward substitute for the changes in [2] process was put forth by [1]. Research on negatively correlated coefficients has been undertaken by [3, 8, 14, 15, 16]. [21] considered the Probability-proportional-to-size Ranked-set Sampling from Stratified Populations. A modification in ratio estimator by using rank set sampling created by [19].

1.1. The superpopulation model

The superpopulation model proposed by [9] must be taken into consideration in order to examine the relative effectiveness of the proposed estimator with regard to PPSWR and SRSWR sampling. A general superpopulation model suitable for our case is

$$yi = Bp_i + e_i,\ i = 1, 2, ..., N$$

Where $p_i = \frac{x_i}{\sum xi}$
where y_i and p_i stand for the relative measure of size p and the value of attributes y, respectively, for the i^{th} unit in the population (i = 1,2,...,N). where errors e_i are such that

$$E(e_i/p_i) = 0, \;\; E(e_i^2/p_i) = \sigma^2 p_i^g$$

$$E(e_i e_j/p_i p_j) = 0$$

Where σ^2, g are a superpopulation model parameters $\left(\sigma^2 > 0, \; g > 0\right)$
For $g = 0$, the variance of e_i constant.
For $g = 1$, variance of $e_i$proportional to p_i.
For $g > 1$, variance e_i increase more rapidly with p_i, where the average overall finite population that can be extracted from the superpopulation is shown by $E(.)$. Several studies, including [4, 5, 7, 11], and many more, successfully compare various sample procedures using the supperpopulation model.

2. Suggested Estimator

Assume that there is a positive correlation between the research variable y and the auxiliary variable $x > \; 0$. Next, we propose to transform x to x^* in such a way that $x^* = \frac{x_i + nX}{N - n}, \; i = 1, 2, \; \ldots, N$, where $x^*$is obviously bigger than zero. Furthermore, it is evident that there is a positive association between y and x^*. Thus, the altered selection probabilities turn into

$$p_i^* = \frac{n + p_i}{Nn + 1}, \; i = 1, 2, \; \ldots, \; N.$$

Then the unbiased estimator of the population total Y is provided by

$$\widehat{Y}_p = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i^*}$$

It is known that the variance of the usual estimator $\widehat{T}_{HH}$ is given by

$$v\left(\widehat{T}_{HH}\right) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{y_i^2}{p_i} - \left(\sum_{i=1}^{n} y_i \right)^2 \right]$$

The corresponding variance of the estimator due to [11] is given by

$$v\left(\widehat{T}_{R}\right) = \frac{N^2}{n} \left[\sum_{i=1}^{N} y_i^2 p_i - \left(\sum_{i=1}^{N} y_i p_i \right)^2 \right]$$

The variance of proposed estimator is obtain by replacing $p\;_i$by p_i^* in(2.2)and is given by

$$v\left(\widehat{Y}_p\right) = \frac{1}{n} \left[\sum_{i=1}^{N} \frac{y_i^2}{p_i^*} - \left(\sum_{i=1}^{N} y_i \right)^2 \right]$$

2.1. RobustnessEstimator

Now, we state two lemmas by [13], which are useful for estimator's comparisons

Lemma 2.1
Let $0 \leq b_1 \leq b_2 \leq ... \leq b_m$ and $c_1 \leq c_2 \leq ... \leq c_m$ satisfying

$$\sum_{i=1}^{m} c_i \geq 0$$

Then

$$\sum_{i=1}^{m} b_i c_i \geq 0$$

Lemma 2.2
Let $b_1 \geq b_2 \geq ... \geq b_m \geq 0$ and $c_1 \geq c_2 \geq ... \geq c_m$satisfying

$$\sum_{i=1}^{m} c_i \geq 0$$

Then

$$\sum_{i=1}^{m} b_i c_i \geq 0$$

Theorem 2.3
Under the super population model, the sufficient condition that $\widehat{T}_{HH}$ has smaller expected variance than $\widehat{Y}_p$ is

$$g \geq 1 + \frac{p_{imax}}{(Nn+1)p^*_{imax}}$$

where the superpopulation model's heteroscedasticity is controlled by g.

Proof
Under the superpopulation model the expected variance of $\widehat{T}_{HH}$ and $\widehat{Y}_p$ are respectively given by

$$nE\left(v\left(\widehat{T}_{HH}\right)\right) = \sigma^2 \sum_{i=1}^{N} p_i^{g-1}\left(1-p_i\right),$$

And the expected value under superpopulation model for the variance of proposed estimator is

$$nE\left(v\left(\widehat{Y}_p\right)\right) = B^2\left[\sum_{i=1}^{N}\frac{p_i^2}{p_i^*} - 1\right] + \sigma^2 \sum_{i=1}^{N} p_i^g\left(\frac{1}{p_i^*} - 1\right).$$

see the appendix.
The difference between them can be written as

$$nE\left(v\left(\widehat{Y}_p\right) - v\left(\widehat{T}_{HH}\right)\right) = B^2\left[\sum_{i=1}^{N}\frac{p_i^2}{p_i^*} - 1\right] + \sigma^2 \sum_{i=1}^{N} p_i^{g-1}\left[p_i\left(\frac{1-p_i^*}{p_i^*}\right) - (1-p_i)\right]$$

$$= B^2\left[\sum_{i=1}^{N}\frac{p_i^2}{p_i^*} - 1\right] + \sigma^2 \sum_{i=1}^{N}\frac{p_i^{g-1}}{p_i^*}\left(p_i - p_i^*\right)$$

$$= B^2\left[\sum_{i=1}^{N}\left(\frac{p_i}{\sqrt{p_i^*}} - \sqrt{p_i^*}\right)^2\right] + \sigma^2 \sum_{i=1}^{N}\frac{p_i^{g-1}}{p_i^*}\left(p_i - p_i^*\right)$$

$$= B^2 \sum_{i=1}^{N} \left(\frac{p_i}{\sqrt{p_i^*}} - \sqrt{p_i^*} \right)^2 + \sigma^2 \sum_{i=1}^{N} b_i \acute{c}_i$$

where $\acute{c}_i = (p_i - p_i^*)$ and $b_i = \frac{p_i^{g-1}}{p_i^*}$. Note that, the above first term of the above expression is always positive. For these condition we observe that $\sum_{i=1}^{N} \acute{c}_i = 0$ and $\acute{c}_i$ is an increasing function of p_i. So in view Royall's lemma 1 it can be shown that $\sum b_i \acute{c}_i > 0$ provided b_i is also increasing function of p_i. By deriving b_i with respect to p_i we get that the sufficient condition that makes $\widehat{T}_{HH}$ has smaller variance than $\widehat{Y}_p$ is

$$g \geq 1 + \frac{p_{imax}}{(Nn+1)p_{imax}^*}$$

Hence the theorem is proved. □

Example 2.4

$$g \geq 1 + \frac{p_{imax}}{(Nn+1)p_{imax}^*}$$

Solution: Suppose that $n = 5$, $p_{imax} = 0.108$ for $x_i = 54$ and $p_{imax}^* = 0.0843$ for $x_i = 54$ then

$$g \geq 1 + \frac{0.108\ f}{(12 * 5 + 1)0.0843} = 1.021$$

The estimator $\widehat{T}_{HH}$ will have smaller expected variance than $\widehat{Y}_p$ if $g \geq 1.021$

Theorem 2.5
Under the superpopulation model the sufficient condition that the proposed estimator $\widehat{Y}$ has smaller expected variance than the estimator $\widehat{T}_{srs}$ is

$$g \geq \frac{p_{imin}}{(Nn+1)p_{imin}^*}$$

Proof

Under the superpopulation model the expected variance of the estimator $\widehat{T}_{srs}$ and $\widehat{Y}_p$ are

$$nEv\left(\widehat{T}_{srs}\right) = B^2 \left[N \sum_{i=1} P_i^2 - 1 \right] + \sigma^2 (N-1) \sum_{i=1} p_i^g$$

and

$$nEv\left(\widehat{Y}_p\right) = B^2 \left[\sum_{i=1}^{N} \frac{P_i^2}{p_i^*} - 1 \right] + \sigma^2 \sum_{i=1}^{N} p_i^g \left(\frac{1}{p_i^*} - 1 \right)$$

Then

$$nEv\left(\widehat{T}_{srs}\right) - nEv\left(\widehat{Y}_p\right) = B^2 \left[\sum_{i=1}^{N} \frac{P_i^2}{p_i^*} (Np_i - 1) \right] + \sigma^2 \left[\frac{p_i^g}{p_i^*} (Np_i^* - 1) \right] = B^2 \sum_{i=1}^{N} b_i c_i + \sigma^2 \sum_{i=1}^{N} \grave{b}_i c_i \tag{1}$$

In view of Roayaii's Lemma 2.1, both parts of Equation 1 are positive. Let

$$c_i = (Np_i - 1), \quad b_i = \frac{P_i^2}{p_i^*}, \quad \text{and} \quad \tilde{b}_i = \frac{p_i^g}{p_i^*}.$$

Now, since

$$\sum c_i = 0$$

and c_i is an increasing function of p_i, and so is b_i. To ensure that

$$n\,\mathrm{Ev}(\hat{Y}_p) < n\,\mathrm{Ev}(\hat{T}_{\text{srs}})$$

for every possible unit, we must find the unit where $\frac{p_i}{p_i^*}$ is larger.

By deriving $\tilde{b}_i$ with respect to p_i, we get

$$g \geq \frac{p_i}{(Nn+1)\,p_i^*}.$$

This must hold for all i. The term $\frac{p_i}{p_i^*}$ is maximized when p_i is smallest. Then, the sufficient condition that makes $\hat{Y}_p$ have smaller variance than $\hat{T}_{\text{srs}}$ is

$$g \geq \frac{p_{i\,\min}}{(Nn+1)\,p_{i\,\min}^*}.$$

Hence, the result.

□

3. Empirical study

To study the behavior of the estimator $\widehat{Y}_p$with the conventional estimator $\widehat{T}_{srs}$,we will consider the four populations, which are given as follow.

Table 1. population.I N=12

x	41	43	54	39	49	45	41	33	37	41	47	39
y	36	47	41	47	47	45	32	37	40	41	37	48

Table 2. population.II N=30

x	3	4	5	8	12	11	8	9	11	10	8	9	7	8	8
y	11	7	9	8	8	9	8	12	10	9	3	14	12	10	10
x	5	6	3	3	9	6	7	8	8	9	11	11	10	5	3
y	10	9	5	7	9	6	12	9	6	9	11	10	14	8	7

Table 3. Population.III N=10

x	25	32	14	70	24	20	32	44	50	44
y	11	7	5	27	30	6	13	9	14	18

Table 4. Population.III N=7

x	428	1177	1869	2544	2618	4113	4567
y	193	819	611	806	1149	1510	1970

Table 5. Result of selection probability and generalized selection probability.

X	Y	p_i	P_i^*
41	36	0.082	0.0838
43	47	0.068	0.0827
54	41	0.108	0.0843
39	47	0.078	0.0831
49	47	0.098	0.0839
45	45	0.090	0.0836
41	32	0.082	0.0832
33	37	0.066	0.0826
37	40	0.074	0.0829
41	41	0.082	0.0832
47	37	0.094	0.0837
39	48	0.078	0.0831
Sum		1	1

From tables 5, 6, 7, and 8 above, we observed that the selection probability p_i and hence, the generalized selection probability p_i^* satisfied the regularity condition of probability normed size measure

1. $0 < p_i < 1$
2. $\sum_{i=1}^{N} p_i = 1$
3. $0 < p_i^* < 1$
4. $\sum_{i=1}^{N} p_i^* = 1$

From table 12 The better precision of the suggested estimator $\widehat{Y}_p$ is confirmed by the fact that it regularly displays the narrowest confidence intervals. As a result of its sensitivity to deviations from the proportionality assumption $y \propto x$, the Hansen-Hurwitz estimator $\widehat{T}_{HH}$ has intervals that are noticeably broader, particularly in Populations I and II. Notably, Population IV exhibits an oddity in which the interval for $\widehat{Y}_p$ is incredibly wide but $\widehat{T}_{HH}$ surprisingly has the narrowest interval.

4. Simulation study

We used the R software to conduct simulation research in which we generated 1000 samples from gamma distribution with scale parameter 2 and shape parameter 10, with varying simple sizes $\mathbf{n = 20, 50, 100, and\ 150}$, to examine the behavior of the variance of $\mathbf{T_{srs}}$, $\mathbf{T_{HH}}$, and the proposed estimator $\mathbf{T_p}$.

Table 13 presents the findings. By looking at these results, we found that all variances decrease with increasing sample size, but the relative efficiency rankings stay the same, with $\widehat{Y}_P$ continuing to be superior.

Conclusion

It is clear from Table 10 that when the correlation is weakly positive, the estimator is more efficient than the estimators and in populations I, II, and III, but in population IV, when the correlation coefficient is strongly positive, the estimator is more efficient than the estimator and since the typical PPS estimator is already quite effective when the correlation is really strong.

In table 13 three estimators $\widehat{T}_{rs}$, $\widehat{T}_{hh}$, and the suggested $\widehat{Y}_p$ were tested in our simulation analysis under weak positive correlation $r = 0.3$ for different sample sizes. The findings validate the robustness of the suggested $\widehat{Y}_p$

Table 6. Result of selection probability and generalized selection probability

Unit No	**X**	**Y**	P_i	P_i^*
1	3 458	11	0.013333	0.033005
2	4	7	0.017778	0.033078
3	5	9	0.022222	0.033151
4	8	8	0.035556	0.033370
5	12	8	0.053333	0.033661
6	11	9	0.048889	0.033588
7	8	8	0.035556	0.033370
8	9	12	0.040000	0.033443
9	11	10	0.048889	0.033588
10	10	9	0.044444	0.033515
11	8	3	0.035556	0.033370
12	9	14	0.040000	0.033443
13	7	12	0.031111	0.033297
14	8	10	0.035556	0.033370
15	8	10	0.035556	0.033370
16	5	10	0.022222	0.033151
17	6	9	0.026667	0.033224
18	3	5	0.013333	0.033005
19	3	7	0.013333	0.033005
20	9	9	0.040000	0.033443
21	6	6	0.026667	0.033224
22	7	12	0.031111	0.033297
23	8	9	0.035556	0.033370
24	8	6	0.035556	0.033370
25	9	9	0.040000	0.033443
26	11	11	0.048889	0.033588
27	11	10	0.048889	0.033588
28	10	14	0.044444	0.033515
29	5	8	0.022222	0.033151
30	3	7	0.013333	0.033005
Sum	225	272	1	1

Table 7. Result of selection probability and generalized selection probability.

X	**Y**	P_i	P_i^*
25	11	0.070423	0.09852
32	7	0.090141	0.09951
14	5	0.039437	0.09716
70	27	0.197183	0.10468
24	30	0.067606	0.09847
20	6	0.056338	0.09791
32	13	0.090141	0.09951
44	91	0.123944	0.10114
50	14	0.140845	0.10195
44	18	0.123944	0.10114
Sum		1	1

Table 8. Result of selection probability and generalized selection probability

X	Y	P_i	P_i^*
428	193	0.02477	0.13491
1177	819	0.06792	0.13785
1869	611	0.10795	0.14059
2544	806	0.14696	0.14318
2618	1149	0.15119	0.14343
4113	1510	0.23756	0.14918
4567	1970	0.26375	0.15096
Sum		1	1

Table 9. The Pearson correlation coefficient $(r)$for populations

Population	**Correlation**
I	0.049988
II	0.338633
III	0.487686
IV	0.996524

Table 10. The Variance of the Estimators

Population	$\hat{T}_{\text{srs}}$	$\hat{T}_{\text{HH}}$	$\hat{Y}_p$
I	3708	6364.892	3667.204
II	5276	12715.85	5201.921
III	6700	7478	6478.15
IV	7334906	2486528	12727941

Table 11. Percentage Variance relative for the Suggested Estimator $\widehat{Y}_p$.

Population	$\hat{T}_{\text{srs}}$	$\hat{T}_{\text{HH}}$	$\hat{Y}_p$
I	98.90	57.62	100
II	98.59	40.91	100
III	96.69	86.63	100
IV	1.735	5.118	100

Table 12. The 95% confidence interval of the Estimators

Pop.	$\hat{T}_{\text{srs}}$	$\hat{T}_{\text{HH}}$	$\hat{Y}_p$
I	[3588.441, 3827.559]	[6208.328, 6521.456]	[3548.311, 3786.097]
II	[5133.633, 5418.367]	[12494.83, 12936.87]	[5060.359, 5343.483]
III	[6539.565, 6860.435]	[7308.30, 7647.69]	[6320.415, 6635.885]
IV	[7281427, 7388385]	[2483431, 2489625]	[12720950, 12734930]

estimator in weak correlation situations by showing that it routinely performs better than $\widehat{T}_{rs}$ and $\widehat{T}_{hh}$ in terms of variance reduction. All variances decline with increasing sample size, but the relative efficiency rankings hold steady, with Y_p continuing to be superior. Because Yp has a reduced variance, practitioners should favor it for data that is weakly correlated. Adaptive estimators that alternate between T_{hh} and Y_p according to the estimated correlation strength may be investigated in future studies. Future work could explore extensions of the proposed

Table 13. The Variance of the Estimators with relative efficiency

Sample size	$\hat{T}_{\text{srs}}$	$\hat{T}_{\text{HH}}$	$\hat{Y}_p$	**Re.ef. (srs vs. p)**	**Re.ef. (hh vs. p)**
20	239988	801636	239939	1	3.4
50	95995	320654	95987	1	3.4
100	47998	160327	47996	1	3.4
150	31998	106885	31998	1	3.4

sampling method under dynamic or feedback-driven settings. In this context, nonlinear models such as the discrete-time FitzHugh–Nagumo neuron model incorporating a memristor [22] may offer inspiration for modeling complex, evolving populations.

Appendix

The variance of proposed estimator is given by

$$v\left(\widehat{Y}_p\right) = \frac{1}{n}\left[\sum_{i=1}^{N}\frac{y_i^2}{p_i^*} - \left(\sum_{i=1}^{N} y_i\right)^2\right]$$

Under the superpopulation model proposed by Cochran (1963)

$$yi = Bp_i + e_i,\ i = 1, 2, ..., N$$

where y_i and p_i stand for the relative measure of size p and the value of attributes y, respectively, for the i^{th} unit in the population (i = 1,2,...,N). where errors e_i are such that

$$E(e_i/p_i) = 0,\ \ E(e_i^2/p_i) = \sigma^2 p_i^g$$

$$E(e_i e_j/p_i p_j) = 0$$

$$\sigma^2 > 0,\ g \geq 0$$

The the expected value of $\widehat{Y}_p$ under superpopulation model is

$$E\left(v\left(\widehat{Y}_p\right)\right) = B^2\left[\sum_{i=1}^{N}\frac{p_i^2}{p_i^*} - 1\right] + \sigma^2\sum_{i=1}^{N} p_i^g\left(\frac{1}{p_i^*} - 1\right)$$

Proof

$$E(v\left(\widehat{Y}_p\right)) = \frac{1}{n}\left[E\left(\sum_{i=1}^{N}\frac{y_i^2}{p_i^*} - \left(\sum_{i=1}^{N} y_i\right)^2\right)\right]$$

$$nE\left(v\left(\widehat{Y}_p\right)\right) = \sum_{i=1}^{N} E\left(\frac{y_i^2}{p_i^*}\right) - E\left(\sum_{i=1}^{N} y_i\right)^2 \tag{2}$$

$$E\left(\frac{y_i^2}{p_i^*}\right) = \frac{1}{p_i^*}E(\beta^2 p_i^2 + 2Bp_i e_i + e_i^2)$$

$$E\left(\frac{y_i^2}{p_i^*}\right) = \frac{1}{p_i^*}\left(\beta^2 p_i^2 + \sigma^2 p_i^g\right) \tag{3}$$

$$E\left(\sum_{i=1}^{N} y_i\right)^2 = E\left(y_i^2\right) + \sum_{i \neq j}^{N} E\left(y_i y_j\right) \tag{4}$$

$$E\left(y_i^2\right) = \left(\beta^2 p_i^2 + \sigma^2 p_i^g\right) \tag{5}$$

And

$$E\left(y_i y_j\right) = E\left[\left(Bp_i + e_i\right)\left(Bp_j + e_j\right)\right] = E\left(\beta^2 p_i p_j + Bp_i e_j + Bp_j e_i + e_i e_j\right)$$

$$E\left(y_i y_j\right) = \beta^2 p_i p_j \tag{6}$$

Substituting Eq 5 and Eq 6 in Eq 4 we get

$$E\left(\sum_{i=1}^{N} y_i\right)^2 = \sum\left(\left(\beta^2 p_i^2 + \sigma^2 p_i^g\right) + \beta^2 p_i p_j\right) = \sum \beta^2\left(p_i^2 + p_i p_j\right) + \sum \sigma^2 p_i^g \tag{7}$$

Substituting Eq 7 and Eq 3 in Eq 2 we get

$$nE\left(v\left(\widehat{Y}_p\right)\right) = \sum_{i=1}^{N} \frac{1}{p_i^*}\left(\beta^2 p_i^2 + \sigma^2 p_i^g\right) - \sum \beta^2\left(p_i^2 + p_i p_j\right) + \sum \sigma^2 p_i^g$$

by simplification we get

$$E\left(v\left(\widehat{Y}_p\right)\right) = B^2\left[\sum_{i=1}^{N} \frac{p_i^2}{p_i^*} - 1\right] + \sigma^2 \sum_{i=1}^{N} p_i^g\left(\frac{1}{p_i^*} - 1\right).$$

□

REFERENCES

1. G. Amahia, Y. Chaubey, and T. Rao, *Efficiency of a new pps sampling for multiple characteristics*, Journal of Statistical Planning and Inference, vol. 21, pp. 75–84, 1989.
2. M. Bansal and R. Singh, *An alternative estimator for multiple characteristics in pps sampling*, Journal of Statistical Planning and Inference, vol. 21, pp. 75–84, 1985.
3. P. Bedi and T. J. Rao, *Pps method of estimation under a transformation*, Journal of the Indian Society of Agricultural Statistics, vol. 54, pp. 184–195, 2001.
4. K. Brewer, *A method of systematic sampling with unequal probabilities*, Australian Journal of Statistics, vol. 5, pp. 5–13, 1963.
5. V. Godambe, *A unified theory of sampling from finite populations*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 17, pp. 269–278, 1955.
6. M. H. Hansen and W. N. Hurwitz, *On the theory of sampling from finite populations*, The Annals of Mathematical Statistics, vol. 14, pp. 333–362, 1943.
7. T. Hanurav, *Optimum utilization of auxiliary information: πps sampling of two units from a stratum*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 29, pp. 374–391, 1967.
8. P. Kumar and P. Bedi, *An alternative estimator in midzuno scheme for multiple characteristics*, Communications in Statistics–Simulation and Computation, vol. 24, pp. 17–30, 1995.
9. J. N. Rao, H. Hartley, and W. Cochran, *On a simple procedure of unequal probability sampling without replacement*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 24, pp. 482–491, 1962.
10. J. N. K. Rao, *Alternative estimators in pps sampling for multiple characteristics*, Sankhya A, vol. 28, pp. 47–60, 1966.
11. J. N. K. Rao, *On the relative efficiency of some estimators in pps sampling for multiple characteristics*, Sankhya A, vol. 28, pp. 61–70, 1966.
12. V. Reddy and T. Rao, *Modified pps method for estimation*, Sankhya C, vol. 39, pp. 185–197, 1977.

13. R. M. Royall, *On finite populations sampling theory under certain linear regression models*, Biometrika, vol. 57, pp. 377–387, 1970.
14. L. Sahoo, G. Mishra, and S. Senapati, *A new sampling scheme with inclusion probability proportional to size*, Journal of Statistical Theory and Applications, vol. 4, pp. 361–369, 2005.
15. L. Sahoo, S. Sc, and M. Ak, *A class of ipps sampling schemes*, Revista Investigación Operacional, vol. 31, pp. 217–224, 2010.
16. L. Sahoo, G. Singh, and B. Das, *A note on an ipps sampling scheme*, Allgemeines Statistisches Archiv, vol. 90, pp. 385–393, 2006.
17. H. Singh and R. Tailor, *Use of known correlation coefficient in estimating the finite population mean*, Statistics in Transition, vol. 6, pp. 555–560, 2003.
18. S. Singh and S. Horn, *An alternative estimator for multi-character surveys*, Metrika, vol. 48, pp. 99–107, 1998.
19. N. Odat, *Modification in ratio estimator using rank set sampling*, European Journal of Science, vol. 29, no. 2, pp. 265–268, 2009.
20. N. A. Alodat, R. M. Hatamleh, and A. M. Qazza, *On the choice of strategy for modified midzuno scheme*, Far East Journal of Mathematical Sciences, vol. 101, no. 7, pp. 1481–1490, 2017.
21. O. Ozturk, *Probability-proportional-to-size Ranked-set Sampling from Stratified Populations*, Survey Methodology, vol. 46, no. 2, pp. 243–265, 2020.
22. M. T. Shatnawi, A. A. Khennaoui, A. Ouannas, G. Grassi, A. V. Radogna, A. Bataihah, and I. M. Batiha, *A Multistable Discrete Memristor and Its Application to Discrete-Time FitzHugh–Nagumo Model*, Electronics, vol. 12, no. 13, article 2929, 2023. https://doi.org/10.3390/electronics12132929.