

RELATIVE EFFICIENCY OF TWO SAMPLE SELECTION PROCEDURES IN PROBABILITY PROPORTIONAL TO SIZE SAMPLING WITHOUT REPLACEMENT

Dawodu, O. O. and Adewara, A. A.

Department of Statistics
University of Ilorin, Ilorin
Nigeria

(e-mail: omotola_dawodu@yahoo.com, aaadewara@gmail.com)

ABSTRACT

In probability proportional to size sampling, units have unequal chances of being selected. It is necessary to consider the unequal probability nature of units. It is important to compare the relative efficiency of a complex design with that of a simple design, so as to know if there would be any gain or loss in efficiency. Here, two sample selection procedures in probability proportional to size sampling without replacement are considered as complex designs. They are: Dawodu and Adewara, and Basit and Shahbaz procedures. Simple random sampling without replacement is used here, as a simple design. The complex sample designs are compared with the simple one. Results are compared with simple random sampling without replacement, in order to get the relative efficiency. It was found that, for this dataset, there was more gain in efficiency when Dawodu and Adewara procedure was used, than when Basit and Shahbaz procedure was used.

1. Introduction

Horvitz and Thompson (1952) gave the estimator of the population total in probability proportional to size sampling without replacement (ppsswor) as:

$$\hat{Y}_{HT} = \sum_{i=1}^N \frac{Y_i}{\pi_i} \quad (1)$$

with the unbiased variance estimator as:

$$\hat{V}(\hat{Y}_{HT}) = \sum_{i=1}^n \left(\frac{1-\pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i=1}^n \sum_{i \neq j} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \right) \left(\frac{y_i y_j}{\pi_i \pi_j} \right) \quad (2)$$

which may give negative results.

Sen (1953), Yates and Grundy (1953) independently gave the expression for the unbiased variance of Horvitz and Thompson (1952) estimator as follows:

Keywords: Probability Sampling; Probability Proportional To Size; Sampling Without Replacement; Inclusion Probability.

$$\hat{V}_{SYG}(\hat{Y}_{HT}) = \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \quad (3)$$

They showed that the variance estimator in (3) produces positive results for most sampling schemes. Y_i represent variables of interest. The probabilities of inclusion of single unit in the sample, which are the total probability assigned to a population unit, for being included in the sample in all draws are represented by π_i . Also, the joint probabilities of inclusion of both units i and j in the sample are represented by π_{ij} , for $i, j = 1, 2, \dots, N$, and N is defined as the total number of units.

Some sample selection procedures that could be used with the Horvitz and Thompson (1952) estimator of the variance as given by Sen (1953), Yates and Grundy (1953) are: Basit and Shahbaz (2007), Alodat (2009), and Dawodu and Adewara (2016).

Rao (2004) considered the Horvitz and Thompson (1952) estimator over the last five decades and reviewed some of its optimality properties. He stressed that Horvitz and Thompson (1952) estimator was of great importance and will continue to play an important role in sampling theory and applications.

2. Sample Selection Procedures In Probability Proportional To Size Sampling Without Replacement

2.1. Dawodu And Adewara Procedure (DAP)

Dawodu and Adewara (2016) proposed a sample selection procedure in ppsswor as follows:

- Select the first unit with probability proportional to $\frac{P_i^\alpha}{2 - P_i^\beta}$ and without replacement,
- Select a random sample of size $n - 1$ from the remaining $N - 1$ units.

Note that α and β are constants and are expressed in powers.

Therefore,

$$\pi_i = \frac{1}{R} \frac{N-n}{N-1} \frac{P_i^\alpha}{2 - P_i^\beta} + \frac{n-1}{N-1} \quad (4)$$

where $R = \sum_{i=1}^N \frac{P_i^\alpha}{2 - P_i^\beta}$

$$\pi_{ij} = \frac{(n-1)(N-n)}{R(N-1)(N-2)} \left[\frac{P_i^\alpha}{2 - P_i^\beta} - \frac{P_j^\alpha}{2 - P_j^\beta} \right] + \frac{(n-1)(n-2)}{(N-1)(N-2)} \quad (5)$$

2.2. Basit And Shahbaz procedure (BSP)

Basit and Shahbaz (2007) proposed a sample selection procedure in ppsswor as follows:

- Select first unit with probability proportional to $\frac{P_i^\alpha (1 - P_i^\beta)}{(1 - 2P_i^\beta)}$ and without replacement,
- Select a random sample of size $(n - 1)$ from the remaining $(N - 1)$ units.

$$\pi_i = \frac{1}{d} \left[\frac{P_i^\alpha (1 - P_i^\beta)}{(1 - 2P_i^\beta)} \left(\frac{N - n}{N - 1} \right) + \frac{(n - 1)}{N - 1} d \right] \quad (6)$$

where $d = \sum_{i=1}^N \frac{P_i^\alpha (1 - P_i^\beta)}{(1 - 2P_i^\beta)}$

$$\pi_{ij} = \frac{(n - 1)(N - n)}{d(N - 1)(N - 2)} \left[\frac{P_i^\alpha (1 - P_i^\beta)}{(1 - 2P_i^\beta)} + \frac{P_j^\alpha (1 - P_j^\beta)}{(1 - 2P_j^\beta)} \right] + \frac{(n - 1)(n - 2)}{(N - 1)(N - 2)} \quad (7)$$

P_i and P_j are called selection probabilities, and are defined as the non-zero known probability values assigned to each and every units of the population at any specific draw with $0 \leq P_i, P_j \leq 1$ for $i, j, = 1, 2, \dots, N$. P_i are calculated from auxiliary variables X_i .

$P_i = \frac{X_i}{X}$, where X_i are the number of items contained in unit i of the auxiliary variable and X is the total number of items contained in the auxiliary variable.

3. Relative Efficiency

DAP and BSP would be used with (3), in order to obtain the variance estimates for specified values of constants α and β . Also, variance estimate using simple random sampling without replacement (SRSWOR) would be obtained. In this study, SRSWOR is considered a simple design, while DAP and BSP are considered complex designs. For DAP and BSP, only the minimum variance estimate in each would be considered. These variance estimates obtained are compared with that of SRSWOR, in order to get relative efficiency.

$$\hat{V}(\hat{Y}_{SRSWOR}) = N^2 \left(\frac{1 - f}{n} \right) S^2 \quad (8)$$

where S^2 is the sample variance, n is the sample size, N is the number of units, and $f = \frac{n}{N}$. The essence of using SRSWOR is to check if there would be any gain or loss in efficiency.

$$\text{Relative efficiency ER} = \frac{\hat{V}(\hat{Y}_{SRSWOR})}{\hat{V}_{SYG}(\hat{Y}_{HT})} \times 100 \quad (9)$$

4. Presentation Of Data

According to 2013 Nigeria Demographic and Health Survey (NDHS), a child's weight at birth is an important indicator of the child's vulnerability to the risk of childhood illnesses and the child's chances of survival. Children who weigh less than 2.5 kilograms have a higher than average risk of early childhood death. The 2013 NDHS, the woman's questionnaire recorded birth weight, if available from written records or the mother's recall, for all births in the five years preceding the survey. Nigeria is divided into six zones as shown in the 2013 NDHS. This shows that the number of units $N = \text{six}$. So, units have different probabilities of selection.

Table 1 : Distribution Of Reported Child's Birth Weight

Units	Births With A Reported Birth Weight		Number Of Sampled Children With Birth Weight Less Than 2.5 kg (Y_i)
	Estimated Percentage Of Children Less Than 2.5 kg At Birth	Number Of Sampled Births	
North Central	7.5	534	41
North East	13.6	302	42
North West	27.2	489	134
South East	4.3	1317	57
South South	11.6	798	93
South West	3.4	1780	61

Source: Final Report, 2013 Nigeria Demographic and Health Survey

5. Analysis

In this study, α and β assume integer values from -5 to $+5$, excluding zero. Here, a sample of size $n = \text{five}$ is selected from the number of units, $N = \text{six}$. The role of Table 2 is to obtain the selection

Table 2 : Distribution Of The Population Of Nigeria Based On Zones

Units	Population (X_i)
North Central	20,369,956
North East	18,894,299
North West	35,915,467
South East	16,395,555
South South	21,044,081
South West	27,722,432
Total	$X = 140,341,790$

Source: 2006 Population and Housing Census of The Federal Republic of Nigeria

probabilities P_i as seen in Table 3, P_i are used to calculate the probabilities of inclusion of units (π_i) in the sample and also the joint probabilities of inclusion (π_{ij}) of both units i and j in the sample. The probability of selection of each zone P_i is shown in Table 3.

Table 3 : Probabilities Of Selection Of Each Unit

Units	Zones	Probabilities of Selection (P_i)
1	North Central	0.1451453
2	North East	0.1346306
3	North West	0.2559143
4	South East	0.1168259
5	South South	0.1499488
6	South West	0.1975351
	Total	1

Source: Researchers' Computation

The variance estimates for both procedures calculated using (3) for all values of α and β under consideration are presented in Table 4.

Table 4 : Comparison Of Variance Estimates Using DAP And BSP

		α									
		-5		-4		-3		-2		-1	
		DAP	BSP	DAP	BSP	DAP	BSP	DAP	BSP	DAP	BSP
β	-5	2335	2973	2060	2922	1779	2848	1534	2734	1347	2566
	-4	2563	2973	2333	2922	2057	2848	1776	2734	1531	2566
	-3	2728	2973	2556	2923	2324	2848	2047	2735	1766	2568
	-2	2834	2974	2712	2924	2533	2851	2293	2739	2012	2573
	-1	2884	2977	2785	2929	2637	2858	2426	2751	2160	2591
	1	2969	2959	2916	2900	2838	2811	2719	2676	2719	2482
	2	2972	2970	2920	2918	2844	2840	2729	2722	2558	2549
	3	2972	2972	2922	2921	2847	2846	2732	2731	2563	2561
	4	2973	2973	2922	2922	2847	2847	2733	2733	2565	2564
	5	2973	2973	2922	2922	2848	2847	2734	2734	2565	2565

Source: Researchers' Computation

Table 4 (Continues)

		α									
		1		2		3		4		5	
		DAP	BSP	DAP	BSP	DAP	BSP	DAP	BSP	DAP	BSP
β	-5	1131	2061	1073	1780	1034	1535	1006	1348	986	1218
	-4	1216	2062	1130	1781	1072	1535	1033	1349	1006	1219
	-3	1339	2065	1212	1783	1127	1537	1070	1350	1032	1220
	-2	1496	2073	1319	1791	1198	1544	1117	1355	1064	1223
	-1	1607	2100	1397	1817	1249	1566	1150	1371	1085	1234
	1	2030	1942	1750	1668	1511	1444	1330	1282	1206	1172
	2	2050	2036	1769	1755	1525	1514	1341	1333	1213	1207
	3	2058	2054	1776	1773	1532	1529	1346	1343	1217	1215
	4	2060	2059	1779	1778	1534	1533	1347	1347	1218	1217
	5	2061	2061	1780	1779	1534	1534	1348	1348	1218	1218

Source: Researchers' Computation

Considering Table 4, the least variance recorded using DAP and BSP are recorded. The probabilities of inclusion of single unit, π_i in the sample, calculated using (4) for DAP and (6) for BSP, for specified values of α and β are also recorded. The results are presented in Table 5.

Table 5 : Probability Of Inclusion Of Single Unit (π_i) While Selecting A Sample Of Size $n = \text{Five}$ From $N = \text{Six}$

Units	Zones	DAP ($\alpha = 5, \beta = -5$)	BSP ($\alpha = 5, \beta = 1$)
1	North Central	0.8006335	0.8066915
2	North East	0.8002986	0.8045171
3	North West	0.9843037	0.9442845
4	South East	0.8000723	0.8021628
5	South South	0.8008773	0.8079377
6	South West	0.8138146	0.8344064
	Sample Size	5	5

Source: Researchers' Computation

Also, the joint probabilities of inclusion of units, π_{ij} in the sample calculated using (5) for DAP and (7) for BSP, for specified values of α and β are recorded. The results are presented in Table 6.

Table 6 : Joint Probability Of Inclusion Of Units (π_{ij}) While Selecting A Sample Of Size $n = \text{Five}$ From $N = \text{Six}$

Units Sampled	DAP ($\alpha = 5, \beta = -5$)	BSP ($\alpha = 5, \beta = 1$)
1,2	0.6009321	0.6112086
1,3	0.7849372	0.7509761
1,4	0.6007057	0.6088543
1,5	0.6015107	0.6146292
1,6	0.6144481	0.6410979
2,3	0.7846024	0.7488016
2,4	0.6003709	0.6066798
2,5	0.6011759	0.6124548
2,6	0.6141132	0.6389235
3,4	0.784376	0.7464473
3,5	0.785181	0.7522223
3,6	0.7981184	0.778691
4,5	0.6009495	0.6101005
4,6	0.6138869	0.6365692
5,6	0.6146919	0.6423441
Total	10	10

Source: Researchers' Computation

Variance estimate is obtained for SRSWOR using equation (8), and relative efficiency obtained using equation (9). Table 7 shows the least variance estimates recorded using DAP and BSP, the variance estimate using SRSWOR, and the relative efficiencies.

Table 7 : The Variance Estimates And The Relative Efficiencies

Country	Variance Estimates			Relative Efficiencies (%)	
	DAP	BSP	SRSWOR	DAP	BSP
Nigeria	986	1172	1557	158	133

Source: Researchers' Computation

6. Discussion Of Results And Conclusion

Here, DAP and BSP are compared by finding the variance estimates using the values of α and β under consideration. For simplicity, only integer values (excluding zero) of α and β are considered. Here, the optimum values of α and β for DAP are 5 and -5 respectively, giving the least value of the variance as 986. For BSP, the optimum values of α and β are 5 and 1 respectively, giving the least variance as 1172.

The variance estimates using SRSWOR, DAP and BSP are used to obtain relative efficiencies. It is realized that the relative efficiencies using DAP and BSP are approximately 158% and 133% respectively. It shows 58% and 33% gains in efficiency when DAP and BSP are used respectively.

It shows that DAP and BSP are better than SRSWOR. This also shows that it is very important to take into consideration the unequal probability nature of the data under consideration.

In conclusion, DAP performs better than BSP for this dataset. Hence, for analyzing data on Child's weight at birth in Nigeria, if a subsample of strata had been required as part of the original design, we would recommend the use of DAP in probability proportional to size sampling without replacement.

REFERENCES

1. Alodat, N. A. : "On Unequal Probability Sampling Without Replacement Sample Size 2". Int. J. Open Problems Comp. Math., 2(1), (2009), 109-112.
2. Basit, A. and Shahbaz, M. Q. : "A General Class Of Selection Procedures And Modified Murthy Estimators". Journal Of Statistics, 14, (2007), 13-19.
3. Dawodu, O. O. and Adewara, A. A. : "Comparison Of Some Sample Selection Procedures In Probability Proportional To Size". ICASTOR Journal Of Engineering, 9(3), (2016), 155-170.
4. Horvitz, D. G. and Thompson, D. J. : "A Generalization Of Sampling Without Replacement From A Finite Population". Jour. Amer. Statist. Ass., 47, (1952), 663-685.
5. Nigeria Demographic And Health Survey: National Population Commission, Federal Republic Of Nigeria, Abuja, Nigeria; ICF Macro Claverton, Maryland, USA: (2013), 155-157.
6. Priority Table: 2006 Population And Housing Census Of The Federal Republic Of Nigeria, Volume 1, National Population Commission.
7. Rao, T. J. : "Five Decades Of The Horvitz - Thompson Estimator And Furthermore...". J. Ind. Soc. Agric. Stat. 58(2), (2004), 177-189.
8. Sen, A. R. : "On The Estimate Of The Variance In Sampling With Varying Probability", Jour. Ind. Soc. Agric. Statist. 5, (1953), 119-129.
9. Yates, F. and Grundy, P. M. : "Selection Without Replacement From Within Strata With Probability Proportional To Size". J. Roy. Stat. Soc. B, 15, (1953), 153-161.