COMPARATIVE ANALYSIS OF SOME SAMPLING TECHNIQUES FOR THE ESTIMATION OF REGISTERED LIFE BIRTH

Supported by tetfund

Akinola Oladiran Adepetun

Department of Statistics, Federal University of Technology, Akure, Nigeria
Email: aoadepetun@futa.edu.ng

Received: February 11, 2023 **Accepted:** April 15, 2023

Abstract

In this paper, comparative analysis of three different sampling techniques which are simple random sampling, stratified random sampling and systematic random sampling for the estimation of registered life birth were investigated. The real data set for the study was obtained from Seychelles National Bureau of Statistics on the registered life birth between 1986 and 2018. The mean and variance of each sampling technique were computed and their efficiency compared using the variance criterion. The study revealed that simple random sampling is the best as it has the least variance compared with two other sampling techniques under study.

Keywords: Sampling techniques, life birth, efficiency, variance criterion.

Introduction

Sampling is the process of selecting a subset of elements from a population. To further explain sampling, the initial step in comprehending the process is to be familiar with the terminology. So we say sample is a representative part of the target population from which the result of the study can be generalised. For example, the outcomes of a study based on number of farmers that are into commercial farming in a major city cannot be generalised to the number of farmers that are into commercial farming in smaller town even though they are both farmers but their environments differ significantly. An important decision that has to be taken in a particular sampling technique is about the size of the sample. The sampling technique used in selecting the sample size goes a long way to determine how reliable the investigation carried out. Adam (2019) argued that systematic sampling is a kind of random sampling technique in which sample members from large population are chosen according to a random start but with fixed interval. Aggarwal (2011) in her research used simple random sampling as a simple method of data collection in the presence of complete sampling frame. Simple random sample is a subset of a statistical population in which each member of the subset has an equal probability of being chosen. A simple random sample is meant to be an unbiased representation of a group (Adam and Eric, 2021) .Akeem *et al*. (2015) posited the different opinions associated with the use of systematic sampling and estimation in stratified survey sampling in terms of the precision of the population mean. Cochran (1977) proved that systematic sample mean has higher level of precision than the simple random sample mean when $S_w^2 < S^2$ and it was concluded that cluster sample implied that systematic sampling has higher precision than simple random sampling if the standard error inherent in the systematic samples is more than the population standard error altogether. Gravetter and Forzano (2011) cited that simple random sampling is the simplest probability sampling technique widely adopted for sample selection. Graham (2014) cited that systematic sampling is a flexible method for choosing a random sample from a finite population. Jambulingam *et al*. (2014) adopted circular systematic sampling for estimation of a finite population mean whenever there exists a linear trend among the population values. Javid and Khan (2015) proposed a Generalised Systematic Sampling design for the estimation of finite population mean. The developed design was found to be more efficient than simple random sampling and other several existing systematic sampling techniques. Llewellyn *et al.* (2015) established that whenever there is linear trend, linear systematic sampling is less precise than stratified random sampling and more precise than simple random sampling. Murthy (1967) early researches on development of systematic sampling theory were stated. Mike (2017) pointed that at the heart of effective survey research is having a representative sample that allows for survey findings to be generalized to the larger population and for the survey research to be repeated. Mukherjee and Singh (2019) in their paper proposed a new sampling scheme for sampling selection for the case of odd samples size such that population size is multiple form of sample size for a population attribute with inherent linear trend. Neil (2010) cited that stratified random sampling is a kind of random sampling that gives room for researchers to enhance relative efficiency with respect to simple random sampling. Patricia and Ulysses (2014) cited the most common problem in their Brazilian research sampling problems with a detailed look on the medicinal plants. Raj and Chandhok (1998) explained systematic sampling as an easier technique of choosing sample when the units are numbered serially from 1 to N assuming N = nk, where *n* and *k* are both the required sample size and an integer respectively. Sayed and Ibrahim (2017) argued that systematic sampling is one of the commonest forms of sampling techniques due to its applicability. Ullah *et al.* (2022) in their study proposed family of estimators of finite population means under both simple random and stratified random sampling techniques using auxiliary information in a more rigorous fashion. The applicability of the proposed family of estimators was demonstrated with real data sets coming from diverse fields of applications.

Therefore, this study involves the comparative analysis of three different sampling techniques with a view to comparing and determining the best sampling technique among the three when applied on registered life birth using the variance criterion.

Methodology
In this section, the established theoretical frameworks of the three different sampling techniques which include simple random sampling, stratified random sampling and the systematic random sampling are presented. Suitable codes were written in R to implement numerically these sampling techniques using real data set obtained from Seychelles National Bureau of Statistics on the registered life birth between 1986 and 2018.

Simple Random Sampling
In this sampling technique, each element of a target population has same chance of selection into the sample. Considering a population total which is defined as

$$Y = \sum_{i=1}^{N} Y_i = Y_1 + Y_2 + \cdots + Y_N \qquad (1)$$

The population mean is defined as

$$\bar{Y} = \frac{Y_1 + Y_2 + \cdots + Y_N}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{Y}{N} \quad (2)$$

The variance of the Y_i in a population of size N is

$$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N} \qquad (3)$$

Since equation (3) is practically impossible, the variance of the simple random sample mean $\bar{y}$ of size n is given as

$$V(\bar{y}) = (1-f)\frac{s^2}{n} \qquad (4)$$

Where $f = \frac{n}{N}$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ are both sampling fraction and sampling variance respectively.

Stratified Random Sampling
In the technique of stratified random sampling, heterogeneous population is divided into sub-populations of homogeneous non-overlapping groups known as strata. A sample is then chosen from each stratum using simple random sampling technique.
For the purpose of simplicity, the notations in stratified random sampling are defined as follows:
N = Total heterogeneous population size
N_h = Total number of units in homogeneous stratum h
$\mathrm{n_h}$ = Number of units in sample stratum h
$\mathrm{y_{hi}}$ = Value obtained from i^{th} sample unit of stratum h
$\mathrm{W_h} = \frac{\mathrm{N_h}}{\mathrm{N}}$ is the homogeneous stratum weight
$\mathrm{w_h} = \frac{\mathrm{n_h}}{\mathrm{n}}$ is the homogeneous sample stratum weight
$\mathrm{f_h} = \frac{\mathrm{n_h}}{\mathrm{N_h}}$ is the homogeneous stratum sampling fraction
$\bar{\mathrm{y}}_\mathrm{h} = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$ is the homogeneous stratum sample mean
$\bar{\mathrm{y}}_{\mathrm{st}} = \sum_{h=1}^{k} \mathrm{w_h}\,\bar{\mathrm{y}}_\mathrm{h}$ is the stratified sample mean
$\mathrm{s_h^2} = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2$ is the homogeneous stratum sampling variance
According to stratified random sampling, the variance of the stratified sample mean is given as

$$V(\bar{y}_{st}) = \sum_{h=1}^{k} w_h^2 (1 - f_h)\frac{s_h^2}{n_h} \qquad (5)$$

Systematic Random Sampling
Systematic random sampling is a random sampling technique which involves a random selection of the first element for the sample and thereafter subsequent elements are chosen in a systematic interval until the required sample size is attained.
In systematic sampling, the variance of the systematic random sample mean $\bar{y}_{sys}$ is defined as

$$V(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{sys} - \bar{y})^2 \qquad (6)$$

Where $\bar{y}_{sys} = \frac{1}{n}\sum_{j=1}^{n} y_{ij} = \bar{y}_i$, $\bar{y} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_i$ and $k = \frac{N}{n}$ is the systematic interval of fixed size respectively.
In this study, the appropriate sample size was generated according to Cochran (1977) as

$$n = \frac{N}{1 + Ne^2} \qquad (7)$$

Where e is the desired level of precision which will be assumed by the investigator.

Analyses and Discussion
The preceding section of this study involves various mathematical expressions for the estimation of mean and variance of each of the three different sampling techniques under study. In this section, the analyses of the three different sampling techniques are demonstrated in order to ascertain their efficiency using real data set on the population of registered life birth.

Table 1: Data with Year and Registered life birth (RLB) in ('000)

Year	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
RLB	1722	1684	1643	1600	1617	1706	1601	1689	1700	1582	1611
Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
RLB	1475	1412	1460	1612	1440	1481	1498	1435	1536	1467	1439
Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
RLB	1546	1580	1504	1625	1625	1566	1557	1592	1645	1651	1650

Source: Seychelles NBS (1986-2018)
The data considered as the population is the registered life birth data between 1986 and 2018 which is the record for 33 years. The sample size to be selected from this population is estimated using $n = \frac{N}{1+Ne^2}$. When $N = 33$, $e = 0.05$, the computed sample size $n = 31$.

Simple Random Sampling
We select 31 observations from the available 33 by the simple random sampling procedure. The sample size selected, its mean and variance are as follows:

Table 2: Table showing sample size selected

Year	RLB	Year	RLB
1986	1722	2003	1498
1987	1684	2004	1435
1988	1643	2005	1536
1989	1600	2006	1467
1990	1617	2007	1499
1991	1706	2008	1546
1992	1601	2009	1580
1993	1689	2010	1504
1994	1700	2011	1625
1995	1582	2012	1645
1996	1611	2013	1566
1997	1475	2014	1557
1999	1460	2015	1592
2000	1512	2016	1645
2001	1440	2018	1650
2002	1481		

Table 3: Mean and variance of simple random sampling

Mean	Variance
1576.387	13.78412

Stratified Random Sampling

The population is split into two strata and the size of samples to be selected from each stratum as well as each stratum computations are enumerated as presented in the tables below:

Table 4: The observations in stratum of '90s

Year	RLB	Year	RLB
1986	1722	1993	1689
1987	1684	1994	1700
1988	1643	1995	1582
1989	1600	1996	1611
1990	1617	1997	1475
1991	1706	1998	1412
1992	1601	1999	1460

Table 5: Samples selected from stratum of '90s randomly

Year	RLB	Year	RLB
1986	1722	1993	1689
1987	1684	1994	1700
1988	1643	1996	1611
1989	1600	1997	1475
1990	1617	1998	1412
1991	1706	1999	1460
1992	1601		

Table 6: Stratum size, stratum weight, sample size and f in stratum of '90s

The mean of the samples selected from the stratum of '90s is computed as 1609.231

Stratum size	Stratum weight	Sample size	F
14	0.4242424	13	0.9285714

Table 7: The observations in stratum of '2000s

Year	RLB	Year	RLB
2000	1512	2010	1504
2001	1440	2011	1625
2002	1481	2012	1645
2003	1498	2013	1566
2004	1435	2014	1557
2005	1536	2015	1592
2006	1467	2016	1645
2007	1499	2017	1651
2008	1546	2018	1650
2009	1580		

Table 8: The samples selected from stratum of '2000s randomly

Year	RLB	Year	RLB
2000	1512	2010	1504
2001	1440	2011	1625
2002	1481	2012	1645
2003	1498	2013	1566
2004	1435	2014	1557
2005	1536	2015	1592
2006	1467	2016	1645
2008	1546	2017	1651
2009	1580	2018	1650

Table 9: The stratum size, stratum weight, sample size and f in stratum of '2000s

Stratum size	Stratum weight	Sample size	F
19	0.5757576	18	0.9473684

The mean of the samples selected at random from the stratum of '2000s is computed as 1551.66

Table 10: The means and variance of stratified random sampling

Means	Variance
1609.231	15.26691
1551.667	

Systematic Random Sampling

In this technique, samples are selected in such a way that every twelfth (12th) element of the population is omitted. Therefore, the selected samples are:

Table 11: Selected Samples

Year	RLB	Year	RLB
1986	1722	2003	1498
1987	1684	2004	1435
1988	1643	2005	1536
1989	1600	2006	1467
1990	1617	2007	1499
1991	1706	2008	1546
1992	1601	2010	1504
1993	1689	2011	1625
1994	1700	2012	1645
1995	1582	2013	1566
1996	1611	2014	1557
1998	1412	2015	1592
1999	1460	2016	1645
2000	1512	2017	1651
2001	1440	2018	1650
2002	1481		

Table 12: Mean and variance of systematic random sampling

Mean	Variance
1576.645	15.23709

Table 13: Variances of the three sampling techniques under study

Sampling Techniques	Variances
Simple Random Sampling	13.78412
Stratified Random Sampling	15.26691
Systematic Random Sampling	15.23709

Conclusion

In this study, the efficiency of three different sampling techniques in estimating registered life birth was compared using variance criterion. From table 13 above, it can be deduced that simple random sampling is the best since it has the least variance value 13.78412 compare with the variances of the other two sampling techniques under study.

References

Adams, H. 2019. Review on systematic sampling. *Journal of the Indian Statistical Association*, 12: 1-2.

Aggarwal, A. 2011. Sampling issues in research methodologies. *Indian Journal of Medical Specialties*, 22: 169-172.

Adam, H. & Eric, E. 2021. Simple Random Sample. Retrieved July 1, 2021, from Investopedia: *https://www.investopedia.com/terms/s/simple-random-sample.asp*

Akeem, O. K., Oshungade, I.O. & Gafar, M. O. 2015. Efficient sampling scheme within Strata. *Journal of Statistics*, 5: 2-4.

Cochran, W. G. 1977. Sampling techniques. Third edition, John Wiley and Sons, New York.

Gravetter, F. J. & Forzano, L. B. 2011. Research Methods for the Behavioural Sciences. Congage Learning, 4: 137-140.

Graham, K. 2014. Statistical Inference. John Wiley and Sons, Ltd.

Jambulingam, S., Gupta, S. N. & Prabavathy, G. 2014. Circular systematic sampling. *America Journal of Mathematical and Management Sciences*, 4: 1-19.

Javid, S. & Khan, K. 2015. Generalised systematic sampling. *Communication in Statistics-Simulation and Computation*, 44(9): 2240-2250.

Llewellyn, N., Delia, N., Temesgen,T. Z. & Raghunath, A. 2015. Balance of modified systematic sampling in the presence of linear trend. *South African Statistical Journal*, 49(2): 187-204.

Murthy, M. N. 1967. Sampling Theory Methods. Second Edition, Statistical Publishing, Calcutta.

Mike, A. 2017. Sampling issues. *The SAGE Encyclopaedia of Communication Research Method*, 6: 18-20.

Mukherjee, A. & Singh, A. 2019. Optimal Systematic Sampling When Sample Size is Odd in the Presence of Linear Trend and Two-Way Linear Trend. *Sankhya A*

doi:10.1007/s13171-019-00172-5

Neil, J. S. 2010. Stratified Sampling. *Encyclopaedia of Research Design Methods*, 3: 25-30.

Patricia, M. M. & Ulysses, P. A. 2014. Sampling Problems on Brazilian Research. *Revista Brasileira de Farmacognosia*, 24: 103-109.

Raj, D. & Chandhok, P. 1998. Sample Survey Theory. Narosa Publishing House, New Delhi.

Sayed, A. M. & Ibrahim, A. A. 2017. Recent development in systematic sampling: A review. Journal of Statistical Theory and Practice, 12: 290-310.

Ullah, K., Hussain, Z., Hussain, I., Cheema, S. A., Almaspoor, Z. & El-Moshedy, M. 2022. Estimation of finite Population Mean in Simple and Stratified Random Sampling by Utilizing the Auxiliary, Ranks and Square of the Auxiliary Information. *Hindawi Mathematical Problems in Engineering*, Volume 2022: 1-14. https://doi.org/10.1155/2022/5263492