Journal of Education and Policy Review
Volume 12, Number 2, 2020
ISSN: 2277-0100
http://www.cenresinjournals.com

COMPARISON ON THE EFFICIENCY OF SIMPLE RANDOM SAMPLING VERSUS SYSTEMATIC RANDOM SAMPLING IN THE ESTIMATION OF AGES OF STUDENTS IN AN EDUCATIONAL SURVEY

Itiveh Francis Enahoro & Aniesedo Johnny M.
Department of Mathematics and Statistics
Delta State Polytechnic, Ozoro
Email: fenoitiveh4@gmail.com

ABSTRACT: This study was intended to ascertain the most efficient sampling techniques in sample survey: Simple random sampling versus systematic random sampling. The target population of study was the Delta State Polytechnic. For the purpose of this study, the school of Engineering Technology was selected randomly using simple random sampling – lottery method. The ages of students in the department of civil engineering technology was randomly selected among the other departments using simple random sampling without replacement (SRSWOR). Eighty (80) questionnaires were distributed using equal allocation to the respective levels in the department. Fifty respondents' questionnaires were retrieved and data set was organized and the normality test was done. Twenty (20) elementary units were drawn from the population of fifty (50) using simple random sampling and systematic random sampling. Sample statistic- mean and variance- were calculated using the two methods compared and their efficiency was tested. Results from the normality test shows that the data was normally distributed using chi-square goodness of fit test at $\alpha = 0.05$ level of significance. From the sample statistic calculated simple random sampling has the smallest variance and as such is more efficient than systematic random sampling. It is hereby recommend that simple random sampling methods should be used in research work.
KEYWORDS: Sampling, statistic, simple random, data, systematic.

INTRODUCTION

Sampling is a process of selecting samples from a group or population to become the foundation for estimating and predicting the outcome of the population as well as to detect the unknown piece of information. It is a scientific method of selecting and using a representative part (samples) of a whole to seek the truth about the whole. Sampling is used extensively, consciously or unconsciously, in everyday life to obtain the required information or to carry out a course of action. A sample survey is therefore defined as the collection and examination of data from a sample in order to make inferences about the whole. Hence, sample survey theory deals with the process of sample selection, data collection and estimation of the population characteristics using the sample data so collected and determining the accuracy of the estimates. A population characteristic is defined as a quantity relating to the population, e.g. mean number of course units taken by students per session in a university, total hectares of land under rice in a country, proportion of female spectators in a football match (Okafor, 2002).

Simple random sampling and systematic sampling scheme used in sample survey. The systematic sampling gives each unit of the population and equal chance of being selected in a sample, this selection can be done in two ways by lottery methods and use of random numbers. Simple random sampling could be done with replacement (SRSR) and without replacement (SRSWOR). The procedure of selecting a sample by SRS scheme is as follows:
Number the whole units in the population serially from 1 to N. then select random numbers between 1 and N inclusive with the aid of table of random numbers, starting from the top of the column or columns, depending on the number of digits that makes up the population size N.

However, in practice it is convenient to adopt whether horizontal or vertical. The units whose serial numbers correspond to the random numbers selected constitute the sample units. Alternatively, the numbers 1 to No could be written out in pieces of paper or any suitable device. A number is drawn one after the other as in a lottery until the required numbers of sample units are obtained. This is called lottery method.

If after each draw, the selected numbers is discarded before the next selection is made we refer to the selection method as simple random sampling without replacement (SRSWOR), generally referred to as simple random sampling (SRS). On the other hand, if the selected number is replaced in the population before the next draw is made we have simple random sampling with replacement (SRSWR). In SRSWR there is the possibility that a unit may appear more than once in the sample. This is not the case with SRSWOR where a unit appears once in the sample. Thus in SRSWR the sample size could exceed the population size. The sample obtained by SRSWR is called an unrestricted or ordered sample. If follows therefore that SRSWOR is called a restricted or unordered sample. If follows therefore that SRSWOR has a hyper geometric probability distribution model and SRSWR a binomial probability distribution model (Okafor, 2002).

A systematic sampling technique is operationally more convenient than simple random sampling. It also ensures at the same time that each unit has equal probability of inclusion in the sample. In this method of sampling, the first unit is elected with the help of random numbers and the remaining unit as selected automatically according to a predetermined pattern.

Suppose the units in the population are numbered 1 to N in some order. Suppose further that N is expressible as a product of two integers n and k, so N = nk. And systematic sampling consists of linearly and circular sampling. Systematic is especially applicable when the population to studied is arranged in some order.
The procedure of selecting a sample by systematic sampling scheme is as follows:
Suppose a sample of n units is to be selected from N units in the population.
Let $k = \frac{N}{n}, K$ is an integer. Select a random number between 1 and K inclusive. Suppose the random number selected is r. add k to the random start r successively until n numbers are obtained. The sample then consists of n units with serial numbers r, r + K, r + 2k, … …, r + (n – 1)k. thus the sample consists of the first unit selected at random

and every k^{th} unit thereafter. This procedure of sample selection is called systematic sample. K is called the sampling interval; $\frac{1}{k} = \frac{n}{N}$ is the sampling fraction (Okafor, 2002).

Pradhan (2004) view it – efficiency of cluster sample on sampling on two occasions – that, even if the intra-class correlation coefficient among the units in the same cluster is positive under certain condition the cluster sampling on two occasions is likely to be more efficient than the simple random sampling to estimate the populations means of the character under study on current (second) occasion. Stephen (1992) is of the view that recommendation for use of a systematic design or simple random depends on the special pattern of misclassification and the objectives of accuracy assessment in a given application. Based on the population studied – comparison of systematic and simple random sampling for estimating the accuracy of maps generated from remotely sensed Data systematic designs are generally more precise and therefore use sampling resource is more efficiently than simple random sample and also offers the greatest potential gains and losses in precision relation to simple random sampling. Brown et al., (2010), opined in their work that all design discussed are remarkably efficient, giving estimates of populations that have lower variance than the conventional design without the adaptive selection. Habib (2014) stated that the method of systematic sampling can be alternative to simple random sampling specially preferred when the information required to construct a sampling is available in a list in any other organized form.

Simple random sampling is an unbiased estimate of the population means and variance estimate than systematic sampling. Thus, when we do not select our sample randomly out of the populations of interest, our sampling result may be biased. Hence, the necessity of simple random sampling arises. And also said; the sampling variance of the mean of a systematic sample from a list can be expected to be less than the means of a simple random sampling.

The most efficient of all unbiased estimators of the parameter θ is the one with the smallest variance for any given sample size Okafor (2002).

Peregrine (2018) in his work concluded that regardless of the sampling technique used, it is important to determine an appropriate sample size. Sample size should balance the cost of collecting data and the accuracy required to evaluate a given hypothesis. By convention, a 10% random sample size is considered for most research problems.
The major aims and objectives of this research work is to establish the sampling method or techniques that is more efficient from sample observation using simple random sampling and systematic random sampling.

STATEMENT OF THE HYPOTHESIS

The following stated hypothesis is used in this study.

Null Hypothesis

Ho: Ages of respondents do not follow a normal distribution

Alternative Hypothesis

H_1: Ages of respondents follows a normal distribution

MATERIALS AND METHODS

The population of study consists of all students of the Delta State Polytechnic, Ozoro, Simple random sampling techniques- lottery methods was used to select the department Civil Engineering technology. Equal allocation of twenty (20) questionnaires was distributed to each levels that make up the department. Data collected from the research instrument was analyzed using normality test, sample statistics of the two sampling methods compared calculated and the efficiency tested. The estimation of population mean of variances was calculated for both sampling scheme.

DATA PRESENTATION AND ANALYSIS

The data below shows the distribution of ages of 50 students who returned their questionnaires in the department of Civil Engineering Technology, Delta State Polytechnic, Ozoro for 2018/2019 session.

Table 1: Distribution of ages of Civil Engineering Technology, Delta State Polytechnic, Ozoro

20	19	24	19	29	18	28	19	18	21
30	31	19	18	22	24	27	26	24	28
23	26	21	27	26	25	22	25	20	19
25	27	20	22	19	28	24	20	25	22
21	18	28	23	18	21	20	30	21	24

Table 2: Grouped Frequencies of ages of respondents

Class Interval (Age)	Frequency
18 – 21	21
22 – 25	15
26 – 29	11
30 – 33	3
Total	50

Source: field survey

Table 3: Calculation of mean and variances of age distribution of respondent

Class interval	Frequency	Class mark	Fx	X^2	Fx^2
18 – 21	21	19.5	409.5	380.25	7985.25
22 – 25	15	23.5	352.5	5552.25	8283.75
26 – 29	11	27.5	302.5	756.25	8318.75
30 – 33	3	31.5	94.5	992.25	2976.75
Total	50	-	1159	-	27564.5

$\bar{x} = 23.18$

$S^2 = 13.9776$

$S = 3.74$

The probability of the expectation is $n\ x\ p\ \left(\frac{x-\mu}{\delta} \leq z \leq \frac{x-\mu}{\delta}\right)$

Normality Test

Ho: Ages of respondents do not follow a normal distribution

H_1: Ages of respondents follows a normal distribution

$\alpha = 0.05$

$\chi^2 = \sum_{i=1}^{k} \frac{(Oi - Ei)^2}{Ei}$

Decision Rule: reject Ho if chi-square calculated is greater than chi-square tabulated

Table 4: The Chi-square calculation

Class interval	Frequency O_i	Expected Frequency E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(Oi - Ei)^2}{Ei}$
18 – 21	21	9.935	11.065	122.4342	12.3235
22 – 25	15	15.6725	-0.6725	0.4523	0.0289
26 – 29+	14	9.8678	4.1322	17.0751	1.7304
Total	50		-		14.08

χ^2

$Chi - square\ tab_{0.05}^{3} = 5.99$

Decision Rule: Reject Ho chi-square calculated is greater than chi-square tabulated

Conclusion: since chi-square calculated is 14.8 and is greater than chi-square tabulated 5.99 we therefore reject the null hypothesis and accept the alternative hypothesis and conclude that age of respondents follows a normal distribution at $\alpha = 0.05$ level of significance.

Empirical Examples and Illustration

Simple random sampling analysis of sample data set

Sample drawn from the population of N = 50 using simple random sampling without replacement/lottery method. Number selected are: 49, 46, 43, 37, 13, 11, 18, 24, 27, 30, 1, 3, 22, 8, 34, 6, 14, 17, 32, 16 corresponding to the ages 21, 2128, 24, 19, 30, 26, 27, 22, 19, 20, 24, 26, 19, 22, 18, 18, 27, 27, 24 respectively

Table 5 Mean and Variance analysis for Simple Random Sampling (Lottery Method)

Number selected	Ages = yi	(yi - $\bar{y}$)	$(yi - \bar{y})^2$
49	21	-2.1	4.41
46	21	- 2.1	4.41
43	28	4.9	24.01
37	24	0.9	0.81
13	19	-4.1	16.81

11	30	6.9	47.61
18	26	2.9	8.41
24	27	3.9	15.21
27	22	-1.1	1.21
30	19	-4.1	16.81
1	20	-3.1	9.61
3	24	0.9	0.81
22	26	2.9	8.41
8	19	-4.1	16.81
34	22	-1.1	1.21
6	18	-5.1	26.01
14	18	-5.1	26.01
17	27	3.9	15.21
32	27	3.9	15.21
16	24	0.9	0.81
Total	462		259.8
Mean ($\bar{y}$)	23.1		
Variance (S^2)	13.69		

Var ($\bar{y}_{srswor}$) = $\frac{N-n}{Nn}S^2 = \frac{50-20}{50\ (20)}$ $(13.67) = 6\ \frac{30}{1000}(13.7) = 0.03\ (13.67)$
= 0.66698

Systematic Sampling Analysis of Sample Data Drawn Using Circular Methods

The procedure is as follow since our k = N/n, next we select the first sample by simple random sampling (lottery method) and is 17 while other units are selected systematically using circular method from the population, the samples are shown below:
17, 20, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 3, 6, 9, 12, 15, 19, 21, 24 corresponding to the ages 27, 28, 21, 22, 20, 27, 19, 20, 21, 23, 20, 24, 24, 18, 18, 31, 22, 26, 23, 27 respectively.

Table 6: Mean and Variance for systematic sampling (Circular Method)

Number selected	Ages = yi	(yi – $\bar{y}$)	$(y - \bar{y})^2$
17	27	3.8	14.44
20	28	4.8	23.04
23	21	-2.2	4.84
26	22	-1.2	1.44
29	20	-3.2	10.24

32	27	3.8	14.44
35	19	-4.2	17.64
38	20	-3.2	10.24
41	21	-2.2	4.84
44	23	-0.2	0.04
47	20	-3.2	10.24
50	24	0.8	0.64
3	24	0.8	0.64
6	18	-5.2	27.04
9	18	-5.2	27.04
12	31	-7.8	60.84
15	22	-1.2	1.44
18	26	2.8	7.84
21	23	-0.2	0.04
24	27	3.8	14.44
Total	464		251.4
Mean $\bar{y}$	23.2		
Variance (S^2)	4.4105		

Var ($\bar{y}_{sy}$) = $\frac{(N-1)S^2}{N} - \frac{k(n-1)}{N}S^2wsy = \frac{(50-1)(13.23)}{50} - \frac{3(20-1)(4.4105)}{50}$

= 12.9654 – 5.02797 = 7.937

The Coefficient of Variation

C.V – Standard deviation (S)/Mean x 100%

Where 'S' is the square root of the variance

For simple Random Sampling without Replacement

Variance = 0.4101

Mean = 23.1

$S = \sqrt{0.410}$

$= 0.6404$

C.V = (0.6404 /23.1) x 100%

= 2. 77%

For Systematic Random Sampling (Circular Method)

Variance = 7.937

Mean = 23.2

$S = \sqrt{7.937}$

$= 2.8173$

C.V = (2.8173/23.2) x 100%
= 12.14%

Efficiency = $\frac{\text{Var}(\bar{y}\text{srswor})}{Var(\bar{y}sy)} = \frac{0.4101}{7.937} = 0.05 < 1$

DISCUSSION OF FINDINGS/CONCLUSION

The result of the analysis and test conducted shows that that data collected are normally distributed at 5% level of significance, the sampling variance in the systematic random sampling is 7.937 and the simple random sampling is 0.4101. Thus, it shows that the simple random sampling is more precise than systematic random sampling. A coefficient of variation shows that systematic random variable has 12.14%, while simple random sampling is 2.77% which corroborate the claim by scholars that simple random sampling is more efficient. Precision helps to compare two or more things to see which one is more precise or efficient and by this findings and according to our data set, the sampling variance simple random sampling of 0.4101 is far less than sampling variance of systematic of 7.937 including their coefficient of variation.

REFERENCES

Brown J.A, Mohammed.S, M., Moradi, M. Panahbehagh, B. and Smith D.R (2012). Adaptive Survey Designs for Sampling Rare and Clustered Populations. *Mathematics and Computer in Simulation, 93:108-116*

Habib Ahmed Elsayir (2014). Comparison of Precision of Systematic Sampling with Some other Probability Samplings, *American Journal of Theoretical and Applied Statistics.* 3(4): 111-116.

Okafor F. C. (2002). Sample survey Theory with Applications. *Afro – Orbis Publishers, University of Nigeria, Nsukka.*

Peregrine P. N. (2018).Sampling Theory. *John Wiley & Sons, Inc*

Pradhan B. K (2004). On Efficiency of Cluster Sampling on Sampling on two Occasions. *Statistica, anno LXIV, n. 1, 183-191*

Stephen V. Stehman (1992). Comparison of Systematic and Random Sampling for Estimating the Accuracy of Maps Generated from Remotely Sensed Data. *Photogrammetric Engineering & Remote Sensing 58(9):1343-1350*