WWJMRD 2024; 10(12): 1-5
www.wwjmrd.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
Impact Factor SJIF 2017: 5.182 2018: 5.51, (ISI) 2020-2021: 1.361
E-ISSN: 2454-6615

MOLUNO, Shedrack Uwadinisu Ph.D
Department of Entrepreneurship,
Faculty of Administration and Management,
Delta State University of Science and Technology.
Ozoro. Delta State, Nigeria.

EDEWHOR, Vincent Ph.D
Department of Entrepreneurship,
Faculty of Administration and Management,
Delta State University of Science and Technology.
Ozoro. Delta State, Nigeria.

Advancing the Argument for Systematic Sampling: Between A Quasi / Pseudo Probability Sampling and Real Systematic Probability Sampling Technique.

MOLUNO, Shedrack Uwadinisu, EDEWHOR, Vincent

Abstract
The paper examines the nature, structure, probability content and application of systematic sampling, a type of probability sampling technique. It seeks to draw a line of divide between systematic sampling, quasi or pseudo-probability systematic sampling and the real systematic probability sampling technique and how to achieve the real systematic probability sampling. The research method adopted here is more of a desk top approach relying on secondary data while deploying very robust literature in statistics, research methods and other related discipline. One of the arguments posed is that systematic sampling-(selection of members from a population on the basis of a predetermined sampling interval *k),* is not enough to ascribe the word probability to the systematic approach. Similarly, the selection of the i^{th} element from the first sampling interval through randomisation and the subsequent addition of *k* to the previous elements selected to determine subsequent elements do not also suffice, as the selection of subsequent elements from the succeeding intervals is predicated on the choice of i^{th} element. The paper posits that the real systematic sampling requires the application of simple probability sampling method(s) in such a manner that every element in the population has a fair chance of being chosen. This is line with Gy's Sampling theory. In the light of the foregoing, study suggests the application of any simple random sampling method in every sampling interval as against the first interval only. This no doubt guarantees every member of the population a fair chance of being selected. The emerging method is the real systematic probability sampling technique. The paper further advance that where there is no randomisation at all in any of the sampling intervals or predetermined sampling interval this will result in ordinary systematic sampling and where randomisation is applied in only one or the first sampling interval, the emerging technique is quasi or pseudo-probability sampling.

Keywords: Systematic Sampling, Quasi-systematic Probability Sampling, Systematic Probability Sampling Technique, skipping interval, randomisation.

1.1 Introduction

In the world of statistics and research, the concept of sampling has become an essential plank in the platform for empirical studies. Conventionally, there are two major approaches to this; the probability and non-probability techniques. The probability sampling techniques is rooted on randomization; allowing every element in the population a fair chance of being selected. In statistics and research literature, there are at least four documented types of this technique; simple random technique comprising of; (lottery method, table of random numbers and computer-generated random numbers), systematic, stratified, cluster and sometimes multi-stage cluster sampling Techniques (Agbonifoh and Yomere, 1999; Asika, 2012, and McBurnley, 1994). Although not the direct subject of discourse of this paper, the non-probability method also comprises of at least four methods; convenience, purposive / judgmental, quota and snowball sampling techniques. The non-probability method neither requires randomization nor guarantees every element in the population a fair chance of being chosen.

1.2 Statement of the Problem

In the domain of statistics and research, it is always necessary to always as much as possible define in clear terms any phenomenon, concept, construct or variable of interest. One of such

Correspondence:
MOLUNO, Shedrack Uwadinisu Ph.D
Department of Entrepreneurship,
Faculty of Administration and Management,
Delta State University of Science and Technology.
Ozoro. Delta State, Nigeria.

- concepts that requires clear definition for purposes of better understanding and application is that of systematic sampling. This concept is sometimes used loosely to mean the same thing as systematic probability sampling. There is some sort of problem when two scientific concepts are used interchangeably as though they are exactly the same. In trying to resolve this problem, the following question becomes pertinent: what exactly is systematic sampling? Is it the same as systematic probability sampling? When and how is it applied? Questions of this nature can only be succinctly answered if the phenomenon is clearly explained and understood and this is what this paper seeks to address.

1.3 Objective of the Study

The objective of this study is to examine the nature, structure, probability content and application of systematic sampling, a type of probability sampling technique. It seeks to draw a line of divide between systematic sampling, quasi or pseudo-probability systematic sampling and the real systematic probability sampling technique and how to achieve the real probability sampling.

1.4 Scope of the Study

The paper is strictly delimited to systematic sampling, a kind of probability sampling technique which is the major issue of discourse.

1.5 Significance of the Study

This study among other things is significant as it stands to benefit academics, researchers, industry and students as well; as it will:

1. contribute to the advancement of systematic sampling discourse by providing a comprehensive analysis and understanding of the concept;
2. unveil the nature, structure, probability content and application of systematic sampling;
3. in clear terms draw a line of divide between systematic sampling, quasi or pseudo-probability systematic sampling and the real systematic probability sampling technique and above all;
4. add the body of knowledge of the discourse and how to achieve the real probability sampling.

2.1 Methodology

The research method adopted here is more of a desk top approach leaning on secondary data while deploying very robust literature in statistics, research methods and other related discipline. It outlined a structured approach to investigate and differentiate, systematic sampling, quasi/pseudo systematic probability sampling and real systematic probability sampling techniques. Various contributions to the sampling theory were synthesised including Gy's sampling theory which provides a strong theoretical foundation upon which the study rests.

2.2 Systematic Sampling Defined

This is a type of probability sampling method in which sample elements are selected from a larger population beginning with a random starting point but in line with a predetermined or calculated, fixed, periodic interval known as the skipping or sampling intervals. This interval is also referred to as the k^{th} or n^{th} term. Thomas (2023) defines it as "a probability sampling method in which researchers select members of the population at a regular interval (or *k*) determined in advance". Brief illustrations stated herein below gives a clearer and near-enough picture of the concept. Suppose a population made up of N units is numbered 1 to N in a given order. Suppose further that N can be expressed as a product of two integers n and k, such that $N = nk$. To draw a sample of size n, a random number between 1 and k is selected. Supposing further that that random number is i, the first unit, whose serial number is i is then selected, thereafter every *kth* unit after *ith* unit is elected. The emerging sample will contain i, $i+k$, $1+2k,...,i+(n-1)k$ serial number units. So the first unit is selected at random and other units are selected systematically. This systematic sample is also called the *k*th systematic sample and *k* is termed as a sampling interval. This is also known as linear Systematic sampling. The following table clearly demonstrates a typical order of arrangements of systematic sampling (Ofo, 1994).

Table 1.1. Demonstration of Systematic Sampling.

Systematic Sample Number		1	2	3	…	*i*	…	*k*
Sample Composition	1	y_1	y_2	y_3	…	y_i	…	y_k
	2	y_{k+1}	y_{k+2}	y_{k+3}	…	y_{k+1}	…	y_{2k}
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	n	$y_{(n-1)k+1}$	$y_{(n-1)k+2}$	$y_{(n-1)k+3}$	…	$y_{(n-1)k+i}$	…	y_{nk}
Probability		1/k	1/k	1/k	…	1/k	…	1/k
Sample Mean		$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	…	$\bar{y}_i$	…	$\bar{y}_k$

Source: Shalabh, K. (nd). Sampling Theory| Chapter 11 | Systematic Sampling | Shalabh, IIT Kanpur https://home.iitk.ac.in/~shalab/sampling/chapter11-sampling-systematic-sampling.pdf

2.3 Procedure for Application of Systematic Sampling

The procedure for the application of systematic sampling is quite simple. It requires the following steps; identification of the population, assignment of numbers to the population, determination of the sample size, determination of the sampling interval, choosing a random starting point and finally, identification of subsequent samples.

i. Identification of the Study Population.

The issue for determination here, is the group from which the researcher is sampling. Their characteristics, attributes and where possible their number, which in the event of a finite population should be known. It is on the basis of this that assignment of numbers to the population in step two below can be achieved. Although some writers like (Frost, 2024) have posited that systematic sampling can be applied to an infinite population, an empirical formula for the determination of the sampling / skipping interval (k^{th} or n^{th} term) has not been provided yet. Proponents of this approach have often resorted to arbitrariness and subjectivity in deciding on an interval as well as starting point.

ii. Assign Numbers to the Population. This requires listing and setting the population of the study in a particular order. It is a list of all members or elements of the population that the researcher is interested in studying. While the target population is the general concept of the group the researcher is assessing, a sampling frame specifically lists all members or elements that make up a population and possibly how to contact them (Asika 1999 & 2006; Frost 2024). In ordering this list, numbers are assigned to all the elements in the population to facilitate their selection. A typical example of this is listing the customers of a business in a spread sheet and numbering them from 1-n, where n is the last member of the population. With an infinite population, a list of this nature is impossible.

iii. Determine Your Sample Size. The key issue here is to know what number to sample and eventually study out of the entire population. This is not just done by predetermining a given percentage or proportion of the population for study. While some researchers do subjectively decide to study certain percentage of the population under the cover of not having enough resource (logistics, money or time) the ideal and most appropriate approach is the deployment of empirical formula for the determination of the study sample. There are a good number of formulae that have been developed for this purpose. They include: Taro-Yamene (1967); Kish (1965); Goddon (2004) among others as in (Israel, 2022; Nanjundeswaraswamy & Divaka, 2021; and Bartlet, Kortrlil, & Higins, 2014). Aside from these, there are also a plethora of online calculators and software that have been developed for calculation of sample size, these includes: Calculator.net, Surveymonkey, Raosoft, Epi-info, Qualtric among others.

iv. Determination of Sampling Interval. This probably could be where the systematic sampling actually begins. The salient issue here is to establish the interval within which samples are selected from the population, given the sampling frame that has been developed. This is derived by the formula N/n; where "N" is the population as in the sampling frame and "n" is the sample size calculated. The product of this operation is the k^{th} or n^{th} term earlier mentioned in the work. For instance, given a population (N) of 5000 and a sample size (n) of 250; the sampling interval will be 5000/250 = 20. Impliedly, for every 20^{th} member of the 5000 population the researcher is required to sample or survey an element. The sapling ratio can also be derived from the formula n/N. Given the forgoing example, the sampling ratio will be 250/ 5000 = 0.05 (Agbonifoh and Yomere 1999). In the absence of a sampling frame, it will be difficult to develop an empirical sampling interval.

v. Choose a Starting Point. This is a very critical stage in systematic sampling. What the researcher does here will determine whether the work is systematic sampling or systematic random sampling. Having determined the intervals from which the sample will be drawn earlier before, here the researcher determines the first and subsequent elements from the population. The first element is usually between 1 and the k^{th} term. In our earlier example above, this will be between 1 and 20. To ensure some level of probability and randomness, the first element is required to be selected using at least one of the three approaches to simple random sampling; the lottery method, table of random numbers and computer-generated random numbers. This now sets the stage for systematic random sampling. The random start using any of these simple random sampling techniques is what makes systematic sampling a random sampling method. If and when the starting point is arbitrarily chosen by the researcher, at best it is a systematic sampling technique and cannot fit into or be addressed as a systematic random sampling method. Impliedly, there is a slim line of divide between systematic sampling and systematic random sampling technique and this lies in the extent to which probability and randomness is contained in the selection process.

vi. Identify Sample Members. The task here is to select other elements of the sample. This is done by adding the k^{th} term to the subsequent element(s) selected. Thus the subsequent elements will be obtained through $i+k$, $1+2k,...,i+(n-1)k$ in the series of number as explained earlier in 1.1.1 herein before. This brings us to the crux and central message of this paper which is to determine whether the selection first or i^{th} element in the first sampling interval is enough condition to classify this process as systematic random sampling. Another issue for determination is whether every other element in the population aside from *i*, had a fair chance of being selected. Conventionally, probability sampling involves a random selection of a predetermined sample size from a larger population. This sample is expected to predict, define or approximate strongly the characteristics, attributes and features of the larger population. While it can be argued the i^{th} or first element selected through the use of simple random sampling (the lottery method, computer generated random and table of random numbers) is a product of probability sampling technique, the same cannot be said of $i+k$, $1+2k,...,i+(n-1)k$ which were obtained by merely adding the k^{th} term to the previous sample. This assumption that the mere application of random sampling in the choice of the i^{th} element or the first member in the sampling interval automatically translates to the randomisation of other elements $i+k$, $1+2k,...,i+(n-1)k$ does not hold water. This so because the inclusion of these elements was predicated on the selection of the member *i* and thus does not hold strong ground. It rather can be argued that the selection of *i* automatically precludes or denies every other member of the population in the subsequent intervals; $i+k$, $1+2k,...,i+(n-1)k$ a fair chance of being chosen aside from *i* alone. Where this is the case, it will also not be very convenient to classify the outcome of this process as systematic probability sampling method but at best a quasi or pseudo-systematic probability sampling method (Shalabh, nd).

2.4 The Real Systematic Probability Sampling Method

Aside from the procedures stated herein before in 1.6.2 i-vi for the application of systematic sampling, which includes: knowing the population, numbering of the population in a given order, calculation of the sample size, calculation of the sampling interval, selecting a random starting point and finally, selection of subsequent samples by addition of *kth* term; the real systematic probability sampling goes one step further. This step is targeted at enriching the probability content of this concept as it is argued earlier that a random starting point followed by a systematic selection of subsequent members in line with the predetermined skipping interval k can at best give rise to a quasi or pseudo-probability sampling technique. As a way forward, the paper therefore proposes the sampling of every element in each sampling interval. In doing this and given our earlier illustration in 1.6.4-iv where the sampling interval was 20; it follows that 20 samples independent of the one sampled earlier will be drawn from the twenty sampling intervals that make up the population. By so doing, every element in the population has a fair chance of being chosen and thus no longer dependent anymore on the choice of the first element i and the subsequent addition of the k^{th} term. In line with the foregoing therefore, the real systematic probability sampling can be defined as the random selection of a predetermined sample size from a larger population on the basis of a calculated sampling interval k. These samples must be randomly selected from every sampling or skipping interval using any of the simple random sampling approaches (lottery method, table of random numbers and computer-generated random numbers) Agbonofoh and Yomere (1999).

3.1 Theoretical Foundation

In order to provide a theoretical groundwork and underpinning upon which the study rests, there is need to expound on the development of sampling theory. Although, contribution to sampling theory are replete in statistical discourse but in order to have a clearer and near-enough understanding of the subject of discourse, there is need to expound on the development of the theory. Sampling theory discusses the principles and techniques used in selecting sample or a subset of individuals or items from a larger population (Pitard, 2019). This is done primarily to estimate characteristics of the whole population. Although, sampling is a fundamental concept in statistics, it widely applies to several fields such as engineering, social sciences, opinion polling, and quality control amongst others. The development of modern sampling theory has its root in the works of some key contributors in statistics, this includes: Thomas Bayes (1701-1761), Bayes made foundational contributions to probability theory, which underpins sampling theory; Pierre-Simon Laplace (1749-1827), Laplace took further the Bayesian ideas and contributed further to the understanding of sampling distributions. Others are; Ronald A. Fisher (1890-1962), who developed the principles of random sampling and experimental design, laying the foundation for present day statistical inference. There is also the work of Jerzy Neyman (1894-1981) and Egon Pearson (1895-1980), that centres on the development of the concept of hypothesis testing. Pearson and Neyman's work also did reasonably advance the theory of sampling distributions as well as confidence intervals. These statisticians and other contributors have together contributed to the development of sampling theory (Catlow, 1993 and Murray & Larry, 2008). This is essential for drawing reliable inferences about populations based on samples. The central message and underlying philosophy behind the works of these contributors and the concept of sampling theory, is the selection of samples in such a manner that allows all elements in the population the same probability of being included in the sample to ensure representativeness.

The contributions of these theories are sythesised as follows: Bayes work also popular through Bayes' theorem, which provides a way of bringing to date the probability of a hypothesis in the event of new or emerging evidence. While this may not be directly tagged as a contributor to sampling theory in the modern sense, this work however, provides the groundwork and base for Bayesian inference, which no doubt has become an essential plank in the platform for modern statistical sampling methods. Aside from Bayes, Laplace is a significant contributor to the development of probability theory and statistics. His work which focused more on celestial mechanics includes; the application of probability to astronomical observations and social sciences as well as the method of inverse probability which laid the foundation and formed basis for the present-day statistical inference and sampling theory. Although, Laplace focused mostly on probability than on specific sampling techniques, his work still serves as a springboard upon which modern sampling techniques rest. Fisher's pioneering work in experimental design and statistical inference remains relevant till date. He pioneered pivotal concepts in sampling such as replication, randomisation which is the core of our current discourse, and blocking in experimental design. These are central to modern sampling theory. Fisher also developed the concept of analysis of variance (ANOVA), a current statistical tool and contributed meaningfully to the fundamentals of modern parametric statistical methods.

Neyman, alongside Egon Pearson, developed the concept of hypothesis testing. They introduced the Neyman-Pearson lemma, which provides a background and framework for constructing optimal tests of hypotheses and laid the foundation for modern statistical hypothesis testing. Their work underscored the importance of stating null and alternative hypotheses, stating the level of significance, and minimizing types I and II errors. Pearson specifically, also contributed to the development of non-parametric statistical methods, which are till date applicable in various sampling scenarios. Beyond the foregoing, he also contributed to the theory of confidence intervals, which are integral elements in sampling theory for estimating population parameters (Pitard, 2019).

Finally, Gy's contribution which was made popular through his work on Gy's sampling theory or theory of sampling (TOS), focused on improving sampling techniques in mineral processing and material science given his background in chemistry. He developed systematic sampling methods through his theory of sampling (TOS). This theory underscores the importance of representativeness of the elements in sampling and error estimation in collecting samples particularly from heterogeneous materials or population. The contributions of this theory have far reached practical implication and applications in academics and industry where accurate and representative sampling is a desideratum to researchers generally. (Esbensen, 2004).

In the final analysis, these works advanced sampling theory by introducing foundational concepts such as randomization, hypothesis testing, confidence intervals, and systematic sampling methods. Their contributions laid the groundwork for both frequentist and Bayesian approaches to inference and continue to shape modern statistical practice across various disciplines (Kubiak and Kawalek, 2022).

Conclusion

The paper concludes by noting that the concept of systematic sampling when used loosely as it is presently will require different approaches to it and may not be helpful. It is therefore posited that where there is no randomisation in any of the sampling intervals or where the intervals are not calculated but assumed such technique is ordinary systematic sampling and in the broader categorization, may not be classified as a probability sampling technique. Furthermore, where randomisation is applied in only one or the first sampling interval, the emerging technique is at best a quasi or pseudo-probability sampling as it does not guaranty a fair chance or equal probability of other elements being selected. Finally, study advances that the real systematic sapling requires the application of simple probability sampling method(s) in within each sampling interval in such a manner that every element in the population has a fair chance of being chosen at least from the skipping interval they belong. In the light of the foregoing, study suggests the application of any simple random sampling method in every sampling interval as against the first or only one interval only. This no doubt guarantees every member of the population a fair chance of being selected from the k^{th} term they fall in. The emergent method is then the real systematic probability sampling technique.

References

1. Agbonifoh, B.A. &.Yomere, G.O.(1999), Research Methodology for Social Sciences and Education.Benin City: Centerpiece Consultants.
2. Asika, N. (1999). Research Methodology in the Behavioural Sciences. Lagos: Longman.
3. Asika, N. (2006). Glossary of Terms and Concepts in Research and Statistics with Technical Notes.Lagos: Christian and Maxwell Publishers.
4. Bartlet, J.E., Kortrlil, J.W. & Higins, C.C. (2014).Organisational Research: Determining Appropriate Sample Size in Survey. https://www.opalco.com/wp-content/uploads/2014/10/Reading-Sample-Size1.pdf.
5. Calculator.net (2024). Sample Size Calculator. https://www.calculator.net/sample-size-calculator.html.
6. Catlow, B. (1993). Quantitative Methods for Computing Students: An Active Learning Approach. London: DP Publishers.
7. Clinicalc (2024). Sample Size Calculator. https://clincalc.com/stats/samplesize.aspx.
8. Cochran, W.G. (1977). Sampling Techniques. 3rd Edition, John Wiley & Sons, New York.
9. Esbensen, K.H. (2004). 50 years of Pierre Gy's "Theory of Sampling"—WCSB1: a tribute. Chemometrics and Intelligent Laboratory Systems. Volume 74, Issue 1.
10. Frost, J. (2024). Systematic Sampling: Definition, Advantages & Examples. http://statisticsbyjim.com/basics/systematic-samling/#respond.
11. Gy, P. (2004), Chemometrics and Intelligent Laboratory Systems, 74, 61-70. https://www.scribbr.com/methodology/systematic-sampling
12. Israel, G.D. (2022). Determining Sample Size. https://www.tarleton.edu/academicassessment/wp content/uploads/sites/119/2022/05/Samplesize.pdf.
13. Kubiak, A.P., & Kawalec, P. (2022). Prior Information in Frequentist Research Designs: The Case of Neyman's Sampling Theory. J Gen Philos Sci **53**, 381–402 (2022). https://doi.org/10.1007/s10838-022-09600-x
14. Mahmutovic, J. (2023). What is Systematic Sampling? Pros, Cons, and Examples. https://www.surveylegend.com/sampling/systematic-sampling/
15. Martin, C. (nd) The Quantitative Sample Calculator. https://blog.flexmr.net/sample-size-calculator.
16. Murray, R.S. & Larry, J.S. (2008). Theory and problems of statistics. New York: McGraw Hill.
17. Nanjundeswaraswamy, T.S. & Divaka, S. (2021). Determination of Sample Size and Sampling Methods in Applied Research. Proceedings on Engineering Sciences, Vol. 03, No. 1 (2021) 25-32, doi: 10.24874/PES03.01.003 https://oaji.net/articles/2021/8593-1615826931.pdf.
18. Nih.gov (2024) Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. www.ncbi.nlm.nih.gov/pmc/articles/pmc4286359.
19. Ofo, J.E. (1994). Research Methods and Statistics in Education and Social Sciences.Lagos: Joja Educational Research and Publishers.
20. Pitard, F. (2019). Theory of sampling and sampling practice (Third ed.).mBoca Raton, FL: Chapman and Hall/CRC.
21. Qualtric (2023). Sample Size Calculator. https://www.qualtrics.com/blog/calculating-sample-size/.
22. Raosoft (2004). Sample Size Calculator by Raosof. http://www.raosoft.com/samplesize.htmlt.
23. Shalabh, (nd). Sampling Theory https://www.studocu.com/row/document/makerere-university/quantitative-economics/chapter-11-sampling-systematic-sampling/37913584
24. Surveymonkey (2024). Sample Size Calculator and Tips for Determining Sample Size. https://www.surveymonkey.com/mp/sample-size-calculator/
25. Thomas, L. (2023). Systematic Sampling | A Step-by-Step Guide with Examples
26. Van-Haute, E. (2021). Sampling Technique. Sample Type and Sample size. https://www.researchgate.net/publication/348805611_Sampling_Techniques_Sample_Types_and_Sample_SizeDOI:10.1093/hepl/9780198850298.003.0057