Review paper

AN EXAMINATION OF REPRESENTATIVE SAMPLE IN DATA-BASED RESEARCH

ABSTRACT

The output of any research work depends, to a reasonable extent, on the adequacy of the sample from which data are obtained for the research. A sample is adequate when it is representative of the population or when it possesses the characteristics that are typical of the population from which it is drawn. The focus of this paper is on how researchers can select samples that are really representatives. Descriptive method was utilized in the writing of the paper. Attempts are made, in the paper, to describe the concepts of sample representativeness, sampling methods and sampling error. Some suggestions are then made on measures to take in carving out truly representative samples.

Keywords: ***Selection, Sample, Representative sample, Descriptive method, sampling error.***

INTRODUCTION

Information Technology has dominated the world today and radical technological change and fusion have changed the way work is organized and performed [1]. To this end, frantic efforts are made to obtain relevant information on different aspects of human life.

Managers are fond of making decisions on different areas of business management in order to achieve organizational objectives. The decision making process requires certain information that will provide essential clues to certain issues. The information so needed could be derived from hundreds or thousands of prospective respondents who could be consumers, suppliers, employees, scholars or government officials. As a result of the largeness of the population being dealt with, a researcher may need to carve out just a fractional segment of the entire population to ensure easy and timely collection of data. Herein lies the relevance of sampling.

It is one thing to create a sample out of a particular population; it is another thing to ensure that the sample is large enough to represent the population. By population, we mean the totality of items in an investigation [2]. In other words, population is a collection of the individual items that are to be observed in a given problem situation; the items could be living or non-living things.

A sample is a selected portion or subset of the population being investigated [3]. The elements in the sample possess the same characteristics with the population but the differences are in the size; that is, sample < population [4].

A quality control manager may decide to take a few cocoa beans, say 50, from a

bag of cocoa to investigate the quality level of the content of the bag. The thinking of the quality control manager is that by examining 50 out of the entire cocoa beans in the bag, he will get a fair notion of the quality of the entire population cocoa beans inside the bag.

The objective of this paper is to find answers to the following questions:

i. What are the different methods of creating samples?

ii. What should be done to ensure selection of a truly representative sample?

iii. How can sampling errors be reduced?

In view of the nature of this study, descriptive method was used in writing, the paper.
Towards this end, efforts were made to describe sampling methods, sample representativeness and sampling error.

SAMPLING METHODS

Sampling is the process of taking any part of a population or universe as representative of that population or universe [5]. Another way of saying the foregoing is that sampling is a means of estimating population parameters from only a few items. The process takes place especially when the population being studied is a large one such that it becomes impossible or costly to investigate each item in it (population). What could be done in that situation is to select a few elements from the whole of materials being investigated and then make a generalization about the population [6]. The method by which sample was drawn determines, to a reasonable degree, the extent to which generalizations about the population can be made [7]). The sampling method so used determines the representativeness of the sample in relation to the population.

Sampling methods can be categorized into two main classes namely, probability sampling techniques and non-probability sampling techniques.

1. PROBABILITY SAMPLING METHODS:

These are sampling techniques for which we can determine the chance of drawing each member of the population to form a sample. The method describes a situation in which each item within the population has equal chance of being chosen to form a sample [8]. Thus, there is no bias in the selection of sample members. One other major advantage of this set of sampling methods is that it is easy to measure the sampling error and interpret sampling results. As a result of these favourable features, conclusions reached by studying a particular sample are considered generalizable to that population or other similar populations. Some of the probability sampling methods are:

Simple Random Sampling: This sampling method Involves selecting a few elements from a total population in such a way that each member of the population has an

equal chance of being selected. Thus, a sample drawn at random is unbiased in the sense that no member of the population has any more chance of being selected than any other member [9].

One means of carving out a sample from the entire population is to name or number each item in the population. Next, each name or number is cut into a small slip and squeezed and then placed in a container. The investigator puts his hand into the container to pick the number of slips required to obtain the required size of the sample.

The selection procedure described above may not be possible in a situation where the studied population is very large. For instance, taking a random sample of the population of Lagos State, a table of random numbers, an example of which is shown in Table 1, becomes useful. The computer is often used to accomplish the selection process at this level.

Table 1

Random Numbers									
63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	954336	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	54358	21625	16999	13385	22782
55363	07449	34826	15290	76616	67194	18277	21151	68684	08263
69393	92785	49902	58477	42048	30378	87618	26933	40640	16281
13186	29431	88130	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84694	48968	75215	75498	49539	74240	03466	49292	36401	45525
63231	11618	12631	75055	43915	26488	41116	64551	56827	30825
70502	53225	03655	05915	37140	57051	28393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	64342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09926	64419	29457
10078	28073	85298	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	76607	52244	63303	10413	63839	74762	50289

Source: Anderson, David R. et al (2003), Pg. 258.

Stratified Sampling: According to [10], stratified sampling method is best suited to populations that have different sets of groups within them. In other words, the sampling method is mostly used when dealing with heterogeneous populations. For instance, if a researcher wants to collect relevant data on a topic that says "Life after death" from a set of people, the best sampling method to use is stratified sampling. The reason being that the population being sampled will comprise people of different religious beliefs who are bound to have different opinions on the subject matter. In this case, the heterogeneous population will have to be divided into three homogeneous groups or strata as follows; Christians, Muslims and Traditional believers.

Stratified sampling is of two types namely, proportionate stratified random sampling (**PSRS**) and disproportionate stratified random sampling (**DSRS**) [7]. In **PSRS,** the population is first stratified in terms of one or more variables of interest to the researcher. Elements are drawn randomly from each stratum in such a way that the relative proportions of the strata in the resultant sample are the same as exist in the parent population. This is saying that the relative contribution of each stratum in the population is exactly its relative contribution in the sample.

Proportionate stratified random sample ensures greater representativeness of the sample relative to the population and guarantees that minority constituents of the population are represented in the sample **[7]** and **[3]**. Table 2 below illustrates **PSRS** with a population of 1220 entrepreneurs.

Table 2: Distribution of a population of 1220 entrepreneurs according to categories/sizes with the proportionate stratified random sample.

Entrepreneurs	Small Scale	Medium Scale	Large Scale	Total
Population Size	549	427	244	1220
Proportion	0.45	0.35	0.20	1.00
Sample size	110	85	49	244

In case of disproportionate stratified random sampling, the relative proportions of the strata in the sample do not correspond to their relative proportions in the population. Some strata may be under-represented or over-represented in the sample. Obviously, this sampling mode allows the researcher the freedom of weighting the various strata in any manner he considers fit. Though the method does not make for proper representativeness, [3] was of the opinion that the **DSRS** method is preferred where the researcher believes that there is likely to be great within-stratum variation in responses or if he has a particular interest in one or more strata.

Systematic Sampling: It is also called quasi-random sampling. In this method, the first sample clement is randomly chosen from numbers I through K and subsequent elements are chosen at every K^{th} interval [5]. K will be determined by the size of the sample required. For example, if the population of Texy University is 16,000 and a

list of all the students making up the population is available. **If** a sample of 200 students is to be taken, the selection of every 80th student will give the required sample.

The value of K above is determined as follows.

$$\frac{Total\ population}{Required\ sample} = \frac{N}{n}$$

Therefore,

$$k = \frac{16{,}000}{200}$$

$$= 80.$$

It must be noted that systematic sampling is used only when the list of all items in a population is available as in the case of a class register. The question as to how to determine the first element that will form the sample is answered by choosing a number at random between 1 and 50.

Cluster Sampling: Cluster sampling is otherwise known as area sampling. It is successive random sampling units or sets and subsets [11]. In selecting a sample using this method, the population (or geographical area) is divided into units or segments with well-spelt-out boundaries. A specified number of these units or a section is drawn. All elements in the units or sections drawn now constitute the sample.

Cluster sampling is used when it can be recognized that some populations are distributed in clusters or groups of settlement and these clusters are to be used as the basis for sampling [3].

Unlike stratified sampling, using cluster sampling does not require a list of the elements in the population before sample can be drawn [12]. As long as these are distinguished clusters or geographical locations, creating samples becomes feasible.

2. NON-PROBABILITY SAMPLING METHODS:

Non-probability sampling is non-scientific approach to sample formation [13]. It is the process of getting samples from populations without following any statistical rules. The researcher or investigator uses only his intuition to select sample members is fraught with bias and partiality simply because each item in the population does not have an equal chance of being selected. The ultimate consequence of this abnormality is that it becomes difficult to measure the sampling error and to interpret results. The following techniques are examples of non-probability sampling methods:

Judgmental Sampling: This involves the use of the researcher's reasoning and judgment to obtain a sample, what determines whether an item in the population will be selected or not into the sample is the personal preference of the investigator.

Sampling here is influenced by the personal bias of the person or group of persons selecting members for the sample. That is not to say, however, that judgment sampling is a complete write-off. It is advantageous in the sense that it saves time as the process does not require any listing or numbering of the population or random number tables [9].

Quota Sampling: According to [7], this involves selecting those elements that have specific characteristics of interest to the researcher and are accessible to him. This type of sampling is used to ensure that specific elements will be included.

Evidently, quota sampling gives room for the researcher or investigator to include any category of the population that is of particular interest to him. It is quicker, easier and cheaper. It is disadvantageous in the sense that the resultant sample is highly biased and thus, cannot be said to be representative [14].

Convenience Sampling: The major consideration of the researcher using this sampling method is the east of selecting the sample. The researcher makes no attempt to bring in any element of randomness. He selects members of the sample in a way that is convenient or easy for him.

As an illustration, a student of Adekunle Ajasin University, Akungba-Akoko, Nigeria, researching on the topic "The problem of small-scale business entrepreneurs in Ondo State" may decide to get the required sample of small-scale business entrepreneurs from Ikare-Akoko and Owo. This is because the two towns are very close to his institution of research and so, can easily or conveniently collect the needed data from respondents in the two towns.

In fact, the researcher does not know how well a Convenient Sample will represent Sample the population regarding the traits or mechanism under research. What makes convenience sample so unpredictable is their vulnerability to severe hidden biases [15].

Therefore, in convenience sampling, the individuals selected by the researcher may not be applicable to the research problem.

Hence, there is a rise of collecting poor quality data due to poor outcomes such as, difficulty to convince others to accept the findings or research based on poor foundation [16].

Snowball Sampling: This method is often used to obtain samples in situations where there is no adequate list which could be used as a sampling frame [9]. Towards this end, the researcher contacts a member of the population of interest or identifies a group of respondents who possess the traits desired for the research

work [3]. This set of people will in turn, identify another set of people suitable for the research work. This chain continues.

SAMPLING ERROR

Sampling error simply means deviation from population values. It is the difference between the result obtained from a sample and the result which would have been obtained from the population. This type of error often occurs when tile complete survey of the population is not carried out, but a sample is taken for estimating the characteristics of the population [17]

When the entire population is considered no sampling error occurs. The occurrence of sampling error when a sample is carved out of a population depends on the size of the sample relative to the population. As shown in Figure 1, the smaller the sample the larger the error, and the larger the sample the smaller the error.

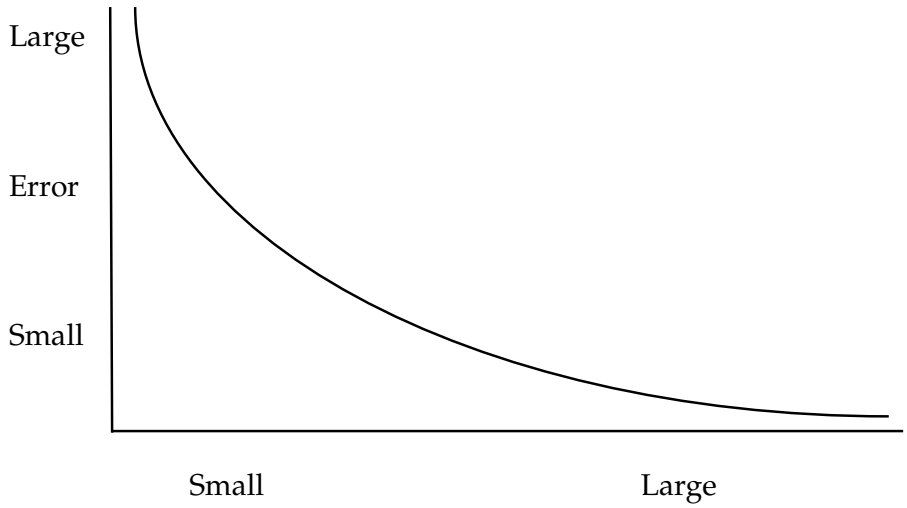

Figure 1: Sample size vis-à-vis sampling error

It is possible to measure the sampling error [8]. The error is measured by the standard error of the statistics in terms of probability under the normal curve. The result of the measurement indicates the precision of the estimates of the population based on the sample study [9].

Sampling error is derived by calculating the standard error of sample mean. It is the standard deviation of the sampling distribution of sample means [6]. It is denoted by δx and is given by

$$\delta x = \frac{\delta}{\sqrt{n}}$$

$\delta =$ Standard deviation of the population

$n =$ Sample size

It is often advocated that large samples should be used. They are advocated in order to give the principle of randomization a chance to work [5].

SAMPLE AND REPRESENTATIVENESS

The word "representative" means to be typical of a population; that is, to exemplify the characteristics of the population. The fact that a segment of a population is taken as representative of that population does not mean that the sample, so taken, is representative. For example, a researcher investigating a banking habit in First Bank Plc, may decide to draw his sample from two branches of First Bank in Lagos, Nigeria, thereby assuming that the two branches represent the total population of the bank's branches. That sample selection may be wrong as data and opinions collected from the two branches may not be representative of what actually happens in the entire First Bank Plc network. That goes to show that too small samples are not good enough as they tend not to represent the characteristics of the population. This will eventually result in getting results that are likely to be lacking in validity.

In research, a "representative sample" means that the sample has approximately the characteristics of the population relevant to the research in question [5]. If sex and socio-economic class are characteristics relevant to the research, a representative sample will have approximately the same proportions of men and women and middle class and working class individuals as the population.

The question as to how large a sample must be to be adequate or to be truly representative is not a simple one. Each situation presents its own problems. If the phenomena under study are homogenous, a small sample is sufficient. If units under study are variable, a much larger sample is needed. The greater the variability of the phenomena, the greater the difficulty of obtaining an adequate sample. But that is not to say that using very large samples is always wise. Making use of samples that are too large amounts to waste resources - money, time and energy.

In order to obtain a representative sample, the following points need to be borne in mind by the researcher.

1. There is need for care and precision on the part of the researcher. He must take care to see that the sample drawn from the population is not biased. Towards this end, it is preferable to use probability sampling methods in selecting sample members since such a procedure guarantees that every population element has equal chance of being selected.

2. The population being investigated must be properly defined. It should be

defined in terms of four things namely, the element, sample unit, extent and time [18]. For instance, a survey of consumers might specify a relevant population as follows:

3. Elements : Male
   Sampling Unit : 27-35 years
   Extent : Kaduna State, Nigeria
   Time : As at 30th October, 2016

In the alternative, the population for a study designed to measure buyers' reaction to a new pharmaceutical item may be the following:

Element : Pharmacists
Sampling Unit : Pharmaceutical companies buying over ₦5 million worth of item per year.
Extent : Enugu State, Nigeria
Time : Year 2014

The essence of defining studied population properly is to ensure that whatever data are obtained from the selected sample meet the researcher's expectation in terms of data currency, respondents/interviewer's level of experience and geographical/professional coverage.

4. The usual thing to do after identifying a population is to obtain a complete and current list of all elements in the population. Nonetheless, in order to get a representative sample, the researcher should avoid drawing a sample by preparing an alphabetical list and proceeding down the list until he has included a sufficient number in the sample. Some letters of the alphabet include more names from certain groups than others, and this may produce bias in the sample on the basis of name alone.
5. The idea of allowing interviewers to select interviewees who are like themselves should be discouraged. Such a habit results in a biased sample. For example, if interviewers or researchers are members of a particular religion, they will likely select many interviewees from the same religion and include just a few individuals from other religions. Samples obtained through such an awkward procedure will produce opinions and data that are not representative of the entire population [14].
6. [3], in his own contribution, opines that samples must be adequate enough to generate generalizable result of findings. There is need for researchers or investigators to ensure that the elements that make up a particular sample are those that possess the features that are typical of the entire population.

CONCLUSION

The validity of the outcome of any research depends, to some extent, on the extent of adequacy of the sample used in the course of the research. [13] states that

the size of sampling error indicates the reliability or precision of the research. According to him, as the sample size increases, the error decreases and hence larger samples are considered more reliable and more representative than smaller samples.

In view of the fact that representative samples invigorate and accelerate the process of research, there is need for researchers' training in the special area of "sample representativeness", According to [19], such training could lead to improved research and accelerated technological innovation.

In this paper, some suggestions are made regarding how representative samples can be carved out from designated universe or populations. If the suggestions are strongly adhered to by researchers, random fluctuations or sampling errors will be minimized. The lower the sample errors, the more representative a sample becomes.

REFERENCES

1. Dauda, Adeleke, Research and Development for Organizational Competitiveness. *Management Discoveries.* 2007:(1), 16-18.
2. Gupta, C.B. & Gupta, Vijay. *An Introduction to Statistical Methods.* New Delhi:Vikas Publishing House PVT Limited. 1987.
3. Oladele, Olajide. *Introduction to Research Methodology.* Lagos: Niyak Print and Publications. 2007.
4. Ayodele, Sam O., Adegbile Joseph A. & Adewale J. Gbenga. *Evaluation Studies,* Ibadan: Powerhouse Press and Publishers. 2003.
5. Kerlinger, Fred N. *Foundation of Behavioural Research* (3rd ed). New York: Harcourt College of Publishers. 1992.
6. Arora, P.N. & Arora. S. *Statistics for Management (4th ed). New Delhi:* S. *Chand and Company Ltd.* 2009.
7. Nworgu, B.G. Educational Research: *Basic Issues and Methodology.* Ibadan: Wisdom Publishers Limited. 1991.
8. Pearson, W.A. Management Development for Scientists and Engineers. Research a*nd Technology Management Journal, January/February,* 1993:19-21.
9. Osuala, E.C. Introduction to Research Methodology (3rd ed). Onitsha: Africana Publishers Limited. 2001.
10. Anderson, David R; Sweeney, Dennis J &Williams, Thomas A. *Modern Business Statistics,* Australia: South Western. 2003.
11. Arikpo, M. *Foundation of Social Science Research: A Methodological Guide for*

406 *Students.* Enugu: ABIC Publishers. 1986.
407 12. Ferguson, G.A. *Statistical Analysis in Psychology and Education*(4th ed). New
408 York: McGraw-Hill.1981.
409 13. Monga, G.S . Mathematics and Statistics for Economics (2nd ed). New
410 Delhi: Vikas Publishing House PVT Ltd. 2000.
411 14. Gupta, S P & Gupta. Archana. *Statistical Methods.* New Delhi: Sultan Chand
412 and Sons. 2007.
413 15. Leineer, D.J. Convenience Samples and Respondent Pools. 2014: 1-36.
414 16. Oppong, S.H. The problem of Sampling in Qualitative research. Asian
415 Journal of Management Services and education, 2013: 1-9.
416 17. Green, Paul E. & Tull, Donald S. *Research for Marketing Decisions*(4th ed).
417 New Jersey: Prentice-Hall, Inc. 1978.
418 18. Moser, A. Claus & Kalton, Graham. *Survey Method in social Investigation*
419 (2nd ed). London: Heinemann Educational Books Ltd. 1979.
420 19. Bushnell. D.S. Input, Process, Output: A model or Evaluating Training.
421 *Training and Development Journal,* London, 1990; 31.
422