EUROPEAN ACADEMIC RESEARCH
Vol. III, Issue 2/ May 2015

Impact Factor: 3.4546 (UIF)
DRJI Value: 5.9 (B+)

ISSN 2286-4822
www.euacademic.org

Essentials of Sampling in Management Research

NIHARIKA SINGH
Research Scholar
Department of Management
Mizoram University, India
Dr. S. N. SINGH
Associate Professor
Department of Library & Information Science
Mizoram University, India

Abstract:

To complete a research, collecting data is a challenge. It would be great and simple if researchers could collect data or study each and every unit of population, which is next to impossible in almost all the field, including management. Hence, only a part of the population (sample) is studied. Sample should be adequate to warrant generalization of the findings to the target population, which depends on how the samples are selected from population. Hence, this chapter deals with process of selecting sample from target population i.e. sampling and related with, in most diluted way possible to match with beginner's knowledge. The paper is based on secondary resouces available on the topic.

Key words: Sampling, Sampling method, probability, non-probability sampling, simple random sampling, quota and convenience sampling.

1. Useful terms and definition

- Census- To collect and analyze data from every possible group member.
- Unit- an element or a group of elements, living or non-living on which observations can be made. E.g.-

Household in the city, an employee, or a branch in a bank etc.

- Population (or Universe)- collection of all the units of a specified type at a particular point or period of time. E.g.- Households in a city, all the branches of a bank. Population can be finite i.e. the number of members of the population can be expressed as a definite quantity or infinite population i.e. there is no limit of number of members of the population. Also, population can be existent i.e. when population really exists, such as income of all persons in a city or hypothetical i.e. when population is hypothetical, such as points obtained in all possible throws of a dice. Designated as "N" in most of the cases.
- Sample- portion of a population that has been selected for study. E.g.- 100 households from a city or 24 branches of a bank.
- Sampling/Sampling Design- the process of selecting a sample from population.
- Sample size- Number of units selected in the sample.

Sample size depends on- a) Population Size- normally bigger the size of population, bigger will be sample size. b) Heterogeneity-In the population concerned characteristic e.g. age or income in the case of a human population or life of electric bulbs in the case of a physical system, more heterogeneity in population is equal to bigger sample. c) Accuracy and reliability- If more accuracy of result is required, sample size has to be bigger.

- Sampling Frame- list of all the populations unit with their identification number, from which sample is to be drawn. Each population unit is given a identification number from 1 to N. Eg.- All houses in a particular locality/colony numbered as 1,2 and so on.
- Sampling with replacement, where once a unit is selected and information is noted down about it, we return it back to the frame. Each unit has the same

probability of getting selected again. E.g.- From a bowl we pick business card of Ram and put it back to the bowl after noting down information about Ram. Population remains the same before each drawing and any of the population unit may appear more than once in the sample.

- Sampling without replacement, where once a unit is selected, cannot put it back and hence cannot select it again. Chance that a card will be selected not previously selected on the second row is now 1 out of N-1 (since one card has already been chosen which cannot be placed back in the frame). Here, either the sample are drawn all at a time, or drawn one by one in such a way that after each drawing selected unit is not returned to the population.

2. Introduction

In most of the researches, it is impossible for the researcher to study all the population. Hence, a sample is selected. There are a number of methods to choose sample from population, known as *samplingmethods/techniques*. In order to make a good estimate of the population characteristics selecting a reasonably good sampling method is of paramount importance.

However it has been proved empirically that the census may not lead to more reliable results as compared to studying only a sample from the population**.**

One important point while sampling is to ensure that the sample is a true representative of the population i.e. if one looks at the sample with a magnifying glass, the sample should appear like the population, or in other words, the sample should be like the photo or image of the population. If due care is taken in selecting a representative sample from the population, the results obtained will generally, not only be more reliable and accurate but also will consume lesser resources in terms of manpower, time and money etc. Collecting data from fewer cases also means that you can collect information that is more detailed. In addition, if you are employing people to collect data (as interviewer) you can afford high quality staff. You can also devote more time to trying to obtain data from more difficult to reach cases.

The possibility of reaching a valid conclusions concerning a population by means of properly chosen sample is based on two important laws- (i) *Law of Statistical Regularity,* means sample of reasonably large size when selected at random is almost sure to represent the characteristics of the population. (ii) Law of inertia of large numbers, says samples of large size shows a high degree of stability i.e. the results obtained theorem are expected to be very close to the population characteristics.

A number of sampling techniques are used by researchers, basically classified under probability/random and non-probability/non-random sampling, will be discussed in forthcoming sections.

3. Objective:

The primary objective of this review paper is to provide a detailed summary of various sampling methods used in management research.

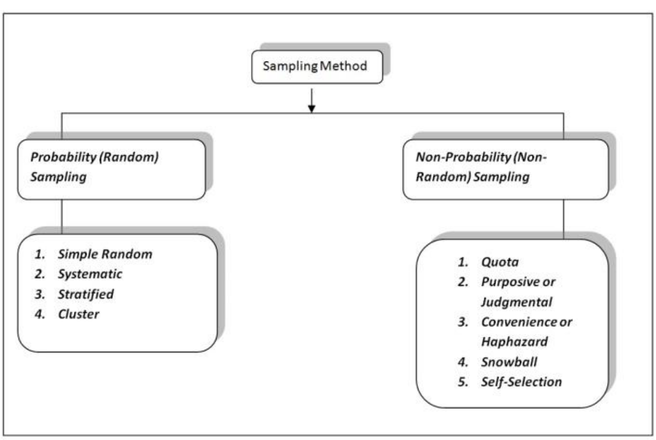
4. Methodology:

The present study is based on descriptive approach, where available literature has been reviewed and summarized in the most diluted way for the understanding of beginners.

5. Sampling Method

To obtain a true representing sample, number of sampling methods / techniques is there with their own advantages and disadvantages. All the techniques are categorized under random and non-random or probability and non-probability sampling, respectively. In random sampling each unit of population has the same probability of being selected as part of the sample and hence also named as probability sampling. Whereas, in non-random sampling every unit of the population is not equally likely to be selected, assigning a probability of occurrence in non-random sampling is impossible. In random sampling it is possible to answer research questions and to achieve objectives that require us to estimate statistically characteristics of the population from the sample. For non-probability sampling, it is impossible to answer research questions or to address objectives that require you to make statistical inferences about the characteristics of the population. We may still be able to generalize from non-probability sample about the population but not on statistical grounds. Under non-probability sampling, methods such as quota, purposive or judgmental, convenience or haphazard, snowball and self-selection falls. And probability sampling method has simple random, systematic, stratified and cluster sampling.(See Figure 1) These are the most vital and highly used techniques of probability and non-probability sampling. For many research projects we need to use a variety of sampling techniques at different stages.

Figure 1: Types of Sampling Method



5.1 Probability Sampling/Random Sampling method can be divided into 4 steps-

a. **Identify a suitable sampling frame based on your research question(s) or objectives:** We need to define our sampling frame and ensure our sampling frame is a complete, accurate and up to date, as possible. Where no suitable list exists we will have to compile our own sampling frame, perhaps drawing upon existing lists. By defining sampling frame we are defining the population about which we want to generalize. This means, if sampling frame is list of employees of organization XYZ, strictly speaking can generalize findings based on our sample to that population only(i.e. employees of the organization XYZ)

b. **Decide a suitable sample size**: As larger our sample size as lower the likely error in generalizing about population. Probability sampling is therefore a compromise between the accuracy of our findings and the amount of time & money we invest in collecting, checking and analyzing the data. Our choice of sample size within this compromise depends on:
   - Confidence we need to have in our data, i.e. the level of certainty that characteristics of the data collected will represent the characteristics of the population.
   - Margin of error that is tolerable, i.e. the accuracy you require for any estimate made from our sample.
   - Size of the population.

Law of inertia of large numbers states that samples of large size shows a high degree of stability i.e. the results obtained are expected to be very close to the population characteristics.

Also, to ensure spurious results do not occur, the data analyzed must be normally distributed. Statisticians have proved that the larger the absolute size of a sample, the more closely its distribution is to the normal distribution and thus more robust it will be. This relationship is called as **Central Limit Theorem.**

Ex-Sample is software which calculates the minimum sample size required for different statistical analysis and maximum sample size in limited resources.

c. **Estimating response rates and actual sample size demanded:** With all probability samples, it is important that your sample size is large enough to provide us with necessary confidence in our data. We therefore need to estimate the likely response rate- that is, the proportion of cases from our sample who will respond and increase the sample size accordingly. Once we have an estimate of the likely response rate and the minimum or the adjusted minimum sample size, the actual sample size demanded can be calculated using the formula:

$$n^a = n * 100 / re\%$$

where, n^a = actual sample size required
n = minimum(or adjusted minimum) sample size
re%= the estimated response rate expressed as a percentage

To estimate the response rate, a small part of the sample can be asked to fill up the questionnaire or interviewed. Other way of obtaining this estimate is to consider the response rates achieved for similar surveys that have been undertaken and base estimate on these.

d. **Selecting the most appropriate sampling techniques and the sample**: Five main techniques can be used are simple random, systematic, stratified random, cluster and multi-stage sampling. Our choice of sampling technique depends on our research question(s) and our objectives. Subsequently, our need for face-to-face contact with respondents, the geographical area over which the population is spread, and the nature of our sampling frame will further influence our choice of probability sampling technique.

5.1.1. Simple Random Sampling (also called lottery sampling): is based on the process of selecting a sample randomly. This does not mean that randomness allows haphazard selection of samples; it means that process of selecting sample should be free from human judgment (bias). Random sampling is most appropriate in case when the population is more or less homogenous with respect to the characteristic under study. There are two methods of drawing a random sample from the population. These two methods are: (1) the lottery method, (2) the use of random number table or random number generator.

In the lottery method, each unit of the population are numbered from 1 to N. The numbers are written on different pieces of paper. The pieces of paper are then folded and mixed in a small box. A sample can be drawn randomly from box by selecting papers randomly. In second case, (a) each element of the population or our sampling frame is given a unique number from 1 to N. (b) n items are selected using random number table or random number generator.MS Excel can generate random numbers and the elements given those numbers are selected for the study.

Whereas, random number table (see Figure 2) consists of a series of digits listed in a randomly generated sequence. Some of the famous series of random numbers are those of Trippett (41,600 digits), Kendall and Babing ton Smith(100,000 digits), Fisher and Yates (15,000 digits), Rand Corporation (1,000,000 digits).For coding elements of population, if say 8 (one digit number) or 800(three digit number) as largest numbers, we code each unit in as many digits as there are in largest number in the population. For 8, we code all the elements as 1, 2, 3 and so on; whereas for 800 we code elements as 001, 002, 003 and so on. Then, from random number table n random numbers are selected, each less than or equal to N, we discard all the values greater than largest number i.e. N. There are a number of ways to choose n random numbers from random number table, one among them is- to choose an arbitrary starting point from the table. One method we can use is to close our eyes and strike the table of random numbers with a pencil, which may give a result as row 06,column 05 as the starting point. We can go in any direction from here, say read from left to right, in sequence of 3 digits, without skipping. Discard numbers higher than N. Also, if number gets repeated-ignore.

Simple Random Sampling is best used when you have an accurate and easily accessible sampling frame that lists the entire population. Consequently, this form of sampling is not suitable if we are collecting data over a large geographical area using a method that requires face to face contact, owing to the associated high travel costs.

Figure 2: Random Number Table

403357	584568	253645	260890	407500	265296	681128	570951	516921
201871	786456	993651	855088	31543	990763	494506	752343	973925
858288	459616	100839	552148	852501	626129	427594	611462	382556
660584	810314	307438	557378	216803	435706	280482	820649	808002
902550	156411	320641	222470	85993	745305	849736	466620	889012
202902	911252	621463	603735	494251	247152	491466	898310	515153
385179	50503	198052	281988	620282	5752	698288	94884	64136
381475	672508	908982	327290	529139	882432	575894	61968	796725
729469	328949	823579	575960	716793	837435	656479	492098	105816
174027	329114	436430	387472	569764	862335	200473	57915	990891
415688	681857	727050	567281	475091	79469	676471	835340	87538
420280	785076	438222	992046	99468	85042	410722	908049	161100
565343	443619	997387	548183	918697	733405	984689	844510	147464
918769	440120	631036	620502	161593	215769	274786	882919	362710
183913	416395	625085	171112	705155	539739	984285	876244	804746
886849	953816	634562	247629	173571	855191	37999	399730	960394
295931	545798	488637	112804	61155	814766	23844	381276	722959
978100	129418	223975	866755	612098	628305	794422	232788	425105
566500	806949	260352	128344	678044	284666	916705	176031	10579
478252	121246	965763	971577	68909	967961	557017	676454	664326
663213	626926	250062	66843	678385	634716	757387	510524	70232
920344	911148	599270	645723	860396	162251	678779	495981	236602
808621	682437	423611	494134	96529	226545	111965	819428	812156

5.1.2. Systematic Sampling (also known as quasi-random sampling): In simple random sampling, we must number the units in population. However, numbering all the population units can be quite time-consuming. As an alternative to simple random sampling, systematic sampling is used.

If the frame consists of a list of pre numbered checks, sales receipts or invoices, a systematic sample is easier to get than simple random sample. Systematic sampling involves selecting the sample at regular intervals from the sampling frame. To select systematic sample- 1. Number each cases (or use pre numbered cases), and 2. Calculate the sample fraction i.e. k= N/n

For eg, N = 900 and n = 30 , then k = 900/30 = 30

At some places, k is given as I and named as sampling interval or sampling cycle. The interval/fraction gives the difference among successive units to be selected in the sample. If I or k is not an integer, then it is rounded off to the nearest whole number, such as 10/3 =3.3. Rounded off to 3.

Now, for obtaining the starting point of sampling process, a random number table can be used. For eg., k= 30 ; a researcher will use random number table to get the first item between 0 and 30. Suppose 5 is selected, hence every 30th item (or ithitem) from 5 will be included in the sample, such as 5th, 35th, 65th, 95th and so on.

Selecting every ithitem in the population is not random. To overcome this boundary, first element is selected randomly, But the other units are still selected on ithitem basis and totally dependent on the selection of first unit. Also, another problem with systematic sampling is if the data is periodic. That is, suppose list from which selection is to be made is combination of 3 lists where item 1-50 is high income group, 51-100 is middle income group and 100-150 is low income group. In such cases, there might be a scenario if using systematic sampling that all the high income group or low income group members are selected, which makes sample biased.

5.1.3. Stratified Random Sampling: is a modification of simple random sampling in which you divide the population or heterogeneous population into homogenous groups, known as *strata*. Next, sample is selected from each stratum (singular of strata) using either simple random or systematic sampling. Stratum is relatively homogenous and strata are heterogeneous among each other. Process of dividing heterogeneous population into homogenous is known as *stratification.*

The stratification variable(s) i.e. the basis on which stratification is done (such as gender, income, age and alike) should represent the discrete characteristics for which we want to ensure correct representation within your sample. That means, if we want to make sure employees from all income groups are included, we use income as variables of interest for stratification.

Dividing the population into a series of relevant strata means that the sample is more likely to be representative, as we ensure that each of the strata is represented proportionally within our sample. For eg., to select 100 customers from population of 1000 customers. If selected randomly, may not be true representative but if divided into strata gives representative from all sections which makes data more reliable.

It follows that a stratified sample provides more accurate information than a random sample of the same size. As a simple example, if all the units in each stratum were exactly the same, then examining only one unit in each stratum would allow us to describe the entire population. Further, stratification can make a sample easier (or possible) to select. Since, in order to take a random sample we must have a frame. Although a frame may not exist for the whole population but it may be there for each stratum.

Stratified random sampling can be either *proportionate* or *disproportionate.*

Proportionate –where percentage(%) of sample taken from each stratum is proportionate to the actual percentage of the stratum within the whole population. (see Table 1a)

Disproportionate- where the percentage of sample taken from each stratum is disproportionate to the actual percentage of the stratum within the whole population.(see Table 1b)

Table 1(a&b): Proportionate and Disproportionate Stratifies Random Sampling

Stratum ⟹		**Lower Level Manager**	**Middle Level Manager**	**Top Level Manager**
(a)	**% of stratum within whole pop.**	75%	15%	10%
	Sample from each stratum	750	150	100
(b)	**Sample from each stratum**	600	200	200

At some places, terms like proportionate allocation or equal allocation are used which signifies equal sample selection from each stratum.

5.1.4: **Cluster (or Area) Sampling**: In cluster sampling, we divide the entire population into non-overlapping areas or clusters. The difference between stratified and cluster sampling is that strata happens to be homogenous but clusters are internally heterogeneous. For cluster sampling, our sampling frame is the complete list of clusters rather than a complete list of individual cases within the population. Sometimes, cluster sampling is the only feasible approach because the sampling frame of the individual elements of the population is unavailable and therefore other random sampling techniques cannot be used. Cluster sampling is very useful in terms of cost and convenience. Compared to strata, clusters are easy to obtain and focus of the study is clusters instead of the entire population, so cost is also reduced.

There are two types of cluster sampling:

5.1.4.1. Single-stage cluster sampling- we divide entire population into clusters, and randomly select few of them and include all the items/elements from selected cluster(s). Selecting clusters randomly make cluster sampling a probability sampling technique. Despite this, the technique normally results in a sample that represents the total population less accurately than stratified random sampling.

5.1.4.2. Multi-stage cluster sampling (or multi-stage sampling)- is a development of cluster sampling. We divide entire population into clusters, and rather than choosing all the elements from the clusters in multi-stage sampling we randomly choose few units from selected clusters. In this way, more number of clusters can be included in the sample for same sample size.

Such a scheme could be extended to 2, 3 or more stages, called as two stage sampling, three stages sampling, respectively. In two-stage sampling, sample is selected in two stages. The clusters selected randomly at first stage –first stage units, and the units or the group of units within clusters which are selected subsequently- second stage units.

5.2: Non-Random/ Non-Probability Sampling: The techniques for selecting samples discussed before have all been based on the assumption that our sample will be chosen statistically at random. However, in some cases, due to inability to specify sampling frame, limited resources such as money, time or to explore something new non probability sampling techniques are demanded.

Non-probability sampling is divided into two stages:

- **Deciding on a suitable sample size:** Unlike probability sampling, there are no rules. Rather the logical relationship between our sample selection technique and the purpose and focus of our research is important. Consequently, our sample size is dependent on our research question(s) and objectives.

- Selecting the most appropriate sampling technique and the sample: Next is to select the most appropriate sampling technique to answer our research question from the range of non-probability sampling technique available. At one end of the range is quota sampling, which, like probability sample tries to represent the total population. At the other end are techniques, where there is no attempt to obtain a representative sample which will allow us to generalize in statistical sense to the population. These include convenience or haphazard sampling & self-selection sampling. Purposive or Judgmental sampling and snowball sampling technique lie between these two extremes.

Some of the techniques are discussed here-

5.2.1. Quota Sampling: is entirely non-random and is normally used for interview surveys. To select a quota sample, we-

(a)Divide the population into specific groups (like strata) based on a particular characteristic, such as gender, age.
(b) Calculate a quota for each group based on relevant and available data.
(c) Interviewers asked to collect data from certain number of cases in each quota from which they must collect data (respondents choice depends on interviewer)
(d) Combine the data collected by interviewers to provide full sample.

Quota sampling is therefore a type of stratified sampling (specifically proportionate stratified sampling) but the selection of cases within subgroups or strata is entirely non-random. Quotas are described by quota controls, which set the size of sample to be obtained from the subgroups or strata. Sample size selected from each subgroup is based on the proportion of subgroup within the population.

Advantages of quota sampling are such as it is less costly and can be set up very quickly. (Since it is based on convenience of researcher which can speed up data gathering). It doesn't require a sampling frame and, therefore, may be the only technique we can use if one is not available. Obtaining a representative sample in quota sampling is difficult because selection largely depends on researcher's convenience. So he tends to choose respondents who are easily accessible and who appear willing to answer the questions. We cannot measure the level of certainty or margin of error as the sample is not probability based.

5.2.2. Purposive Sampling (or Judgmental Sampling): enables us to use our own judgment to select cases that will best enable us to answer our research question(s) and to meet our objectives.

This form of sample is often used when working with very small samples such as in case study research and when we wish to select cases that are particularly informative than others. The judgment of the researcher makes the sampling process non-random and hence, determining sampling error is difficult. The researcher tends to make error of judgment in one direction. These systematic errors lead to what are called biases.

The logic on which we base our strategy for selecting cases for a purposive sample should be dependent on our research question(s) and objectives.

5.2.3. Convenience or Haphazard Sampling: involves selecting haphazardly those cases that are easiest to obtain for our sample. For example, if a researcher wants to survey 1000 consumers door-to-door in a particular locality, samples can be selected from houses which are nearby, houses where people are responsive and friendly and houses which are on the first floor. The sample tends to be less variable than the population because in many environments the extreme elements of the population are not readily available.

5.2.4. Snowball Sampling: is commonly used when it is difficult to identify members of the desired population. For example, people who are working, while claiming for unemployment benefits. We need to-

(a) Make contact with one or two cases in the population

(b) Ask these cases to identify further cases and so on.

(c) Stop when either no new cases are given or the sample is as large as is manageable.

5.2.5. Self-Selection Sampling: occurs when we allow each case, usually individuals, to identify their desire to take part in the research. We –

(a) Publicize our need for cases, either by advertising through appropriate media or by asking them to take part.

(b) Collect data from those who are interested to respond.

6. Conclusion

Choice of appropriate sampling technique is basically dependent on our research question(s) and objectives. Also, geographically dispersed sample, resource constraints and such factors play role in final selection of sampling technique to be used.

Research question(s) and objectives that need us to estimate statistically the characteristics of the population from a sample require probability sampling. Whereas, research question(s) and objectives does not require such generalizations can alternatively make use of non- probability sampling technique.

Non-probability sampling techniques also provide us with the opportunity to select our sample purposively and to reach difficult-to-identify members of the population.

In many cases, we need to use a combination of different sampling techniques.

REFERENCES

Bajpai, Naval. *Business Research Methods*. New Delhi: Pearson, 2011.

Black, Ken. *Business Statistics-For Contemporary Decision Making-* Fifth Edition. New Delhi: Wiley-India, 2011, 219-228.

Bowerman, Bruce L., Rochard T. O'Connell and Emily S. Murphree. *Business Statistics in Practice*. New Delhi:Tata McGraw Hill, 2010.

Das, N.G. *Statistical Methods*(Vol II). New Delhi: Tata McGraw Hill, 2009.

Dunn, Olive J. and Virginia A. Clark. *Basic Statistics-A Primer for the Biomedical Science.* New Jersey: John Wiley & Sons, 2009.

Gupta, Santosh. *Research Methodology and Statistical Techniques*. New Delhi: Deep & Deep Publications Pvt. Ltd, 2010.

Kumar, C. Rajendra. *Research Methodology*. New Delhi: APH Publishing Corporation, 2008.

Kothari, C.R. *Research Methodology-Methods & Techniques*. New Delhi: New Age International (P) Limited, 2009.

Levine, David M., Timothy C. Krenhbiel, Mark L. Berenson and P. K. Vishwanathan.*Business Statistics-A First Course-* Fifth Edition. New Delhi: Pearson, 2010, 220-237.

Saunders, Mark, Philip Lewis and Adrian Thornhill. Research Methods for Business Student-Fifth Edition. New Delhi: Pearson, 2011.

Shrivastava, T.N. & Shailaja Rego. *Statistics for Management*. New Delhi:Tata McGraw Hill, 2010.

Spatz, Chris. *Basic Statistics-Tales of Distribution*. USA:Thomson Wadsworth, 2008.

Taylor, Bill, Gautam Sinha and Taposh Ghoshal. *Research Methodology- A Guide for Research in Management and Science*. New Delhi: Prentice Hall of India, 2007.