|  |
| --- |
| **Topic 6: Selecting Samples (Sampling)** |

|  |
| --- |
| **Introduction**  |

This chapter will explain about sampling and the process of selecting a sufficient number of participants from a population so that by studying the sample and understanding the characteristics of the sample subjects it would be possible to generalise the characteristics of the population.

The reasons for using a sample rather than collecting data from the entire population:

1. It would be practically impossible to collect data several hundreds or thousands of people

 in a population. Even if it were possible, it would be prohibitive in terms of cost, time and

 other human resources.

2. Studying a sample rather than the entire population is more likely to lead to more reliable

 results, mostly because fatigue is reduced, resulting in fewer errors in collecting data.

* 1. **Representativeness of samples**

The need for choosing the right sample for a research investigation is imperative. This is because a sample will rarely be the exact replica of the population from which it is drawn. However, it is possible to choose a sample in such a way that it is representative of the population as there is a small probability that the sample values might fall outside the population parameters.

The characteristics (height, weight) in the population are generally normally distributed. These attributes are such that most people will be clustered around the mean and there will only be a small number at the extremes that are very tall or very short, very heavy or very light. If we are able to estimate the population characteristics from the characteristics represented in a sample with reasonable accuracy, the sample then has to be chosen such that the distribution of the characteristics of participants follows the same type of normal distribution in the sample as it does in the population. The sampling distribution of the sample mean is then normally distributed. As the sample size ‘n’ increases the means of the random samples taken from practically any population approach a normal distribution with mean ‘U’ and the standard of deviation ‘ ‘. If we take a sufficiently large number of samples and choose them with care, we will have a sampling distribution of the means that has normality.

In scientific research, the need of a sample to be representative of the population is important for generalizability. However in certain research, this generalizability of the sample may not be that important. This is seen in the case study situation.

(*Parameter refersto the characteristics of a population. On the other hand, statistics refers*

 *to the characteristics of a sample and it is used to estimate the value of the population.)*

* 1. **Sampling Frame**

Sampling frame is basically a complete list of all the cases/members in the population. E.g. in a Lion Club the complete sampling frame is the complete list of all its members. It is important to have the complete list of all members in order to avoid biasness and to make your sample a representative of the population.

Checklist for selecting a sampling frame

* Are cases listed in the sampling frame relevant to your research topic, for example are they current?
* Does the sampling frame include all cases in other words is it complete?
* Does the sampling frame exclude irrelevant cases, in other words is it precise?
* (For purchased list/s) Can you establish and control precisely how the sample will be selected?
	1. **Deciding on a Suitable Sample Size**

A sample can be a representative of a population i.e. it has generalizability, if it is appropriate chosen. A large sample size reduces error. Sampling is like a compromise between the accuracy of the findings and the amount of time and money invested. Such a compromise is governed by:

1. The confidence you need to have in your data i.e. the level of certainty that the characteristics of the data collected will represent the characteristics of the population.
2. The margin of error that you can tolerate i.e. the accuracy you require for any estimate made from your samples.
3. The types of analyses you are going to make - each type of analysis may require a minimum threshold number.
4. The size of the total population from which your sample is being drawn.

As a rule of thunb, the Economist’s (197) advised that for any statistical analysis, a minimum number of **30** is required for a sample size.

As for researchers they normally work to a **95%** level of certainty i.e. if your sample is selected 100 times at least 95 of these samples would be certain to represent the characteristics of the population.

**Margin of Error**

Margin of error describes the precision (accuracy) of the estimates of the population. A larger sample size normally reduces the margin of error.

The table below provides a rough guide to the different sample sizes required for different sizes of population at 95% level of certainty.

|  |
| --- |
|  Margin of Error |
| Population | 5% | 3% | 2% | 1% |
| 50 | 44 | 48 | 49 | 50 |
| 100 | 79 | 91 | 96 | 99 |
| 150 | 108 | 132 | 141 | 148 |
| 200 | 132 | 168 | 185 | 198 |
| 500 | 217 | 340 | 414 | 475 |
| 1000 | 278 | 516 | 706 | 906 |
| 10000 | 370 | 964 | 1936 | 4899 |
| 100000 | 383 | 1056 | 2345 | 8762 |
| 1000000 | 384 | 1066 | 2395 | 9511 |
| 10000000 | 384 | 1067 | 2400 | 9595 |

**The Importance of High Response Rate**

A sample that is selected must be a perfect representative sample of the population. This means that every member in the sampe is able to make the same response. If a sample is made up of 100 members in the small service business, it means the 100 members must be really involved in the small service business and they could make the necessary responses. This implies a high response rate because the sample provides a true representative of the population. It will then reduce the biasness because of the absence of non-responses and the cost of the survey.

Non-response can be due to 4 reasons:

1. Refused to respond.
2. Ineligibility to respond (not meeting the requirements).
3. Inability to locate respondent.
4. Respondent located but unable to make contact.

Calculation of response rate:

 Total number of responses

 Total response rate =

 Total number in sample - ineligible

Calculation of active response rate

 Total number of responses

 Active response rate =

 Total number in sample - (ineligible + unreachable)

Calculate Actual Sample Size from Estimated Response Rate

The sample size must be big enough to provide the necessay confidence in you data i.e. as a representative of the population.

First estimate the response rate from the sample and then calculate to increase the size of the sample using the formula:

 n x 100

 n2 =

 re%

 n2 = actual sample size

 n = minimum sample size

 re% = estimated response rate in percentage

Example:

John wanted to make a survey of customers and calculated that an adjusted minimum sample size of 429 was required. He estimated the response rate would be 30%. What is the actual sample size John should have?

Applying the formula:

 n x 100

 n2 =

 re%

 439 x 100

 =

 30

 = 1463

John’s actual sample size should be 1463.

It is important to know the response rate when selecting the sampe size. E.g. using post survey, it is found that only 30% response rate is achievable and for interview only 50%.

* 1. **Sampling Designs (Techniques)**

There two types of sampling designs:

1. Probability sampling - the persons in the population have the known chance or probability

 of being selected. This design is used when the representativeness of the sample is of

 importance in the interests of widdr generalizability.

2. Nonprobability sampling - not every person in the population has the equal chance of

 being selected. This design is not for generalizability but critical for other factors.

Each of these two designs has different sampling strategies.

***Probability Sampling***

Types:

1. Simple random
2. Systematic
3. Stratefied random
4. Cluster

1. Simple random - Every person in the population has a ‘known-and-equal’ chance of being selected to make up a sample. E.g. there are 1000 people in the population and we need a sample of 100 persons. We can pieces of paper, each bearing the name of one of the persons into a hat and we can draw 100 of them from the hat. Each of those names has a 100/1000 chance of being drawn. The probability of any one of them being chosen as a subject is 1 and that everyone has the same or equal probability of being chosen. The sampling design is known as ***simple random sampling***. It has the least bias and offers the most generalizability. However, it is cumbersome and expensive and updating of the listing of the population is not possible. As a consequence complex probability sampling designs are used such as:

1. Systematic sampling
2. Stratified random samplinng
3. cluster sampling
4. Area sampling
5. Double sampling

2. Systematic sampling - This involves drawing every nth element in the population starting

 with a randomly chosen element between 1 and n.

 E.g. There are 260 houses and we want to have 35 houses for a sample. We can draw

 every 7th house starting from a random number from 1 to 7. Let us say that the random

 number is 7, then houses numbered 7, 14, 21, 28 and so on would be sampled until the 35

 houses are selected.

 However this systematic sampling design may suffer from systematic biasness. E.g. every 7th house is a corner house.

 For market surveys, consumer attitude surveys and the like, the systematicsampling design

 is often used.

3. Stratified random sampling - there are subgroups of elements existing within the

 population. E.g. the HR manager wants to train everyone employee in the organisation. Within the population of the organisation there different levels of people such as top management managers, middle-level managers and lower-level managers, first live supervisors, computer analysts, clerical workers and so on. The training requirements for different levels will be different. Data will need to be collected from each subgroup level in the population to assess the needs at each subgroup. In such a situation, stratified random sampling is used.

 This stratified random sampling is used to study buying habits of customers on the basis of life stages, income levels or geographical areas. Stratification ensures homogeneity within each subgroup or stratum.

4. Cluster sampling - intact grops, not individuals are randomly selected. All the members of selected groups hhave similar characteristics. E.g. instead of randomly selecting all fifthgraders in a large school district, you can randomly select fifthgrade classrooms and use all the students in each classroom.

 Cluster sampling is more convenient when the population is very large or spread out over a wide geographical area.

 It may be the only possible method of selecting a sample when the researcher is unable to obtain a list of all members of the population.

 Any location within which we find an intact group of similar characteristics is a cluster. Examples of clusters are classrooms, schools, city blocks, hospitals and department stores.

Benefits of cluster sampling:

 1. Involves less time and expense

 2. Generally it is more convenient than other techniques.

 3. It is easier to get permission to work with all members (students) in several classrooms

 than to work with a few students in many classrooms or to survey all the people in a

 limited number of city blocks than a few people in many city blocks.

Drawbacks of Cluster Sampling

1. The chances are greater of selecting a sample that is not representative of the population.
2. The smaller the sample size, the more likely that the sample selected may not represent the population.
3. Cluster samples offer more heterogeneity within groups and more homogeneity among groups.
4. Cluster sampling lends itself to greater biases and is the least generalizable of all the probability sampling designs (techniques) because most naturally occurring clusters in the organisational context do not contain heterogeneous elements. In other words, the conditions of intracluster heterogeneity and inter-homogeneity are often not met. For this reason cluster sampling is not well accepted in organisational research.

Single Stage and MultiStage Cluster Sampling

Cluster sampling can also be done in several stages and is then called ‘multistage cluster sampling. E.g. A national survey is to be done to determine the monthly bank deposits. Cluster sampling would first be used to select the urban, semiurban and rural geographical locations for study.

At the next stage particular areas in each of these locations would be chosen.

At the third stage banks in each area would be chosen

***Nonprobability Sampling***

In nonprobability sampling designs,the members in the population do not have the known and equal chance of being selected as sample subjects. This implies that the findings from the study of the sample cannot be confidently generalized to the population. This is the reason for using the nonprobability sampling designs. However, some nonprobability sampling designs are more dependable than others and could offer some important leads to potentially useful information with regard to the population.

Types:

1. Quota sampling
2. Purposive sampling
3. Snowball sampling
4. Self-selection sampling
5. Convenience sampling

1. Quota Sampling

It is a form of proportionate stratified sampling in which a predetermined proportion of people are sampled from differentgroups, but on a convenience basis.

E.g. It is assumed that the work attitudes of blue-colar workers are quite different from those of white-collar workers. If there are 60% blue-colar workers and 40% white-collar workers in the organisation and that a total of 30 people are to be interviewed to find the answer to the research question. Then a quota of 18 blue-collar workers and 12 white-collar workers will form the sample because these numbers represent 60 and 40 percent of the sample size.

It is obvious that the sample will not be totally representative of the population and therefore the generalizability of the findings will be restricted.

However it is a convenience that the type of sampling technique can provide in terms of effort, costs and time. This technique becomes necessary when a subset of the population is underrepresented in the organisation e.g. minority groups, foremen and so on. Its good point is that it ensures that all the subgroups in the population are adequately represented in the sample. It also indicates that *quota samples are basically stratified samples from which subjects are selected nonrandomly.*

As the workplace or the society becomes more heterogeneous because of the changing demographics, quota sampling can be expected to be used more frequently in the future. E.g. quota sampling can be used to have some idea of the buying behaviour of various ethnic groups, for getting an understanding as to how employees from different nationalities perceive the organisational culture and so on.

Although quota sampling is not as generalizable as stratified random sampling, it does offer some information based on which further investigation can be done. This means that at the first stage of research the nonprobability design of quota sampling can be used and once some information has been attained, a probability design will follow. The opposite approach is also possible, by starting the research with probability sampling design and from the information received a new area for researh is indicated and a nonprobability sampling design might be used to explore further findings.

2. Purposive sampling

This approach is to obtain from specific target groups. Here, the sampling is confined to specific types of people who can provide the desired iinformation, either because they are the only ones who possess it or conform to some criteria set by the researcher. The two major types of purposive sampling are judgement sampling and quota sampling.

In judgement sampling, the choice of subjects are those who are in the best position to provide the information required. E.g. if a research wants to find out what it takes for women mangers to make it to the top, the only people who can provide firsthand information are the women who have risen to the positions of presidents, vice presidents and important top-level executives in work organisations. Therefore this sampling design is used when a limited number or category of people have the information that is sought. Obviously judgement sampling may curtail the generalizability of the findings. It is, however, the only viable sampling method for obtaining the type of information that is required from very specific pockets of people who alone possess the needed facts and can give the information sought. It requires the researcher to locate them.

3. Snowball sampling

This is normally used when it is difficult to identify members of the desired populationl e.g. people who are workingwhile claimingunemployment benefit. You need to:

1. Make contact with one or two persons in the population.
2. Ask these persons to identify further people.
3. Ask these new persons to identify further new people and so on.
4. Stop wheneither no new ones are givenor the sampe is as large as is manageable.

The main problem in the sampling method is making initial contact. It is used when there are difficulty to identify the people for the sample.

4. Self-selection sampling

This occurs when you allow an individual to identify his/her desire to take part in the research. You therefore:

1. Publicise your need for cases, either by advertising through appropriate media or by asking them to take part.
2. Collect data from those who respond.

Cases that self-select often do so because of their feelings or opinionsabout the research question(s) or stated objectives.

E.g. Sila’s research was concerned with teleworking. She had decided to administer hor questionnaire using the Internet. She published her research on a range of bulletin boards and through the teleworkers’ association asking for volunteers to fill in a questionnaire. Those who responded were sent a short questionnaire by email.

5. Convenience sampling

Convenience or haphazard sampling involves selectinghaphazardly those cases that are easiest toobtain for your sample, such as the person interviewed at random in a shopping centre for a television programme. The sample selection process is continued until your required sample size has been reached. This technique may be widely used but it is prone to bias and influneces that are beyond your control, as the individuals only appear in the sample because of the ease of obtaining them. Often the sample is intended to represent the total population, for example managers taking an MBA course as a surrogate for all managers! In such instances the choice of sample is likely to have biased the sample, meaniing that subsequent generalisations are likely to be at best flawed. These problems are less important where there is little variation in the population, and such samples often serve as pilots to studies using more structured samples.