Topic ▶ Sampling
10

10.1 CONCEPT OF SAMPLING

Sampling is a process of selecting samples from a group or population to become the foundation for estimating and predicting the outcome of the population as well as to detect the unknown piece of information. A sample is the sub-unit of the population involved in your research work. There are a few advantages and disadvantages associated with the sampling process.

(a) **Advantages of Sampling**

> Among the advantages are that sampling can save cost and human resources during the process of research work. In ICT, sampling does not cause much constraint such as heavy use of tools and technology in predicting the research output.

(b) **Disadvantages of Sampling**

> A researcher may not find the information about the population being studied especially on its characteristics. The research can only estimate or predict them. This means that there is a high possibility of error occurence in the estimation made. Sampling process only enables a researcher to make estimation about the actual situation instead of finding the real truth. If you take a piece of information from your sampling population, and if your reasoning is correct, your findings should also be accurate to a certain degree.

When selecting a sample, it is very important for a researcher to consider the possibility of error during the selection process. In the field of ICT, sampling has little significance because the main purpose of ICT research is to explore or describe diversity in technology, phenomenon and issues. Another factor is the nature of ICT research which focuses on qualitative approach. Qualitative approach does not make an attempt to quantify or determine the extent of diversity. A researcher can select a sample and describe his/ her inquiry based on the research problem. Then, the study proceeds based upon the obtained sample.

You must always remember that qualitative research has a characteristic called saturation point. Saturation point is where a researcher reaches the limit of obtaining information after many attempts to get new information. When you find you are not obtaining new information, it is assumed you have reached the saturation point. Again, saturation point is subjective judgement which a researcher always decide about it in the entire research process.

10.2 SAMPLING TERMINOLOGIES

In sampling, there are a few terminologies that a researcher should be familiar with. For example, let's say you are working in a research project on computing implementation for elderly and disabled citizens for a smart home system. You are supposed to find out the average age of senior and disabled citizens involved in your study.

(a)   The community, families living in the town with smart homes form the **population** or **study population** and are usually denoted by the letter **N.**

(b)   The sample group of elderly people or senior citizens and disable people in the vicinity of the smart home community is called **sample**.

(c)   The number of elderly people or senior citizens and disabled people you obtain information to find their average age is called the **sample size** and is usually denoted by letter **n**.

(d)   The way you select senior citizens and disabled people is called the **sampling design** or **strategy**.

(e)   Each citizen or disabled people that becomes the basis for selecting your sample is called the **sampling unit** or **sampling element**.

(f)   A list identifying each respondent in the study population is called **sampling frame**. In case when all elements in a sampling population cannot be individually identified, you cannot have a sampling frame for the study population.

(g)   Finally, the obtained findings based on the information of the respondents are called **sample statistics**.

> **ACTIVITY 10.1**
>
> Identify the importance of sampling in a research work.

10.3   SAMPLE SIZE AND SELECTION

Most of the new researchers always wonder about the sample size that needs to be selected. You must remember that the larger the sample for your research, the better outcome you can evaluate at the end of the research process. The larger the sample, the more likely the sample mean and standard deviation will become a representation of the population mean and standard deviation. For instance, in IT survey, the sample size required depends on the statistical outcome needed for the findings. The following are some guidelines to decide on how large a sample should be:

- When the selected sample needs to be segregated into smaller clusters involving comparisons of clusters, a large sample would be appropriate.

- The longer the duration of a study, the higher the number of subjects that will drop out. To reduce attrition, a researcher should keep demands on subjects to the minimum, to fully inform the subject about the study and research, and make frequent communication with subjects to maintain the interest.

- A larger sample is needed when the population is highly heterogeneous on the variables being studied so that different characteristics can be identified. If members of the population is less, then a small sample size would do to obtain the necessary characteristics.

Gay and Airasian (2003) also offered similar guidelines in their work for selecting sample size in a research study. In their work, it is mentioned that beyond a certain point (at about 5,000 respondents or more), the population size is almost irrelevant and a sample size of 400 should be adequate. To some extent, the size of sample depends on how homogenous or heterogenous the population is and how alike or different its members are with respect to the characteristics of particular research interest.

10.3.1   Selecting a Sample

The objective of selecting a sample is to achieve maximum accuracy in your estimation within a given sample size and to avoid bias in the selection of the sample. This is important as bias can attack the integrity of facts and jeopardise your research outcome.

Table 10.1 explains how bias can occur in sample selection.

Table 10.1: Reasons Bias Occurs in Sample Selection

No.	Reason(s)
1.	Sampling done using non-random method (we will see sampling designs in the next section) which can be influenced by human choices.
2.	Sampling frame like list, indexing and records which serve as the platform of selection does not cover the sampling population accurately or completely.
3.	A section of sampling population refuses to co-operate.

There are also factors that may influence the degree of certainty in inferences drawn from a sample for research study. As we know, the size of samples influence findings such that large samples have more certainty than those based on smaller ones. Therefore, the larger the sample, the researcher will obtain more accurate findings.

Another factor is the extent of variation in the sampling population where the greater the variation in the population will have greater uncertainty with respect to its characteristics. Therefore, it is crucial for a researcher to bear these in mind especially when selecting a sample for her/his respective research work.

SELF-CHECK 10.1

1.   Describe the guidelines on selecting samples.

2.   Identify how bias can occur during sample selection process.

10.4 SAMPLING TECHNIQUES

Sampling techniques often depend on research objectives of a research work. Generally there are two types of sampling techniques that are widely deployed. These techniques are:

(a)   **Probability Sampling**

> This sampling technique includes sample selection which is based on random methods. The techniques that are based in this category are **random sampling, stratified sampling, systematic sampling** and **cluster sampling.**

(b) **Non-probablity Sampling**

> This sampling techniques is not based on random selection. Some examples are **quota sampling, purposive sampling** and **convenience sampling**.

10.4.1 Probability Sampling

The techniques in probality sampling are as follows:

(a) **Random Sampling**
Random sampling is used to increase the probability of the sample selected. By deploying this technique, each member of a population stands a chance to be selected. Let's say you are interested to survey the usage of e-commerce application in business-to-consumer (B2C).

The sample you select needs to represent the types of e-commerce application and its usage. Due to financial and time constraints you are unable to survey the usage of all types of e-commerce application across the Malaysian network (N= 100,000). Therefore you decide to confine the study to e-commerce application for merchandise products in Malaysia (n=10,000) which is called the accessible population.

From this accessible population, a sample of 100 e-commerce application is retrieved. How do we randomly select sample? It is understood that random sample is a procedure in which all individuals in the defined population have an equal and independent chance to be selected in the sample design. In the above example, the number of e-commerce application on merchandise products across Malaysian network is 10,000 and you may intend to draw a sample of 100. When you select the first application, it has 1:10,000 chances of being selected. Once the first application selected, the remaining will be 9,999 so that each application has 1:9,999 of being selected as second case. Therefore, once each case is selected, the probability of being selected next changes because the population of selection has become one case smaller each time.
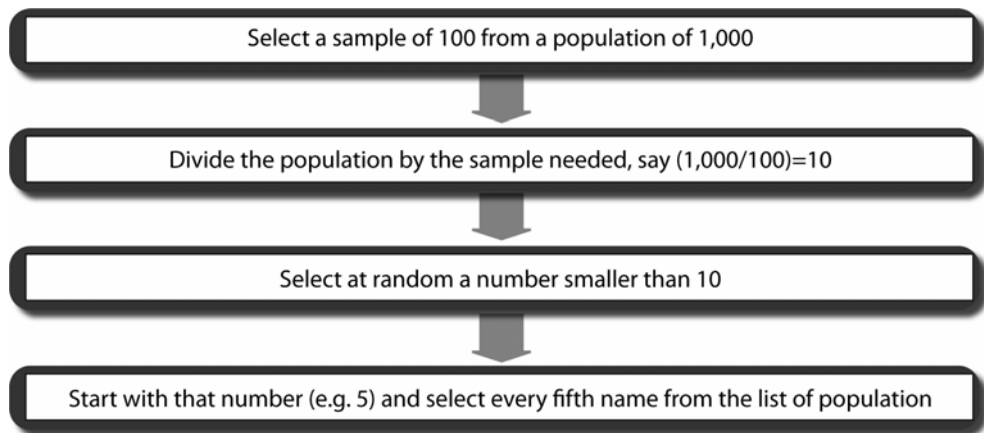
(b) **Stratified Sampling**
In some IT surveys, a researcher may want to ensure individuals with certain characteristics are included in the sample to be studied. For such cases, stratified sampling is used. In this sampling design, a researcher will attempt to stratify population in such a way that the population within a

stratum is homogeneous with respect to the characteristics on the basis of which it is being stratified. You must bear in mind that it is important for the characteristics chosen as the basis of stratification, are clearly identifiable in the population. For example, it is much easier to stratify the population on the basis of gender rather than age or income group.

(c) **Systematic Sampling**
Systematic sampling also known as 'mixed sampling' category since it has both random and non-random sampling designs. A researcher has to begin by having a list names of members in the population, in random approach. Figure 10.1 below shows the order of the sampling.

Select a sample of 100 from a population of 1,000

Divide the population by the sample needed, say (1,000/100)=10

Select at random a number smaller than 10

Start with that number (e.g. 5) and select every fifth name from the list of population

Figure 10.1: Example of systematic sampling

This sampling method is good as long as the list does not contain any hidden order. Systematic sampling is frequently used in ICT research and survey, especially in selecting specified number of records from computer documents.

(d) **Cluster Sampling**
In cluster sampling, the unit of sampling is not referring to an individual entity but rather a group of entities. For example, in an organisation there are 25 departments and in each department there are an estimated 20 IT administrators. You need a sample of about 100 staff but this would mean going to many departments if random sampling approach is used. Using cluster sampling, you may select 5 departments randomly from a total of 25 departments. You study all the staff in the 5 departments you chose. The advantage that can be highlighted here is: it saves cost and time especially if the population is scattered. The disadvantage is that it is less accurate compared to other techniques of sampling discussed.

10.4.2   Non-Probability Sampling

In some research scenarios, it is not possible to ensure that the sample will be selected based on random selection. Non-probability sampling is based on a researcher's judgement and there is possibility of bias in sample selection and distort findings of the study. Nonetheless, this sampling technique is used because of its practicality. It can save time and cost, and at the same time, it is a feasible method given the spread and features of a population. Some common sampling methods are quota sampling, purposive sampling and convenience sampling.

(a)   **Quota Sampling**
The main reason directing quota sampling is the researcher's ease of access to the sample population. Similar to stratified sampling, a researcher needs to identify the subgroups and their proportions as they are represented in the population. Then, the researcher will select subjects based on his/ her convenience and judgement to fill each subgroup. A researcher must be confident in using this method and firmly state the criteria for selection of sample especially during results summarisation.

(b)   **Purposive Sampling**
This sampling method is selected on the basis that members conform to certain stipulated criteria. You may need to use your own judgement to select cases to answer certain research questions. This sampling method is normally deployed if the sample population is small and when the main objective is to choose cases that are informative to the research topic selected. Purposive sampling is very useful in the early stages of an exploratory study. One of the disadvantages of this technique is that the sample may have characteristics different from population characteristics.

(c)   **Convenience Sampling**
Using this sampling method, a researcher is free to use anything that they could find in the research outline. The sample is selected based on preferences and ease of sampling respondents. This sampling is easier to conduct and less expensive. However, it has poor reliability due to its high incidence of bias. In ICT, convenience sampling seems to be dominant especially in cases of organisations that conduct web surveys, mail their responses to a survey questions and SMS their opinions to a question. Although convenience sampling can cater to a lot of data, it is not reliable in terms whether the sample represents the real population or not.

SELF-CHECK 10.2

1. What are the sampling techniques of probability sampling?

2. Distinguish between quota and purposive sampling techniques.

SUMMARY

- **Sampling** is a process of selecting samples from a group or population to become the foundation on estimating and predicting outcome of the population.

- **Two main techniques** of sampling: **probability** and **non-probability.**

- **Probability sampling** is based on random selection while non-probability sampling is not based on random selection.

- Probability sampling consists of **random sampling, stratified sampling, systematic sampling and cluster sampling**.

- Non-probability sampling consists of **quota sampling, purposive sampling and convenience sampling**.

- In cluster sampling, the unit of sampling does not refer to an individual entity but a group of entities.

KEY TERMS

Non-probability sampling	Sampling
Probability sampling	Sampling design
Purposive sampling	Sampling population
Random sampling	Sampling statistics

DISCUSSION

You wish to study the impact of Service Oriented Architecture adoption among software architects for software testing methodology. There are 300 software architects in 15 departments in the unit you choose. You plan to obtain 1,000 software architects using the cluster sampling technique. Describe the steps you would take in selecting the sample.

REFERENCES

Internet Resources

Easton, V. J. & McColl, J. H. (2007) Statistics Glossary: Sampling [Electronic Version]http://www.stats.gla.ac.uk/steps/glossary/sampling.html

Galloway, A. (1997). Sampling: A Workbook [Electronic version] http://www.tardis.ed.ac.uk/~kate/qmcweb/scont.htm

Trochim, W. K. (2007). *Research method tutorials* [Electronic version] http://www.socialresearchmethods.net/kb/sampling.php