

PROBABILIY AND STATISTICS I

LECTURE TWO

Data collection methods

Lecturer: Dr. Emily Roche

INTRODUCTION

This lecture will focus on sampling methods together with methods of collecting primary and secondary data.

Intended learning outcomes

At the end of this lecture, you will be able to explain the key principles of sampling, distinguish different sampling and data collection methods.

References

These lecture notes should be supplemented with relevant topics from the book listed in the Bibliography at the end of the lecture.

Sampling

This is the process of selecting a sample (subset of the population) from a population so that it is studied in order to understand the population.

Reasons for sampling

1. To save time and money.
2. For safety purposes especially when a research involves the justification of a product.
3. To increased response rate.
4. To increase the degree of accuracy

Sampling methods

There are two major types of sampling designs:

A. Probability sampling techniques

This is used when the elements in the population have some known chance or probability of being selected as sample subjects.

- **Simple Random Sampling**

In this method, all members of the population have an equal chance of being selected.

There are two basic methods of selecting a simple random sample:

- The urn method.
- Use of computer generated random numbers or random number tables to pick n items.

- **Systematic Random Sampling**

This involves selecting n items following a certain interval given by the sampling fraction (k).

- **Stratified Random Sampling**

This involves dividing the population into groups called strata then selecting a sample from each stratum using a random technique

Points to note:

- i. The strata must be mutually exclusive i.e. no member of the population should belong to more than one stratum.
- ii. There should be maximum homogeneity within each strata and maximum heterogeneity between strata.
- iii. Sampling within each stratum is done using a random based technique i.e. either systematic or simple random sampling. Each stratum must contribute to the final sample

- **Cluster Sampling**

The population is divided into groups called clusters, then randomly selecting the cluster(s). The cluster(s) chosen make up the sample.

Note:

- i. It should be possible to clearly separate the population into clusters.

- ii. Clusters must contain maximum heterogeneity within each cluster and maximum homogeneity between the clusters.

B. Non probability sampling techniques

These include the following

- **Purposive Sampling (Judgment Sampling)**

In this method, the researcher deliberately picks elements of the population that are responsible or relevant to the purpose of the investigation.

- **Convenience Sampling**

Sample elements are readily available persons on the road or shop, office or anywhere.

- **Snow-Ball (Chain referral) Sampling**

The researcher first chooses respondents believed to have the required attributes and gets their responses. These respondents are then responsible for recruiting the other members of the sample whose responses are also recorded and so on.

- **Voluntary Sampling**

The need for participation is made public then the sample elements self - select.

Data collection methods

Primary Data collection methods

1. Experiments

It is a method where one or more variables are manipulated to determine their effect on a dependent variable.

Types of experiments

- **Controlled**

It is commonly referred to as true experiment. It is considered as the most accurate experimental research. A researcher creates a hypothesis and then tries to prove or

disprove it. It uses control groups where the subjects have been randomly assigned, then the researcher only tests one effect at a time on the groups.

An advantage of this type of experiment is that it is true and reliable. One major disadvantage is that the results may be impractical due to their 'almost perfect' conditions and also they can be difficult and expensive to set up.

➤ Quasi experimental design.

This is an empirical study in where the experimental conditions are determined by other factors out of the control of the experimenter. It deals with social sciences and psychology. It is considered as unreliable since it does not use scientific methods.

This experiment is very useful in generating results for general trends. It also uses figures and results from individual case studies to reinforce its findings, thus reducing the amount of time and resources needed for experimentation.

➤ Field experimental design

This is an experiment that is performed in the 'real' world hence it does not need confounding variables.

Field experiments may have little validity due to the fact that there are many variables affecting a given experiment.

They are often used in social sciences especially in economic analyses of educational and health interventions.

2. Focus Group Discussion

This involves a group of people of 6-12 who are through a discussion for their input toward a given idea, product, service, concept or advertisement. Questions are asked in an interactive manner allowing participants freedom to talk to each other within the group.

Types of Focus Group Discussion

- a. Two-way focus group – One group watches the activities of the other group and later on they discuss the observed interactions and conclusions made.
- b. Dual moderator focus group – One particular group makes sure that the session progresses smoothly while another group ensures that the topics of discussion are completely covered.
- c. Respondent moderator focus group – Only one of the respondents is asked to act as the moderator temporarily.
- d. Dueling moderator focus group – Two moderators deliberately take two opposite sides on the issue under discussion.

Advantages

1. It is a quick method which enables a researcher to produce a lot of information quickly.
2. It is less costly unless it involves gathering people from different parts which would result into transport costs.
3. It produces data and insights that would be less accessible without interaction.
4. It is great tool for identifying and exploring beliefs, ideas or opinions of a community.
5. It is used in identifying relevant and appropriate questions for individual interviews.

Disadvantages

1. Focus group has the incapability to tell the definite frequency or distribution of beliefs and behaviors in a population. Information is gathered from a small

number of people and it would be difficult to project the results to the whole population.

2. Analysis of results are harder than individual interviews due to the complexities involved.
3. Its facilitation and conduction requires considerable skills.
4. The results obtained are influenced by the researcher or his own reading of the group's discussion.
5. Participants may hold back their views or give what the moderator wants to hear if the setting is formal or conducted by a superior moderator.
6. There is lack of confidentiality which is a very important factor in obtaining information in research. It becomes very difficult when tape recorders are used.

3. **Observation**

A qualified person gathers data by walking through the actual process associated with a system and notes the physical characteristics in their natural setting.

When should you use observation for evaluation?

- When you are trying to understand an on-going process or situation.
- When you are gathering data on individual behaviours or interactions between people.
- When you need to know about a physical setting.
- When data collection from individuals is not a realistic option.

Types of Observation

There are basically two types of observation

- Participant Observation:

It requires that the researcher become a participant in the context being observed.

➤ **Direct Observation:**

The direct observer tries to be unobtrusive as possible so as not to be biased in the observations.

Advantages

1. Data is accurate (first-hand information) since it is collected at the time it occurs.
2. Individuals do not have to rely on other people's information since they can gather it on their own.
3. One is able to capture the whole event as it occurs in its natural environment.
4. Subjects are more perceptive to observation than questioning, since they come to trust the observer.

4. Questionnaire

This is a set of related questions based on a specific topic of study asked to respondents by the researcher in order to get people's opinions.

Types of Questionnaires

➤ **Open Ended Questionnaires**

This questionnaire asks respondents to answer questions in their own words.

Advantages

1. It allows respondents to include more information thus enabling the researchers to better assess the respondents' true feelings.
2. It makes the respondents read through the questions first before answering, hence, it is accurate.
3. It shows interest and respect in others as you value their ideas/ responses.

Disadvantages

1. Other respondents don't take their time to answer all the questions hence the information will not be valid enough.

2. The respondents may explain a short straight forward answer in many unnecessary words making data collection difficult.

➤ Closed Questionnaires

A closed ended questionnaire is one which limits the respondent to the answers listed in the questionnaire for the question.

Advantages

1. They are more easily analyzed
2. Answers can be statistically analyzed and assessed as they are given numerical in nature
3. They are easy to answer as they take less time
4. They are more specific, thus communicate similar answers.
5. Higher response rate as the respondents do not have to put in much thought for the answers.

Disadvantages

1. They limit the respondents' answers hence they cannot fully elaborate what they feel. They cannot write their own opinions.
2. They collect limited information on the area of survey as the answers are limited to the choices given.
3. The choices given are based on the researchers' opinion hence, they are biased.

General Advantages of questionnaires

1. They are cheap.
2. They do not require as much effort from the questioner as in interviews
3. It is relatively quick to collect information using questionnaires.

4. Information can be collected from a large portion of a group.

General Disadvantages of questionnaires

1. The research done using questionnaires is limited to a small geographical region thus the researcher won't know what other people in other locality have to say.
2. Conducting the research is tiresome as the researcher has to move from place to place and convince people on the importance of the study.
3. Data collection may be difficult as one has to go through each script to sample the reports.

5. Interview

This is a tool used to gather information through verbal interaction with a person (interviewee).

Types of interviews:

Oral Interview - The interviewer and interviewee communicate via word of mouth. The common types include

1. Informal Interviews
2. Screening/Telephone Interviews
3. Individual Interview
4. Small Group or Committee Interview
5. Behavioral-Based Interview
6. Task Oriented/Testing Interview
7. Stress Interview

Written interview/Questionnaire - The interviewee fills in a particular form and sends it back to the interviewer.

Advantages

1. Efficient in getting information on topics that require detailed elaboration

2. Dynamic - The interviewer has the ability to ask questions as he/she thinks of them.
3. Have a much higher response rate
4. Reliable because the respondents own words can be recorded.

Disadvantages

1. The interviewer can influence the data gathered if they are not careful, in the case that notes are being made during the interview.
2. Interviews are limited to a small number of people at a time.
3. The questions asked during the interview may be biased.
4. Interviews are very time consuming.
5. Can be expensive

Collecting secondary data

This is data collected from any published source. The sources include:

- Published Statistics: National Government Sources(Census, surveys, planning documents), professional bodies, Academic and research institutes
- Non-Published / Electronic Sources: Data Archives, International Sources on Internet & Web
- Biographies - subjective interpretation involved
- Diaries - more spontaneous, less distorted by memory lapses
- Newspapers - public interest & opinion

Bibliography

Gupta, SP (Dr.), (2014). *Statistical methods* (43rd Ed.). Sultan Chand & Sons.

S. C. Gupta and V. K. Kapoor, (2020). *Fundamentals of mathematical Statistics* (12th Ed). Sultan Chand & Sons.