ଓଡ଼ିଶା ରାଜ୍ୟ ମୁକ୍ତ ବିଶ୍ୱବିଦ୍ୟାଳୟ, ସମ୍ବଲପୁର, ଓଡ଼ିଶା
Odisha State Open University, Sambalpur, Odisha
Established by an Act of Government of Odisha.

Journalism and Mass Communication (JMC)

JMC-08

Block -03

Research Tools and Data Analysis

Expert Committee

Professor. Mrinal Chatterjee
Professor, IIMC, Dhenkanal –**Chairman**

Dr. Asish Kumar Dwivedy
Asst. Professor, Humanities and Social Science (Communication Studies), SoA University, BBSR -**Member**

Sudhir Patnaik
Editor, Samadrusti - **Member**

Sujit Kumar Mohanty
Asst. Professor, JMC, Central University of Orissa, Koraput - **Member**

Dr. Dipak Samantarai
Director, NABM, BBSR- **Member**

Jyoti Prakash Mohapatra
Faculty
JMC- **Convener**

Course Writer

Sujit Kumar Mohanty
Asst. Prof, JMC

Course Editor

Jyoti Prakash Mohapatra
OSOU, Sambalpur

Material Production

Dr. Jayanta Kar Sharma
Registrar
Odisha State Open University, Sambalpur

Printers by : Sri Mandir Publication, Sahid Nagar, Bhubaneswar

UNIT 1: DATA AND ITS TYPES

Structure

1.1 LEARNING OBJECTIVES

After studying this lesson, you will be able to:

- Understand the concept of primary and secondary data
- Understand the various advantages and disadvantages of Primary and Secondary Data
- Trace some other methods of data collection

Any good research is incomplete without the collection of relevant data. Thus, the task of data collection begins after the research problem has been defined and research design/plan chalked out. Primarily, there are two methods of data collection i.e. primary and secondary.

1.2 What is Primary Data

Primary data are information collected by a researcher specifically for a research assignment. In other words, primary data are information that a company/researcher must gather because no one has compiled and published the information in a forum accessible to the public. Companies generally take the time and allocate the resources required to gather primary data only when a question, issue or problem presents itself that is sufficiently important or unique that it warrants the expenditure necessary to gather the primary data. Primary data are original in nature and directly related to the issue or problem and current data. Primary data are the data which the researcher collects through various methods like interviews, surveys, questionnaires etc. Primary data has its own advantages and disadvantages:

1.2.1 Advantages of primary data Disadvantages of primary data

Advantages of primary

Some of the advantages of primary data are as follows:

i) The primary data are original and relevant to the topic of the research study, so the degree of accuracy is very high.

ii) Primary data is that it can be collected from a number of ways like interviews, telephone surveys, focus groups etc. It can be also collected across the national borders through emails and posts. It can include a large population and wide geographical coverage.

iii) Moreover, primary data is current and it can better give a realistic view to the researcher about the topic under consideration.

iv) Reliability of primary data is very high because these are collected by the concerned and reliable party.

Disadvantages of primary data

Some of the disadvantages of primary data are as follows:

i) For collection of primary data where interview is to be conducted the coverage is limited and for wider coverage a more number of researchers are required.

ii) A lot of time and efforts are required for data collection. By the time the data collected, analysed and report is ready the problem of the research becomes very serious or out dated. So the purpose of the research may be defeated.

iii) It has design problems like how to design the surveys. The questions must be simple to understand and respond.

iv) Some respondents do not give timely responses. Sometimes, the respondents may give fake, socially acceptable and sweet answers and try to cover up the realities.

v) With more people, time and efforts involvement the cost of the data collection goes high. The importance of the research may go down.

vi) In some primary data collection methods there is no control over the data collection. Incomplete questionnaire always give a negative impact on research.

vii) Trained persons are required for data collection. In experienced person in data collection may give inadequate data of the research.

1.2.2 Collection of Primary Data

Primary data can be collected through some of the following methods:

i) Observation method
ii) Interview method
iii) Questionnaires
iv) Schedules

i) Observation Method

The observation method is the most commonly used method specially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability.

Observation method often requires the researcher to play a number of roles and to use a number of techniques; including her/his five senses, to collect data. The observer puts himself in the actual situation and watch carefully. On the basis of his knowledge, skills and experience he collects the data without contacting the respondents. The results of observation entirely depend on the talents of the researcher. This method can be used only by expert persons in the research. Observation methods have been developed with the objective of 'observing people in their natural setting - as they go about their everyday lives. Observation methods can overcome some of the criticisms of quantitative research methods (Validity, bias etc.) and can be useful when its subject can't provide information, or can only provide inaccurate information.

However, observation method has various limitations.

i) It is an expensive method.
ii) The information provided by this method is very limited.
iii) Sometimes unforeseen factors may interfere with the observational task. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

ii) Interview method

In this method the interviewer personally meets the informants and asks necessary questions to them regarding the subject of enquiry. Usually a set of questions or a questionnaire is carried by him and questions are also asked according to that. The interviewer efficiently collects the data from the informants by cross examining them. The interviewer must be very efficient and tactful to get the accurate and relevant data from the informants. Interviews like personal interview/depth interview or telephone interview can be conducted as per the need of the study.

Advantages of interview

Some of the advantages of interview method are as follows:

i) In this method information can be gathered from illiterate people too.
ii) There are no chances of non-response as the interviewer personally collects data.
iii) The collected data is very reliable since the interviewer tactfully collects the data by cross examining the responders.

Disadvantages of interview

Some of the disadvantages of interview method are as follows:

i) There is a chance of bias.
ii) The informants may not answer some personal questions.
iii) It is a time-consuming process.
iv) Money and manpower requirements are very high.
v) Some time the interviewers are involved in pressurising respondents to share their personal information.

iii) Questionnaire method

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organisations and even by governments.

Questionnaire is a series of questions asked to individuals to obtain statistically useful information about a given topic.

A number of questions usually in printed or electronic form are to be answered by the individuals. The forms often have blank spaces in which the answers can be written. Sets of such forms are distributed to groups and the answers are collected relating to research topic.

When properly constructed and responsibly administered, questionnaires become a vital instrument by which statements can be made about specific groups or people or entire populations. Inappropriate questions, incorrect ordering of questions, incorrect scaling, or bad questionnaire format can make the survey valueless, as it may not accurately reflect the views and opinions of the participants.
In a research or survey, questions are asked to respondents, and designed to extract specific information. It serves four basic purposes:

i) collect the appropriate data,
ii) make data comparable and amenable to analysis,
iii) minimize bias in formulating and asking question,
iv) to make questions engaging and varied.

Advantages of questionnaire method

Some of the advantages of questionnaire method are as follows:

i) There is low cost even when the universe is large and is widely spread geographically.
ii) It is free from the bias of the interviewer; answers are in respondents' own words.
iii) Respondents have adequate time to give well thought out answers.
iv) Respondents, who are not easily approachable, can also be reached conveniently.
v) Large samples can be made use of and thus the results can be made more dependable and reliable.

Disadvantages of questionnaire method

Some of the disadvantages of questionnaire method are as follows:

i) Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
ii) It can be used only when respondents are educated and cooperating.
iii) The control over questionnaire may be lost once it is sent.
iv) There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.

v) There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.

vi) It is difficult to know whether willing respondents are truly representative.

vii) This method is likely to be the slowest of all.

iv. Schedules

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose.

These enumerators along with schedules, go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma. In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms.

This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected. The enumerators should be trained to perform their job well and the nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule. Enumerators should be intelligent and must possess the capacity of cross examination in order to find out the truth.

Above all, they should be honest, sincere, hardworking and should have patience and perseverance. This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

1.3 Some other methods of data collection

Apart from the above mentioned popular methods of data collection there are quite a number of other methods which researchers/big business houses use in recent times. Let us look at some of them.

Warranty cards

Warranty cards are usually postal sized cards which are used by dealers of consumer durables to collect information regarding their products. The information sought is printed in the form of questions on the 'warranty cards' which is placed inside the package along with the product with a request to the consumer to fill in the card and post it back to the dealer.

Distributor or store audits

Distributor or store audits are performed by distributors as well as manufacturers through their salesmen at regular intervals. Distributors get the retail stores audited through salesmen and use such information to estimate market size, market share, seasonal purchasing pattern and so on. The data are obtained in such audits not by questioning but by observation.

Projective techniques

Projective techniques (or what are sometimes called as indirect interviewing techniques) for the collection of data have been developed by psychologists to use projections of respondents for inferring about underlying motives, urges, or intentions which are such that the respondent either resists to reveal them or is unable to figure out himself. In projective techniques the respondent in supplying information tends unconsciously to project his own attitudes or feelings on the subject under study. Projective techniques play an important role in motivational researches or in attitude surveys.

Depth interviews

Depth interviews are those interviews that are designed to discover underlying motives and desires and are often used in motivational research. Such interviews are held to explore needs, desires and feelings of respondents. In other words, they aim to elicit unconscious as also other types of material relating especially to personality dynamics and motivations. As such, depth interviews require great skill on the part of the interviewer and at the same time involve considerable time. Unless the researcher has specialised training, depth interviewing should not be attempted.

Content analysis

Content-analysis consists of analysing the contents of documentary materials such as books, magazines, newspapers and the contents of all other verbal materials which can be either spoken or printed. Content-analysis prior to 1940's was mostly quantitative analysis of documentary materials concerning certain characteristics that

can be identified and counted. But since 1950's content-analysis is mostly qualitative analysis concerning the general import or message of the existing documents.

1.4 Secondary data

Secondary data are the data collected by a party not related to the research study but has been collected for some other purpose and at different time in the past. If the researcher uses these data then these become secondary data for the current users. These may be available in written, typed or in electronic forms. A variety of secondary information sources is available to the researcher gathering data on an industry, potential product applications and the market place. Secondary data is also used to gain initial insight into the research problem. Secondary data is classified in terms of its source – either internal or external. Internal, or in-house data, is secondary information acquired within the organization where research is being carried out. External secondary data is obtained from outside sources.

A researcher must be very careful in using secondary data. He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study. In this connection Dr. A.L. Bowley very aptly observes that it is never safe to take published statistics at their face value without knowing their meaning and limitations and it is always necessary to criticise arguments that can be based on them.
There are various advantages and disadvantages of using secondary data.

1.4.1 Advantages and Disadvantages of secondary data

Advantages of secondary data

Some of the advantages of secondary data are as follows:

i) The primary advantage of secondary data is that it is cheaper and faster to access.
ii) It provides a way to access the work of the best scholars from all over the world.
iii) Thirdly, secondary data gives a frame of mind to the researcher that in which direction he/she should go for the specific research.
iv) Fourthly secondary data saves time, efforts and money and adds to the value of the research study.

Disadvantages of secondary data

Some of the disadvantages of secondary data are as follows:

i) The data collected by a third party may not be reliable and hence the accuracy of the data may go down.

ii) Data collected from one location may not be suitable for the other due to variable environmental factor.

iii) With the passage of time the data becomes obsolete and very old

iv) Secondary data collected can distort the results of the research. For using secondary data, special care is required to amend or modify for use.

v) Secondary data can also raise issues of authenticity and copyright.

1.5 CHECK YOUR PROGRESS

Q1. What is primary data? How is it different from secondary data?
Q2. What are the advantages of interview method?
Q3. What are the disadvantages of questionnaire method?
Q4. What are the limitations of observation method?
Q5. What is content analysis?

Unit 2: Census and Sampling

Structure

2.1 Learning Objectives

After studying this lesson, you will be able to:

- Understand what is census
- Understand sampling and its advantages
- Identify different types of sampling
- Differentiate between census and sampling

2.2 What is census?

A census is the procedure of systematically acquiring and recording information about the members of a given population. It is a regularly occurring and official count of a particular population. The term is used mostly in connection with national population and housing censuses; other common censuses include agriculture, business, and traffic censuses.

The word is of Latin origin: during the Roman Republic, the census was a list that kept track of all adult males fit for military service. The modern census is essential to international comparisons of any kind of statistics, and censuses collect data on many attributes of a population, not just how many people there are but now census takes its place within a system of surveys where it typically began as the only national demographic data collection. Although population estimates remain an important function of a census, including exactly the geographic distribution of the population, statistics can be produced about combinations of attributes e.g. education by age and sex in different regions.

2.2.1 Purpose of census

The purpose of census is to:

i) Provide the facts essential to government for policy-making, planning and administration.

ii) Decision-making that facilitates the development of socio-economic policies -enhance the welfare of the population.

iii) Provides important data for the analysis and appraisal of the changing patterns of rural/urban movement and concentration, the development of urbanized areas, geographical distribution of the population according to such variables as occupation and education, as well as the socio-economic characteristics of the population and the labour force.

iv) Aids in the decision-making processes of the private sector. Population size and characteristics influence the location of businesses and services that satisfy the needs of the target population.

v) Population censuses also constitute the principal source of records for use as a sampling frame for the household surveys during the years between censuses.

2.2.2 Users of census data

Since census provides rich and valuable information about the population, there are many users of such data. Some of them are:

- Central and local government
- Interest and representative groups
- Businesses
- Local community
- Ethnic groups
- Students
- Media

2.3 What is Sampling

Sampling may be defined as the selection of some part of an aggregate or totality on the basis of which a judgement or inference about the aggregate or totality is made. In other words, it is the process of obtaining information about an entire population by examining only a part of it. In most of the research work and surveys, the usual approach happens to be to make generalisations or to draw inferences based on samples about the parameters of population from which the samples are taken. The researcher quite often selects only a few items from the universe for his study purposes. All this is done on the assumption that the sample data will enable him to estimate the population parameters. The items so selected constitute what is technically called a sample, their selection process or technique is called sample design and the survey conducted on the basis of sample is described as sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

2.3.1 Importance of sampling

A large population cannot be studied in its entirety for reasons of size, time, cost or inaccessibility. Limited time, lack of large amount of funds, and population scattered

in a very wide geographical area often make sampling necessary. Sarantakos (1998:140) has pointed out the following purposes of sampling:

1. Population in many cases may be so large and scattered that a complete coverage may not be possible. Suppose, the Patanjali Ayurved Limited wants to find the reactions of customers who have used its noodles and honey. For this, lakhs of customers would have to be contacted in different cities. Some of these customers might be residing in inaccessible areas and hence it would be impossible to collect information within a short span of time.
2. It offers a high degree of accuracy because it deals with a small number of persons.
3. In a short period of time valid and comparable results can be obtained. A lengthy period of data collection generally renders some data obsolete by the time the information is completely in hands.
4. Sampling is less demanding in terms of requirements of investigators since it requires a small portion of the target population.
5. It is economical since it contains fewer people. Large populations would involve employing a large number of interviewers which will increase the total cost of the research.

2.3.2 Key terms in Sampling

Let us familiarise ourselves with some fundamental definitions concerning sampling concepts and principles.

a). Universe/Population:

From a statistical point of view, the term 'Universe'refers to the total of the items or units in any field of inquiry, whereas the term 'population' refers to the total of items about which information is desired. The attributes that are the object of study are referred to as characteristics and the units possessing them are called as elementary units. The aggregate of such units is generally described as population. Thus, all units in any field of inquiry constitute universe and all elementary units (on the basis of one characteristic or more) constitute population.

Quite often, we do not find any difference between population and universe, and as such the two terms are taken as interchangeable. However, a researcher must necessarily define these terms precisely.

The population or universe can be finite or infinite. The population is said to be finite if it consists of a fixed number of elements so that it is possible to enumerate it in its totality. For instance, the population of a city, the number of workers in a factory are examples of finite populations. The symbol 'N' is generally used to indicate how

many elements (or items) are there in case of a finite population. An infinite population is that population in which it is theoretically impossible to observe all the elements. Thus, in an infinite population the number of items is infinite i.e., we cannot have any idea about the total number of items. The number of stars in a sky, possible rolls of a pair of dice are examples of infinite population.

b). Sampling frame

The elementary units or the group or cluster of such units may form the basis of sampling process in which case they are called as sampling units. A list containing all such sampling units is known as sampling frame. Thus sampling frame consists of a list of items from which the sample is to be drawn. If the population is finite and the time frame is in the present or past, then it is possible for the frame to be identical with the population. In most cases they are not identical because it is often impossible to draw a sample directly from population. As such this frame is either constructed by a researcher for the purpose of his study or may consist of some existing list of the population. For instance, one can use telephone directory as a frame for conducting opinion survey in a city. Whatever the frame may be, it should be a good representative of the population.

c). Sampling design

A sample design is a definite plan for obtaining a sample from the sampling frame. It refers to the technique or the procedure the researcher would adopt in selecting some sampling units from which inferences about the population is drawn. Sampling design is determined before any data are collected.

d). Statisitc(s) and parameter(s)

A statistic is a characteristic of a sample, whereas a parameter is a characteristic of a population. Thus, when we work out certain measures such as mean, median, mode or the like ones from samples, then they are called statistic(s) for they describe the characteristics of a sample. But when such measures describe the characteristics of a population, they are known as parameter(s).

e). Sampling error

It is the difference between total population vale and the sampling value, ot it may be said that it is the degree to which the ‘sample characteristics’ approximate the “characteristics the total population”.

Suppose one parameter of the population pertaining to age is that average age is 20 years. Now, suppose, we have drawn three samples from that population and have calculated the average age for these samples (statistic). In the first sample, the average age is 21 years, in the second it is 24 years and in the third is 26 years. Thus, the sampling error in the first sample would be one year, in the second sample would be four years and in the third would be six years.

2.3.3 Advantages of sampling

Sampling has a number of advantages and some of them are listed below.

1. It is not possible to study large number of people scattered in wide geographical area. Sampling will reduce their number.
2. It saves time and money.
3. It saves destruction of units.
4. It increases accuracy of data (having control on the small number of subjects).
5. It achieves greater response rate.
6. It achieves greater cooperation from respondents.
7. It is easy to supervise few interviewers in the sample but difficult to supervise a very large number of interviewers in the study of total population.
8. The researcher can keep a low profile.

2.3.4 Sample size

2.3.5 Considerations in sample size

A question is often asked: how many persons should be included in the sample, i.e., how large or small must the sample be to be representative? Some people say, the most common size is one-tenth of the total population. Some other say that a minimum of 100 subjects is required to allow statistical inferences. However these estimates are not always correct. The sample size has to be based on the following considerations:

1. The size of the population, i.e., whether the total population to be studied is very large, large or small.
2. Nature of population, i.e., whether the population is homogenous. In the former, a small sample may suffice but in the latter, a larger sample is required.
3. Purpose of study, i.e., whether the study is descriptive, exploratory or explanatory.
4. Whether the study is qualitative or quantitative. In qualitative studies, sampling does not resort to numerical boundaries to determine the size of

sample. Similarly, when purposive or accidental sampling are employed, the researcher himself, can decide the 'sufficient' number of respondents. In such cases, generalisations are concerned with quality rather than with quantity.

5. Accessibility of the elements. Many a time it is difficult to contact respondents at time and place convenient to the researchers.
6. Cost of obtaining elements. With more resources, an adequate number of investigators can be appointed and a large sample may be considered.
7. Variability required. Sometimes the respondents required have to be persons of different groups, e.g., of different age, different income, different educational background, different occupations and so on.
8. Desired accuracy or confidence level. For high degree of accuracy, large samples need to be drawn. One has to think of the level at which one will be confident that his sample is representative.
9. Sampling error or desired risk level. The minimum the sample error, maximum will be the sample's representativeness. For example, the study of parents (with children of school-going age) who want to send their children to English medium private schools or to government schools. If the average annual family income of parents in the area in which the study is to be conducted is Rs. 40,000, then, the researcher should make sure that his sample's average income is as close to Rs. 40,000 as possible. Smaller the percentage error, greater the chance of proving (through the selected sample) that income is one factor that affects parent's choice.
10. Stratification, i.e., how many times the sample has to be sub-divided during the data analysis. This is to ensure an adequate size for each sub-division. In stratified sampling, the researcher should draw a sample having some characteristics as population. In the study of parents willing to send their children to private or government schools, of the total population (of parents), if 75 percent have annual income of over Rs. 40,000 and 25 percent less than Rs. 40,000, the researcher should be sure that his sample also has the same distribution of income.

2.3.6 Mathematical formulas for sample size

Some scholars have suggested mathematical formulas for determining the sample size.

For example, Tara Yamane (1970) has given the following formula:

$$n = \frac{N}{1 + n\ (e)^2}$$

where, N is total population and 'e' is 'error' or confidence level.

Fink and Kosecoff (1995) have presented the following formula to determine the sample size:

$$N = (Z / e)^2 (-p)$$

Where, N is sample size, Z is standard score corresponding to a given confidence level, e is the proportion of sampling error and p is estimated proportion or incidence of cases.

2.3.7 Different types of sampling

There are basically two types of sampling: probability sampling and non-probability sampling. Probability sampling is one in which every unit of the population has an equal probability of being selected for the sample. It offers a high degree of representativeness. However, this method is expensive, time-consuming and relatively complicated since it requires a large sample size and the units selected are usually widely scattered. Non-probability sampling makes no claim for representativeness, as every unit does not get the chance of being selected. It is the researcher who decides which sample units should be chosen.

a). Probability sampling

Probability sampling today remains the primary method for selecting large, representative samples for social science and business researches. According to Black and Champion (1976), the probability sampling requires following conditions to be satisfied:

1) Complete list of subjects to be studied is available
2) Size of the universe must be known
3) Desired sample size must be specified, and
4) Each element must have an equal chance of being selected.

The six forms of probability sampling are: simple random, stratified random, systematic (or interval), cluster, multi-stage and multi-phase.

a) Simple random sampling
   In this sampling, the sample units are selected by means of a number of methods like lottery method, pricking blind foldedly, tippert's tables, computer, personal identification number (PIN) or by first letter.
b) Stratified random sampling
   This is the form of sampling in which the population is divided into a number of strata or sub-groups and a sample is drawn from each stratum. These sub-samples make up the final sample of the study. It is defined as "the method involving dividing the population in homogenous strata and then selecting simple random samples from each of the stratum". The division of the

population into homogenous strata is based on one or more criteria, e.g., sex, age, class, educational level, residential background, family type, religion, occupation and so on.

c) Systematic (or interval) sampling

This sampling is obtaining a collection of elements by drawing every n^{th} person from a pre-determined list of persons. In simple words, it is randomly selecting the first respondent and then every n^{th} person after that; 'n' is a number termed as sampling interval.

d) Cluster sampling

This sampling implies dividing population into clusters and drawing random sample either from all clusters or selected clusters. This method is used when (a) cluster criteria are significant for the study, and (b) economic considerations are significant.

Initial clusters are called primary sampling units; clusters within the primary clusters are called secondary sampling units; and clusters within the secondary clusters are called multi-stage clusters. When clusters are geographic units, it is called area sampling. For example, dividing one city into various wards, each ward into areas, each area into each neighbourhood and each neighbourhood into lanes.

e) Multi-stage sampling

In this method, sampling is selected in various stages but only the last sample of subjects is studied. For example, for studying the panchayat system in villages, India is divided into zones (say, four zones, viz., North, South, East and West), one state is selected from each zone (say, Punjab, Rajasthan, Andhra Pradesh and Assam), one district is selected from each state, one block is selected from each district, and three villages are selected from each block. This will help us in comparing the functioning of panchayats in different parts of India.

f) Multi-phase sampling

The process in this type of sampling is same as in multi-stage sampling, i.e., primary selection, secondary selection, and so on. However, in a multi-phase sampling procedure, each sample is adequately studied before another sample is drawn from it. Consequently, while in multi-stage sampling, only the final sample is studied, in multi-phase sampling, all samples are researched. This offers an advantage over other methods because the information gathered at each phase helps the researcher to choose a more relevant and more representative sample.

b). Non-probability sampling

In many research situations, particularly those where there is no list of persons to be studied (e.g., wife battering, widows, Maruti car owners, consumers of a particular type of detergent powder, alcoholics, migrant workers and so on), probability sampling is difficult and inappropriate to use. In such researches, non-probability sampling is the most appropriate one.

Non-probability sampling procedures do not employ the rules of probability theory, do not claim representativeness, and are usually used for qualitative exploratory analysis. The five types of non-probability sampling are: convenience, purposive, quota, snowball and volunteer.

a) Convenience sampling

This is also known as 'accidental' or 'haphazard' sampling. In this sampling, the researcher studies all those persons who are most conveniently available or who accidentally come in contact during a certain period of time in the research.

The most obvious advantage of convenience sample is that it is quick and economical. But it may be a very biased sample. The possible sources of bias could be: (i) the respondents may have a vested interest to serve in cooperating with the interviewer, and (ii) the respondents may be those who are vocal and/or want to brag. Convenience samples are best utilised for exploratory research when additional research will subsequently be conducted with a probability sample.

b) Purposive sampling

In this sampling, also known as judgemental sampling, the researcher purposely chooses persons who, in his judgement about some appropriate characteristic required of the sample members, are thought to be relevant to the research topic and are easily available to him. For example, suppose the researcher wants to study beggars. He knows the three areas in the city where the beggars are found in abundance. A visit will be made only to these three areas and beggars of researchers choice will be interviewed.

c) Quota sampling

This is a version of stratified sampling with the difference that instead of dividing the population into strata and randomly choosing the respondents, it works on 'quotas' fixed by the researcher.

Determining quotas depends on a number of factors related to the nature and type of research.

Quotas can also be fixed according to their proportion in the entire population. The advantages of quota sampling are: (1) it is less costly than other techniques. (2) It does not require sampling frames. (3) It is relatively effective. (4) It can be completed in a very short period of time.
It also has some limitations. (1) It is not representative. (2) It has interviewer's bias in the selection. (3) Estimating sampling error is not possible. (4) Strict control of fieldwork is difficult.

d) Snowball sampling
In this technique, the researcher begins the research with the few respondents who are known and available. Subsequently, these respondents give other names who meet the criteria of research, who in turn give more names. This process is continued until 'adequate' number of persons are interviewed or until no more respondents are discovered.
This method is employed when the target population is unknown or when it is difficult to approach the respondents in any other way. Reduced sample sizes and costs are a clear advantage of snowball sampling. Bias enters because a person known to someone (also in the sample) has a higher probability of being similar to the first person. If there are major differences between those who are widely known by others and those who are not, there may be serious problems with snowball sampling.

e) Volunteer sampling
This is the technique in which the respondent himself volunteers to give information he holds.

2.4 Difference between Census and Sampling

The paramount differences between census and sampling are discussed below:

1. The census is a systematic method that collects and records the data about the members of the population. The sampling is defined as the subset of the population selected to represent the entire group, in all its characteristics.
2. The census is alternately known as a complete enumeration survey method. In contrast, sampling is also known as a partial enumeration survey method.
3. In the census, each and every unit of population is researched. On the contrary, only a handful of items is selected from the population for research.
4. Census, is a very time-consuming method of survey, whereas, in the case of sampling, the survey does not take much time.

5. The census method requires high capital investment as it involves the research and collection of all the values of the population. Unlike sampling which is a comparatively economical method.
6. The results drawn by conducting a census is accurate and reliable while there are chances of errors in the results drawn from the sample.
7. The size of the sample determines the probability of errors in the outcome, i.e. the larger the size of population the less are the chances of errors and the smaller the size; the higher are the chances of errors. This is not possible with census as all the items are taken into consideration.
8. Census is best suited for the population of heterogeneous nature. As opposed to sampling which is appropriate for homogeneous nature.

2.5 CHECK YOUR PROGRESS

Q1. What is a census?
Q2. What are the various purposes of census?
Q3. What is a sampling design?
Q4. List out the five types of non-probability sampling.
Q5. What is the difference between census and sampling?

UNIT 3: DATA EVALUATION TECHNIQUES

Structure

3.1 Learning Objectives

After studying this lesson, you can:

- Understand objects and types of classification
- Identify the various types of tables
- Understand diagrammatical and graphical representations
- Define Hypothesis
- Appreciate the need for interpreting data

Introduction

Everybody collects, interprets and uses information, much of it in a numerical or statistical forms in day-to-day life. It is a common practice that people receive large quantities of information everyday through conversations, televisions, computers, the radios, newspapers, posters, notices and instructions. It is just because there is so much information available that people need to be able to absorb, select and reject it. In everyday life, in business and industry, certain statistical information is necessary and it is independent to know where to find it how to collect it. As consequences, everybody has to compare prices and quality before making any decision about what goods to buy. As employees of any firm, people want to compare their salaries and working conditions, promotion opportunities and so on. In time the firms on their part want to control costs and expand their profits. One of the main functions of statistics is to provide information which will help on making decisions. Statistics provides the type of information by providing a description of the present, a profile of the past and an estimate of the future.

The following are some of the objectives of collecting statistical information.

1. To describe the methods of collecting primary statistical information.
2. To consider the status involved in carrying out a survey.
3. To analyse the process involved in observation and interpreting.
4. To define and describe sampling.
5. To analyse the basis of sampling.
6. To describe a variety of sampling methods.

Statistical investigation is comprehensive and requires systematic collection of data about some group of people or objects, describing and organizing the data, analyzing the data with the help of different statistical method, summarizing the analysis and using these results for making judgements, decisions and predictions. The validity and accuracy of final judgement is most crucial and depends heavily on how well the data was collected in the first place. The quality of data will greatly affect the

conditions and hence at most importance must be given to this process and every possible precautions should be taken to ensure accuracy while collecting the data.

3.2 Classification and tabulation of data

The collected data, also known as raw data or ungrouped data are always in an unorganised form and need to be organised and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible form.

The process of grouping into different classes or sub classes according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data. Thus classification is the first step in tabulation. For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore, Mumbai etc.,

3.3 Objects of Classification

The following are main objectives of classifying the data:

1. It condenses the mass of data in an easily assimilable form.
2. It eliminates unnecessary details.
3. It facilitates comparison and highlights the significant aspect of data.
4. It enables one to get a mental picture of the information and helps in drawing inferences.
5. It helps in the statistical treatment of the information collected.

3.4 Types of classification

Statistical data are classified in respect of their characteristics. Broadly there are four basic types of classification namely

a) Chronological classification
b) Geographical classification
c) Qualitative classification
d) Quantitative classification

a). Chronological classification

In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc., The data is generally classified in ascending order of time. For example, the data related with population, sales of a

firm, imports and exports of a country are always subjected to chronological classification.

Example

The estimates of birth rates in India during 1990-96 are

Year	1990	1991	1992	1993	1994	1995	1996
Birth Rate	36.8	36.9	36.6	34.6	34.5	35.2	34.2

b). Geographical classification

In this type of classification the data are classified according to geographical region or place. For instance, the production of paddy in different states in India, production of wheat in different countries etc.,

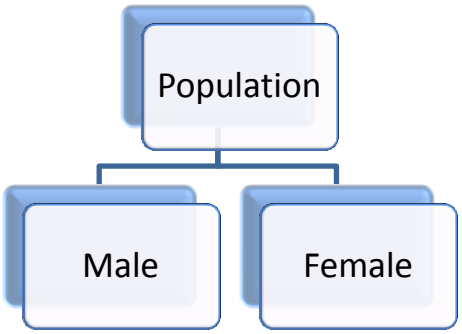
Example

Country	America	China	Denmark	France	India
Yield of wheat in (kg/acre)	1925	893	225	439	862

c). Qualitative classification

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc., Such attributes cannot be measured along with a scale. For example, if the population to be classified in respect to one attribute, say sex, then we can classify them into two namely that of males and females. Similarly, they can also be classified into ‘ employed’ or ‘ unemployed’ on the basis of another attribute ‘ employment’ . Thus when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is called simple or dichotomous classification.

A simple classification may be shown as under

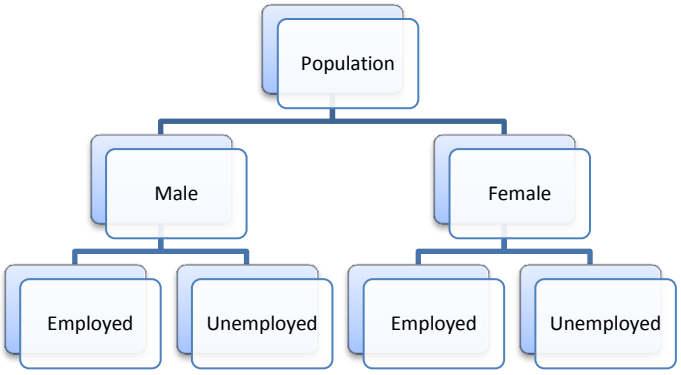


The classification, where two or more attributes are considered and several classes are formed, is called a manifold classification. For example, if we classify population

simultaneously with respect to two attributes, e.g sex and employment, then population are first classified with respect to ‘ sex’into ‘males’ and ‘ females’. Each of these classes may then be further classified into ‘ employment’ and ‘ unemployment’ on the basis of attribute ‘ employment’ and as such Population are classified into four classes namely,

i) Male employed
ii) Male unemployed
iii) Female employed
iv) Female unemployed

Still the classification may be further extended by considering other attributes like marital status etc. This can be explained by the following chart



d). Quantitative classification

Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc., For example the students of a college may be classified according to weight as given below:

Weight (In kg)	No. of students
40-50	50
50-60	200
60-70	260
70-80	360
80-90	90
90-100	40
Total	1000

In this type of classification there are two elements, namely (i) the variable (i.e) the weight in the above example, and (ii) the frequency in the number of students in each class. There are 50 students having weights ranging from 40 to 50 kgs, 200 students having weight ranging between 50 to 60 kgs and so on.

3.5 Tabulation

Tabulation is the process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information.

A table is a systematic arrangement of classified data in columns and rows.
Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious. Classification and 'Tabulation', as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation data are classified and then displayed under different columns and rows of a table.

a). Advantages of Tabulation

Statistical data arranged in a tabular form serve following objectives:

1) It simplifies complex data and the data presented are easily understood.
2) It facilitates comparison of related facts.
3) It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
4) It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided. Moreover, the needed information can be easily located.
5) Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams.

b). Preparing a Table

The making of a compact table itself an art. This should contain all the information needed within the smallest possible space. What the purpose of tabulation is and how the tabulated information is to be used are the main points to be kept in mind while preparing for a statistical table. An ideal table should consist of the following main parts:

1) Table number
2) Title of the table
3) Captions or column headings
4) Stubs or row designation
5) Body of the table
6) Footnotes
7) Sources of data

Table Number

A table should be numbered for easy reference and identification. This number, if possible, should be written in the centre at the top of the table. Sometimes it is also written just before the title of the table.

Title

A good table should have a clearly worded, brief but unambiguous title explaining the nature of data contained in the table. It should also state arrangement of data and the period covered. The title should be placed centrally on the top of a table just below the table number (or just after table number in the same line).

Captions or column Headings

Captions in a table stands for brief and self explanatory headings of vertical columns. Captions may involve headings and sub-headings as well. The unit of data contained should also be given for each column. Usually, a relatively less important and shorter classification should be tabulated in the columns.

Stubs or Row Designations

Stubs stands for brief and self explanatory headings of horizontal rows. Normally, a relatively more important classification is given in rows. Also a variable with a large number of classes is usually represented in rows. For example, rows may stand for score of classes and columns for data related to sex of students. In the process, there will be many rows for scores classes but only two columns for male and female students.

A model structure of a table is given below

Table Number Title of the Table

Sub Heading	Caption Headings	Total
	Caption Sub-Headings	
Stub Sub-Headings	Body	
Total		

Footnotes:

Sources Note:

Footnotes

Footnotes are given at the foot of the table for explanation of any fact or information included in the table which needs some explanation. Thus, they are meant for explaining or providing further details about the data, that have not been covered in title, captions and stubs.

Sources of data

Lastly one should also mention the source of information from which data are taken. This may preferably include the name of the author, volume, page and the year of publication. This should also state whether the data contained in the table is of 'primary or secondary' nature.

c) Requirements of a Good Table

A good statistical table is not merely a careless grouping of columns and rows but should be such that it summarizes the total information in an easily accessible form in minimum possible space. Thus, while preparing a table, one must have a clear idea of the information to be presented, the facts to be compared and the points to be stressed. Though, there is no hard and fast rule for forming a table, yet a few general points should be kept in mind:

1) A table should be formed in keeping with the objects of statistical enquiry.
2) A table should be carefully prepared so that it is easily understandable.
3) A table should be formed so as to suit the size of the paper. But such an adjustment should not be at the cost of legibility.
4) If the figures in the table are large, they should be suitably rounded or approximated. The method of approximation and units of measurements too should be specified.
5) Rows and columns in a table should be numbered and certain figures to be stressed may be put in ' box' or ' circle' or in bold letters.
6) The arrangements of rows and columns should be in a logical and systematic order. This arrangement may be alphabetical, chronological or according to size.
7) The rows and columns are separated by single, double or thick lines to represent various classes and sub-classes used. The corresponding proportions or percentages should be given in adjoining rows and columns to enable comparison. A vertical expansion of the table is generally more convenient than the horizontal one.

8) The averages or totals of different rows should be given at the right of the table and that of columns at the bottom of the table. Totals for every sub-class too should be mentioned.
9) In case it is not possible to accommodate all the information in a single table, it is better to have two or more related tables.

3.6 Type of Tables

Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:

1) Simple or one-way table
2) Two way table
3) Manifold table

a). Simple or one-way Table

A simple or one-way table is the simplest table which contains data of one characteristic only. A simple table is easy to construct and simple to follow. For example, the blank table given below may be used to show the number of adults in different occupations in a locality.

The number of adults in different occupations in a locality

Occupations	No. Of Adults
Total	

b).Two-way table

A table, which contains data on two characteristics, is called a two-way table. In such case, therefore, either stub or caption is divided into two co-ordinate parts. In the given table, as an example the caption may be further divided in respect of ‘ sex’ . This subdivision is shown in two-way table, which now contains two characteristics namely, occupation and sex.

The number of adults in a locality in respect of occupation and sex

Occupation	No. of adults		Total
	Male	Female	
Total			

c). Manifold table

Thus, more and more complex tables can be formed by including other characteristics. For example, we may further classify the caption sub-headings in the above table in respect of "marital status", " religion" and "socio-economic status" etc. A table ,which has more than two characteristics of data is considered as a manifold table. For instance , table shown below shows three characteristics namely, occupation, sex and marital status.

Occupation	No. of adults						Total
	Male			Female			
	M	U	Total	M	U	Total	
Total							

Footnote: M stands for married and U stands for unmarried

Manifold tables, though complex are good in practice as these enable full information to be incorporated and facilitate analysis of all related facts. Still, as a normal

practice, not more than four characteristics should be represented in one table to avoid confusion. Other related tables may be formed to show the remaining characteristics.

3.7 Diagrammatic and graphical representations

Introduction

Earlier, we have discussed the techniques of classification and tabulation that help in summarising the collected data and presenting them in a systematic manner. However, these forms of presentation do not always prove to be interesting to the common man. One of the most convincing and appealing ways in which statistical results may be presented is through diagrams and graphs. Just one diagram is enough to represent a given data more effectively than thousand words. Moreover even a layman who has nothing to do with numbers can also understands diagrams. Evidence of this can be found in newspapers, magazines, journals, advertisement, etc.

a). Diagrams

A diagram is a visual from for presentation of statistical data, highlighting their basic facts and relationship. If we draw diagrams on the basis of the data collected they will easily be understood and appreciated by all. It is readily intelligible and save a considerable amount of time and energy.

Significance of Diagrams and Graphs

b) Diagrams and graphs are extremely useful because of the following reasons.

1) They are attractive and impressive.
2) They make data simple and intelligible.
3) They make comparison possible
4) They save time and labour.
5) They have universal utility.
6) They give more information.
7) They have a great memorizing effect.

c). General rules for constructing diagrams

The construction of diagrams is an art, which can be acquired through practice. However, observance of some general guidelines can help in making them more attractive and effective. The diagrammatic presentation of statistical facts will be advantageous provided the following rules are observed in drawing diagrams.

1) A diagram should be neatly drawn and attractive.
2) The measurements of geometrical figures used in diagram should be accurate and proportional.
3) The size of the diagrams should match the size of the paper.

4) Every diagram must have a suitable but short heading.
5) The scale should be mentioned in the diagram.
6) Diagrams should be neatly as well as accurately drawn with the help of drawing instruments.
7) Index must be given for identification so that the reader can easily make out the meaning of the diagram.
8) Footnote must be given at the bottom of the diagram.
9) Economy in cost and energy should be exercised in drawing diagram.

3.8 Types of diagrams

In practice, a very large variety of diagrams are in use and new ones are constantly being added. For the sake of convenience and simplicity, they may be divided under the following heads:

1. One-dimensional diagrams
2. Two-dimensional diagrams
3. Three-dimensional diagrams
4. Pictograms and Cartograms

a). One-dimensional diagrams

In such diagrams, only one-dimensional measurement, i.e height is used and the width is not considered. These diagrams are in the form of bar or line charts and can be classified as

1. Line Diagram
2. Simple Diagram
3. Multiple Bar Diagram
4. Sub-divided Bar Diagram
5. Percentage Bar Diagram

Line Diagram

Line diagram is used in case where there are many items to be shown and there is not much of difference in their values. Such diagram is prepared by drawing a vertical line for each item according to the scale. The distance between lines is kept uniform. Line diagram makes comparison easy, but it is less attractive.

Simple Bar Diagram

Simple bar diagram can be drawn either on horizontal or vertical base, but bars on horizontal base is more common. Bars must be uniform width and intervening space between bars must be equal. While constructing a simple bar diagram, the scale is

determined on the basis of the highest value in the series. To make the diagram attractive, the bars can be coloured. Bar diagram are used in business and economics. However, an important limitation of such diagrams is that they can present only one classification or one category of data.

For example, while presenting the population for the last five decades, one can only depict the total population in the simple bar diagrams, and not its sex-wise distribution.

Multiple Bar Diagram

Multiple bar diagram is used for comparing two or more sets of statistical data. Bars are constructed side by side to represent the set of values for comparison. In order to distinguish bars, they may be either differently coloured or there should be different types of crossings or dotting, etc. An index is also prepared to identify the meaning of different colours or dottings.

Sub-divided Bar Diagram

In a sub-divided bar diagram, the bar is sub-divided into various parts in proportion to the values given in the data and the whole bar represent the total. Such diagrams are also called Component Bar diagrams. The sub divisions are distinguished by different colours or crossings or dottings. The main defect of such a diagram is that all the parts do not have a common base to enable one to compare accurately the various components of the data.

Percentage bar diagram

This is another form of component bar diagram. Here the components are not the actual values but percentages of the whole. The main difference between the sub-divided bar diagram and percentage bar diagram is that in the former the bars are of different heights since their totals may be different whereas in the latter the bars are of equal height since each bar represents 100 percent. In the case of data having sub-division, percentage bar diagram will be more appealing than sub-divided bar diagram.

b). Two-dimensional Diagrams:

In one-dimensional diagrams, only length is taken into account. But in two-dimensional diagrams the area represent the data and so the length and breadth have both to be taken into account. Such diagrams are also called area diagrams or surface diagrams.

The important types of area diagrams are:

1) Rectangles

2) Squares
3) Circles or Pie-diagrams

Rectangles

Rectangles are used to represent the relative magnitude of two or more values. The area of the rectangles are kept in proportion to the values. Rectangles are placed side by side for comparison. When two sets of figures are to be represented by rectangles, either of the two methods may be adopted. We may represent the figures as they are given or may convert them to percentages and then subdivide the length into various components. Thus the percentage sub-divided rectangular diagram is more popular than sub-divided rectangular since it enables comparison to be made on a percentage basis.

Squares

The rectangular method of diagrammatic presentation is difficult to use where the values of items vary widely. The method of drawing a square diagram is very simple. One has to take the square root of the values of various item that are to be shown in the diagrams and then select a suitable scale to draw the squares.

Pie Diagram or Circular Diagram

Another way of preparing a two-dimensional diagram is in the form of circles. In such diagrams, both the total and the component parts or sectors can be shown. The area of a circle is proportional to the square of its radius. While making comparisons, pie diagrams should be used on a percentage basis and not on an absolute basis. In constructing a pie diagram the first step is to prepare the data so that various components values can be transposed into corresponding degrees on the circle. The second step is to draw a circle of appropriate size with a compass. The size of the radius depends upon the available space and other factors of presentation. The third step is to measure points on the circle and representing the size of each sector with the help of a protractor.

c). Three-Dimensional Diagrams

Three-dimensional diagrams, also known as volume diagram, consist of cubes, cylinders, spheres, etc. In such diagrams three things, namely length, width and height have to be taken into account. Of all the figures, making of cubes is easy. Side of a cube is drawn in proportion to the cube root of the magnitude of data. Cubes of figures can be ascertained with the help of logarithms. The logarithm of the figures can be divided by 3 and the antilog of that value will be the cube-root.

d). Pictograms and Cartograms

Pictograms are not abstract presentation such as lines or bars but really depict the kind of data we are dealing with. Pictures are attractive and easy to comprehend and as such this method is particularly useful in presenting statistics to the layman. When Pictograms are used, data are represented through a pictorial symbol that is carefully selected. Cartograms or statistical maps are used to give quantitative information as a geographical basis. They are used to represent spatial distributions. The quantities on the map can be shown in many ways such as through shades or colours or dots or placing pictogram in each geographical unit.

3.9 Graphs

A graph is a visual form of presentation of statistical data. A graph is more attractive than a table of figure. Even a common man can understand the message of data from the graph. Comparisons can be made between two or more phenomena very easily with the help of a graph.

Some important types of graphs which are quite popular is being discussed below:

1) Histogram
2) Frequency Polygon
3) Frequency Curve

Histogram

A histogram is a bar chart or graph showing the frequency of occurrence of each value of the variable being analysed. In histogram, data are plotted as a series of rectangles. Class intervals are shown on the 'X-axis' and the frequencies on the 'Y-axis' . The height of each rectangle represents the frequency of the class interval. Each rectangle is formed with the other so as to give a continuous picture. Such a graph is also called staircase or block diagram. However, we cannot construct a histogram for distribution with open-end classes. It is also quite misleading if the distribution has unequal intervals and suitable adjustments in frequencies are not made.

Frequency Polygon

If we mark the midpoints of the top horizontal sides of the rectangles in a histogram and join them by a straight line, the figure so formed is called a Frequency Polygon. This is done under the assumption that the frequencies in a class interval are evenly distributed throughout the class. The area of the polygon is equal to the area of the histogram, because the area left outside is just equal to the area included in it.

Frequency Curve

If the middle point of the upper boundaries of the rectangles of a histogram is corrected by a smooth freehand curve, then that diagram is called frequency curve. The curve should begin and end at the base line.

3.10 Hypothesis Testing

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing.

In social science, where direct knowledge of population parameter(s) is rare, hypothesis testing is the often used strategy for deciding whether a sample data offer such support for a hypothesis that generalisation can be made. Thus hypothesis testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing.

Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

A)What is A Hypothesis

Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intends to resolve. Thus, a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts.

Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variable to some dependent variable. For example, consider statements like the following ones: “Students who receive counselling will show a greater increase in creativity than students not receiving counselling” Or “the automobile A is performing as well as automobile B.” These are hypotheses capable of being objectively verified and tested. Thus, we may

conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity.

B) Characteristics of Hypothesis

Hypothesis must possess the following characteristics:

i) Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.

ii) Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis "is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation."

iii) Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.

iv) Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.

v) Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.

vi) Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.

vii) Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.

viii) Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original

problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

C) Procedure for Hypothesis Testing

To test a hypothesis means to tell (on the basis of the data the researcher has collected) whether or not the hypothesis seems to be valid. In hypothesis testing the main question is: whether to accept the null hypothesis or not to accept the null hypothesis? Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between the two actions i.e., rejection and acceptance of a null hypothesis.

The various steps involved in hypothesis testing are stated below:

1) Making a formal statement: The step consists in making a formal statement of the null hypothesis (H0) and also of the alternative hypothesis (Ha). This means that hypotheses should be clearly stated, considering the nature of the research problem.

2) Selecting a significance level: The hypotheses are tested on a pre-determined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% level is adopted for the purpose.

3) Deciding the distribution to use: After deciding the level of significance, the next step in hypothesis testing is to determine the appropriate sampling distribution. The choice generally remains between normal distribution and the t-distribution.

4) Selecting a random sample and computing an appropriate value: Another step is to select a random sample(s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution. In other words, draw a sample to furnish empirical data.

5) Calculation of the probability: One has then to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true.

3.11 Data Interpretation

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings. In case of hypotheses testing studies, if hypotheses are tested and upheld several times, the researcher may arrive at generalizations. But in case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. This may at times result in new questions, leading to further researches. All this analytical information and consequential inference(s) may well be communicated, preferably through research report, to the consumers of research results who may be either an individual or a group of individuals or some public/private organisation.

3.12 Need for Data Interpretation

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

1) It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete world of events. Fresh inquiries can test these predictions later on. This way the continuity in research can be maintained.

2) Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.

3) Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.

4) The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such

a study have to be interpreted on a post factum basis in which case the interpretation is technically described as ‘post factum’ interpretation.

3.13 Check Your Progress

Q1. What are the objectives of collecting statistical information?

Q2. What are the various types of data classification?

Q3. What are the advantages of tabulation?

Q4. What is a histogram?

Q5. What is a hypothesis?

Q6. List out some characteristics of hypothesis.

Q7. What is data interpretation?

Q8. What are some reasons for interpreting data?

UNIT 4: LEVEL OF MEASUREMENT

Structure:

4.1 Levels of Measurement

In our daily life we are said to measure when we use some yardstick to determine weight, height, or some other feature of a physical object. We also measure when we judge how well we like a song, a painting or the personalities of our friends. We, thus, measure physical objects as well as abstract concepts.

Measurement is a relatively complex and demanding task, specially so when it concerns qualitative or abstract phenomena. By measurement we mean the process of assigning numbers to objects or observations, the level of measurement being a function of the rules under which the numbers are assigned. It is easy to assign numbers in respect of properties of some objects, but it is relatively difficult in respect of others. For instance, measuring such things as social conformity, intelligence, or marital adjustment is much less obvious and requires much closer attention than measuring physical weight, biological age or a person's financial assets. In other words, properties like weight, height, etc., can be measured directly with some standard unit of measurement, but it is not that easy to measure properties like motivation to succeed, ability to stand stress and the like. We can expect high accuracy in measuring the length of pipe with a yard stick, but if the concept is abstract and the measurement tools are not standardized, we are less confident about the accuracy of the results of measurement.

Technically speaking, measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence. In measuring, we devise some form of scale in the range (in terms of set theory, range may refer to

some set) and then transform or map the properties of objects from the domain (in terms of set theory, domain may refer to some other set) onto this scale.

4.2 Measurement scales

From what has been stated above, we can write that scales of measurement can be considered in terms of their mathematical properties. The most widely used classification of measurement scales are:

a) nominal scale;
b) ordinal scale;
c) interval scale; and
d) ratio scale.

4.2.1 Nominal scale:

Nominal scale is simply a system of assigning number symbols to events in order to label them. The usual example of this is the assignment of numbers of basketball players in order to identify them. Such numbers cannot be considered to be associated with an ordered scale for their order is of no consequence; the numbers are just convenient labels for the particular class of events and as such have no quantitative value. Nominal scales provide convenient ways of keeping track of people, objects and events. One cannot do much with the numbers involved. For example, one cannot usefully average the numbers on the back of a group of football players and come up with a meaningful value. Neither can one usefully compare the numbers assigned to one group with the numbers assigned to another. The counting of members in each group is the only possible arithmetic operation when a nominal scale is employed. Accordingly, we are restricted to use mode as the measure of central tendency. There is no generally used measure of dispersion for nominal scales. Chi-square test is the most common test of statistical significance that can be utilized, and for the measures of correlation, the contingency coefficient can be worked out. Nominal scale is the least powerful level of measurement. It indicates no order or distance relationship and has no arithmetic origin. A nominal scale simply describes differences between things by assigning them to categories. Nominal data are, thus, counted data. The scale wastes any information that we may have about varying degrees of attitude, skills, understandings, etc. In spite of all this, nominal scales are still very useful and are widely used in surveys and other ex-post-facto research when data are being classified by major sub groups of the population.

4.2.2 Ordinal scale

The lowest level of the ordered scale that is commonly used is the ordinal scale. The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales

and are frequently used in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Ram's position in his class is 10 and Mohan's position is 40, it cannot be said that Ram's position is four times as good as that of Mohan. The statement would make no sense at all. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made. Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. A percentile or quartile measure is used for measuring dispersion. Correlations are restricted to various rank order methods. Measures of statistical significance are restricted to the non-parametric methods.

4.2.3 Interval scale

In the case of interval scale, the intervals are adjusted in terms of some rule that has been established as a basis for making the units equal. The units are equal only in so far as one accepts the assumptions on which the rule is based. Interval scales can have an arbitrary zero, but it is not possible to determine for them what may be called an absolute zero or the unique origin. The primary limitation of the interval scale is the lack of a true zero; it does not have the capacity to measure the complete absence of a trait or characteristic. The Fahrenheit scale is an example of an interval scale and shows similarities in what one can and cannot do with it. One can say that an increase in temperature from 30° to 40° involves the same increase in temperature as an increase from 60° to 70°, but one cannot say that the temperature of 60° is twice as warm as the temperature of 30° because both numbers are dependent on the fact that the zero on the scale is set arbitrarily at the temperature of the freezing point of water. The ratio of the two temperatures, 30° and 60°, means nothing because zero is an arbitrary point. Interval scales provide more powerful measurement than ordinal scales for interval scale also incorporates the concept of equality of interval. As such more powerful statistical measures can be used with interval scales. Mean is the appropriate measure of central tendency, while standard deviation is the most widely used measure of dispersion. Product moment correlation techniques are appropriate and the generally used tests for statistical significance are the 't' test and 'F' test.

2.2.4 Ratio scale

Ratio scales have an absolute or true zero of measurement. The term 'absolute zero' is not as precise as it was once believed to be. We can conceive of an absolute zero of length and similarly we can conceive of an absolute zero of time. For example, the zero point on a centimetre scale indicates the complete absence of length or height. But an absolute zero of temperature is theoretically unobtainable and it remains a concept existing only in the scientist's mind. The number of minor traffic-rule violations and the number of incorrect letters in a page of type script represent scores on ratio scales. Both these scales have absolute zeros and as such all minor traffic violations and all typing errors can be assumed to be equal in significance. With ratio scales involved one can make statements like "Jyoti's" typing performance was twice as good as that of "Reetu." The ratio involved does have significance and facilitates a kind of comparison which is not possible in case of an interval scale. Ratio scale represents the actual amounts of variables. Measures of physical dimensions such as weight, height, distance, etc. are examples.

Generally, all statistical techniques are usable with ratio scales and all manipulations that one can carry out with real numbers can also be carried out with ratio scale values. Multiplication and division can be used with this scale but not with other scales mentioned above. Geometric and harmonic means can be used as measures of central tendency and coefficients of variation may also be calculated.

Thus, proceeding from the nominal scale (the least precise type of scale) to ratio scale (the most precise), relevant information is obtained increasingly. If the nature of the variables permits, the researcher should use the scale that provides the most precise description. Researchers in physical sciences have the advantage to describe variables in ratio scale form but the behavioural sciences are generally limited to describe variables in interval scale form, a less precise type of measurement.

4.3 Check Your Progress

Q1. What is nominal scale?

Q2. What is ratio scale?

References

Books

Ahuja, R. (2002). Research Methods. Jaipur: Rawat .
Kothari, C. R. (2004). Research Methodology: Methods & Techniques . New Delhi: New Age.

Varalakshmi, V., Suseela, N., Sundaram, G. G., Ezhilarasi, S., & Indrani, B. (2017).

Collection of data, classification and tabulation. In Statistics. Chennai: Tamil Nadu Textbook and Educational Services Corporation.

Wimmer, R., & Dominick, J. (2011). Mass Media research: An Introduction. Boston: Cengage.

Web Resources

(n.d.). Retrieved from http://shodhganga.inflibnet.ac.in/bitstream/10603/3704/12/12_chapter%202.pdf
(n.d.). Retrieved May 20, 2017, from http://keydifferences.com/difference-between-census-and-sampling.html