

Block 2
Data Collection and Presentation of
Data

UNIT 4 DATA COLLECTION: METHODS AND SOURCES

Structure

- 4.0 Objectives
- 4.1 Introduction
- 4.2 Methods of Data Collection
 - 4.2.1 Census Method
 - 4.2.2 Survey Method
 - 4.2.3 The Observation Method
 - 4.2.4 The Experimental Method
- 4.3 Planning and Organisation of Census and Surveys
- 4.4 Errors in Data or Data Collection
 - 4.4.1 Non-Sampling Errors
 - 4.4.2 Sampling Errors
- 4.5 Cost of the Enquiry
- 4.6 Census or Survey?
- 4.7 Sources of Secondary Data
- 4.8 Let Us Sum Up
- 4.9 References
- 4.10 Answers or Hints to Check Yours Progress Exercises

4.0 OBJECTIVES

After going through this unit, you will be able to :

- state the methods of data collection;
- discuss the various stages involved in planning and organising census and surveys;
- explain the sources of various types of errors arising in data collection;
- identify the various official sources of secondary data collected and compiled by the various data compiling agencies; and
- identify the considerations which enable us to choose census or survey as a method of data collection.

4.1 INTRODUCTION

An essential input for the data analysis is availability of reliable data on the different variables involved in the issues of any research or project study. Data are needed not only for development of theory but also for evolving economic and social policy and evaluating the impact of such policies. Collection of data has to be carried out on a scientific basis so as to ensure

that the data collected on various variables do reflect, their essential features as accurately as possible. This is because Data analysis involves the process for testing the hypothesis and assessing the efficacy of models for which data are to be collected on scientific basis. This calls for objectivity and integrity on the part of data collector. In the process for data collection several questions arise: how to gather data on scientific basis? Which stages are involved in planning and organising census and surveys? Which errors may arise in data collection? All these questions have been addressed in this unit. Further, Indian statistical system provides data on variety of economic and social phenomenon. Availability of secondary data on various micro and macro variables compiled by various government organisations is very useful for data analysts. Hence you will find the details of websites containing the different types of data in this Unit.

4.2 METHODS OF DATA COLLECTION

Gathering data can be accomplished through a primary or a secondary source.

The data, which are collected from the units or individual respondents directly for the purpose of certain study or information, are known as primary data. For instance, an enquiry is made from each tax payer in a city to obtain their opinion about tax collecting machinery. The data obtained in a study by the investigator are termed as primary data. If an experiment is conducted to know, the observations taken on each plot or patient constitute the primary data.

On the other hand the data, which had been collected by certain people or agency, and statistically treated, is secondary data. Now the information contained in it is used again from records, processed and statistically analysed to extract some information for other purpose, is termed as secondary data. For instance, if the data given in different census years is again processed to obtain trends of population growth, professional changes, changes in sex ratio, mortality rate etc., it is termed as secondary data. Usually secondary data is obtained from year books, census reports, survey reports, official records, or reported experimental findings. Different organisations and government agencies publish such information in the forms of reports, periodicals, journals etc. We shall take up this issue in Section 4.7.

There are broadly four different methods of Collection of Primary data—Census Method, Survey Method, Observation Method, and Experimental Method. Let us discuss these methods in detail.

4.2.1 Census Method

The complete set or collection of animate beings or inanimate objects that are the subject of an enquiry or study is called the population or the universe. The population or universe can be finite, or infinite. An infinite universe is one in which it is theoretically impossible to observe all the units. There are universes that are very large, say, the population of a country but a census does attempt to cover every unit of it. A truly infinite population of

physical objects does not seem to exist. The individual members of the population or the universe are referred to as units. A list of units is called the frame. The units are also called respondent units (or respondents /informants) of the enquiry because these provide the required information to the investigators employed for the enquiry.

A complete enumeration or collection of details from, or about, in the Universe is called a Census. Some of the well-known examples of Censuses in India are:

- i) The population census that is conducted in the first year of every decade since 1872, enumerating the demographic, economic, social and cultural characteristics of the entire population of the country by the Registrar General of India; the last population census was conducted in 2011.
- ii) The economic census conducted from time to time since 1977, enumerating every enterprise engaged in any economic activity other than crop production and plantation (the last one was in 2015) by the Central Statistical Organisation; and
- iii) The livestock census conducted (the last one was in 2003) by the Ministry of Agriculture enumerating cattle, poultry, etc, in the country.

4.2.2 Survey Method

A subset of units selected from the universe is called a **sample** and an enquiry that is confined to the selected sample units is called a **survey or sample survey**. A sample is thus a fraction of the population or the universe. The number of units in a sample is called the **sample size**. The ratio of the **size of the sample** to the size of the population is called **the sampling fraction**. The frame or the list of units of the universe is used for selecting a sample from the universe and is therefore also referred to as the **Sampling Frame**. There are two ways of choosing a sample from the universe – one is random sampling and the other non-random sampling. **Random Sampling** is a method of selection of units from the universe in which every unit in the universe has a specified chance of getting included in the sample. **Non-random sampling** is a method in which such a statement about the chance of any unit in the universe getting included in the sample cannot be made. It may be mentioned here that it is possible with a random sample to draw conclusions regarding the population from which the random sample has been drawn (**parent population**) and also indicate the magnitude of the error to which these conclusions are subject. A random sample survey is thus a cost effective alternative to a census. Some of the well-known random sample surveys are:

The random sample surveys conducted by the National Sample Survey Organisation (NSSO) and the Annual Survey of Industries (ASI) conducted by the Central Statistical Organisation (CSO).

4.2.3 The Observation Method

The Observation Method records data as things occurs, making use of an appropriate and accepted method of measurement. An example is to record

the body temperature of a patient every hour or a patient's blood pressure, pulse rate, blood sugar levels or the lipid profile at specified intervals. Other examples are the daily recording of a location's maximum and minimum temperatures, rainfall during the South West/North East monsoon every year in an area, etc.

4.2.4 The Experimental Method

The Experimental Method collects data through *well designed and controlled* statistical experiments. Suppose for example, we wish to know the rate at which manure is to be applied to crops to maximise yield. This calls for an experiment, in which all variables other than manure that affect yield, like water, quality of soil, quality of seed, use of insecticides and so on, need to be controlled so as to evaluate the effect of different levels of manure on the yield. Other methods of conducting the experiment to achieve the same objective without controlling "all other factors" also exist. Two branches of statistics – The Design and Analysis of Experiments and Analysis of Variance – deal with these.

Check Your Progress 1

- 1) What is census?

.....

.....

.....

.....

.....

- 2) What do you understand by the term 'sampling fraction'?

.....

.....

.....

.....

.....

- 3) State the name of two agencies conducting random sample surveys.

.....

.....

.....

.....

.....

.....

.....

4.3 PLANNING AND ORGANISATION OF CENSUS AND SURVEYS

1) Scope and Objectives of the Enquiry

A successful conduct of a census or a survey, that is, an enquiry, calls for careful **planning** and a good deal of preparatory work. The first is of course to be clear about why the effort is being made or **the scope and objective(s) of the census or survey**. That is, what do we want from the enquiry? We should then translate these objectives into the kind of data that would be necessary to achieve these objectives. This is best done in the form of a set of tables – **the tabulation plan**. The tabulation plan will set out the variables on which data should be collected, the time point or the time period to which the data should relate (**reference period of the enquiry**) the manner in which these will be presented and the different cross-classifications and combinations of the variables involved that would be of interest. **The tabulation plan simply sets out in detail the analytical dimensions of the objectives of the enquiry.**

2) Preparation of the (Sampling) Frame of the Enquiry

From whom do we collect information? This is important while taking a view on the kind of information we wish to collect as this has a bearing on the details to be collected. We, therefore, have to **decide on the respondent units** from which we should collect the data. This would of course depend on the enquiry in question and the choice would be such as to ensure a complete coverage of the universe in question (or the sample to be selected from it) and the kind of **data** we need.

We can then move on to the task of preparing the list of the responding units, called **the frame of the enquiry**. This is also called the **sampling frame**, as this will be used to draw the sample required for a survey. This list may be readily available and can be updated – for example, the house list from the Census Organisation, if the Population Census has recently been undertaken. Otherwise, the sampling frame will have to be prepared by the organisers of the survey. This frame needs to be up to date. Otherwise, the accuracy of the results of the enquiry will be seriously compromised.

3) Finalisation of the Schedule or Questionnaire of the Enquiry

We then move on to the task of translating the requirements of **data** and the tabulation plan to items of information to be collected from the respondent units and organise these items of information into a **logical format** that *can* be **used** to collect information from each of the respondent unit covered by the enquiry. Such a format, **setting** out the items of information to be collected from each unit, is called the **schedule or questionnaire of the enquiry**. We shall discuss the details about the questionnaire/ schedule in next Unit.

4) Manual of Instructions to Field Staff and Training

The schedule is not complete without an accompanying document, namely, Manual of Instructions to Field Staff that contains, besides the scope and objectives of the enquiry;

- Concepts and definitions of the terms used in the schedule (or questionnaire), those relevant to the enquiry and items of information sought in the schedule, (for example, if we are collecting data on income of the household, we have to explain clearly about what all receipts of the household should be treated as income and what should not be so treated. If data on age of individual members of the household is to be collected, we should clearly define 'age' – that it is to be measured as “age in completed years as on, say, 01-01-2018”, so that variations in the manner in which people state their age does not introduce errors in the age data being collected in the schedule.);
- Instructions regarding identification of (responding) units;
- Instructions about the manner in which information should be collected from the responding units; and
- Methods of cross checking answers provided by the units for different items in the schedule to ensure accuracy of the information recorded in the schedule.

An intensive training programme for the field staff-separate programmes for different levels – is another important component of the preparatory phase of the enquiry. The training programme should include training in the field in enquiry conditions.

5) Scrutiny Note and Coding Plan

Other items of preparatory work relate to scrutiny and coding of the filled-in schedules as soon as these reach the processing centers from the field. **A Scrutiny Note** has to be prepared for the guidance of those employed for scrutiny work. The scrutiny staff has to be trained in scrutiny work. The Scrutiny Note indicates (a) the manner in which the schedules have to be checked for errors, (b) the different types of mistakes that may be found in the filled up details against different items and (c) the interrelationships between different items in the schedule and so on.

6) Recruitment and Training

The conduct of an enquiry calls for the deployment of a large number of people with professional expertise and technical skills. The manpower requirements will be low in the initial stages of planning and the number deployed will rise to a peak during the fieldwork and taper off to lower levels after the fieldwork and data processing operations are completed. The manpower requirements will have to be estimated by taking into account the workload involved in the enquiry at different stages and at different levels of supervision and control.

7) Organisation of the Fieldwork

This is the next task. The allocation of the workload, in terms of the number of responding units and time required for identification and collection of information from these units, among investigators is one aspect.

The second is a similar allocation of the workload of supervision and allocation of the same among supervisors.

The third is the organisation of the supply of the requisite number of schedules and related stationery to the field staff in such a way that there is zero idle time in the field. (While the requisite material is supplied to the field staff initially while they set out to the field, the needs arising while they are in the field have to be met immediately.)

The fourth relates to arrangements being in place to ensure quick response to quests from the field for clarifications, on technical, administrative, personnel and personal issues that come up during fieldwork.

The fifth is the provision for a leave reserve for personnel and an administrative system that ensures replacements taking up, with the least possible delay in the fieldwork, field positions falling vacant due to members of the field staff leaving for any reason.

The sixth is the safe dispatch of the filled-in schedules, after the necessary checking up in the field, to the main office for processing of the schedules.

Check Your Progress 2

1) What do you mean by the term sampling frame?

.....

.....

.....

.....

.....

2) List the various stages involved in planning and organising the censuses and surveys.

.....

.....

.....

.....

.....

3) Identify the various steps that need to be taken in organising the fieldwork.

.....

.....

.....

.....

.....

.....

4.4 ERRORS IN DATA OR DATA COLLECTION

Data collected would be useful in achieving the objectives for which these are collected only if they are not subject to errors. How do errors arise in data? Let us consider two classes of errors – (i) **non-sampling errors**, and (ii) **sampling errors**.

4.4.1 Non-Sampling Errors

Let us enumerate the ways in which non-sampling errors arise and how these could be tackled.

- i) **Errors of measurement:** Faulty measurement of the variable involved leads to errors of measurement. For instance, there *can* be errors in the milk example in measuring the milk yield of cows. This type of error is best avoided by clearly defining the variable to be measured and the method of measurement and training the investigators in the method of measurement. While this kind of an error is committed by investigators, errors arise also in the way people report information. For example, people, especially the rural folk, report their age incorrectly. The preference of people to report their ages in digits ending in zero or in five is well-known. Similarly, incomes reported by individuals and households are generally less than what they actually receive. The investigator can prevent these errors by suitable additional questions like age at marriage, ages of the children of the person, the number of children born and the number alive and so on. Statements on income have to be cross-checked with broad items of expenditure of the household and so on. Intensive training of investigators plays an important part in minimising such errors.
- ii) **False or cooked up data:** An investigator may have filled up the details relating to a respondent unit, say, a household, from *out of his own imagination, without even visiting the household*. The only way of preventing such situations is to tighten up the supervision of the field work.
- iii) **The attitude or the mindset of the investigator or ‘investigator bias’ as well of the ‘respondent bias’:** A typical example of this is of a male head of the household providing information on the workforce status or the consumption status or the health status or the views (on gender issues covered in the enquiry) of the womenfolk in his household. If the investigator also has the same mindset, he or she might record in the schedule the information as given by the male head of the household. If the investigator is a gender-sensitive male, he would try to put some probing questions to arrive at a response as near to the true state of affairs in the household as possible. The training programme for the field staff should pay due attention to these matters to free them of attitudes and biases likely to affect the data collected.
- iv) **‘Recall lapse’ on the part of respondents:** An informant may be asked whether he was not well (suitably defined) during the preceding 365 days (reference period), and if so, the nature of the illness and the medicine he took to get well. The accuracy of the informant's response

to these questions will depend on how well he is able to recall his health/medical events in the last year. He might also be responding to such questions in respect of the members of his family and these responses would also depend on his ability to recall these events accurately. The solution for this is not to keep the reference period too long. The investigator should also help refresh the memory of the informant by references to common events like reports in the media of the spread of certain diseases during that period or the events that trigger one's memory.

- v) **Errors arising from a faulty or incomplete frame** distort the results of the enquiry. Careful preparation of an up to date frame is the only solution to overcome this problem.
- vi) **Errors of non-response** arise when information is not available from some of the respondent units either due to the absence of the members of the household at the time of the investigator's (repeated) visits or when the members of the household refuse to part with the information called for in the enquiry. Several units may not respond in a mail enquiry. Such errors can be tackled in the following ways:
 - In the case of a mail enquiry, repeated requests by mail and personal visits may help in collecting data from **all** non-responding units;
 - In the case of a mail enquiry, **a sample** may be selected from among the non-responding units and data collected from such units; data may be collected from the selected non-responding units by another request by mail and through personal visits to the units by investigators; and
 - In the case of a sample survey (through data collection in the field), a method of replacing any non-responding unit by another unit from the universe/population should be decided upon for uniform adoption and suitable instructions given to fieldworkers. For instance, the replacement could be the unit adjacent to the non-responding unit (not already included in the sample).
- vii) **Processing and presentation errors:** These arise due to mistakes in scrutiny, coding, data entry, (manual) tabulation and presentation and printing of tables. Quality checking and supervision is the remedy for this problem.

4.4.2 Sampling Errors

In a sample survey, a fraction of a population or a universe is chosen for conducting the survey. As the survey thus leaves out of its consideration information available in a large part of the universe, the conclusions regarding the parent population drawn from the results based on the chosen sample is subject to error. What is the magnitude of this error? This cannot be estimated in the case of a non-random sample because we know nothing of the relationship between the non-random sample and its parent universe. However, the situation is different if we choose a random sample from the universe. It is then possible to estimate the error, called the

sampling error, to which our conclusions about the universe, derived from the random sample, are subject. In fact, we can even decide the extent of such error that can be tolerated by us and derive from this objective the size of the random sample that should be drawn from the universe.

Some important non-random sampling methods with their advantages and disadvantages can be given here, very briefly.

4.5 COST OF THE ENQUIRY

Time and funds, which are not unlimited resources, act as constraints on the scope and size of an enquiry. Cost and time considerations enter at almost every stage of the enquiry and determine the scope of every aspect of the study. Preparing estimates of the cost of conducting the enquiry is part of the preparatory work relating to the enquiry. After all, the sponsor or the agency launching the enquiry should be aware of the financial implications of the enquiry and mobilise the resources needed for it. Let us list out the factors that impact the cost and duration of the enquiry:

- i) The decision whether the enquiry would be a census or a survey;
- ii) The number of respondent units to be covered and their spread over space and time;
- iii) The amount of data to be collected from each of the respondent units; and
- iv) The **manner** of collecting **data** from the respondent units— by visiting respondent units, by mail, by telephone or by extraction of information from sampled **records**.

4.6 CENSUS OR SURVEY?

Suppose we wish to launch an enquiry. Shall we go in for a census or a sample survey? Let us consider the advantages of one over the other. A sample survey is much less costly than a census. The saving in cost could be used to collect more data through a sample survey. The time required to complete a sample survey and make results available for further use will be shorter than in the case of a census. Sample surveys can, therefore, be conducted at more frequent intervals than censuses. A sample survey is relatively easier to organise than a census. A sample survey being an operation conducted on a much smaller scale than a census, it can

- organise a longer and more intensive training programme for the field staff than is possible in a census;
- concentrate on collecting more detailed and more complex **data** with well trained investigators than is possible in a large scale operation like a census; and
- exercise more effective supervision and control over non-sampling errors than is possible in a census.

As explained in an earlier paragraph, it is relatively less difficult to tackle errors due to non-response in a sample survey than in a census.

The next question that confronts us is – what about the choice between a sample survey based on a random sample and the one based on a non-random sample? The crucial points to be considered in this connection are:

- Can we draw valid conclusions about the universe from the results of the sample drawn from the universe.
- Such conclusions are subject to error (sampling error) as these are based on a sample. Can we find out the size of this error?

These questions are also important in the context of the choice between a census or sample survey. **The answer to these questions can be "yes" only if the sample survey is based on a random sample drawn from the universe.** As mentioned earlier, **the adoption of the random sampling method even makes it possible to choose the size of the sample in such a way as to restrict this error, the sampling error, to the desired level.** A sample survey based on a random sample is thus a better choice than a sample survey based on a non-random sample because of these properties. Non-random sample surveys are generally used for exploratory research studies.

On the other hand a census can provide data for every administrative unit, however small, but a random sample survey cannot because the size of the sample may not be sufficient to provide valid estimates below a certain size of administrative units.

4.7 SOURCES OF SECONDARY DATA

The collection of primary data require a lot of time, money and human resources, etc. Some times either resources are not sufficient to go for the collection of primary data or it may not be feasible to collect primary data. In such situations secondary data are the most available solutions.

The Indian Statistical System with the **Ministry of Statistics and Programme Implementation (MoSPI)** at its apex collects, compiles and disseminates an enormous amount of data on diverse aspects of the Indian economy following, to the extent feasible, international standards prescribed by the **United Nations** and other international bodies. **MoSPI** sets standards to be followed by statistical agencies in the country and coordinates statistical activities carried out by these agencies. It also publishes a comprehensive **Statistical Abstract annually and every month**, bringing together data collected by different agencies at one place, besides a **Guide to Official Statistics**. It has also developed a standardised classification of economic activities, namely, **NIC**, revising it from time to time to keep pace with the structural changes taking place in the economy. Other agencies also bring together and publish comprehensive data on some or all sectors of the economy, these publications being, the annual **Economic Survey** of the Ministry of Finance, the Budget documents, the Five-Year Plan documents, **RBI's Statistical Handbook on the Indian Economy, its monthly bulletin and the website on the database of the Indian Economy**, and the **Human**

Development Reports prepared by the Planning Commission and the State Governments. Most of these data are available in electronic format – in CD ROMS, floppies/diskettes or on the websites of the agencies concerned. The **CMIE (EIU), Mumbai** and the **EPWRF, Mumbai** also bring together and publish comprehensive data on different sectors/aspects of the economy regularly. The **Statistical Handbooks of the State/Union Territory DES** and the **District Census Handbooks** published by the Census Directorates/Registrar general of India provide comprehensive information across sectors at the State and sub-State levels. Finally, the **Report of the National Statistical Commission**, which submitted its report in September, 2001 gives a comprehensive account of the data base of the economy. The search for any kind of data can thus very well start from one of these sources and move on, if necessary, to the primary source.

For your convenience and ready reference, the details of different types of the data available on different websites have been provided in Table 4.1 below:

Table 4.1: The Websites and Data Source

Nature of Data	Data Compiling Agencies	Details of Data Type	Data Websites
Macro-Variable Data: National Income, Saving and Investment	CSO	Input Output Transaction Tables	http://mospi.nic.in/Mospi_New/upload/iott-07-08_6nov12.htm
	RBI	The handbook of Statistics on the Indian Economy	rbidocs.rbi.org.in/rdocs/Publications/PDFs/000HSE13120914FL.pdf
	Private Websites	Macro Variable Data	www.indiastat.com
	System of National Accounts (SNA) 2008	Macro Variable Data	http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf
Agriculture and Industrial Data	Directorate of Economics and Statistics Dept. of Agriculture and Cooperation	Agricultural Data	http://eands.dacnet.nic.in
	Animal Husbandry Statistics Division, Dept of Animal Husbandry, Dairy and Fisheries	Animal Husbandry	Dahd.nic.in/dahd/statistics/animal-husbandry-statistics.aspx

Ministry of Agriculture (Dept. of Agriculture.	Statistics on Agriculture and Allied Activities: Agricultural Statistics at a Glance, Agricultural Situation in India	www.data.gov.in
Dept. of Agriculture	Agriculture Census in India	http://agcensus.dacnet.nic.in
Dept. of Agriculture	Agriculture Input Survey	http://inputsurvey.dacnet.nic.in/nationalsummary.aspx
CACP	The Cost of Cultivation Studies Reports and Related Data	http://cacp.dacnet.nic.in/ViewContents.aspx?Input=1&PageId=36&KeyId=0A
Private Websites	The Data related to the Terms of Trade between Agriculture and Industry	http://articles.economictimes.indiatimes.com/2002-02-04/news/27340862_1_bananas-issue-prices-trade
Dept. of Agriculture	The Pocket Book on Agriculture Statistics	http://eands.dacnet.nic.in/latest_2006.htm
NABARD	Cooperative Movement and Agriculture Credit	http://www.nabard.org/english/home.aspx
Ministry of Finance and RBI	Data on Area, Production and Yield of Crops	http://rbidocs.rbi.org.in/rdocs/Publications/PDFs/000HSE13120914FL.pdf
RBI	The RBI Handbook of Statistics on Indian Economy – 2013 – 14 00	http://dbie.rbi.org.in
NSSO	Periodic Unorganised Sector Surveys	http://mospi.nic.in/Mospi_New/upload/nssoratelists_UnitData.pdf
CSO	Annual Survey of Industries: CMI and SSMI Reports	http://www.csoisw.gov.in/CMS/En/1024-asi-manual.aspx http://www.csoisw.gov.in/

			cms/cms/Files/572.pdf
	Indian Bureau of Mines (IBM)	Mineral Data	http://ibm.nic.in/msmpmar13.htm
	UNSD	Energy Statistics of 224 countries of the world	http://unstats.un.org/unsd/energy/yearbook/2011/004-10.pdf
	ILO	Surveys on Employment and Child Labour related issues	http://labordoc.ilo.org/
	International Energy Agency (IEA)	World Energy Outlook	http://iea.org/statistics
	Food and Agriculture Organisation (FAO)	Food and Agricultural Related Data	www.fao.org
Trade and Finance	DGCI & S	Detailed Data on India's Foreign Trade	http://www.dgciskol.nic.in/new_registration.asp
	RBI	FDI Data	www.rbi.org.in
	Ministry of Commerce and Industries	Monthly Data on FDI	http://dipp.gov.in/English/Publications/FDI_Statistics/FDI_Statistics.aspx
	SEBI	SEBI Handbook of Statistics of Indian Securities Market	http://sebi.gov.in/sebiweb/home/list/4/32/0/0/Handbook%20of%20Statistics
	RBI	Report on Currency and Finance, Report on Trends in Banking	http://dbie.rbi.org.in/DBIE/dbie.rbi?site=home
	SEBI	Handbook of Statistics on the Indian Security Market – 2014	www.sebi.gov.in
Social Sector	Census of India	Data on Employment and Unemployment	http://www.census.nic.in

	MoSPI	NSSO quinquennial surveys on Employment and Unemployment	www.mospi.gov.in
	MoSPI	The Unit Level Data collected in the Surveyed Households. Also available in CD form	http://mospi.nic.in/Mospi_New/upload/nssoratelistsurveyedhouseholdunitdata.pdf
	Ministry of Human Resource Development	Educational Data	http://mhrd.gov.in/sites/upload_files/mhrd/files/statistics/InSCED2014_0.pdf
	NUEPA	Report of the eighth AIES	http://aises.nic.in/surveyoutputs
	Ministry of Environment and Forests	Environment Statistics	http://www.moef.nic.in/sites/default/files/annual_report/AR-2013-14-Eng.pdf

About secondary data, it is necessary to determine and record (i) Who did collect the data? (ii) When was it collected? You should note the following details about every source of information:

- i) Full title of the document, book, journal, map, etc.
- ii) Department/agency/author who has prepared the document, book, journal etc.
- iii) Date/year when the document was prepared, and
- iv) Edition, place of publication and publisher for books and articles.

Check Your Progress 3

- 1) Identify the factors which influence the cost and duration of the enquiry.

.....

.....

.....

.....

- 2) Explain the different types of non-sampling errors that arise in an enquiry.

.....

.....

.....

- 3) Which crucial points are taken into consideration while making a choice between random sample survey and non-random sample survey?

.....

.....

.....

.....

.....

- 4) Which document gives a comprehensive account of the data base of the Indian Economy?

.....

.....

.....

.....

.....

4.8 LET US SUM UP

There are four ways of collecting data, namely, census and sample surveys, observation method and experimental method. A census covers all the units in a population or universe while a sample survey confines its attention to a subset of the population. Sample surveys fall into two categories – random sample surveys and non-random sample surveys. A census or a survey (enquiry) has to be carefully planned and organised. The different stages of the enquiry are: translation of the objectives of the enquiry into data needs, preparation of the sampling frame, the designing and pretesting of the schedule or questionnaire of the enquiry, assessment of the manpower needs of the enquiry, recruitment and training of the field and other staff, organisation of field work, processing of filled-in schedules, analysis and drafting of the report etc.

Estimates of the cost of the enquiry are important as this indicates the requirement of funds so that steps could be taken well in time for mobilising the resources needed.

The data collected in an enquiry could be subject to various types of errors and steps are necessary to minimise these. There are two kinds of errors – non-sampling errors and sampling errors.

Since sample survey is based on information provided by only a subset of a universe, conclusions drawn about the parent universe from the results derived from the sample are subject to error – the sampling error. This error cannot be estimated as we do not know anything about the relationship between the sample and the universe when the sample is a non-random sample. If the sample is drawn randomly from the universe, it is possible to

estimate the sampling error to which the conclusions derived from the random sample about the parent universe are subject.

The decision whether a census or a survey should be the method of conducting an enquiry has to be based on a number of considerations. A survey is relatively much less costly than a census. It takes relatively less time than a census to give results, its scale of operations is smaller than a census and can, therefore, facilitate better control and supervision; and the saving in cost in a survey can be used to collect more data than is possible in a census. A survey based on a random sample can also provide estimates of the characteristics of the parent population.

Censuses can provide information on all administrative units, however small sample surveys cannot. Censuses are also necessary at regular intervals in order to provide benchmark data on a number of variables.

The Indian Statistical System collects, compiles and disseminates an enormous amount of data on diverse aspects of the Indian economy. Secondary data compiled by various data compiling agencies, which is accessible on various websites provided in the Unit, is very useful for data analysts.

4.9 REFERENCES

- 1) Murthy, M.N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, 20411, Barrackpore Trunk Road, Calcutta - 700035.
- 2) Nagar A.L. & Das R. K. (1983). *Basic Statistics*, Second Edition, Oxford University Press, New Delhi.
- 3) Richard Levin & David S. Rubin (1991). *Statistics for Management*, Fifth Edition, Prentice Hall International, Hall of India Private Ltd., M-97, Connaught Circus, New Delhi-110001.

4.10 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

- 1) Read Sub-section 4.2.1 and answer
- 2) Read Sub-section 4.2.2 and answer
- 3) National Sample Survey Organisation (NSSO) and Annual Survey of Industries (ASI)

Check Your Progress 2

- 1) The list of responding units is called the sampling frame.
- 2) See Section 4.3
- 3) See Section 4.3 under point 7

Check Your Progress 3

- 1) See Section 4.5
- 2) See Section 4.4
- 3) See Section 4.6
- 4) See Section 4.7



UNIT 5 TOOLS OF DATA COLLECTION

Structure

- 5.0 Objectives
- 5.1 Introduction
- 5.2 Quantitative and Qualitative Research
- 5.3 Tools of Data Collection: Quantitative Research
 - 5.3.1 Questionnaire
 - 5.3.2 Schedule
 - 5.3.3 Interview
- 5.4 Tools of Data Collection: Qualitative Research
 - 5.4.1 Participant Observation
 - 5.4.2 Non-Participant Observation
 - 5.4.3 Focused Interview
 - 5.4.4 Oral Histories
 - 5.4.5 Case Study Method
 - 5.4.6 Group Discussion
 - 5.4.7 Focus Group Discussion
 - 5.4.8 Key Informants
 - 5.4.9 Narratives
- 5.5 Let Us Sum Up
- 5.6 References
- 5.7 Answers or Hints to Check Your Progress Exercises

5.0 OBJECTIVES

After going through this unit, you will be able to:

- explain the concept and meaning of data;
- differentiate between quantitative and qualitative research in the context of data analysis;
- discuss the various tools used in collection of quantitative data;
- explain the various tools often used in generation of qualitative data in the domain of social sciences; and
- state the distinction between data collection and data generation.

5.1 INTRODUCTION

How do we collect data? This involves two aspects: the methods of collecting information/ data requirement and the tools to be used. You have studied the methods used to collect quantitative data and the various stages involved in planning and organising census and surveys in Unit 4. The data requirement of any proposed study is reflected into items of information to

be collected from the respondents through various instruments. In quantitative research, questionnaire, schedules, interview, observation are frequently used to set out the items to be collected from the respondent units. In qualitative research, data is generated/ compiled by way of participant observation. Semi-structured interviews, life histories, experiments, case study, focus group discussion and narratives etc.

Thus, tools for gathering data vary in quantitative and qualitative research. Let us, therefore, first understand and make the distinction between quantitative research and qualitative research.

5.2 QUANTITATIVE AND QUALITATIVE RESEARCH

Quantitative Research is the systematic and scientific investigation of quantitative properties and phenomena and their relationships. The objective of quantitative research is to develop mathematical models, and to test hypotheses. Based on ontological and epistemological premises, quantitative research is characterised by the following attributes:

- A belief in a single reality. The pursuit of identifying universal laws.
- Separation of knower (researcher) from known (researched).
- The possibility and necessarily of value free research.
- Pursuit of universal laws (findings) beyond the limit of research/social context.
- The tendency to work with large, representative sample.
- An emphasis on deductive research via falsifiable hypothesis.
- Formal hypothesis testing.

In contrast, **Qualitative Research** captures a set of purposes associated with meaning and interpretation. Stress is laid on the socially constructed nature of reality, the intimate relationship between the research and researched and situational constraints that shape the enquiry. The key attributes associated with qualitative research are:

- A belief in a constructed reality, multiple realities, co-existence realities.
- Interdependence between the knower and the known.
- Impossibility to separate the researcher from the research subject.
- Heavy role of context in research process.
- The impossibility to generalise research findings beyond the limits of the immediate context.
- Non-separation of cause and effect.
- The explicit focus on inductive, exploratory research approaches.
- The tendency to work with small and purposely chosen sampling.
- Analyses holistic system.

At this stage let us understand the distinction between quantitative data and qualitative data. Quantitative data are kind of data which are based on numbers and frequencies. On the other hand qualitative data (variable) refer to the attributes on which choice can be expressed (and a score assigned) like one's feeling or opinions (e.g. good/bad, male/female, agreeing/disagreeing, yes/no). Even variables which are expressed in qualitative terms can be counted where upon their total numbers become quantified expressions. A popular example of developmental economics which ranks different countries based on their status attained in a number of areas/variables (some of which like the political freedom/ choice enjoyed by the country, though expressed in some quantitative measure, is qualitative in its nature) is the human development index.

In actual research (particularly research in the field of social sciences), it is important to note that variables of both the nature normally co-exist.

5.3 TOOLS OF DATA COLLECTION: QUANTITATIVE RESEARCH

5.3.1 Questionnaire

The questionnaire is an important tool for data collection in the field of social sciences particularly in case of big enquiries. Basically, the questionnaire is a research instrument consisting of a series of questions related to the research topic in written or printed form, and respondents have to answer these questions in the space meant for it. Sir Francis Galton first invented this method. It is used where factual information from the respondents is desired.

Questionnaire is mainly suitable for survey type of research, where large numbers of respondents have to be contacted and large amount of data has to be gathered. Therefore, questionnaire as a tool of data collection appears as the most suitable and appropriate for survey.

Main Aspects of a Questionnaire: The main aspects of a questionnaire consist of the general form, question sequence, question formulation and wording.

- 1) **General Form:** Questionnaire can either be structured or unstructured. Structured questionnaires include the questions which are definite and the respondents come out with concrete and pre-determined answers. The questions are presented with exactly the same wording and in the same order to all respondents. The questions are framed in advance and may be either closed or open. Such questionnaires may also have fixed alternative answers in which responses of the informants are limited to the stated alternatives.

The questionnaires not having the above properties are termed as unstructured questionnaires. Unstructured questionnaires are generally used in exploratory qualitative research studies. Such questionnaires are used to probe the person's attitude and reasons for certain actions or feelings.

2) Question Sequence: In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because such questions influence the attitude of respondent in seeking the desired information. The following type of questions may be avoided in the formulation of a questionnaire:

- i) Questions that put strain on the memory or intellect of the respondent.
- ii) Questions of a personal character.
- iii) Questions related to personal wealth/ income.

Every question should have some relevance to the research problem and the research questions to be answered. The investigator should look at the framed questions how they will add the analysis of the study.

3) Question Formation and wording: Every question in the questionnaire should be very clear. Questions should be impartial in order to give an unbiased picture of the true state of affairs. Questions should be constructed taking into consideration the logical part of a well thought out tabulation plan. The question should be (a) easily understood (b) simple i.e. should convey only one thought at a time (c) be concrete.

The questions can be either multiple choice question or open end questions. The multiple choice questions have the advantage of easy handling, simple to answer, quick and easy to analyse. They are most amenable to statistical analysis. However, such questions are not appropriate when the issue under consideration is complex. In such situations open-ended questions are useful. However open-ended questions are more difficult to handle raising the problem of analysis and interpretation.

In practice, questions of different forms are included in one single questionnaire.

Characteristics of a good questionnaire

A good questionnaire must have following characteristics.

- a) **Covering letter:** Questionnaire should contain a covering letter highlighting the purpose of the study and the confidentiality of the responses.
- b) **Length of the questionnaire:** As far as possible, questionnaire should be kept short and number of the questions to the minimum. Less number of questions lead to the greater in chances of getting a better response.

- c) Simplicity in language: Questions should be simple, short and direct. In choosing the language for questions, researcher must keep in view the population for which the questionnaire is constructed. Technical terms should be avoided and words should convey the exact meaning.
- d) Questions of sensitive or personal nature should be avoided: The question should not be such kind of that require the respondents to disclose any private, present or confidential information.
- e) Double negatives: Questions containing double negatives should be avoided. Such negatives create confusion for the respondents.
- f) Logical arrangement: The questions should be logically arranged to maintain the continuity of responses. It should proceed from general to specific questions and from simple to more complex one.
- g) The questions should be objective and should not provide any hints or suggestions regarding possible answers.
- h) A questionnaire must be attractive, neatly printed, and clearly arranged.

Guidelines for formulation of a questionnaire have been provided in Appendix A of this Unit.

5.3.2 Schedule

The schedule is a sort of proforma, which contains a list of questions filled by the researchers. Goode and Hatt (1987) has defined Schedule as *“Schedule is the name usually applied to a set of questions which are asked and filled in by interviewer in a face-to-face situation with another person.”* Researchers go to the informants with the schedule, and the researchers ask them from set of questions in the sequence and record the responses in the space provided. In certain situations, researchers may distribute the schedule to the respondents and assist them in answering the questions.

Investigators play a major role in the collection of data using schedules. Investigators explain the aims and objectives of the research to the respondents, and sometime interpret the questions to the respondents whenever required. This method is little expensive as the selection, appointment and intensive training of the researchers are essential. Most common example of data collection through schedule is population census and other extensive enquiries conducted by the government agencies.

Features of Schedules

The features of schedules are:

- The interviewer or researcher asks the items of schedule to the respondents and notes down the answers.
- The schedule can be used in a very narrow sphere of social research.

Purpose of Schedule

The main purposes of schedule are three fold:

- 1) It is a standardised tool for observation or interview, so it helps in attaining objectivity in the scientific research.
- 2) It works as a reminder for the researcher and it keeps the memory of the interviewer/ observer refreshed. Schedule reminds the researcher about the different aspects/dimensions that are to be observed.
- 3) Schedule facilitates the work of tabulation and analysis.

Types of Schedule

There are several kinds of schedule:

Rating Schedule: Rating schedule is a kind of schedule which is used to obtain opinions, preferences etc. from the respondents regarding the phenomenon studied. The schedule comprises of positive and negative statements of opinion on the phenomenon.

Documents Schedule: Document schedule is used to collect data/information from recorded evidences and/or case histories e.g. census data, government policy documents etc. Document schedule comprises of blanks and its related functional issues which are likely to be filled up from the records and documents are present.

Survey Schedules: Survey Schedules are like questionnaires.

Observation Schedule: Observation Schedule is used for data collection when observational method is being used.

Essentials of a Good Schedule

A good schedule must have the following features:

- **Content:** A good schedule comprises of questions or statements related to all essential aspects of the study.
- **Dissectional:** A good schedule not only includes all aspects of problem analytically but also dissecting every major and significant components of the problem.
- **Context:** A good schedule goes well with the context in which the schedule is applied. Different studies require to use different appropriate contextual schedules.
- **Criterion:** A schedule should use sound logic in classifying respondents based opinion expressed.
- **Construction:** A schedule should be constructed in such a way that the difficulty level of questions/statements progress gradually in ascending order. Most appropriately, it should be divided into sub-parts, and each part is inquiring about one specific sub-theme.
- **Qualities related words should be avoided:** Good schedule should never include long and complex items or statements with the quality like- presumptuous, personal, embarrassing, hypothetical, morality disoriented, etc.
- **Language:** Good schedule should have clear and straightforward language.

- **Reliable:** A schedule should be reliable, it means whenever the schedule is used in a specific sample results should remain same.

Advantages

- The schedule helps the researcher to collect maximum responses.
- It helps to ascertain accurate information.
- Proper administration of schedules leads to bias free research.
- It helps to conduct research in more difficult situation.
- It is useful and applicable for the educated as well as uneducated participants.

Disadvantages

- It is more expensive and costly.
- Schedule is more time consuming than questionnaires and other methods of research investigation
- Its administration requires more skillful and experienced investigators.
- Wide-range coverage is not possible through schedule.

Difference between Questionnaire and Schedule

Questionnaire and schedule are most popular methods of data collection in social sciences. There are lots of similarities between questionnaire and schedule but they differ on following dimensions:

- 1) A questionnaire comprises of a series of written questions along with alternative answers. Whereas, schedule is a formalised set of questions, statements, and spaces for answers, provided to the investigators who ask questions to the participants and note down the responses.
- 2) Questionnaires can be delivered to the participants by post or mail and answered sheets could be received by the same. On the other hand, schedules are filled by the researchers, who interpret the questions to the participants if it is required so.
- 3) The non-response rate would be high in questionnaires as compared to schedule.
- 4) A questionnaire can be circulated among large number of people at the same time, and even to the sample those are not approachable. Whereas, a schedule can be administered by the trained and skilled investigators only, and it can not be conducted through mail or post.
- 5) A questionnaire method is comparatively cheaper and economical than the schedule method.
- 6) The administration of questionnaires is applicable for literate and co-operative respondents only; whereas, schedule can be used for data collection from all classes of participants.

5.3.3 Interview

The interview is a kind of research method that permits an exchange of ideas and information between the interviewer and the interviewee. Goode and Hatt (1952) defined interview as a fundamental process of social interaction. According to Young (1966), "The interview may be regarded as a systematic method by which a person enters more or less imaginatively into the inner life of a comparative stranger."

In the initial phase of the interview, a proper rapport should be established between the interviewer and the interviewee. At the beginning of the interview, proper physical proximity (such as eye contact, nearness, etc.) is maintained, and other social and cultural barriers are also removed, if any. It consists of the oral/verbal presentation of stimuli and response in the same mode.

Types of Interview

In the Interview method generally interviewer asks questions generally in a face-to-face situation with the interviewee. It is a direct form of personal investigation. On the basis of form, interviews can be of two types – structured and unstructured.

Structured Interviews

Structured interviews comprise of a fixed set of predetermined questions and standardised methods to record the responses. The interviewer follows a firm procedure while asking questions during the interview. In quantitative research structured interviews are resorted.

Unstructured Interviews

Unstructured interviews comprise of the flexible approach in terms of asking questions. It does not follow a system of predetermined questions and standardised techniques of recording responses. The interviewer has sufficient freedom and flexibility in asking supplementary questions. The interviewer is free to answer the questions in his/her own way. There is no fixed response. The interviewer may omit some questions if the situation does not require so. However, this sort of flexibility results in lack of comparability between two interviewers. For effective conduction of unstructured interviews, it requires deeper knowledge and skills in conducting the interview.

In explanatory and qualitative research, unstructured interview method can be the central technique of collecting information. However, in case of descriptive studies, structured interviews are used because of it being more economical, providing a safer basis for generalisation, and requiring relatively lesser skill on the part of the interviewer to conduct the interview.

Advantages of Interview

- In-depth and insightful information can be obtained through the interview method.
- In case, the respondent displays resistant behaviour the interviewer can use different interviewing skills to overcome it.

- This method of research has a lot of flexibility and freedom to restructure/modify the questions as per the requirement of situations especially with unstructured interview method.
- Interview method is very useful in revealing personal information of the respondents.
- The language difficulty level of the interview can be adopted according to the ability and educational level of the interviewee.
- The interview method is also helpful in collecting supplementary information about the interviewee's personal characteristics and environment that is often of great value in interpreting results.

Disadvantages

- Interview method could be financially very expensive especially when large and widely spread geographical sample is taken for the study.
- In spite of several advantages, there is the possibility of the bias from both the interviewer and interviewee.
- A group of participants such as executives and officials or people belongs to higher socio-income groups may not be easily approachable.
- Conducting interviews itself is a time-consuming process especially when the participant is not expressive in nature i.e. shy, hesitant, and introvert people.
- Selection, training and supervision of interviewers are more complex formidable task; especially, when working with multiple interviewers at a time.
- Effective interview presupposes proper rapport with the respondents that would facilitate free and frank responses. This is often a very difficult requirement.

In addition to the above traditional important tools, several mechanical, electro-mechanical or electronic devices like audio recorder, video camera, emails, whatsapps, websites, etc. are also used for collection of data.

Guidelines for conducting the interview have been provided in Appendix B of this Unit.

Check Your Progress 1

- 1) State the two important characteristics of qualitative research.

.....

.....

.....

- 2) What is the distinction between quantitative and qualitative variable?

.....

.....

.....

- 3) State whether following statements are true or false:
 - i) In quantitative research knower and known are inter-dependent.
 - ii) The study which pre-supposes that reality is socially constructed is quantitative in nature.
 - iii) In qualitative research, cause and effect are separate to each other.
 - iv) Questions that put strain on the memory or intellect of the respondent need to be avoided.
- 4) Distinguish between questionnaire and schedule.

.....

.....

.....

.....

5.4 TOOLS OF DATA COLLECTION: QUALITATIVE RESEARCH

5.4.1 Participant Observation

Uncontrolled Participant observation is defined by Goode and Hatt (1981: 121) as a procedure in which “the investigator can so disguise himself as to be accepted as a member of the group”. When researchers live at a field site for long duration, the assumption is that they are regarded as member of the community or the group. A researcher wanting to study consumer behaviour may not disclose his identity as a researcher and pretend to be a consumer, sharing their experiences and concerns. While doing participant research, researchers often assume insider identity. Sometimes researchers take neutral positions without disclosing the real purpose of their being there.

Advantages of Method of Participant Observation

Researcher can record all the information that is relevant for his work without disturbing the community and intruding upfront into their privacy. Instead of sitting in a corner in the market, an organisation or any other work or field situation writing notes, the researcher participates in these activities. In doing so, he collects relevant information without disturbing the normal activities and records his observations in a work diary there and then makes a mental note of it and returns to his/her workplace and records these observations immediately. The community or people may not be aware that they are being researched. This helps obtain unhindered and unbiased information.

Participant observation helps obtain detailed information without any subject bias. Many researchers regard it as a better tool for data gathering vis-à-vis questionnaire and in-depth interviews.

Drawbacks of Method of Participant Observation

One of the major drawbacks of participant observation is its inability to develop procedures for standardisation. Researcher doing participant observation acquires a specific position in the community and observes from this vantage point. His recordings are personal or individual specific. Another observer observing the same situation may not be able to view it with the same perspective.

In participant observation method, the researcher may get emotionally involved with the issues and tends to lose his/her objectivity. This often happens on issues of dowry, bride burning, female foeticide, student agitations, price rise, public-private partnership, trade union activities etc.

To come over some of these drawbacks, researchers proposed method of non-participant observations.

5.4.2 Non-Participant Observation

The qualitative researchers innovate in several ways to do Non-Participant observations. Many of them would participate in some activities and observe others from the outside. Some others would use video-recordings and analyse these recordings later for generalisations. While doing so, it was possible to take into account people's gestures and non-verbal movements for detailed scrutiny. However, critics of the method thought that as one is not fully participating in this situation, he/she is not in a position to attribute right meaning to these observations.

Quasi-Participant Observation

Most researchers these days prefer to use quasi-participant method of observations. In this method of research, the researcher can follow the dual entry system. In this system for instance, if you are trying to examine labour management relations, then it is imperative that you obtain permission from the management and also establish rapport with the trade union leaders to have access to the union meetings. Many a times, one has to conduct interviews in different locations. In one such study in Punjab and Uttar Pradesh, we found that the workers preferred to talk to the researcher outside the factory. The researcher would talk to the managers and supervisors in the factory and had lunch in the workers canteen. Appointments were fixed with the workers while having lunch and then researcher went to their houses, often spent hours there and after completing the interviews with the supervisors and managers hired a small room in the same slum area, where most of the workers lived and shared their daily experiences.

In addition to the important research tool of observations, qualitative researchers often use Focused interview, in-depth interviews, narratives and oral /Life case histories, Focus group discussion/interviews', and content analysis.

5.4.3 Focused Interview

In a focused interview, the researcher and the respondent are engaged in face-to-face or one-to-one conversation in which the researcher poses

questions and the respondent provides detailed or precise answers. Talking to people or a specified sample, using an interview schedule, interview guide or a structured questionnaire is standard method of data collection. For qualitative research, focused interviews are often unstructured and questions are invariably open ended. This gives liberty to the respondent to provide extensive answers.

It is important to remember that interview is “fundamentally a process of social interaction” (Goode and Hatt, 1981: 186) and involves developing a social relationship. In an interview situation it is not only the interviewer but also the interviewee make an opinion about the researcher. Respondent’s responses can be determined by what he thinks researcher may expect from him and also the amount of trust and faith he invests in the researcher. We will repeatedly stress the importance of confidentiality and researcher’s ability to convey the same as hall marks of participatory research.

5.4.4 Oral Histories

Uninterrupted long interviews giving space to the respondents to express his opinion or experiences at length is the preferred method by several qualitative researchers. The interviewee is encouraged to share his life experiences, to express his opinion, to recollect and recount their antecedents or lives of their contemporaries, “and to discuss their perceptions of the processes involved and the changes they have seen”(Babbie, 2010:207). Collecting oral histories or narratives requires skill of being a good listener. Many a times, you may assume that the information provided by the respondent is not directly related to the area of your focused interest. But that subsidiary information can provide certain reference for a better understanding of the problem, issues or question that you are exploring. In oral histories, one may pose probe questions in the beginning and then occasionally intervene to ask a lead question for seeking any detailed answer to specific query.

LIFE HISTORY: Another important tool that ethnographic and qualitative researchers have in their kit is called the life history method. In this method life histories of individuals are reconstructed to understand the historical events of that period. Many a times researchers would use diaries and autobiographies to construct chronology of events of that period. Experience of individuals facilitates understanding of how and what happened. Life history approach is largely dependent on the ability of the respondent to recall and share that memory with the researcher. With the help of the respondent, researcher gradually determines the social processes. In the context of economics, rise and fall of a particular model is determined by prevalent social processes of that period, e.g. popularity and weakening of models of conservative and liberal economics in the last fifty years. It also helps to understand different perspectives based on gender and class. While recollecting their life history, many a times respondents may fall silent. This silence also provides data as it tells the researcher, how that particular event may have caused anxiety, anguish or an intense emotional experience to the respondent.

5.4.5 Case Study Method

To study firms and organisation behaviour, Business economists use the case study method. The researcher can focus on one company or organisation and collect a detailed case history of the organisation. Yin (1984:23) defines case study research method “as an empirical inquiry that investigates a contemporary phenomenon within its real life context; when the boundaries between phenomenon and context are not clearly evident and in which multiple sources of evidence are used”. In an organisation, there are multiple players and a detailed case study will take into account perspective of each stakeholder there. You can use more than one method to collect evidence from different stake holders e.g. to collect information from the supervisors, you can use a schedule or a questionnaire, to talk to the workers follow the interview method, to collect history of the organisation and organogram depicting position and functions of each member of the organisation, use content analysis by examining available records of the organisation. This provides a holistic picture of the organisation. Case study as an approach is holistic, inductive and idiographic. It takes complete account of the organisation from all perspectives; theorising moves from general to specific and individual cases are located in their particular context.

5.4.6 Group Discussion

Group discussions have a special advantage over personal interviews in the sense that a larger body of information can be collected, covering larger groups of people, in a short time. Group discussions are also useful to cross check information. Also, certain information that may be sensitive is more easily obtained in larger gatherings where the source cannot be pin-pointed to one individual alone. Such information often relates to misuse of funds and resources or to maltreatment or violence directed against certain groups of people. However, selections of the group should be done carefully and it must be homogeneous in nature.

5.4.7 Focus Group Discussion

One of the methods of qualitative research that found ready acceptance by economists was that of focus group. It is frequently used in market surveys, for quick appraisal of response to an economic programme, for evaluation studies and for media and health research. It is also useful to elicit information regarding social customs, food patterns and behaviour, infant feeding practices, and also local food and fads and taboos. In focus group research, a small group of six to eight (market researchers prefer groups of 10-12 participants in each group) targeted audiences is collected in a suitable location. They are encouraged to have a frank discussion on a given topic and a recorder sitting outside the group will either audio record or take notes on the discussion. The researcher acts as a prompter posing questions on the subject of inquiry. Kruger (1988: 18) defines focus group as a “carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment”.

Bloor et al. (2001) caution that 'focus groups are the method of choice only when the purpose of the research is to study group norms, group meanings and group processes'. (cf. Barbour, 2008:133)

Steps for focus group

- Topic for focus group discussion should be of general interest to all the participants in the group.
- Number of participants can vary depending on the requirements of your research. Sampling is the key to data generation.
- 'Stimulus material' for initiating the discussion should be brief and should encourage a free flowing conversation. Use of video clips, leaflets, cartoons, brief recollect of event or topic being discussed etc.
- Moderator must manage the discussion in a manner that without sounding intrusive, long extensive talks should be avoided.
- Barbour (2008) suggests 'anticipate the analysis through attentive moderating'.
- Locate themes visible in the form of repetitive patterning of the responses.
- Second stage sampling could be considered to further explore any emerging theories, paradigms or even trends from the data.

These are some of the tools that participatory researchers frequently use to gather primary data. In this brief discussion, there are two more important steps that a participatory researcher must remember. First is **grounded theory** and second is **universe selection and sampling** procedure.

5.4.8 Key Informants

Key informants can be a major source of information. People from the community who, because of official position or informal leadership, have access to information about the community rather than individual problems, are good resources. Key informants can be government officials, local health service personnel, traditional healers, community leaders (elected or self-appointed), local shop owners and members of non-governmental organisations. However, leaders having biases against a certain group of people need to be excluded among key informants.

5.4.9 Narratives

Social science research in anthropology, sociology, political science, history and now economics is moving from the positivist position of research to narrative approach. The positivist position also popularly known as scientific approach was the tradition in which researchers from 50's 80's were generally trained. The standardised tools of survey research, questionnaire approach or even interviews with a standardised set of questions was expected to produce if not same, similar responses. Narrative research emphasised that this is not true. Different respondents may interpret each question asked in the same words differently. In structured interviews both meaning and plurality of responses is lost, researcher-using method of

narrative understands that there can be standardised questions but there are no 'standardised meanings'.

Check Your Progress 2

- 1) What do you mean by participant observation?

.....

.....

.....

.....

- 2) How non-participant observation removes – drawback of participant observation?

.....

.....

.....

.....

- 3) How does focused group discussion is useful in generating data for in-depth qualitative research?

.....

.....

.....

.....

- 4) State the features of case study method as a tool for data generation.

.....

.....

.....

5.5 LET US SUM UP

The data requirement of any study is translated in terms of items of information to be collected from the respondents through various tools. The tools to be used depend on the type and nature of research. In quantitative research apart from the questionnaire, schedule and interview, various mechanical, electro-mechanical or electronic devices are used. In qualitative research, in addition to semi-structured questionnaire and interviews, various other tools like participant and non-participant observation, group discussion, focus group discussion. Oral histories, case study etc are resorted for data/ information generation.

However, in practice, mixed methods research are increasingly being used wherein different tools are used in single study.

5.6 REFERENCES

- 1) William J. Goode Paul K. Hatt (1952). *Methods in Social Research*. McGraw-Hill Book Co., New York.
- 2) Babbie, E. (2010). *The Practice of Social Research*. Wadsworth Cengage Learning, USA.
- 3) Kerlinger, F. N. (1986). *Foundations of Behavioral research*. New york : Holt Rinehart and Winston.
- 4) Kothari, C. R. (2009). *Research Methodology : Methods and Techniques*. New Age International Publication, New Delhi.
- 5) Kumar, Ranjit (2011). *Research Methodology (A step by step guide for beginners)*. Sage publication.
- 6) Singh, Y. K. (2006). *Fundamentals of Research Methodology and Statistics*. New age International, New Delhi
- 7) Young, P.V. (1988). *Scientific Social and Survey Research*. Prentice Hall of India, New Delhi.
- 8) IGNOU (2006), MEC-009, *Research Methods in Economics*, Block-5, Unit-14, Qualitative research and techniques.
- 9) IGNOU (2015), MEC-109, *Research Methods in Economics*, Block-1, 2 & 5, Unit-1, 2, 6, 7 & 18.
- 10) Yin, R.K. (1984). *Case Study Research: Design and Methods*. Sage Publications, Beverly Hills, California.
- 11) Krueger, R. A. (1988). *Focus groups: A practical guide for applied research*. Thousand Oaks, CA, US: Sage Publications, Inc.
- 12) Michael Bloor, Jane Frankland, Michelle Thomas & Kate Robson (2001). *Focus Groups in Social Research*. London: Sage.

5.7 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

- 1) Read Section 5.2
- 2) Read Section 5.2
- 3) (i) False (ii) False (iii) False (iv) True
- 4) See Sub-section 5.3.2

Check Your Progress 2

- 1) Read Sub-section 5.4.1
- 2) Read Sub-section 5.4.2
- 3) See Sub-section 5.4.7
- 4) See Sub-section 5.4.5

UNIT 6 DATA PRESENTATION

Structure

6.0 Objectives

6.1 Introduction

6.2 Classification of Data

6.2.1 Objectives of Classification

6.2.2 Types of Classification

6.2.2.1 Simple Array

6.2.2.2 Discrete Frequency Distribution

6.2.2.3 Grouped Frequency Distribution

6.2.2.4 Types of Grouped Frequency Distribution

6.2.2.5 How to Use Spreadsheet Software for Frequency Distribution?

6.3 Tabulation of Data

6.3.1 Components of a Statistical Table

6.3.2 General Rules for Preparing a Table

6.3.3 Importance of Tabulation

6.4 Diagrammatic Presentation of Data

6.4.1 Bar Diagram

6.4.2 Multiple Bar Diagram

6.4.3 Overlapping Bar Diagram

6.4.4 Component Bar Diagram

6.4.5 Pie Chart

6.4.6 Pictogram

6.5 Graphical Representation of Data

6.5.1 Line Graph

6.5.2 Histogram

6.5.3 Frequency Polygon

6.5.4 Frequency Curve

6.5.5 Cumulative Frequency Curve or Ogive

6.5.6 Importance of Graphical Presentation of Data

6.5.7 Misuse of Graphical Presentations

6.6 Let Us Sum Up

6.7 References

6.8 Answers or Hints to Check Your Progress Exercises

6.9 Exercises

6.0 OBJECTIVES

After going through this unit, you will be able to:

- explain the process of grouping data into classes and sub-classes based on their characteristics;

- enter the data in a spread sheet;
- state the general rules for constructing the table;
- represent the data through various types of bar diagrams;
- draw the line graphs, histogram, frequency polygon and the frequency curves from the given data set by using spread sheets.

6.1 INTRODUCTION

Data are set of facts and raw data provide a partial picture of reality. Data collected either from primary or secondary source are often a huge mass of haphazard numerical figures. Whether data are being collected with certain purpose or collected data are being utilised, questions regarding what information the data are conveying? how the data can be used? and what should be done to include more useful information must constantly kept in mind.

The successful use of data collected depends to a great extent upon the manner in which it is arranged, displayed and summarised. Planning how the data will be presented is essential before appropriately processing raw data. In this unit we shall be mainly interested in the presentation of data. It can be in the form of classification, tabulation, diagrammatic and graphical representation.

6.2 CLASSIFICATION OF DATA

After data collection we face the problem of arranging them into a format from which we will be able to draw some conclusions. The arrangement of these data in different groups on the basis of some similarities is known as classification. According to Tuttle, “ a classification is a scheme for breaking a category into a set of parts, called classes, according to some precisely defined differing characteristics possessed by all the elements of the category.” Thus, classification is the process of grouping data into sequences according to their common characteristics, which separate them into different but related parts. Such classification facilitates analysis of the data and consequently prepares a foundation for absolute interpretation of the obtained scores.

6.2.1 Objectives of Classification

The major objectives of the classification system in research are concerned with reducing trouble and complexities with raw scores by grouping them into some classes, which provides a comprehensive insight into the data. Some of the objectives of classification are listed below.

Presenting data in a condensed form: The prime objective of classification is to put the scores into few classes/groups so that they can be easily analysed in an economical way with due precision.

Explaining the affinities and diversities of data: Classification system helps in understanding the affinities and diversities of the data in a way that observations having similar characteristics are put in one class and scores

with different characteristics are kept in another class. There may be some data with extreme low frequencies. It is not justified to maintain these data separately and classification system enables the researcher to overcome this problem. The criterion for classification depends on the objectives of the study. In one classification scores may be classified on the basis of gender and in another they may be grouped on the basis of age or education and so forth.

Facilitating comparisons: Classification facilitates comparison which is either very difficult or not possible in raw scores or in unclassified data.

Over all classification procedure in statistics enables the investigator to manage raw scores in such a way that they can proceed with ease in a systematic and scientific manner. Along with that it helps understand the range and trend of data. Finally, it can be stated that classification is an essential and foremost step that enables the investigator to summarise the results in a meaningful and convenient way.

6.2.2 Types of Classification

In social sciences, we come across various kinds of data like discrete, continuous, frequency data (when objects are counted), or metric data (when objects are measured). Looking towards the diversity of data different types of classification systems are in use.

6.2.2.1 Simple Array

The simple array is one of the simplest ways to present data. It is an arrangement of given raw data in ascending or descending order. In ascending order the scores are arranged in increasing order of their magnitude. For example, numbers 2,4,7,8,9,12, are arranged in ascending order. In descending order the scores are arranged in decreasing order of their magnitude. For example, numbers 12,9,8,7,4,2, are arranged in descending order. Simple array has several advantages as well as disadvantages over raw data. Using Simple array, we can easily point out the lowest and highest values in the data and the entire data can be easily divided into different sections. Repetition of the values can be easily checked, and distance between succeeding values in the data can be observed on the first look. But sometimes a data array is not very helpful because it lists every observation in the array. It is cumbersome for displaying large quantities of data.

6.2.2.2 Discrete Frequency Distribution

Here different observations are not written as in simple array. Here we count the number of times any observation appears which is known as frequency. The literary meaning of frequency is the number or occurrence of a particular event/ score in a set of sample. According to Chaplin (1975) “frequency distribution shows the number of cases falling within a given class interval or range of scores.” A frequency distribution is a table that organises data into classes, i.e., into groups of values describing one characteristic of the data. It shows the number of observations from the data set that fall into each of the classes.

The process of preparing a frequency distribution is very simple. In the case of discrete data, place all possible values of the variable in ascending order in one column, and then prepare another column of 'Tally' mark to count the number of times a particular value of the variable is repeated. To facilitate counting, block of five 'Tally' marks are prepared and some space is left in between the blocks. The frequency column refers to the number of 'Tally' marks against an observation. Similarly in case of a grouped frequency distribution, this column contains the tally marks a particular class will contain.

To illustrate the construction of a discrete frequency distribution, consider a sample study in which 40 families were surveyed to find the number of children per family. The data obtained are:

Table 6.1 : Frequency distribution of no. of children in 40 families

No. of Children	No. of Families	Frequency
0	III	03
1	IIII II	07
2	IIII III	08
3	IIII IIII I	11
4	IIII I	06
5	IIII	04
6	I	01
Total		40

6.2.2.3 Grouped Frequency Distribution

To prepare a frequency distribution, we first decide about the number of classes into which data are to be grouped. There are no hard and fast rules but insight of the investigator works. If there are very few scores it is useless to have a large number of class intervals. Ordinarily, the number of classes should be between 5 and 30. The number of classes, would also depend on the number of observations. More the number of observations, more will be the classes. With less number of classes, accuracy is lost and with more number the computation becomes tiresome. Another factor used in determining the number of classes, is the size/width or range of the class which is known as *class interval* and is denoted by *i*. Class interval should be of uniform width resulting in the same-size classes of frequency distribution. The width of the class should be a whole number and conveniently divisible by 2, 3, 5, 10, or 20.

After deciding the class interval, the range of scores should be decided by subtracting the highest value, the lowest value of the data array. To determine the number of classes, following formula is used.

$$\text{Number of Classes} = \frac{\text{Range of scores}}{\text{Class interval}}$$

Now, the next step is to decide from where the class should be started. There are three methods for describing the class limits for distribution—

- Exclusive method
- Inclusive method
- True or actual class method

Exclusive method

In this method of class formation, the classes are so formed that the upper limit of one class also becomes the lower limit of the next class. Exclusive method of classification ensures continuity between two successive classes. In this classification, it is presumed that score equal to the upper limit of the class is exclusive, i.e. a score of 40 will be included in the class of 40 to 50 and not in a class of 30 to 40. Finally we count the number of scores falling in each class and record the appropriate number in frequency column. The number of scores falling in each class is termed as class frequency. Tally bar is used to count these frequencies (Table 6.1).

Example: Scores of 30 students are given below. Prepare the frequency distribution by using exclusive method of classification.

3, 30, 14, 30, 27, 11, 25, 16, 18, 33, 49, 35, 18, 10, 25, 20, 14, 18, 9, 39, 14, 29, 20, 25, 29, 15, 22, 20, 29, 29

The above ungrouped data do not provide any useful information about observations rather it is confusing to understand.

Solution:

Step 1: First of all arrange the raw scores in ascending order of their magnitude.

3, 9, 10, 11, 14, 14, 14, 15, 16, 18, 18, 18, 20, 20, 20, 22, 25, 25, 25, 27, 29, 29, 29, 29, 30, 30, 33, 35, 39, 49

Step 2: Determine the **range** of scores by adding 1 to the difference between largest and smallest scores in the data array. For above array of data it is $49 - 3 = 46 + 1 = 47$.

Step 3: Decide the number of classes. Say 5 for present array of data.

Step 4: To decide the approximate size of class interval, divide the range with the decided number of classes (5 for this example). If the quotient is infraction, accept the next integer. For example, $47/5 = 9.4$. Take it as 10

Step 5: Find the lower class-limit of the lowest class interval and add the width of the class interval to get the upper class-limit. (e.g. 0 – 10)

Step 6: Find the class-limits for the remaining classes (10–20), (20–30), (30–40), (40–50)

Step 7: Pick up each item from the data array and put the tally mark (I) against the class to which it belongs. Tallies are to be marked in bunch of five, four times in vertical and fifth in cross-tally on the first four. Count the

number of observations, i.e., frequency in each class. (an example is given in Table 6.2)

Table 6.2: Frequency distribution of scores of 40 students by using exclusive method of class interval

Class Interval	Tallies	Frequency
40–50	I	1
30–40		5
20–30		12
10–20		10
0–10		2
Total		30

Note: The tallying of the observations in frequency distribution may be checked out for any omitted or duplicated one because the sum of the frequencies should be equal to the total number of scores in the array.

Inclusive method

In this method, classification includes scores, which are equal to the upper limit of the class. Inclusive method is preferred when measurements are given in whole numbers. Above example may be presented by using inclusive method of classification as shown in Table 6.3.

Table 6.3: Frequency distribution of scores of 40 students by using inclusive method of class interval

Class Interval	Tallies	Frequency
40–49	I	1
30–39		5
20–29		12
10–19		10
0–9		2
		30

True or actual class method

In inclusive method upper class limit is not equal to the lower class limit of the next class. Therefore, there is no continuity between the classes. However, in many statistical measures continuous classes are required. To have continuous classes it is assumed that an observation or score does not just represent a point on a continuous scale but an internal unit length of which the given score is the middle point. Thus, mathematically a score is internal when it extends from 0.5 units below to 0.5 units above the face value of the score on a continuum. These class limits are known as true or actual class limits (Table 6.4).

Table 6.4 : An example of preparing the class interval in the True method

Exclusive Method	Inclusive Method	True or Exact Method
70–80	70–79	69.5–79.5
60–70	60–69	59.5–69.5
50–60	50–59	49.5–59.5
40–50	40–49	39.5–49.5
30–40	30–39	29.5–39.5
20–30	20–29	19.5–29.5

6.2.2.4 Types of Grouped Frequency Distribution

There are various ways to arrange frequencies of a data array based on the requirement of the statistical analysis or the study. A few of them are discussed as follows.

Open End Frequency Distribution: Open end frequency distribution is one which has at least one of its ends open. Either the lower limit of the first class or upper limit of the last class or both are not specified. Example of such frequency distribution is given in Table below.

Table 6.5: Open end class frequency

Class	Frequency
Below 10	4
10-20	9
20-30	12
30-40	10
40-50	8

Relative frequency distribution: This kind of distribution indicates the proportion of the total number of cases observed at each score value.

Cumulative frequency distribution: Sometimes investigator is interested in knowing the number of observations less than a particular value. This is possible by computing the cumulative frequency. A cumulative frequency corresponding to a class interval is the sum of frequencies for that class and of all classes prior to that class.

Cumulative relative frequency distribution: A cumulative relative frequency distribution is one in which the entry of any score of class interval expresses that score's cumulative frequency as a proportion of the total number of cases. Table 6.6 shows frequency distributions for ability scores of 100 students.

Table 6.6: A representation of different kinds of frequency distributions

Class Interval	Frequency	Relative F	Cumulative F	Cumulative relative F
95–99	7	0.07	100	1.00
90–94	3	0.03	93	0.93
85–89	5	0.05	90	0.90
80–84	6	0.06	85	0.85
75–79	4	0.04	79	0.79
70–74	5	0.05	75	0.75
65–69	10	0.10	70	0.70
60–64	7	0.07	60	0.60
55–59	3	0.03	53	0.53
50–54	10	0.10	50	0.50
45–49	15	0.15	40	0.40
40–44	10	0.10	25	0.25
35–39	5	0.05	15	0.15
30–34	10	0.10	10	0.10
	100	1.00		

6.2.2.5 How to Use Spreadsheet Software for Frequency Distribution?

In this sub-section 6.2.2, you have gone through various types of classification processes. This sub-section explains how can you create these using a spreadsheet software. For this purpose, you may use the same data as given in the example discussed in sub-section 6.2.2.3. The process is explained with the help of following activities:

Activity 1: Data Input to Spreadsheet Software

You need to enter the following data into the spreadsheet. You may assume that Marks are in Economics Subject, and Maximum Marks are 50.

3, 30, 14, 30, 27, 11, 25, 16, 18, 33, 49, 35, 18, 10, 25, 20, 14, 18, 9, 39, 14, 29, 20, 25, 29, 15, 22, 20, 29, 29

How to Do it?

- Open a Spreadsheet Package, for example MS-Excel
- Enter the heading in A1 cell (How do you know which cell is current? Check Fig. 6.2.2.5.1 which shows current cell is B3) as "Marks in Economics (Maximum Marks 50)".

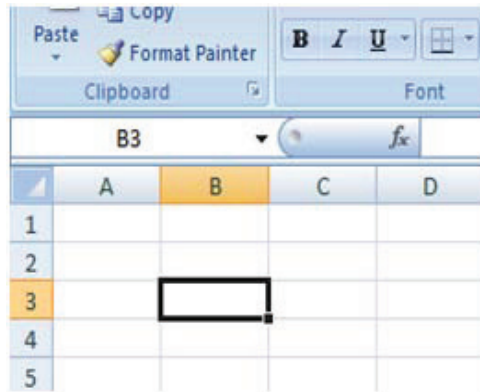


Fig. 6.2.2.5.1: Current Cell - B3

- Enter all the marks in cells A2 to A31. You may use number keys to enter the marks and arrow keys and Enter key for moving to appropriate cell.
- Save the worksheet with a name say "Activity 1". The Fig. 6.2.2.5.2 shows this data:

	A	B	C
1	Marks in Economics (Maximum Marks 50)		
2	3		
3	30		
4	14		
5	30		
6	27		
7	11		
8	25		
9	16		
10	18		
11	33		
12	49		
13	35		
14	18		
15	10		
16	25		
17	20		
18	14		
19	18		
20	9		
21	39		
22	14		
23	29		
24	20		
25	25		
26	29		
27	15		
28	22		
29	20		
30	29		
31	29		
32			
33			

Fig. 6.2.2.5.2: Data of the Activity 1

Activity 2: Discrete Frequency Distribution

To find the frequency of marks you can simply sort the data, create a list of unique discrete value from the sorted data and count the occurrence of each discrete value.

How to Do it?

- Open the spreadsheet file created by you in Activity 1
- Select the cells A1 to A31 which contains the data and heading.
- Select the **Data** tab from Menu bar and From the Ribbon select Sort Button to display various options (please note that you may directly use AZ or ZA buttons too)

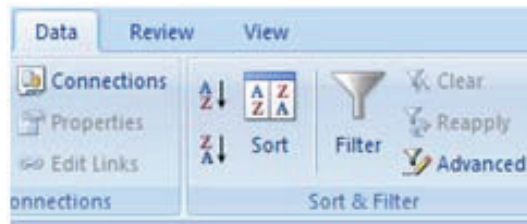
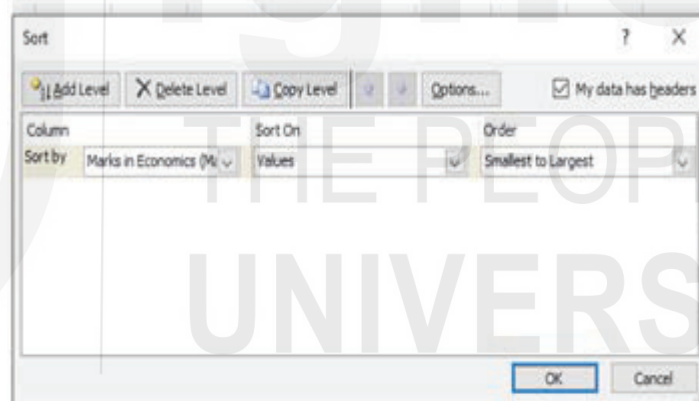
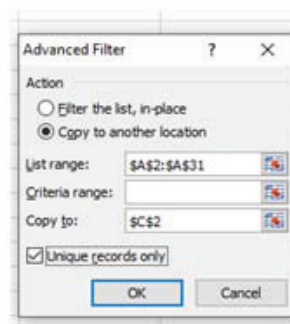


Fig. 6.2.2.5.3: Sort and Filter in the Ribbon of Data Tab

- From the option box select the options as shown below (please note that My Data has headers box has a tick in it as your first row contains the heading "Marks in Economics (Maximum Marks 50)")



- When you press OK button as shown above, the data in the range A2 to A31 will get sorted.
- Now, to extract list of unique discrete values from the list, you can use Advanced Filter Option from the Data Tab (Fig. 6.2.2.5.3). The following options should be selected from the Advanced Filter dialog box:



(please select option *Copy to another Location* and select the *List range* from A2 to A31 and the cell where the unique values are to be copied as C2. YOU MUST also check the Check box *Unique records only* and Press OK button.

- Now you will see a list of unique values in cell C2 to C19. To find the frequency of each discrete value of this list enter the following formula in Cell D2:

=COUNTIF(\$A\$2:\$A\$31,"=" & C2)

But what this formula means. This formula is using the Function called COUNTIF(range, criteria) which counts the values that fulfill the criteria in the specified range. By selecting the range as \$A\$2:\$A\$31 we are specifying the range as the marks entered by you (please note \$ in cell addressing indicates absolute addressing explained in first block). The criteria that is entered is very specific here, it is "=" & C2 which indicates the values should be equal to values in the cell C2 (this relates to concept of using operators = and concept of referencing a cell using &. A detailed discussion on them is beyond the scope of this Unit.

- Now, simply copy this formula from cell D2 to Cells D3 to D19. You will find the frequency of each discrete value as shown in Fig. 6.2.2.5.4
- Just to cross check you can enter a formula =SUM(D2..D19) in cell D20 to sum the frequencies as calculated in D2 to D19. What do you expect this sum to be? of course 30 and that is what it is (please refer to Fig. 6.2.2.5.4).

	A	B	C	D	E	F
1	Marks in Economics (Maximum Marks 50)					
2		3	3	1		
3		9	9	1		
4		10	10	1		
5		11	11	1		
6		14	14	3		
7		14	15	1		
8		14	16	1		
9		15	18	3		
10		16	20	3		
11		18	22	1		
12		18	25	3		
13		18	27	1		
14		20	29	4		
15		20	30	2		
16		20	33	1		
17		22	35	1		
18		25	39	1		
19		25	49	1		
20		25		30		
21		27				
22		29				
23		29				
24		29				
25		29				

Fig. 6.2.2.5.4: Output of Activity 2

Activity 3: Grouped Frequency Distribution

Suppose you want to find the frequency of marks in the classes:

≤ 5 ; > 5 and ≤ 10 ; > 10 and ≤ 15 and so on.

This activity is to be performed on the data of Activity 1 (let us assume sorted data).

How to Do it?

- Open the spreadsheet file created by you in Activity 1. Cells A1 to A31 contains the data and heading.
- You can create classes on the basis of class requirements as given above. Fig. 6.2.2.5.5 shows exclusive, inclusive and true and exact class intervals for the desired frequency ranges.

<i>Exclusive method</i>	<i>Inclusive method</i>	<i>True or exact method</i>
0-6	0-5	0-5.5
6-11	6-10	5.5-10.5
11-16	11-15	10.5-15.5
16-21	16-20	15.5-20.5
21-26	21-25	20.5-25.5
26-31	26-30	25.5-30.5
31-36	31-35	30.5-35.5
36-41	36-40	35.5-40.5
41-46	41-45	40.5-45.5
46-51	46-50	45.5-50.5

Fig. 6.2.2.5.5: Class interval using Exclusive, Inclusive and Exact methods

G4		f _x [=FREQUENCY(A2:A31,F4:F13)]							
A	B	C	D	E	F	G	H	I	
1	Marks in Economics (Maximum Marks 50)								
2	3								
3	9	<i>Exclusive method</i>	<i>Inclusive method</i>	<i>True or exact method</i>	Bin1	Frequency	Bin2	Frequency	
4	10	0-6	0-5	0-5.5	5	1	5.5	1	
5	11	6-11	6-10	5.5-10.5	10	2	10.5	2	
6	14	11-16	11-15	10.5-15.5	15	5	15.5	5	
7	14	16-21	16-20	15.5-20.5	20	7	20.5	7	
8	14	21-26	21-25	20.5-25.5	25	4	25.5	4	
9	15	26-31	26-30	25.5-30.5	30	7	30.5	7	
10	16	31-36	31-35	30.5-35.5	35	2	35.5	2	
11	18	36-41	36-40	35.5-40.5	40	1	40.5	1	
12	18	41-46	41-45	40.5-45.5	45	0	45.5	0	
13	18	46-51	46-50	45.5-50.5	50	1	50.5	1	
14	20				TOTAL	30		30	
15	20								

Fig. 6.2.2.5.6: Grouped Frequency Calculations

- You need to use FREQUENCY(data array, bin_array) function to calculate the grouped frequency. The data array is simple pointing to data so it will be A2..A31 cells. Bin array shows the upper inclusive limit of class array. Therefore, for the defined classes the bin array will be either the upper limit of inclusive classes or upper limit of true or exact class intervals (Refer to Fig. 6.2.2.5.6). In this figures two bin arrays are shown **Bin1** in cells F4 to F13 (based on Inclusive values) and **Bin2** in cells H4 to H13 (based on exact values).
- Please note since frequencies are to be calculated for a number of ranges, you need to use ARRAY formulas. You will enter Array formula for both the bin values (Bin1 and Bin2) separately. To enter the array formula do the following steps:
 - Select Cells G4 to G13
 - In the Formula bar enter **=FREQUENCY(A2..A31, F4..F13)** and PRESS the two Keys CTRL+SHIFT+ENTER together.
 - It will enter the array formula in all the cells viz. G4 to G13 showing you frequency. You may check the formula as shown in the Figure in Cell G4, it is {=FREQUENCY(A2..A31, F4..F13)}
 - Notice that the formula has { } around it, indicating it is an array formula.
 - Now, you can enter array formula in the range I4 to I13 using the same data range but bin ranges as H3..H13. You can also check the total of all the frequencies using sum formula of excel in the cell G14 as **=SUM(G4:G13)**

Check Your Progress 1

- 1) What do you understand by classification of data?

.....

.....

.....

.....

- 2) Why classification is necessary?

.....

.....

.....

.....

.....

- 3) When both the lower and upper limits are considered, such classification is known as

(i) Exclusive, (ii) Inclusive, (iii) None of these

- 4) In “less than ” cumulative frequency distribution, the omitted limit is
(i) lower , (ii) upper , (iii) last , (iv) none of these
- 5) Consider the marks of 10 students in Physics (out of 20)
3, 5, 17, 9, 13, 20, 11, 16, 14, 17
Perform the following tasks using a spreadsheet:
 - i) Enter the data in a spreadsheet
 - ii) Sort the values in descending order
 - iii) Create a discrete frequency distribution
 - iv) Create a Group frequency division in 4 ranges: ≤ 5 ; > 5 and ≤ 10 ;
 > 10 and ≤ 15 ; > 15 and ≤ 20 .

6.3 TABULATION OF DATA

Tabulation is the process of presenting the classified data in the form of a table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. A table is a systematic arrangement of classified data in rows and columns with appropriate headings and sub-headings.

6.3.1 Components of a Statistical Table

The main components of a table are given below—

Table number: A table should be marked with a number for reference and identification. The number should be written at the top of the table.

Title of the table: Every table should have an appropriate title, which describes the content of the table. The title should be clear, brief, and self-explanatory. Title of the table should be placed either centrally on the top of the table or just below or after the table number.

Caption: Captions are brief and self-explanatory headings for columns. Captions may involve headings and sub-headings. The captions should be placed in the middle of the columns.

Stub: Stubs stand for brief and self-explanatory headings for rows. A relatively more important classification is given in rows.

Body of the table: This is the real table and contains numerical information or data in different cells. This arrangement of data remains according to the description of captions and stubs.

Head note: This is written at the extreme right below the title and explains the unit of the measurements used in the body of the tables.

Footnote: This is to be written below the table explaining certain points related to the data which have not been covered in title, caption, and stubs.

Source: The source from which data have been taken is to be mentioned at the end of the table. Reference of the source must be complete so that if the readers want to consult the original source, they may do so.

Table No. 6.7: (—————TITLE OF THE TABLE —————)

Stub Head	Caption			
Stub Entries	Column Head I		Column Head II	
	Sub Head	Sub Head	Sub Head	Sub Head
	MAIN BODY	OF	THE TABLE	
Total				

Footnote(s):

Source:

6.3.2 General Rules for Preparing a Table

There are no hardcore rules for preparing a table. But tabulation requires a lot of skills and common sense on the part of the researcher. Though specific rules have not been provided for tabulation, some general rules are in tradition. They are as follows:

- i) Table should be compact and readily comprehensible besides being complete and self-explanatory.
- ii) It should be free from confusions.
- iii) The table should be arranged in the space provided. It should neither be too small nor too large.
- iv) Items in the table should be placed logically and related items should be placed nearby.
- v) All items should be clearly stated.
- vi) If an item is repeated in the table, its full form should be written.
- vii) The unit of measurement should be explicitly mentioned, preferably in the form of a head note.

6.3.3 Importance of Tabulation

Tabulation is the process of condensation of data for convenience in statistical processing, presentation and interpretation of the information combined therein. Importance of tabulation are given below:

Simplifies complex data: If data are presented in tabular form, these can be readily understood. Confusions are minimum when one is going through the data for further analysis or drawing conclusions about the observation.

Facilitates comparison: Data in the statistical table are arranged in rows and columns very systematically. Such an arrangement enables the researcher to compare the information in an easy and comprehensive manner.

Presents data in true perspective: With the help of tabulation, the repetitions can be dropped out and data can be presented in true perspective highlighting the relevant information.

Figures can be worked out more easily: Tabulation also facilitates further analysis and finalisation of figures for understanding the data.

Check Your Progress 2

1) Differentiate between following pairs of statistical terms

i) Column and row entry

.....

.....

.....

ii) Caption and sub head

.....

.....

.....

iii) Head note and foot note

.....

.....

.....

2) Distinguish between classification and tabulation.

.....

.....

.....

.....

3) State the general rules for constructing a table.

.....

.....

.....

.....

6.4 DIAGRAMMATIC PRESENTATIONS OF DATA

A diagram is a visual form for the presentation of statistical data. They present the data in simple, readily comprehensible form. Diagrammatic presentation is used only for presentation of the data in visual form, whereas graphic presentation of the data can be used for further analysis. There are different forms of diagram e.g., Bar diagram, Sub-divided bar diagram, Multiple bar diagram, Pie diagram and Pictogram.

6.4.1 Bar Diagram

Bar diagram is most useful for categorical data. This is known as dimensional diagram also. A bar is defined as a thick line. It is drawn from the frequency distribution table representing the variable on the horizontal axis and the frequency on the vertical axis. The height of each bar will be corresponding to the frequency or value of the variable. However, width of the rectangles is immaterial but proper and uniform spacing should be between different bars. It is different from the histogram where both the height and width of the bar are important and even bars are placed adjacent to one another without any gap.

Table 6.8: In a study on causes of strikes in a shoe-making company the data are given below.

Causes of strikes:	Economic	Personal	Political	Rivalry	Others
Occurrence of strikes:	40	15	25	15	10

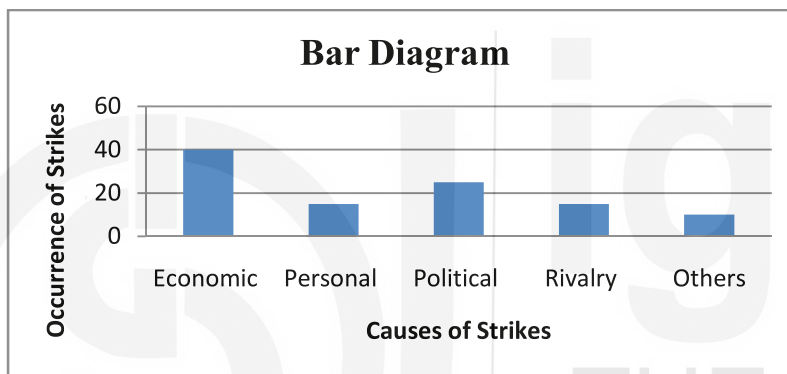


Fig.6.1: Bar diagram showing the causes of strikes with their occurrence in a shoe-making company.

Activity 4: Creating a Bar Diagram using Spreadsheet

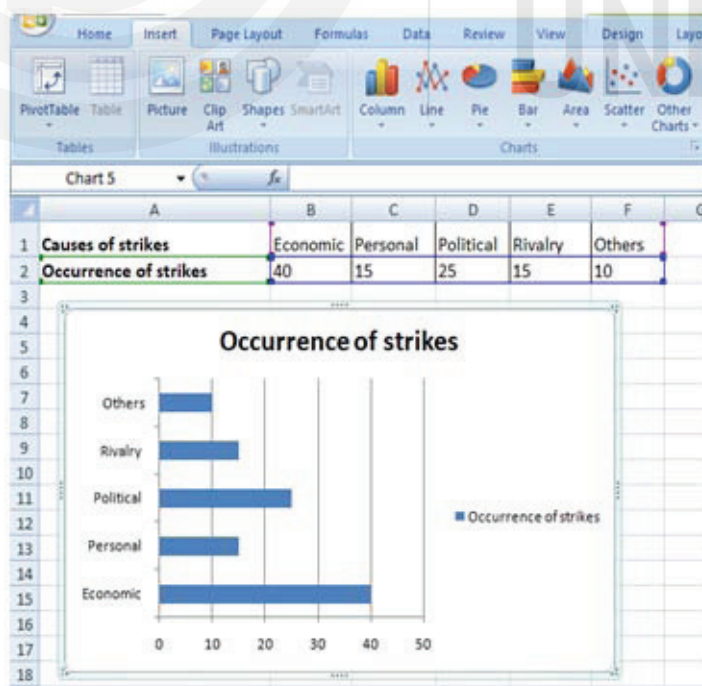


Fig. 6.1.1: Clustered Bar Chart with single data bar

Use the data as given in this section

How to Do it? (Refer to Fig. 6.1.1)

- Enter the Data in the spreadsheet
- Select the data A1..F2 as entered in the Cells
- From the Insert tab Select the Bar from Charts and select clustered bar

6.4.2 Multiple Bar Diagram

This type of diagram is appropriate to present the information regarding two or more related facts of different places or periods. It is known as compound bar diagram also. The technique of drawing this type of diagram is same as that of bar diagram.

Table 6.9: A group of three students were assessed on three different psychological parameters like, anxiety, adjustment, and stress

Students	Anxiety	Adjustment	Stress
X	20	15	30
Y	12	25	16
Z	18	13	25

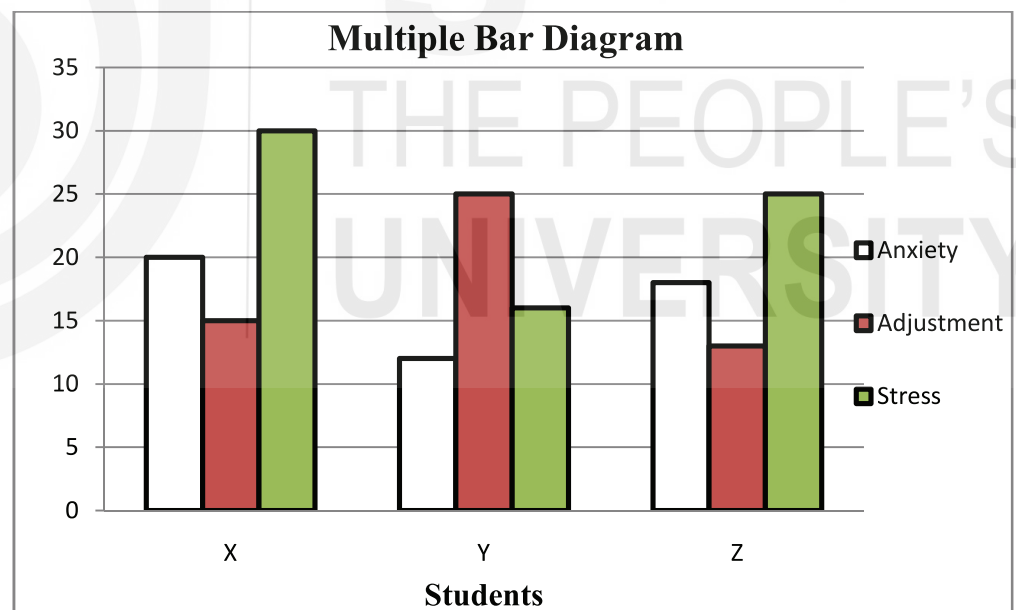


Fig.6.2 : Multiple bar diagram showing the scores of three students on anxiety, adjustment and stress.

Activity 5: Creating Clustered Bar Diagram using Spreadsheet

Use the data as given in this section

How to Do it? (Refer to Fig. 6.2.1)

- Enter the Data in the spreadsheet
- Select the data A1..D4 as entered in the Cells

- From the Insert tab Select the Bar from Charts and select Clustered Bar

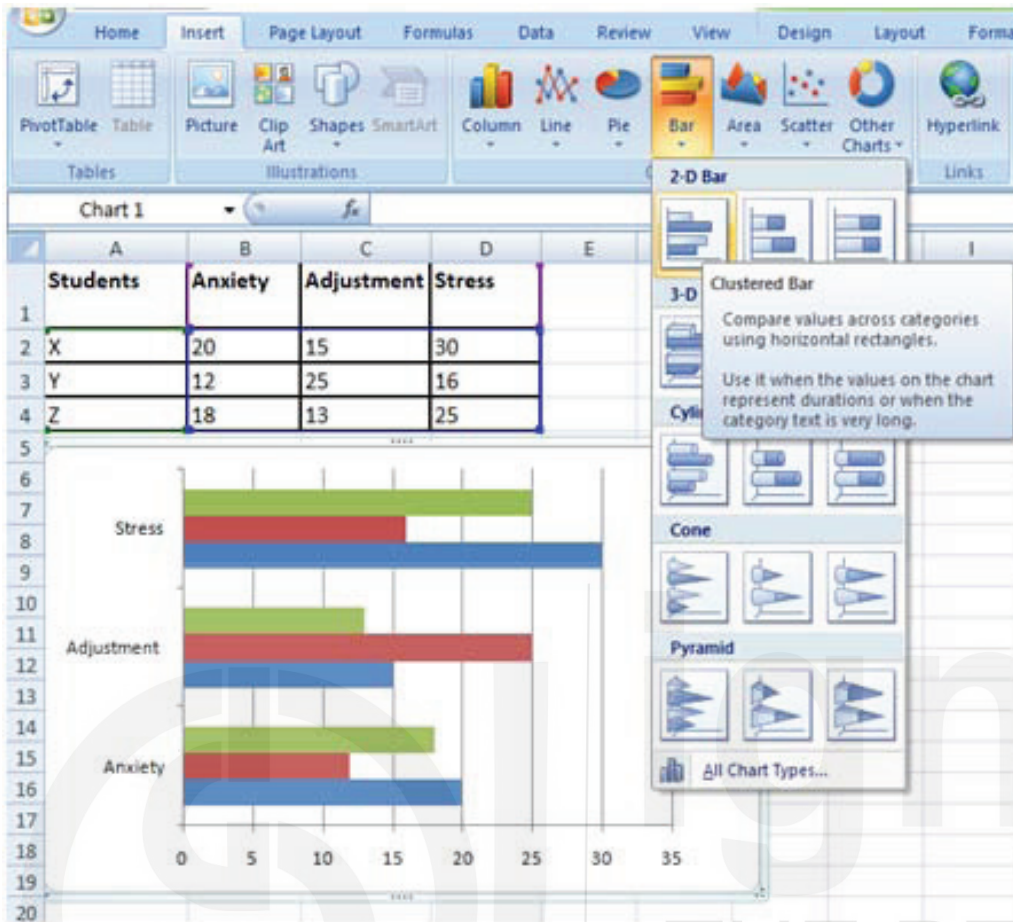


Fig. 6.2.1: Clustered bar graph with three bars

6.4.3 Overlapping Bar Diagram

In multiple bar diagram, bars for some phenomenon are erected side by side for each fact in a set. If the number of bars in each set is more than two, then it occupies wide space. So for the sake of brevity, overlapping bar diagram is used because it gives a better presentation of the data. In this diagram one bar of fact penetrates into last bar by half of its width. Each bar is distinctly shown by different pattern of colour in a set. But same colour or pattern is maintained for a fact in all the sets for various years or places, etc.

Table 6.10: Following table provides the data for net production of cereals and net availability of cereals and pulses in million tones from 1997 to 2000

	Years			
	1997	1998	1999	2000
Net production of cereals	162	157	165	171
Net availability of cereals	152	147	156	159
Net availability of pulses	13	12	13	12

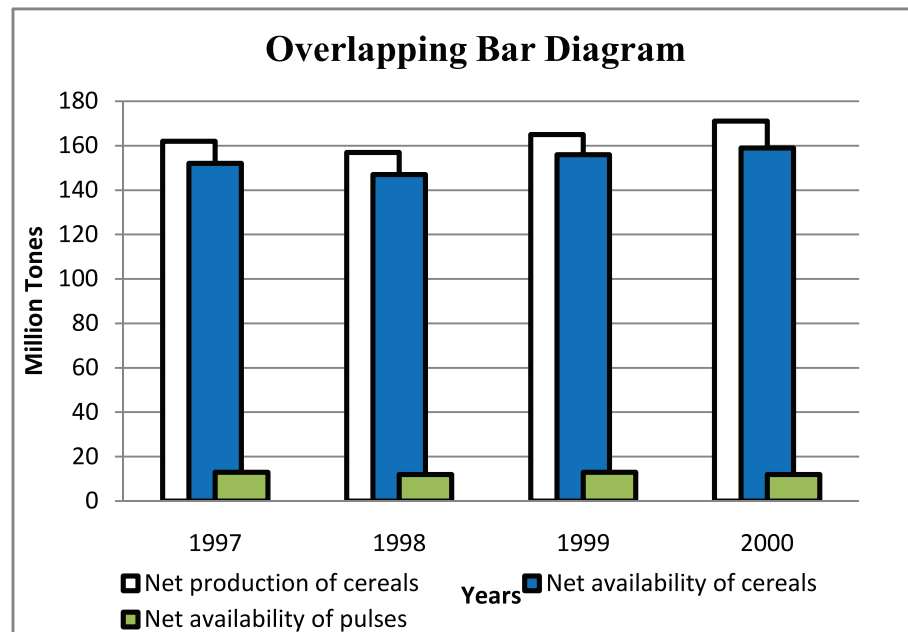


Fig. 6.3: Overlapping bar diagram showing net production of cereals and net availability of cereals and pulses in million tones from 1997 to 2000.

There is no such type in spreadsheet. You can make custom charts for such type. This is beyond the scope of this Unit.

6.4.4 Component Bar Diagram

The bar diagram shows an aggregate values whereas the component bar diagram gives break-up in parts which constitutes the aggregate in a year, place or sector. Such a diagram makes it possible to compare changes occurring in parts and in aggregate as well. In these types of diagrams a bar is further sub-divided into parts in proportion to the size of the sub-division. These sub-divided rectangles are shaded differently by lines, dots, colours etc. Such diagram is more useful than simple bar diagram.

Table 6.11: Following table provides the data for number of students passed with first, second and third division during 1990 to 1992

Result	No. of Students		
	1990	1991	1992
First division	8	13	11
Second division	22	26	35
Third division	45	36	29
total	75	75	75

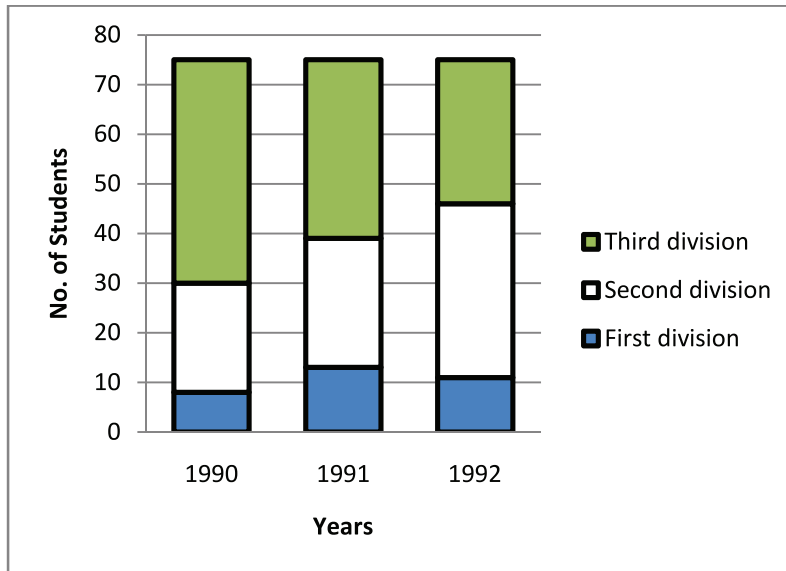


Fig.6.4: Component bar diagram showing number of students passed with first, second and third division during 1990 to 1992.

Activity 6: Creating Stacked Bar Diagram using Spreadsheet

Use the data as given in this section

How to Do it? (Refer to Fig. 6.4.1)

- Enter the Data in the spreadsheet
- Select the data A2..D5 only from the Cells (leaving the First and Last rows)
- From the Insert tab Select the Bar from Charts and select Stacked Bar

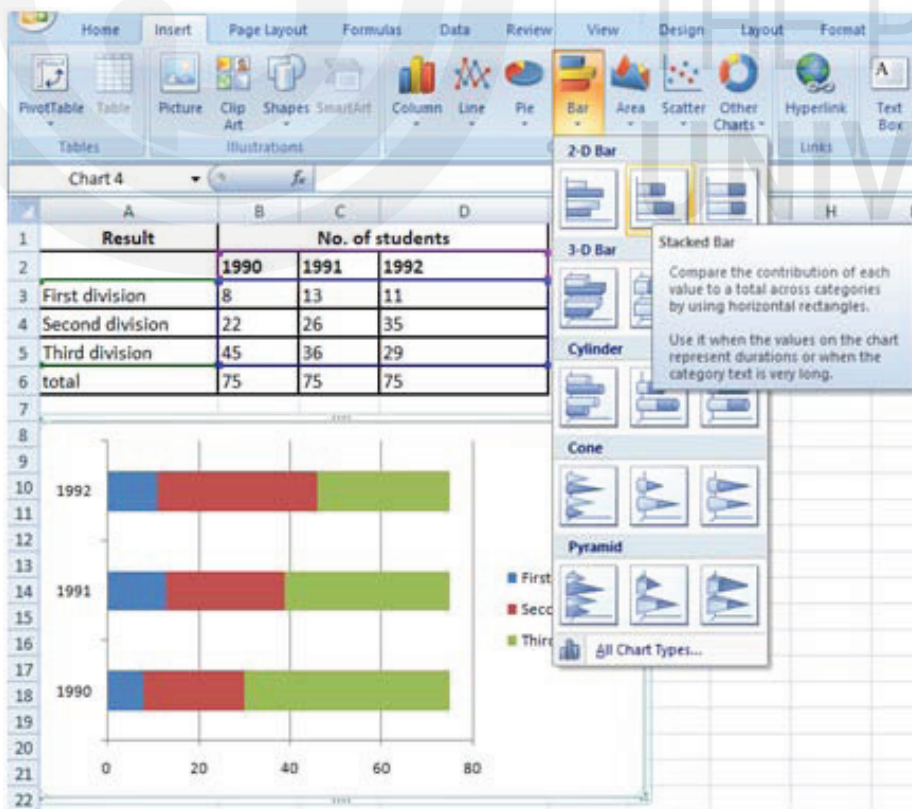


Fig. 6.4.1: Stacked Bar Graph

6.4.5 Pie Chart

It is also known as angular diagram. A pie chart or diagram is a circle divided into component sectors according to the break up of components given in percentage. Each sector will be proportional to the frequency of the variable in the group. A circle represent 360° . So 360° angle is divided in proportion to percentages. The degrees represented by the various component parts of given magnitude can be obtained by using this formula.

$$\text{Degree of any component part} = \frac{\text{Component Value}}{\text{Total Value}} \times 360^\circ$$

After the calculation of the angles for each component, segments are drawn in the circle in succession corresponding to the angles at the center for each segment. Different segments are shaded with different colour, shades or numbers. A pie chart is good to represent the component break up of a thing or commodity.

Table 6.12: 1000 software engineers pass out from an institute X and they were placed in four different company in 2009

Company	Placement
A	400
B	200
C	300
D	100

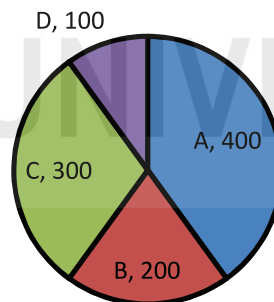


Fig. 6.5: Pie chart showing the placement of 1000 software engineers in four different companies.

Activity 7: Creating a Pie chart using Spreadsheet

Use the data as given in this section

How to Do it? (Refer to Fig. 6.5.1)

- Enter the Data in the spreadsheet
- Select the data A1..B5 which contains the data

- From the Insert tab Select the Pie from Charts. A pie chart without label will be displayed as shown in Fig. 6.5.1. To display the labels, select Layout tab from Charts tools and select Data Label from the ribbon. (Refer to Fig. 6.5.1)

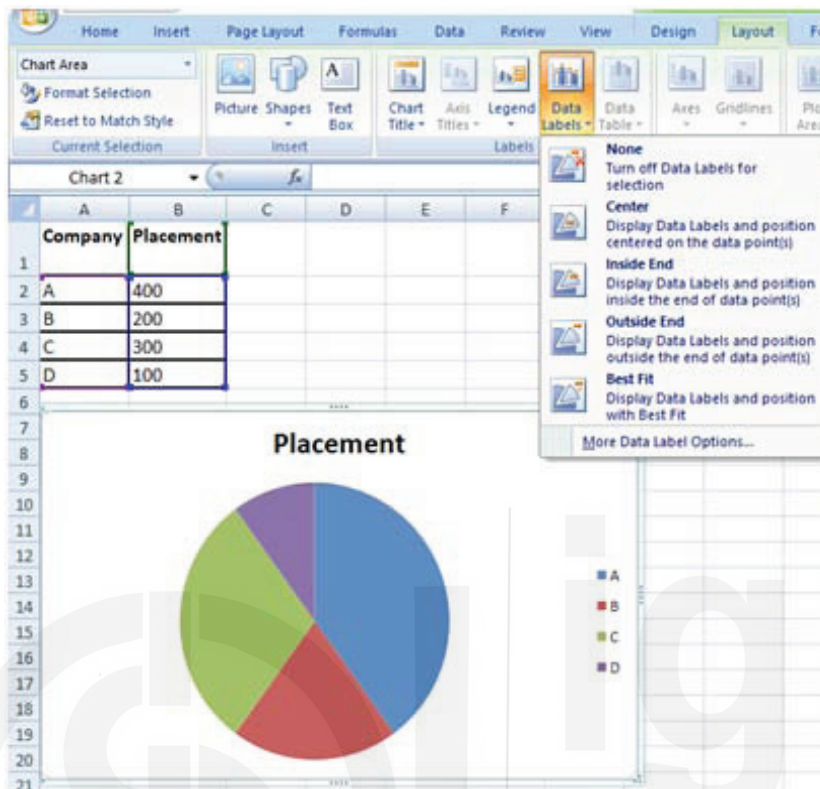


Fig. 6.5.1: Pie Chart with Data Label Option

- Select More Data Label Options from the Data Label Options, the dialog box as shown in Fig. 6.5.2 will appear. Check the boxes in this dialog box as shown in Fig. 6.5.2. The final graph is also shown in this Figure. Press close button to see the final output.

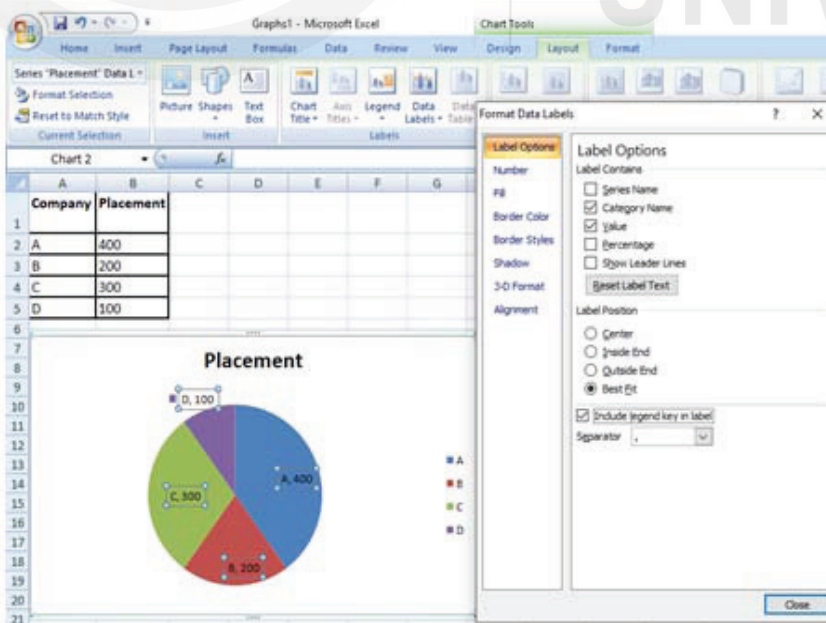


Fig. 6.5.2: Selecting Labels in Pie Chart

6.4.6 Pictogram

It is known as cartographs also. In pictogram we used appropriate picture to represent the data. The number of picture or the size of the picture being proportional to the values of the different magnitudes to be presented. For showing population of human beings, human figures are used. We may represent 1 Lakh people by one human figure. Pictograms present only approximate values.

6.5 GRAPHICAL REPRESENTATION OF DATA

Visual presentation also helps in comparison of data or determining the trends of the past performance. We have already discussed about one of the techniques of visual presentation of data i.e., diagrammatic presentation. Another important technique of visual presentation of data is the presentation in the form of graphs. To elaborate the data array after constructing the frequency distribution, it is a scientific tradition to plot the frequencies on a pictorial platform formed of horizontal and vertical lines, known as *graph*.

A graph is created on two mutually perpendicular lines called the *X* and *Y* axes on which appropriate scales are indicated. The horizontal line is called *abscissa* and the vertical one is *ordinate*. Like different kinds of frequency distributions, there are many kinds of graphs too, which enhance the scientific understanding of the reader. Here we will discuss some of the important types of graphical patterns used in statistics i.e. Line graph, Histogram, Frequency polygon, Frequency curve, Cumulative frequency curve.

6.5.1 Line Graph

In the line graph we have two variables under consideration. One variable is taken along *X* axis and the other along *Y* axis. The variate values are suitably scaled along the axes and all distances are measured from the origin. If the smallest value is at distance from zero, the origin is shifted suitably to a value other than zero. The independent variable should be taken on *X* axis and the dependent variable on *Y* axis. The points are plotted and joined by line segments in order. These graphs depict the trend of the data. Two or more graphs maybe plotted on the same graph paper with same scale to facilitate the comparison.

Line graph is appropriate when we need to present the movement or variation in a variable. It is quite simple to draw and indicates the increase or decrease in a variable over time or across observations. Line graphs can be used for discrete data. Recall that in the case of continuous data we assume that the average value of each class is its mid-point. Thus we can plot the frequencies for each class against its mid-point and join these points to obtain a line graph.

Suppose you are provided with data on number of books issued in a library (monthwise for the year 2010) as given in Table 6.13.

Table 6.13: Number of Books Issued in a Library in the year 2010

Month	No. of Books Issued
January	75
February	86
March	79
April	100
May	125
June	90
July	150
August	165
September	175
October	100
November	110
December	145

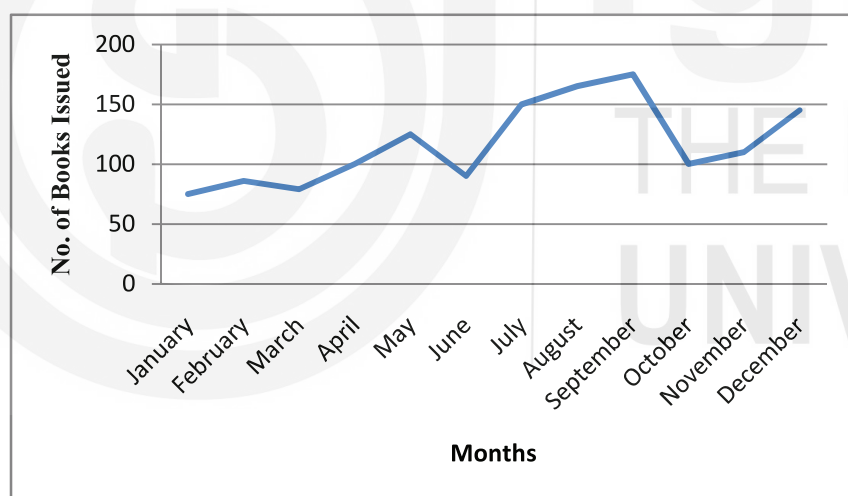


Fig.6.6 : Line graph showing month wise number of books issued in a library.

Activity 8: Creating a Line Graph using Spreadsheet

Use the data as given in this section

How to Do it? (Refer to Fig. 6.6.1)

- Enter the Data in the spreadsheet
- Select the data A1..B13 which contain the data
- From the Insert tab Select the **Line** from Charts and select Line from the list of different line charts. Fig. 6.6.1 shows a line chart

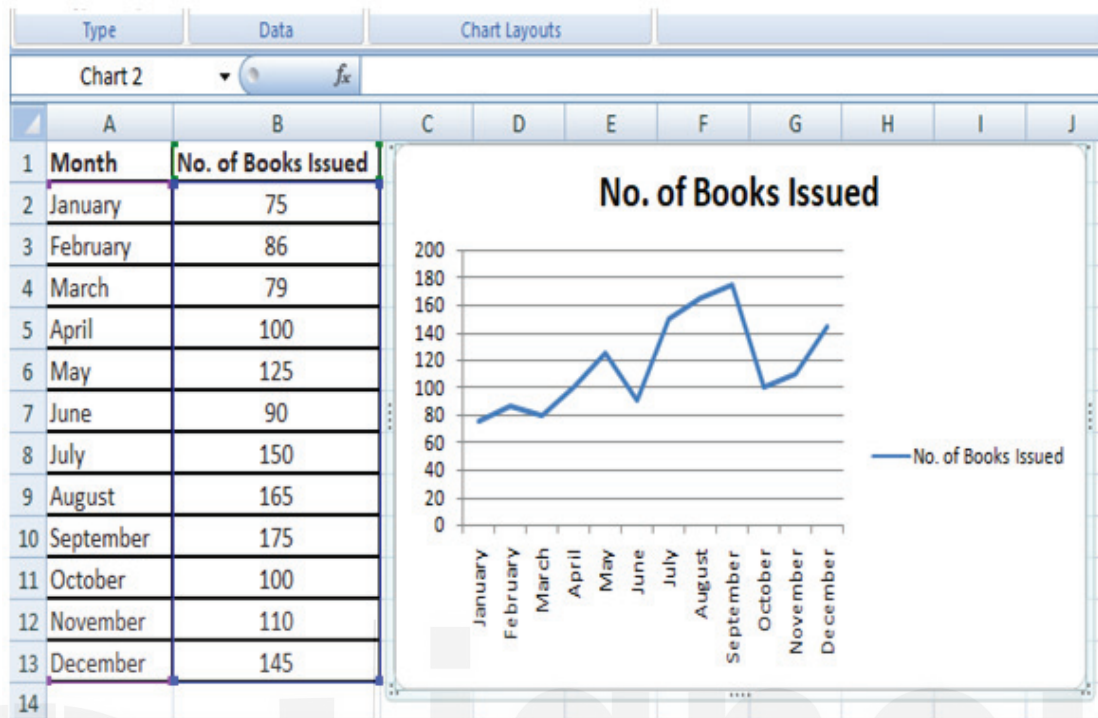


Fig. 6.6.1: Line Chart

6.5.2 Histogram

It is one of the most popular method for presenting continuous frequency distribution in a form of graph. In this type of distribution, the upper limit of a class is the lower limit of the following class. The histogram consists of series of rectangles, with its width equal to the class interval of the variable on horizontal axis and the corresponding frequency on the vertical axis as its heights. The steps in constructing a histogram are as follows:

Step 1: Construct a frequency distribution in table form.

Step2: Before drawing axes, decide on a suitable scale for horizontal axis then determine the number of squares (on the graph paper) required for the width of the graph.

Step 3: Draw bars of equal width for each class interval. The height of a bar corresponds to the frequency in that particular interval. The edge of a bar represents both the upper real limit for one interval and the lower real limit for the next higher interval.

Step 4: Identify class intervals along the horizontal axis by using either real limit or midpoint of class interval. In case of real limits, these will be placed under the edge of each bar. On the other hand, if you use midpoint of class interval, it will be placed under the middle of each bar.

Step 5: Label both axes and decide appropriate title to the histogram.

Table 6.14: Results of 150 students on numerical ability test

Scores	Frequency
10-20	15
20-30	30
30-40	50
40-50	35
50-60	20

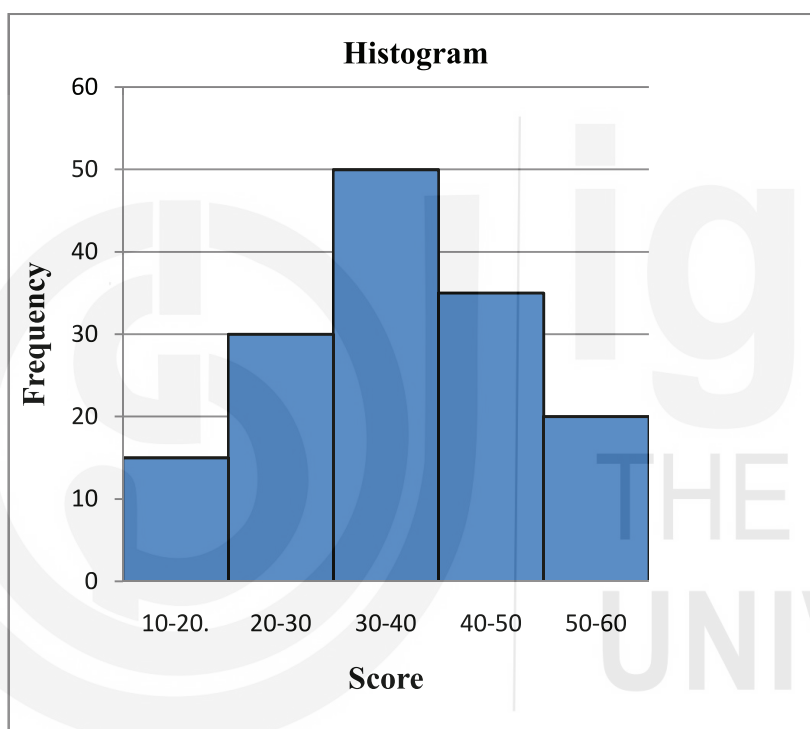


Fig. 6.7: Histogram showing the scores of 150 students on numerical ability test.

Activity 9: Creating a Histogram

Use the data used for Activity 2

How to Do it?

- Enter the Data and Bin values in the spreadsheet (Refer to Fig. 6.7.1)
- Select Data Menu and Data Analysis in the Ribbon (if this option is not installed on your machine, you may need to install Data Analysis Pack as explained in Unit 3) (Refer to Fig. 6.7.1)

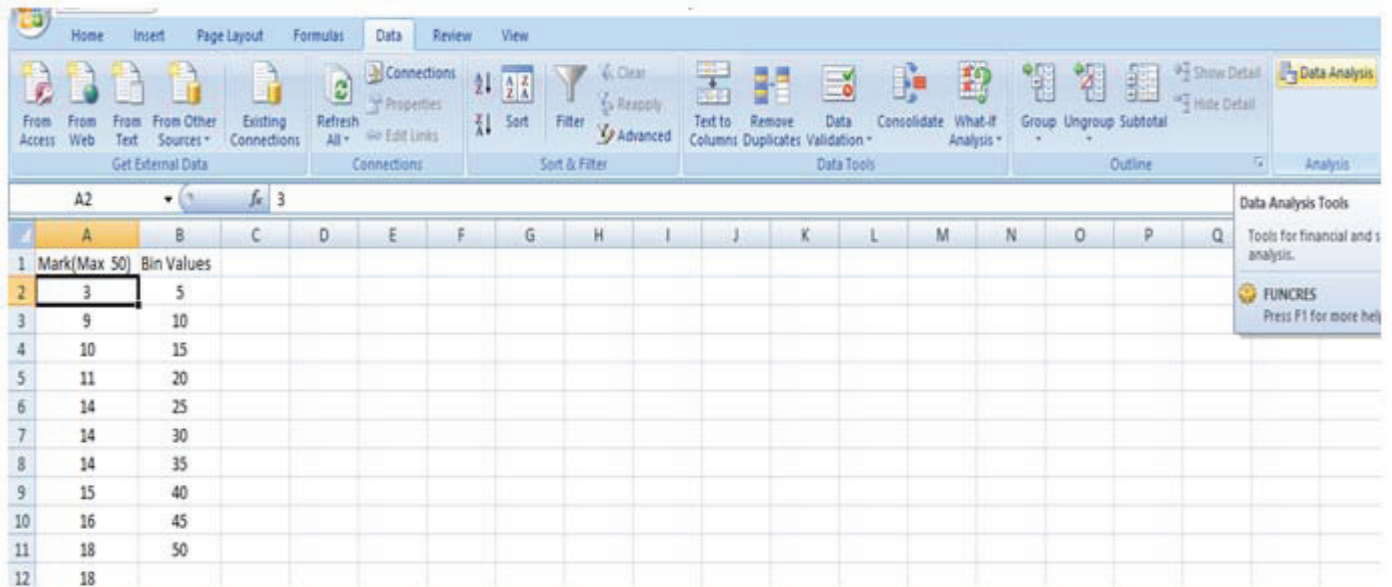
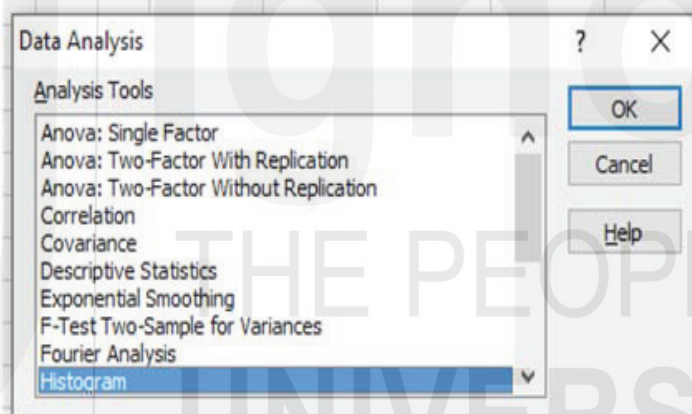
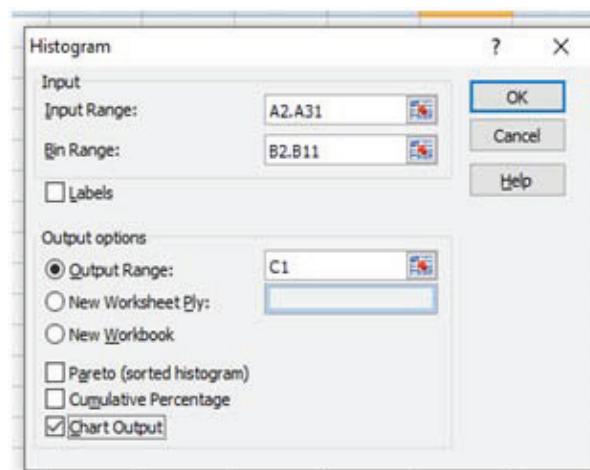


Fig. 6.7.1: Data for Histogram and Data Analysis Option

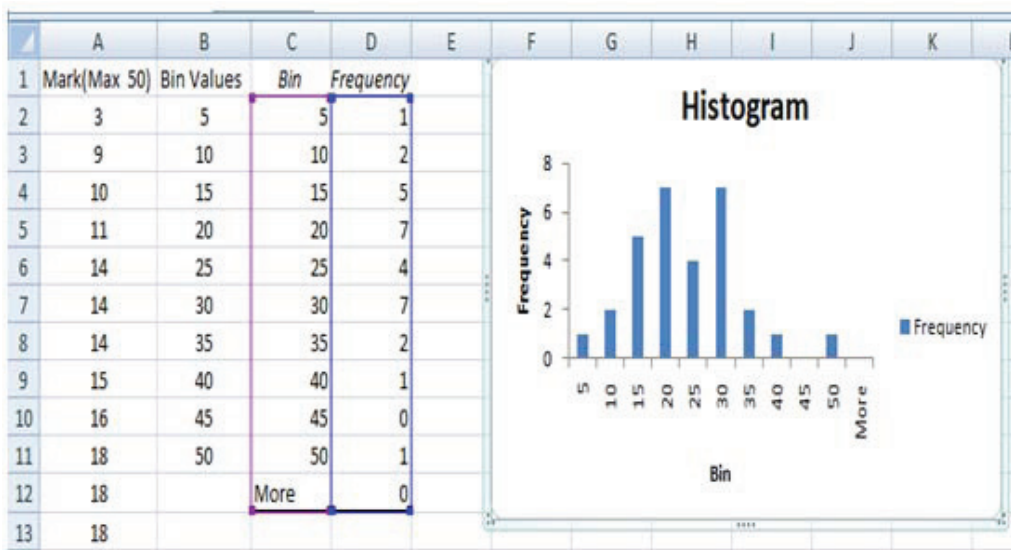
- From the dialog box select Histogram (the Highlighted option) and press OK button.



- Now, select the following ranges (Input Range A2.A31; Bin Range B2.B11 and in the Output option Selection Output Range as C1) in the resulting dialog box. Please do not forget to tick the Chart Output option as shown in the following figure.



- Now, press OK button. The following histogram can be seen in the spreadsheet.



6.5.3 Frequency Polygon

Frequency polygon is another way of showing frequency distribution graphically. There are two ways of constructing a frequency polygon. In the first method, prepare an abscissa originating from 'O' and ending to 'X'. Again construct the ordinate starting from 'O' and ending at 'Y'. Now label the class-intervals on abscissa stating the exact limits or midpoints of the class-intervals. There is also a fashion to add one extra limit keeping zero frequency on both side of the class interval range. The size of measurement of small squares on graph paper depends upon the number of classes to be plotted. Next step is to plot the frequencies on ordinate using the most comfortable measurement of small squares depending on the range of whole distribution. To obtain an impressive visual figure it is recommended to use the 3:4 ratio of ordinate and abscissa though there is no tough rules in this regard. To plot a frequency polygon you have to mark each frequency against its concerned class on the height of its respective ordinate. After putting all frequency marks draw line segments joining them. This is the polygon. A polygon is a multi-sided figure and various considerations are to be maintained to get a smooth polygon in case of smaller N or random frequency distribution. The very common way is to compute the smoothed frequencies of the classes by having the average of frequencies of that particular class along with upper and lower classes' frequencies. For instance, the frequency 55 of class-interval 40-50 might be smoothed as $35+55+45/3=45$ (Refer Table 6.15).

Another way of preparing a frequency polygon is to construct a histogram in a normal way. Here put a dot against the mid points of the tops and the first and last vertical lines of the histogram. Then join the different dots by straight line segments to get the frequency polygon. These methods point out that the frequency polygon originates at a point lying to the left of the lower limit of the first class interval equal to one-half of the width of the class interval. Similarly, it is extended at the point lying at half-the class

interval width away to the right of the upper boundary of the last class interval. It should be noted that the area under the two histograms and the frequency polygon is always the same and they represent the total frequency.

Table 6.15: Results of 200 students on Academic achievement test

Class Interval	Frequency
10-20	12
20-30	10
30-40	35
40-50	55
50-60	45
60-70	25
70-80	18

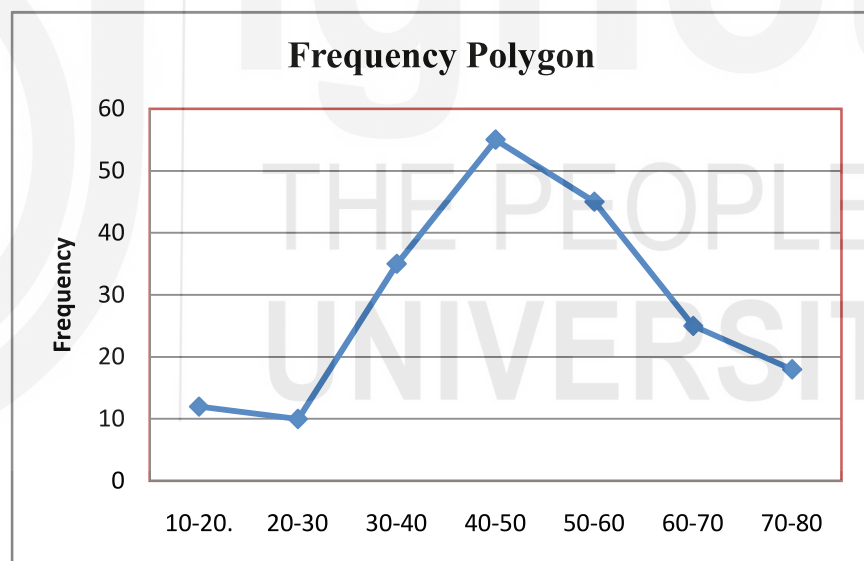


Fig.6.8 :Frequency polygon showing the scores of 200 students on Academic achievement test.

As far as spreadsheet package is concerned, you can draw it as a line graph with markers in a similar way as you have drawn a line graph. This is left as an exercise for you.

6.5.4 Frequency Curve

A frequency curve is constructed to smoothen the frequency polygon. It is prepared by drawing a free hand curve by the side of the frequency polygon in such a manner that the area under polygon and the area under curve drawn is the same. The total area of the frequency curve also represents the

total frequency. Frequency curve represents the generalisation of frequency polygon. This means that if the data given is a small sample out of a very large group, then frequency curve gives the general tendency as exhibited by the big group.

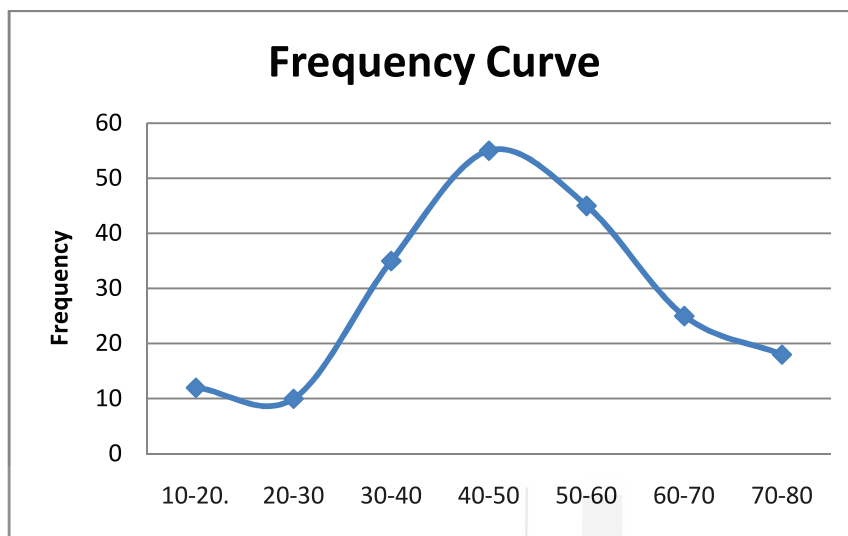


Fig.6.9: Frequency curve showing the scores of 200 students on Academic achievement test.

There are no direct method to draw such curve, you can actually draw these curve using Scatter plot but for that you need to keep frequency range (bin values) as numeric data. Try making such plots yourself.

6.5.5 Cumulative Frequency Curve or Ogive

When the points are plotted for variate values and their corresponding cumulative frequencies are joined by a free hand smooth curve, the curve is known as cumulative frequency curve or ogive. Ogive helps us to determine median, quartiles, percentages, etc. There are two types of cumulative frequency distribution e.g., “less than” and “more than” cumulative frequencies, we can have two types of ogives.

- i) *‘Less than’ Ogive:* In ‘less than’ ogive, the less than cumulative frequencies are plotted against the upper class boundaries of the respective classes. It is an increasing curve having slopes upwards from left to right.
- ii) *‘More than’ Ogive:* In more than ogive, the more than cumulative frequencies are plotted against the lower class boundaries of the respective classes. It is decreasing curve and slopes downwards from left to right.

Example of ‘Less than’ and ‘more than’ cumulative frequencies based on data.

Table 6.16: Following data show the marks obtained by 200 students in numerical ability test

Class Interval	Frequency	Less than Cumulative Frequency	More than Cumulative Frequency
10-20	12	12	200
20-30	10	22	188
30-40	35	57	178
40-50	55	112	143
50-60	45	157	88
60-70	25	182	43
70-80	18	200	18

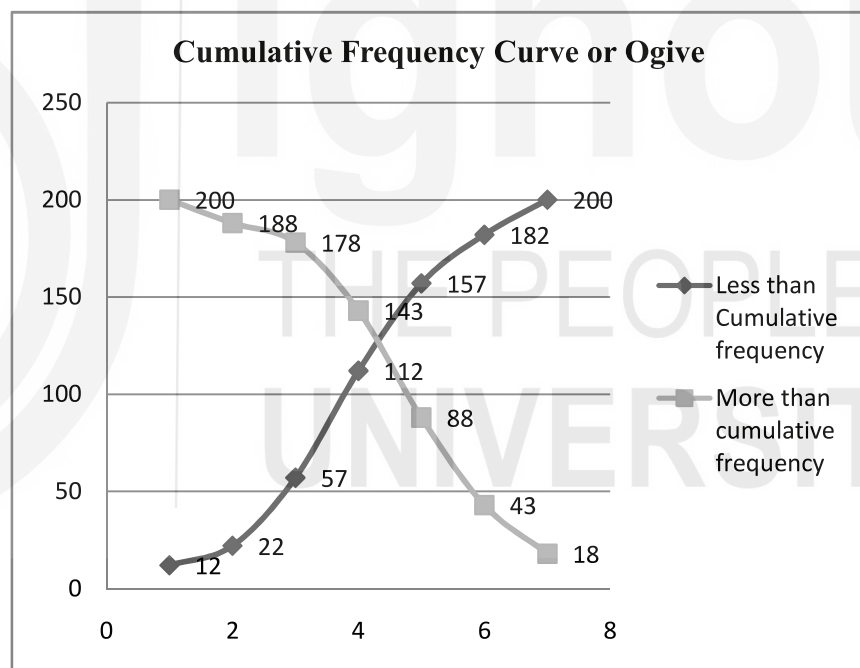


Fig. 6.10: Less than and more than cumulative frequency curve showing the scores of 200 students on numerical ability test.

Activity 10: Creating graph for Cumulative frequencies

Use the data shown in this section.

How to Do it?

- Enter the Data in the spreadsheet (Refer to Fig. 6.10.1)
- Select the entire data range A1..D8 and Insert Line with Markers Graph.
- The following graph will be displayed.

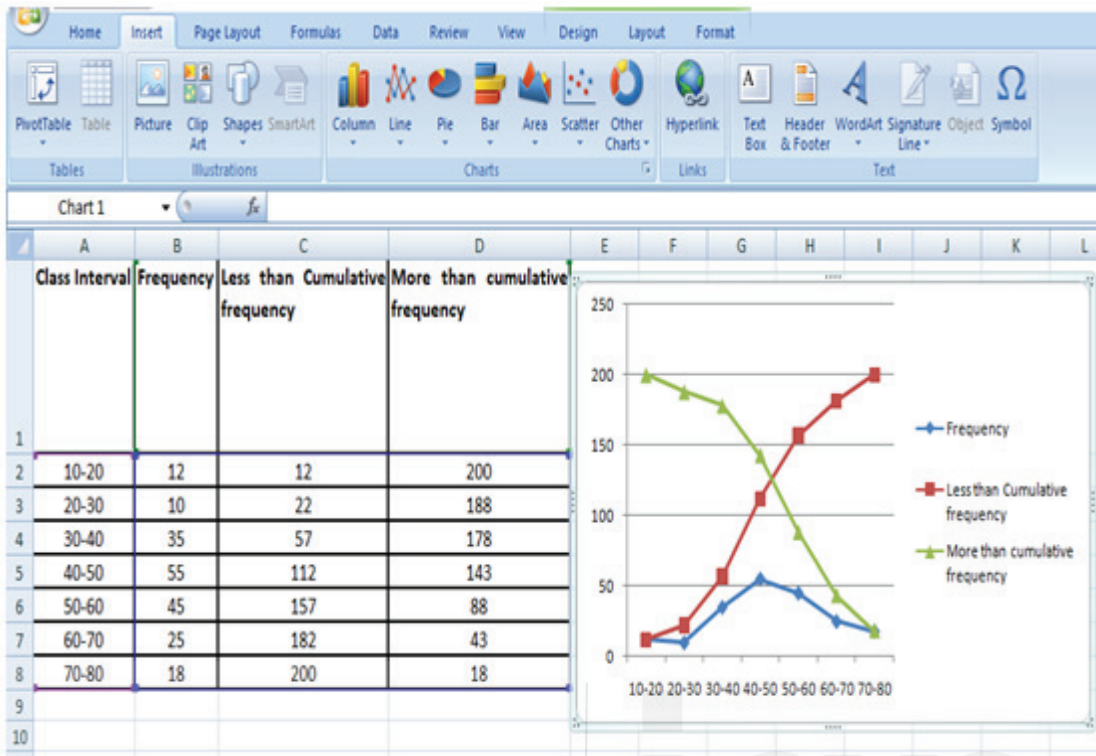


Fig. 6.10.1: The Cumulative frequency line Graph

6.5.6 Importance of Graphical Presentation of Data

Graphic presentation of data leaves a strong impact on the mind and it is easier to draw trends of data from it. The graphic presentation of data has the following advantages:

- i) Graphic presentation of data renders comparison of data much easier. The direction of curves or straight lines on the graphs makes it very simple to draw comparisons.
- ii) Graphic presentation of data helps in establishing trends of the time series data which makes it possible to interpolate or extrapolate the values. Thus it helps in forecasting.
- iii) Graphic presentation of data makes it possible to determine the values of the positional averages such as median, quartiles, mode, etc.
- iv) Through graphic presentation it is also possible to establish correlation between variables. Scatter diagram is a graphic presentation technique to determine the degree of correlation.

6.5.7 Misuse of Graphical Presentations

It is possible to mislead the observer or reader in a pictorial data presentation by manipulating the vertical (ordinate or Y-axis) and horizontal (abscissa or X-axis) lines of a graph. Elimination of zero frequency on ordinate difference among bars or ups and downs in a curve line can be highlighted in a desired way distorting the real findings of the study. Therefore, utmost care should be taken while presenting the findings graphically.

Check Your Progress 3

- 1) Draw a histogram from the following data

Value	5-10	10-15	15-20	20-25	25-30
Frequency	12	25	30	40	20

- 2) What is the procedure of making a frequency polygon ? Illustrate with the help of suitable example.

.....

.....

.....

.....

- 3) Explain the concept of “less than ogive” and “more than ogive” with the help of suitable example.

.....

.....

.....

.....

.....

6.6 LET US SUM UP

Data collected either from primary or secondary source are often in rudimentary form. These unsystematic raw data would fail to reveal any meaningful information to us. To draw any conclusions from these data, it must be arranged or organised in a standard way. This can be done with the help of classification. There are various types of classification e.g., data array, discrete frequency distribution, grouped frequency distribution, cumulative frequency distribution, cumulative relative frequency distribution. After classifying the raw data, its good presentation is also equally important. A good presentation enables us to highlight important features of the data and makes it fit for comparison and further statistical analysis. This can be achieved through statistical table, bar diagram, multiple bar diagram, and pie diagram and graphical presentations like histogram, frequency polygon, cumulative frequency curve, etc.

6.7 REFERENCES

- 1) Aron, A., Aron, N., and Coups, E. (2014). *Statistics for the Behavioral and Social Sciences : A brief course (Fifth Edition)*. Pearson Education Ltd., USA.
- 2) Asthana, H. S. and Bhushan, B. (2016). *Statistics for Social Sciences (with SPSS Application)*. Second Edition, Prentice Hall of India, New Delhi.

- 3) Cohen, B.H., and Lea, R.B. (2014). *Essential of Statistics for Social and Behavioural Sciences*. John Willey & Sons Inc.
- 4) Elhance, D. N., and Elhance, V. (1988). *Fundamentals of Statistics*. Kitab Mahal, Allahabad.
- 5) Garret, H. E. (2005). *Statistics in Psychology and Education*. Jain publishing, India.
- 6) Gupta, S.P. and M.P. Gupta, (1988). *Business Statistics*, Sultan Chand & Sons.: New Delhi.
- 7) Mohanty, B. (2015). *Statistics for Behavioural and Social Sciences*. Sage Publishing ,India.
- 8) Nagar, A. L., and Das, R. K. (1983). *Basic Statistics*. Oxford University Press, Delhi.
- 9) Rajaretnam, T (2016). *Statistics for Social Sciences*. Sage Publishing, India.
- 10) Sani, F., and Todman, J. (2006). *Experimental Design and Statistics for Psychology*. A first course book. Blackwell Publishing.
- 11) Yale, G. U., and M.G. Kendall (1991). *An Introduction to the Theory of Statistics*. Universal Books, Delhi.

6.8 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

- 1) Arrangement of data on the basis of some similarities is know as classification.
- 2) In order to get compressive insight of data, classification is needed.
- 3) Exclusive
- 4) Upper

Check Your Progress 2

- 1) Read Sub-section 6.3.1
- 2) Read Section 6.2 and Section 6.3
- 3) Read Sub-section 6.3.2

Check Your Progress 3

- 2) Read Sub-section 6.5.3
- 3) Read Sub-section 6.5.5

6.9 EXERCISES

- 1) Following are the marks obtained by 24 students in population studies examination. Classify them in a frequency table.

37	35	49	60	56	63	35	36
56	50	70	65	59	50	32	58
51	42	25	43	38	30	54	61

- 2) Convert the following inclusive series into an exclusive series.

Class Interval	1–10	11–20	21–30	31–40	41–50	51–60
Frequency	7	9	12	8	13	15

- 3) Tabulate the given data for sex education at high school level.

Town A	55% were male. 40% were in favour of sex education. 25% males were in favour of sex education.
Town B	50% were male. 35% were in favour of sex education. 40% males were in favour of sex education.

- 4) Discuss the meaning, utility, and limitations of the graphical presentation of statistical data.
- 5) Describe the general rules of constructing a graph.
- 6) Discuss the merits and demerits of graphical presentation.
- 7) Marks of 100 students in an examination are given below. Draw a histogram and a frequency polygon from the following data.

Marks	No. of Students
Less than 10	5
Less than 20	10
Less than 30	19
Less than 40	30
Less than 50	44
Less than 60	55
Less than 70	70
Less than 80	85
Less than 90	100

8) Draw a histogram from the following data.

Value	5–10	10–15	15–20	20–25	25–30	30–35
Frequency	12	25	35	42	20	16

9) Draw a frequency polygon from the data given below.

Age	0–20	20–40	40–60	60–80	80–100
No. of persons	15	30	42	28	10

10) Use a spreadsheet package to attempt question 7-9.



ignou
THE PEOPLE'S
UNIVERSITY

