

## UNIT -3

### COLLECTION OF DATA, CLASSIFICATION

The different types of data can be collected for different purposes. The data can be collected in connection with time or geographical location or in connection with time and location. The following are the three types of data:

1. Time series data.
2. Spatial data
3. Spacio-temporal data.

#### **Time series data:**

It is a collection of a set of numerical values, collected over a period of time. The data might have been collected either at regular intervals of time or irregular intervals of time.

#### **Spatial Data:**

If the data collected is connected with that of a place, then it is termed as spatial data.

#### **Spacio Temporal Data:**

If the data collected is connected to the time as well as place then it is known as spacio temporal data.

#### **Categories of data:**

Any statistical data can be classified under two categories depending upon the sources utilized.

These categories are,

1. Primary data
2. Secondary data

#### **Primary data:**

Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.

If a researcher is interested to know the impact of noon- meal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

The primary data can be collected by the following five methods.

1. Direct personal interviews.
2. Indirect Oral interviews.
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.

### **1. Direct personal interviews:**

The persons from whom informations are collected are known as informants. The investigator personally meets them and asks questions to gather the necessary informations. It is the suitable method for intensive rather than extensive field surveys. It suits best for intensive study of the limited field.

### **2. Indirect Oral Interviews:**

Under this method the investigator contacts witnesses or neighbours or friends or some other third parties who are capable of supplying the necessary information. This method is preferred if the required information is on addiction or cause of fire or theft or murder etc., If a fire has broken out a certain place, the persons living in neighbourhood and witnesses are likely to give information on the cause of fire. In some cases, police interrogated third parties who are supposed to have knowledge of a theft or a murder and get some clues. Enquiry committees appointed by governments generally adopt this method and get people' s views and all possible details of facts relating to the enquiry. This method is suitable whenever direct sources do not exists or cannot be relied upon or would be unwilling to part with the information.

The validity of the results depends upon a few factors, such as the nature of the person whose evidence is being recorded, the ability of the interviewer to draw out information from the third parties by means of appropriate questions and cross examinations, and the number of persons interviewed. For the success of this method one person or one group alone should not be relied upon.

### **3. Information from correspondents:**

The investigator appoints local agents or correspondents in different places and compiles the information sent by them. Informations to Newspapers and some departments of Government come by this method. The advantage of this method is that it is cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are likely to be negligent, prejudiced and biased. This method is adopted in those cases where informations are to be collected periodically from a wide area for a long time.

### **4. Mailed questionnaire method:**

Under this method a list of questions is prepared and is sent to all the informants by post. The list of questions is technically called questionnaire. A covering letter accompanying the questionnaire explains the purpose of the investigation and the importance of correct informations and request the informants to fill in the blank spaces provided and to return the form within a specified time. This method is appropriate in those cases where the informants are literates and are spread over a wide area.

### **5. Schedules sent through Enumerators:**

Under this method enumerators or interviewers take the schedules, meet the informants and filling their replies. Often distinction is made between the schedule and a questionnaire. A schedule is filled by the interviewers in a face-to-face situation with the informant. A questionnaire is filled by the informant which he receives and returns by post. It is suitable for extensive surveys.

### **Secondary Data:**

Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency. According to W.A.Neiswanger, ' A primary source is a publication in which the data are published by the same authority which gathered and analysed them. A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible' .

## **Sources of Secondary data:**

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. The sources of secondary data can broadly be classified under two heads:

1. Published sources, and
2. Unpublished sources.

### **1. Published Sources:**

The various sources of published data are: Clinical and other personal records, death certificates, published mortality statistics, census publications, etc. Examples include:

1. Official publications of Central Statistical Authority
2. Publication of Ministry of Health and Other Ministries
3. News Papers and Journals.
4. International Publications like Publications by WHO, World Bank, UNICEF
5. Records of hospitals or any Health Institutions.

### **2. Unpublished Sources**

All statistical material is not always published. There are various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc. Such sources can also be used where necessary

## **Precautions in the use of Secondary data**

The following are some of the points that are to be considered in the use of secondary data

1. How the data has been collected and processed
2. The accuracy of the data
3. How far the data has been summarized
4. How comparable the data is with other tabulations
5. How to interpret the data, especially when figures collected for one purpose is used for another

Generally speaking, with secondary data, people have to compromise between what they want and what they are able to find.

### **Classification:**

The collected data, also known as raw data or ungrouped data are always in an unorganised form and need to be organised and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible and assimilable form. The process of grouping into different classes or sub classes according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data. Thus classification is the first step in tabulation. For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore, Mumbai etc.,

### **Objects of Classification:**

The following are main objectives of classifying the data:

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

### **Types of classification:**

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

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

**a) Chronological classification:**

In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc., The data is generally classified in ascending order of time.

**b) Geographical classification:**

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

**c) Qualitative classification:**

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc., Such attributes cannot be measured along with a scale.

For example, if the population to be classified in respect to one attribute, say sex, then we can classify them into two namely that of males and females. Similarly, they can also be classified into 'married or ' single' on the basis of another attribute 'marital status' .

Thus when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is called simple or dichotomous classification.

The classification, where two or more attributes are considered and several classes are formed, is called a manifold classification. For example, if we classify population simultaneously with respect to two attributes, e.g sex and marital status, then population are first classified with respect to ' sex' into ' males' and ' females' . Each of these classes may then be further classified into ' maaarried' and single on the basis of attribute ' employment' and as such Population are classified into four classes namely.

- (i) Male married
- (ii) Male single
- (iii) Female married
- (iv) Female single

Still the classification may be further extended by considering other attributes like marital status etc.

#### **d) Quantitative classification:**

Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc.,

#### **What is a Basemap?**

A basemap provides a user with context for a map. You can add information to a basemap by overlaying other information on top of it. Basemaps contain reference information that may provide different geospatial information based on what the cartographer is trying to communicate. Foreexample, the NHD basemap has Watershed Boundaries, country and state boundaries, city names, and generalized rivers. Someone interested in viewing hydrography may not be interested in major highways or National Park boundaries so these layers are excluded from the basemap.

A base map is a layer with geographic information that serves as a background. A base map provides context for additional layers that are overlaid on top of the base map. Base maps usually provide location references for features that do not change often like boundaries, rivers, lakes, roads, and highways. Even on base maps, these different categories of information are in layers. Usually a base map contains this basic data, and then extra layers with a particular theme, or from a particular discipline, are overlaid on the base map layers for the sake of analysis.

#### **Compilation**

Maps may be compiled from other maps, usually of larger scale, or may be produced from original surveys and photogrammetric compilations. The former are sometimes referred to as derived maps and may include information from various sources, in addition to the maps from which they are principally drawn. Most small-scale series, such as the International Map of the World and World Aeronautical Charts, are compiled from existing information, though new data are occasionally produced to strengthen areas for which little or doubtful information exists. Thus compiled maps may contain fragments of original information while those representing original surveys may include some existing data of higher order, such as details from a city plat.

## **Changing of Scale**

In the process of compiling maps cartographers are often required to either reduce or enlarge maps. Reduction or enlargement involves change in the size.

- An enlargement provides the same map but proportionally larger than the original.
- A reduction gives the same map that is proportionally smaller than the original.

The above image or map has been reduced by  $\frac{1}{2}$ . The amount that an original image has been enlarged or reduced is called a scale factor, or an enlargement or reduction factor. It is the constant factor by which all dimensions of an object are enlarged or reduced in a map. If shapes have been reduced by half, the scale factor is  $\frac{1}{2}$ .

The ratio between the area of a map on one scale and its area to another scale is equal to the square of the ratios between the scales of the original and enlarged or reduced maps.

*Two methods:*

1. Graphical
2. Mechanical

### **Graphical Method**

Graphically maps can be enlarged or reduced with the help of similar squares.

The square method is the most common and simplest method for enlargement and reduction of maps. In order to enlarge a map, cover the original map with a set of squares of equal sides. The side of the squares has to be enlarged proportionally to that the original map. The side of the square of the new map has to be determined using the formula.

Scale of the new map = New scale/old scale x Side of the square of the original map.



1. Square method
2. Method of similar triangles

### **Mechanical method**

1. Pantograph
2. Proportional Compass
3. Eidograph
4. Camera Lucides

Road maps, produced by the millions, are compiled from road surveys, topographic maps, and aerial photography. City maps often represent original surveys, made principally to control engineering plans and construction. Some are, however, compiled from enlargements of topographic maps of the area.

Notations regarding the sources from which they were drawn are usually carried on compiled maps. This sometimes includes a reliability diagram showing the areas for which good information was available and those that may be less dependable. Comments regarding certain features or areas, which the editor may deem helpful to the user, may be made in the map itself.

Maps reflecting original surveys, such as a national topographic map series, carry standard marginal information. Date of aerial photography, process and instrumentation employed, notes regarding control and projection, date of field edit, and other information may be included. References to the availability of adjoining maps and those of other scales or series may also be included. Marginal ticks for intervals of plane coordinate systems, military grids, and other reference features are also shown and appropriately labeled.

### **Generalization**

The generalization of detail is a problem that frequently confronts the cartographer in original mapping and in reducing the scale of existing maps. There are two principal reasons for

taking such liberties (or topographic license in the case of the original mapping). The primary purpose is to avoid overcrowding and the resulting poor legibility. In addition, the degree of generalization or detail should be as consistent as possible throughout the map. Generalizations in some parts and excessive detail in others confuse the user and make the map's reliability suspect. Effective generalization requires good judgment based on seasoned knowledge and experience.

In approaching such problems as the thousands of islets in the Stockholm archipelago or the thousands of small lakes in the Alaskan tundra areas, when the map scale will accommodate only a small number, the cartographer may decide to draw the features in groupings that reflect the patterns shown in the large-scale source maps or aerial photos. This is difficult and at best presents the nature of the respective areas rather than a literal portrayal. There is also the possibility that the source maps may already have been generalized by some omissions to accommodate to their own scales. Another device is to note, in appropriate text or marginal references, that many minor lakes or islets are omitted because of scale. Such areas may also be symbolized and explained. The "pattern" representation noted above is actually a form of symbolization.

Intricate coastlines are also extremely difficult to generalize consistently. Here again, the purpose is to omit minor details while retaining the main features and their distinguishing characteristics. These and many equally perplexing questions arise in preparing maps of very small scale from any source. The problems of equalization of detail are also present in such cases. The topographer of earlier days had the equalization problem between areas close at hand and those viewed distantly. In addition, the topographer had to deal with terrain on the far sides of obscuring features.

Photogrammetrists—that is, persons who compile original maps from aerial photos—have similar problems when, for example, one side of a ridge is seen in more detail than the opposite side. Indeed, in steep terrain, parts of the far sides of some mountains are not seen at all. Appropriate steps must be taken in such cases to avoid differing renditions on opposite sides of the mountain. This may be accomplished by adding, in field completion of the manuscript map,

the segments not seen by the photogrammetrist; or additional aerial photography, patterned to cover the obscured sectors, may be requested.

Methods of generalization • Selection • Simplification – attribute generalization • classification – geometric generalization • simplification, enhancement, displacement, exaggeration, • areal editing (ambulation, elimination,...) – generalization of labels • Change of graphic representation.