# UNIT 9: DATA PROCESSING AND ANALYSIS

## UNIT STRUCTURE


9.1 Learning Objectives

9.2 Introduction

9.3 Processing of Data

9.4 Editing of Data

9.4.1 Types of editing of Data

9.5 Coding of Data

9.6 Tabulation of Data

9.7 Let Us sum Up

9.8 Further Reading

9.9 Answers to Check your progress

9.10 Model Questions


## 9.1 LEARNING OBJECTIVES

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

- learn the concept of editing
- discuss the meaning coding
- outline the process of tabulation
- describe the computer processing of data

## 9.2 INTRODUCTION

In this unit we are going to discuss the process and analysis of data. In the earlier units we discussed about the collection of data through primary and secondary methods, also we discussed on the concept and approaches of sampling and questionnaire design. Let us now discuss some of the basic terms and concepts related to research i.e. data analysis , data processing and data estimation. In our day to day life, we come across several information through print, audio and visual media, social gatherings and discussions, but we rarely think about that how these informations are being

collected, processed and analyzed. As discussed in the previous unit, the collection of data refers to a plan for gathering data, information from field situations. All these data collection effort will be become worthless, if the collected data are not properly processed and analyzed. **Data processing** is concerned with editing, coding, classifying, tabulating and charting and diagramming research data. Let us now discuss the approach in detail in the following sections and subsections.

## 9.3 PROCESSING OF DATA

The processing of data is an essential approach of managing and systemizing the facts and writing of a research report. Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use. Methods of processing must be rigorously documented to ensure the utility and integrity of the data. Processing of data is classified in the following five steps:

1. Editing of Data
2. Coding of Data
3. Classification of Data
4. Tabulation of Data
5. Diagrammatic representation of Data

Let us discuss these steps briefly:

1. **Editing of Data:** Editing is the first step in data processing. Editing is the process of examining the data collected in questionnaires/ schedules to detect errors and omissions and to see that they are corrected and the schedules are ready for tabulation. When the whole data collection is over a final and a thorough check up is made. (will be discussed in the following section 9.4)
2. **Coding of Data:** Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis.

Coding decisions should usually be taken at the designing stage of the questionnaire. (will be discussed in the following section 9.5)

3. **Classification Data:** Classification or categorization is the process of grouping the statistical data under various understandable homogeneous groups for the purpose of convenient interpretation. A uniformity of attributes is the basic criterion for classification; and the grouping of data is made according to similarity. Classification becomes necessary when there is a diversity in the data collected for meaningless for meaningful presentation and analysis.

4. **Tabulation of Data:** Tabulation is the process of summarizing raw data and displaying it in compact form for further analysis. Therefore, preparing tables is a very important step. Tabulation may be by hand, mechanical, or electronic. The choice is made largely on the basis of the size and type of study, alternative costs, time pressures, and the availability of computers, and computer programmes.

5. **Presentation of data:** The presentation of data could be tabular, statistical and cartographic. In case of tabular form of presentation, data related to different variables should be classified and compared. Various statistical techniques are available to derive accurate and precise results. Since techniques have a large range coupled with the limitations of their own, selection of appropriate technique needs to be made for the purpose. The construction of graphs, charts, diagrams and maps are the various forms of cartographic presentations. The data is transformed into cartographic system which is used for visual presentation. A brief account of tabular, statistical as well as cartographic presentation of data is discussed below.

   *i. Tabular Presentation:* It is used for summarization of data in its micro form. It helps in the analysis of trends, relationship and other characteristics of a given data. Simple tabulation is used to answer question related to one characteristic of the data whereas complex tabulation is used to present several interrelated characteristics. Complex tabulation results in two way, three way tables which give

information about two or three inter-related characteristic of data. The following points may be kept in mind while constructing a table.

*ii. Statistical Presentation of data:* The data collected through various sources needs to be processed statistically for precise explanations. Very often it becomes necessary to obtain a single representative value for the whole data set. The statistical measures that enable us to work out a single representative figure for the entire data distribution, is known as central tendency. Measures of central tendency help us to compare different distributions besides being representative for each distribution. These measures normally denote the central points of values, distance and occurrence in a distribution. The commonly used measures of central tendency are:

*a.* Arithmetic mean or average

*b.* Median

*c.* Mode

*iii. Cartographic Presentation of Data:* The primary data collected through the field survey may be presented cartographically. The representation of data in visual form refers either to time, space or to both. The cartographic presentation refers to the display of data by constructing graphs, diagrams and maps. The set of data is transformed into some form of figure which is used for illustrations. These figures could be graphic, geometric or theme specific maps.

## 9.4 EDITING OF DATA

Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding. When the researcher collects the data it is in a raw form and it needs to be edited, organized and analyzed. The raw data needs to be transformed into a comprehensible form of data. Editing of data is the process of examining errors and omissions in the collected data and making necessary correction in the same. Editing of data is very important when there is some inconsistency in the responses entered in the questionnaire or any half done or partial information in the

questionnaire. The editing of data can be done at two stages: field and post-field editing. The field editing is a review of reporting by the researcher for completing what has been written in an abbreviated form during interviewing the respondent. The post-field editing is carried out when field survey is completed and all the forms of schedule have been collected together.

A researcher should take care of the following pints while collecting the data:

1. Accuracy of the table
2. Consistent with other facts secured,
3. Unimormity of data
4. Completeness of the data,
5. Acceptable for tabulation and arranged to facilitate coding tabulation.

### 9.4.1 Types of Editing of Data

Let us now discuss the different types of editing of data:

**1. Editing for quality:** Quality is very important aspect while formulating and collecting data. To ensure the quality of the data the following should be taken care of:

**i.** the data forms should be complete,

**ii.** the data should be free from bias,

**iii.** the recordings are free from errors,

**iv.** the should be inconsistencies in responses within limits,

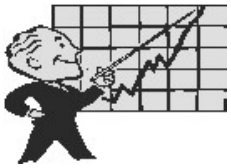
**v.** there should be evidences to show dishonesty of enumerators or interviewers and

**vi.** there should not be any willful manipulation of data.

**2. Editing for tabulation:** Editing for tabulation is done for certain accepted modification to data or even rejecting certain pieces of data in order to facilitate tabulation

**3. Field Editing:** The questionnaire filled up by the researcher or the respondent might have some abbreviated writings, illegible writings and the like. These are rectified by the researcher. This should be done soon after the researcher or interview before the loss of memory. The field editing should not extend to giving some guess data to fill up omissions.

**4. Central Editing:** It is done by the researcher after getting all schedules or questionnaires or forms from the enumerators or respondents. Obvious errors can be corrected. For missed data or information, the editor may substitute data or information by reviewing information provided by likely placed other respondents. A definite inappropriate answer is removed and “no answer” is entered when reasonable attempts to get the appropriate answer fail to produce results.



## CHECK YOUR PROGRESS

Q1: What is Processing of Data

.....

.....

Q.2: What is editing of Data?

.....

.....

## 9.5 CODING OF DATA

Coding is the procedure of classifying the answers to questions into meaningful categories. Coding is necessary for efficient analysis and through coding many responses may be reduced to a small number of classes which contain the critical information required for analysis.

Coding is basically done at the designing phase of the questionnaire. This makes it possible to pre-code the questionnaire choices and which in turn is helpful for computer tabulation.Coding is the process by which data are organized into classes or categories and numerals or other symbols are given to each item according to the class in which it falls. So, coding involves two steps:

1. To specify the different categories or classes into which the responses are to be classified
2. To allocate respondents responses to different categories.

While allocating the coding to the categories, it should be taken care of that the categories must be 'all inclusive' and 'mutually exclusive'. The all inclusive aspects can be taken care by adding one or more categories as "other", 'no information' and 'none'. The other aspect is that categories must be 'mutually exclusive' i.e. they must note be overlapping and ambigious.

To keep the response with in limited alternatives, the researcher need to assign some alphabetical or numerical symbols or both to the answers. For example in a question of educational qualifications alternative choices given are: Uneducated; Below Matriculation; Matriculation & above but below Graduate; Graduate & above; Technical Diploma; Technical Degree The alphabetical codes assigned to these alternatives could be A,B,C,D,E and F. Similarly, numerical codes to these alternatives could be 1,2,3,4.,and 5 respectively. It is necessary for the efficient analysis.

Let us now discuss the process of coding:

- Firstly, the study of the responses is very important. As discussed, coding begins at the preparation of interview schedules.

- Secondly, coding frame is developed by listing the possible answers to each question and assigning code numbers or symbols to each of them which are the indicators used for coding.
- Thirdly, after preparing the sample frame, the gradual process of fitting the answers to the questions must be begun.
- Lastly, transferring of the information from the schedules to a separate sheet called transcription sheet. Transcription sheet is a large summary sheet which contain the answer/codes of all the respondents. Transcription may not be necessary when only simple tables are required and the number of respondents are few.
- Coding is necessary to carry out the subsequent operations of tabulating and analyzing data.

## 9.6 TABULATION OF DATA

After the collection of data is completed, it is prepared for analysis. As the data is raw, it needs to be transformed in such a way, that it is appropriate for analysis. Tabulation comprises of sorting of data into different categories and counting the numbers of cases that belong to each category. Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.

In other words, it is a systematic arrangement of data in columns and rows that represents data in concise and attractive way. One should follow the given guidelines for tabulation.

The simplest way to tabulate is to count the number of responses to one question, which is called univariate tabulation. Where two or more variables are involved in tabulation , it is called bivariate or multivariate tabulation.

Tabulation can be done by hand or by mechanical or computer aided. The choice depends on the size and type of study, cost considerations, time

pressures and the availability of tabulating machines or computers. In relatively large inquiries, we may use mechanical or computer tabulation.

Let us now discuss the different types of tabulation:

There are generally two types of tables simple and complex. They are discussed following:

**Simple Tabulation:** Simple tabulation, gives information regarding one or more independent questions.Under it , the different attribute are stated in the left hand column and the frequency or extend of occurrence of each of these class are written in another column. In this three things are essential a) the classes made must be mutually exclusive, b) the tabulation must have internal logic and order, and c) the class intervals must carefully and reasonably selected. Following is an illustration of the same:

| No of visit to joggers park per month | No. of Respondents | % of Respondents |
|---|---|---|
| Below 5 | 2 | 7 |
| 6-10 | 15 | 50 |
| 12-16 | 13 | 43 |
| | 30 | 100 |

**Complex Tabulation or Cross Tabulation**: Complex tabulation gives information regarding two mutually dependent questions. It may often be necessary to tabulate responses to two or more questions simultaneously. Such tabulations are known as bivariate or multivariate tabulations, depending upon whether two or more than two variables are involved. In constructing cross tabulation tables, one has to first determine which data should be given primary emphasis and which should be given secondary emphasis. Data with primary emphasis are normally given in columns while those with secondary emphasis are shown in raws.

The following table is an example of cross tabulation.

| Income Level | Shopping Centres | | Total nos. of Households |
|---|---|---|---|
| | A | B | |
| Upto to Rs. 1000 | 100 | 200 | 300 |
| Rs. 1000+ | 120 | 80 | 200 |
| Total | 220 | 280 | 500 |

In the above table there are three variants i.e income level, shopping centres and households are being studied and tabulated.

Let us now discuss the steps are that are necessary in the **preparation of table:**

1. **Title of table:** The table should be first given a brief, simple and clear title which may express the basis of classification.
2. **Columns and rows:** Each table should be prepared in just adequate number of columns and rows.
3. **Captions and stubs:** The columns and rows should be given simple and clear captions and stubs.
4. **Ruling:** Columns and rows should be divided by means of thin or thick rulings.
5. **Arrangement of items;** Comparable figures should be arranged side by side.
6. **Deviations:** These should be arranged in the column near the original data so that their presence may easily be noted.
7. **Size of columns:** This should be according to the requirement.
8. **Arrangements of items:** This should be according to the problem.
9. **Special emphasis:** This can be done by writing important data in bold or special letters.

**10. Unit of measurement:** The unit should be noted below the lines.

**11. Approximation:** This should also be noted below the title.

**12. Foot – notes:** These may be given below the table.

**13. Total:** Totals of each column and grand total should be in one line.

**14. Source** : Source of data must be given. For primary data, write primary data.



**CHECK YOUR PRIGRESS**

Q3: What is Coding of Dat.....

.....

Q4: What is Tabulation of Data?

.....

.....



## 9.7 LET US SUM UP

In this unit we discussed the following:

- Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.
- Processing of data is classified in the following five steps:
  a. Editing of Data
  b. Coding of Data
  c. Classification of Data

d. Tabulation of Data

e. Diagrammatic representation of Data

- Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding.
- The editing of data can be done at two stages: field and post-field editing.
- The different types of editing of data are :Editing for quality, Editing for tabulation, Field Editing and Central Editing
- Coding is the procedure of classifying the answers to questions into meaningful categories.
- Coding is the process by which data are organized into classes or categories and numerals or other symbols are given to each item according to the class in which it falls. So, coding involves two steps:

  a. To specify the different categories or classes into which the responses are to be classified

  b. To allocate respondents responses to different categories.

- Tabulation comprises of sorting of data into different categories and counting the numbers of cases that belong to each category. Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.



## 9.8 FURTHER READING

1. Kothari, C.R. and Garg, G.(2014), 'Research Methodology - Methods and Techniques', New Age International Publishers
2. Malhotra. N.K.(2011), Marketing Research – An Applied Orientation, Pearson



## 9.9 ANSWERS TO CHECK YOUR PROGRESS

**Ans to Q1:** Data processing is a series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.

**Ans to Q2:** Editing of data is the first task of data processing. It is the process where the data are prepared for ensuing coding.

**Ans to Q3:** Coding is the procedure of classifying the answers to questions into meaningful categories.

**Ans to Q4:** Tabulation is a process of summarizing data and presenting it in a compact form, by putting data into statistical table.



## 9.10 MODEL QUESTIONS

1. What is Data Processing?
2. Write the classification of processing of data
3. What is presentation of Data?
4. Why editing of Data is important. Discuss.
5. Discuss the different types of editing of data
6. Explain Coding of Data
7. What is tabulation of data? Write the types of tabulation data.

*****