Chapter Seven: Data Analysis and Presentation

After collecting data from the field, the researcher has to process and analyze then in order to arrive at certain conclusions, which may or may not support the hypothesis, which he had formulated towards the beginning of his research work. Planning for data processing must be done well in advance of fieldwork as an integral part of the research design. Following are the stages through which the raw data must be processed in order ultimately to deliver the final products.

Editing: Editing means to look for and remove any errors, incompleteness or in consistency in the data, if the raw data are erroneous in compete, or inconsistent, these deficiencies will be carried through all subsequent stages of processing and will greatly distort the results of any inquiry. Therefore, at this stage, certain questions are specified for 100 percent editing because they are known to be especially troublesome or particularly critical to study objectives. The editor is responsible for seeing that the data are as accurate as possible; Consistent with other facts secured; uniformly entered; as complete as possible; Acceptable for tabulation; and Arranged to facilitate coding and tabulation. With regard to points or stages at which editing should be done one can talk of field editing and control editing.

a. <u>Field Editing:</u> It consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the later written in abbreviated and/or in illegible from at the time of recording the respondent's responses. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for other to decipher. This sort of editing should be done as soon as possible after the interview, preferable on the very day or on next day. While doing field auditing, the investigator must restrain himself and must not correct errors of omissions by simply guessing what the informant would have said if the question had been asked.

b. Central Editing: This should take place when all forms or schedules have been completed and returned to the office. This type of editing implies that all forms should get although editing by a single editing by a single editing a small study and by a team of editors in case of a large enquiry. Editors must keep in view several points while performing their work.

   (i) They should be familiar with instructions given to the interviewers and coders as well as with the editing instructions supplied to them for the purpose.

(ii) While crossing out an original entry for one reason or another, they should draw a single line on it so that same may remain legible.

(iii) They most make entries (if any) on the form in some distinctive color and that too in a standard form

(iv) They should initial all answers which they change or supply

(v) Editor's initials and date of the editing should be pace on each completed form or schedule.

1. **Coding:** Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put in to a limited number of categories or classes. Coding is necessary for classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. Following are important guidelines for coding:

    i. Coding should be mutually excusive

    ii. Set of categories should be collectively exhaustive so that all responses should be classified in one or the other category.

    iii. Separate categories should be created for recording 'non-response' and no knowledge response.

    iv. Inter coder and intera-coder agreement tests should be conducted through out the entire coding process to check its reliability.

    v. To help ensure that responses are being coded systematically.

2. **Classification:** Most research studies result in a large volume of raw data which must reduced in to homogenous groups for getting meaningful relationships. In this step data having common characteristic are placed in one class and in this way the entire data get divided in to a number of groups or classes. Classification can be of like following two types, depending upon the nature of the phenomenon involved:

(i) Classification according to Attributes: Data are classified on the basis of common characteristics which can either be descriptive or numerical. Descriptive characteristics refer to qualitative phenomenon, which can not be measured quantitatively; only their presence or absence in an individual item can be noticed. Data obtained this way on the basis of certain

attributes are known as statistics of attributes and their classification is said to be classification according to attributes. Such classifications can be simple or manifold classification: In *simple classifications* we consider only one attribute and divided the universe in to two classes- one consisting of items possessing attributes and the other class consisting of items which do not possess the given attribute. *Manifold classification* we consider two or more attributes simultaneously, and divided the data in to number of classes.

(ii) Classification according to class-intervals: Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomena which can be measured through some statistical units. Data relating to income production, age, weight etc. Come under this Category. Such data are known as statistics of variables and are classified on the basis of class intervals.

3. **Tabulation:** When a mass data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of concise and logical order. This process of summarizing raw data and displaying the same in compact form for the further analysis. In the broader sense, tabulation is the process of summarizing data and displaying the same in compact from for the further analysis. In the broader sense, tabulation is an orderly arrangement of data in columns and rows. Tabulation is essential because of the following reasons: It conserves space and reduces explanatory and descriptive statement to a minimum; It facilitates the process of comparison; It facilitates the summation of items and the detection of errors and omissions; and It provides a basis for various statistical computations.

Tabulation can be done by hand or by mechanical or electronic devices. The choice depends on the size and type of study, study, cost conditions time pressures and the availability of tabulating machines or computers. Generally accepted principles of tabulation are:

(i) Every table should have clear concise and adequate title so as to make the table intelligible with out reference to the text.

(ii) Every table should be given distinct number to facilitate easy reference

(iii) The column heading s and roe headings of the table should be clear and brief

(iv) Units of measurement under each heading or sub heading must always be indicated

(v) Explanatory foot notes, if any, concerning the table should be placed directly beneath the table.

(vi) Source of data must be indicated below the table.

(vii) It is generally consider better to approximate figures before tabulation as the same would reduce unnecessary details in the table it self.

(viii) In order to emphasize the relative significance of certain categories, different kinds of type, spacing and indentations may be used.

(ix) Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table.

(x) Miscellaneous and exceptional items, if any should be usually placed in the last row of the table.

(xi) Table should be made as logical clear, accurate and simple as possible, very large data should not be crowded in a single table.

(xii) The arrangement of the categories in the table may be chronological, geographical, and alphabetical or according to must suit the need and requirements of an investigation.

Elements/Types of Analysis

The manner in which data can be analyzed depends largely up on the measurement and sampling procedure followed in their collection. Depending up on two components the analysis is always more precise and objective. It also enables the readers of the study to evaluate the quality of the research. This is not possible in the case of non-statistical analysis, which is always qualitative and therefore less accurate. Analysis may therefore be categorized as descriptive analysis and inferential analysis.

I. Descriptive Analysis: is largely the study of distributions of one variable. This study provides us with profile of companies work group, persons and other subjects on any of a multitude of characteristics such as size composition, efficiency, preferences etc. This sort of analysis may be in respect of one variable (Uni-dimensional analysis), in respect of the two variables (bivariate analysis), or in respect of more than two variables (multivariate analysis)

Statistical Measures for single variate Analysis: Where the data consists of measurement of only one variable, they are often presented either in the form of a frequency table or a time series. In a frequency table, one column gives observed values of a random variable X and the other gives the frequency of each value. In a time series, one column gives certain units of time and the other gives observed values of a variable as it varies from one time to another. **Frequency table** is

commonly analyzed in commonly analyzed in terms of its four important characteristics viz-Central tendency, dispersion, skew ness and kurtosis. **Time series** is analyzed in terms of four important components Viz, Trend, seasonal variations, cyclical variations and irregular variations.

1. **Measures of central Tenancy or Averages:** These measures are so called because they show a tendency of the distribution to concentrate at certain values, somewhere in the center of the distribution. These include mean, mode, G.M, H.M, Quadratic mean index number.

2. **Measures of Dispersion:** How the items in a series are distributed and how they scatter. Measures of Dispersion are:
   a. Range as a measure of dispersion represents a difference between the values of entrée items i.e., largest and like smallest items of the data under review.
   b. Semi-inter quartile range is the difference between the upper and lower quarrel divided by 2.
   c. Mean Deviation is the average distance of the items in a series from their average
   d. Standard Devaluation

3. **Measures of Skew ness:** The data in a frequency distribution may fall in to symmetrical and asymmetrical patterns. The measures of the direction and degree of asymmetry are called measures of skew ness. In a symmetrical distribution the mean medium and mode between any two of these values indicates the extent of skew ness. Thus the measures of skew ness are: Mean- Mode; Mean-Median; and Median- Mode.

4. **Measure of Kurtosis:** These measure show the extent to which the distribution (when method and graph) is more peaked or less peaked than the normal curve. If the terms are more closely bunched around the mode than normal, making the curve unusually peaked, we say that curve is lep to kurtic. If on the other hand, the curve is more flat-topped than normal, we say it is platy Kurtic. The condition of peaked ness or of flat-topped ness itself is known as kutosis or excess.

II. **Statistical measures for Bivariate Analysis:** If we have the data on two variables i.e. for every measurement of a variable x, we have corresponding value of a second variable Y; the resulting pairs of values are called a bivariate population.
   i. Correction techniques
      a, karl person's coefficient of correlation

(simple linear correlation)

b, Partial correlation

c, Charles superman's coefficient of correlation

ii, Association of Attributes

iii. Simply Regression

III. Statistical Measures for Multivariate Analysis: Much behavioral research is multivariate nature and cannot be done with a bivariat approach. The determinants of phenomena like achievement, learning aggression, intelligence, certainly, risk taking organizational productivity, group cohesiveness etc. is complex. Many variables influence such phenomena, and multi-variables on one or more dependents variables.

Various statistical measures for multivariate analysis are