DATA PROCESSING AND ANALYSIS

Processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis. The term **analysis** refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups.

Processing operations

Processing operations are:

1. **Editing:** Editing of data is a process of examining the collected raw data (specially in surveys) to **detect errors and omissions and to correct these when possible.** As a matter of fact, editing involves a careful scrutiny of the completed questionnaires and/or schedules. Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered, as completed as possible and have been well arranged to facilitate coding and tabulation.

2. **Coding:** Coding refers to the process of **assigning numerals** or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. 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

3. **Classification:** Most research studies result in a large volume of raw data which must be reduced into **homogeneous** groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristics.

    4. **Tabulation:** When a mass of data has been assembled, it becomes necessary for the researcher to arrange the same in some kind of **concise and logical order.**

    Some problems in processing

1. The problem concerning "**Don't know" (or DK) responses** While processing the data, the researcher often comes across some responses that are difficult to handle

2. Use or percentages: Percentages are often used in data presentation for they simplify numbers, **reducing all of them to a 0 to 100 range**.

Data analysis embraces a whole range of activities of both the qualitative and quantitative type. It is usual tendency in behavioral research that much use of quantative analysis is made and statistical

methods and techniques are employed. The statistical methods and techniques are employed. The statistical methods and techniques have got a special position in research because they provide answers to the problems.

Quantitative analysis tools

Different types of quantitative analysis tools include **graphs, linear regressions and hypothesis testing**. These tools provide analysts with statistical methods of organizing and examining data. These tools are useful for analyzing **survey results, historical data or financial numbers**.

Qualitative analysis tools

Some popular types include **interviews, focus groups, surveys, e-mails, customer feedback, customer service tickets, observation notes, and phone calls**. This data, when it comes back to you, can be immense. Qualitative data analysis tools can help you organize, process, and analyze data for actionable insights

The following are the main purposes of data analysis:

(i) Description:

It involves a set of activities that are as essential first step in the development of most fields. **A researcher must be able to identify a topic about which much was not known; he must be able to convince others about its importance and must be able to collect data.**

(ii) Construction of Measurement Scale:

The researcher should construct a measurement scale. All numbers generated by measuring instruments can be placed into one of four categories

(a) Nominal: The number serves as nothing more than labels. For example no 1 was not less than no 2 .Similarly no 2 was neither more than no 1 and nor less than no 3.

(b) Ordinal: Such numbers are used to designate an ordering along some dimensions such as from less to more, from small to large, from sooner to later.

(c) Interval: The interval provides more précised information than ordinal one. By this type of measurement the researcher can make exact and meaningful decisions. For example if A,B and C are of 150cm, 145cm and 140 cm height, the researcher can say that A is 5 cm taller than B and B is 5 cm taller than C.

(d) Ratio Scale: It has two unique characteristics. The intervals between points can be demonstrated to be precisely the same and the scale has a conceptually meaningful zero point.

(iii) Generating empirical relationships:

Another purpose of analysis of data is identification of **regularities and relationships among data.** The researcher has no clear idea about the relationship which will be found from the collected data. If the data were available in details it will be easier to determine the relationship. The researcher can develop theories if he is able to recognize pattern and order of data. The pattern may be showing association among variables, which may be done by calculating correlation among variables or showing order, precedence or priority. The derivation of empirical laws may be made in the form of simple equations relating one interval or ratio scaled variable to a few others through graph methods.

(iv) Explanation and prediction:

Generally knowledge and research are equated with the identification of causal relationships and all research activities are directed to it. But in many fields the research has not been developed to the level where causal explanation is possible or valid predictions can be made. In such a situation explanation and prediction is construct as enabling the values of one set of variables to be derived given the values of another.

Functions:

The following are the main functions of data analysis:

(i) The researcher should analyze the available data for examining the statement of the problem.

(ii) The researcher should analyze the available data for examining each hypothesis of the problem.

(iii) The researcher should study the original records of the data before data analysis.

(iv) The researcher should analyze the data for thinking about the research problem in lay man's term.

(v) The researcher should analyze the data by attacking it through statistical calculations.

(vi) The researcher should think in terms of significant tables that the available data permits for the analysis of data.

Methods of data analysis

How you will analyze the data from the field e.g

- Percentages and frequencies
- Bar charts
- Tesing hypothesis
- Descriptions
- histogram

Statistical Calculations:

The researcher will have to use either descriptive statistics or inferential

statistics for the purpose of the analysis

Statistics

They are in two forms.

Descriptive statistics

Summarizes large amount of data so as to draw the conclusion by the researcher.

Inferential statistics

Is when the researcher is testing hypothesis so as to infer and make a decision.

(i) The descriptive statistics may be on any of the following forms:

(a) Measures of Central Tendency:

These measures are **mean, median, mode geometric mean and harmonic mean.** In behavioral

statistics the last two measures are not used. Which of the first three will be used in social

statistics depends upon the nature of the problem.

(b) Measures of Variability:

These measures are **range, mean deviation, quartile deviation and standard deviatio**n. In

social statistics the first two measures are rarely used. The use of standard deviation is very

frequently made for the purpose of analysis.

(c) Measures of Relative Position:

These measures are standard scores **(Z or T scores), percentiles and percentile ranks** .All of

them are used in educational statistics for data analysis.

(d) Measures of Relationship:

There measures are **Co-efficient of Correlation, partial correlation and multiple correlations.**

All of them are used in educational statistics for the analysis of data. However the use of rank method is made more in comparison to Karl pearson method.

(ii)The inferential statistics may be in any one of the following forms:

(a) Significance of Difference between Means:

It is used to determine whether a true difference exists between

population means of two samples.

(b) Analysis of Variance:

The Z or t tests are used to determine **whether there was any significant difference between the means of two random samples.** The F test enables the researcher to determine whether the sample means differ from one another to a greater extent then the test scores differ from their own sample means using the F ratio.

(c) Analysis of Co-Variance:

It is an extension of analysis of variance to test the significance of **difference between means of final experimental data** by taking into account the Correlation between the dependent variable and one or more Co-variates or control variables and by adjusting initial mean differences in the group.

(d) Correlation

The correlation can be used for the purpose of calculating the significance of the **difference between Co-efficient of correlatio**n.

(e) Chi Square Test:

It is used to estimate **the like hood that some factor other than chance accounts to the observed relationship**. In this test the expected frequency and observed frequency are used for evaluating Chi Square.

(f) Regression Analysis:

For calculating the probability of occurrence of any phenomenon or for predicting the phenomenon or relationship between different variables regression analysis is done.

The importance of statistics

1. Statistical knowledge helps you use the proper methods to collect the data
2. Employ the correct analyses and effectively present the results.
3. Statistics is a crucial process behind how we make discoveries in science, make decisions based on data. Helps to make predictions

Presentation of results

Results: A detailed presentation of the findings of the study, with supporting data in the form of **tables and charts together with a validation of results,** is the next step in writing the main text of the report. This **generally comprises the main body of the report, extending over several chapters.** The result section of the report should contain statistical **summaries and reductions** of the **data rather than the raw data**. All the results should be presented in logical sequence and splitted into readily identifiable sections**. All relevant results must find a place in the report.** But how one is to decide about what is relevant is the basic question. Quite often guidance comes primarily from the research problem and from the hypotheses, if any, with which the study was concerned. **But ultimately the researcher must rely on his own judgment in deciding the outline of his report.** "Nevertheless, it is still necessary that the states clearly the problem with which he was concerned, the procedure by which he worked on the problem, the conclusions at which he arrived, and the bases for his conclusions.

Interpretation of the results is a process of suggesting what results of a study might mean for theory, for practice, and for future research.

The interpretation of data **helps researchers to categorize, manipulate, and summarize the information in order to answer critical questions**. The importance of data interpretation is evident and this is why it needs to be done properly.