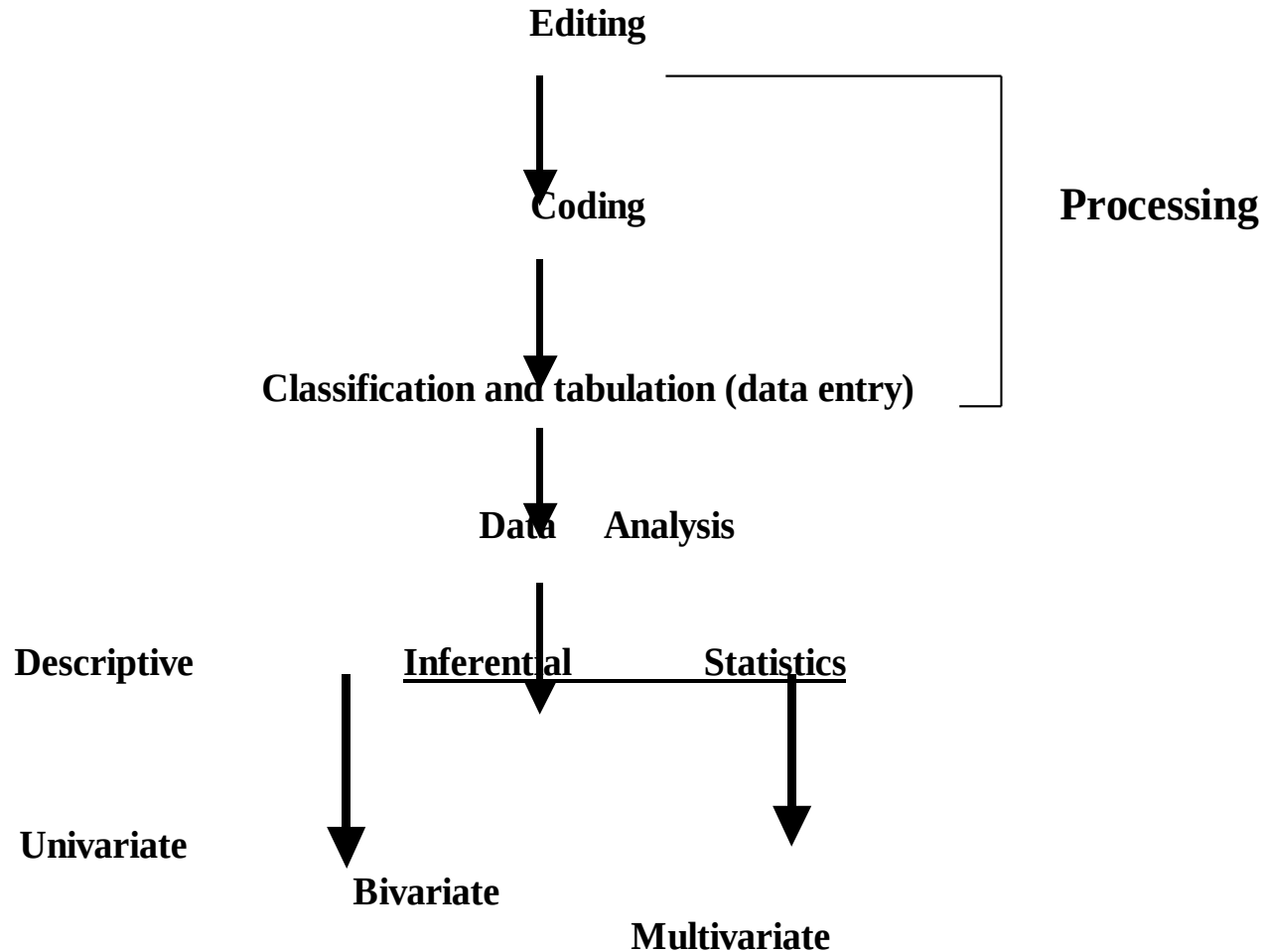


CHAPTER: –7

Overview of Data Processing and Analysis



7.1. Data processing

Data possessing implies

- **Editing:-** examining the collected raw data to detect errors and omission to correct those when possible
 - Field editing:- completing what has been written in abbreviation and/ or in illegible form at a time of recording the respondents' response
 - Central editing (to correct errors such as entry in the wrong place, omission)
- **Coding** (assigning numerical or other symbols to answers so that responses can be put into a limited number of classes)

Continued...

- **Classification:-** arranging data in groups or classes on the basis of common characteristics.

Classifications:

- According to attributes which is descriptive in nature (such as literacy, sex, honesty, etc) or numerical (such as weight, age, height, income, expenditure, etc.)

Continued...

- **According to class interval** -Data relating to income, production, age, weight, come under category. Such data are known as statistics of variables and are classified on the basis of class interval
- **Tabulation:-** arrangement of data in to rows and columns so that it becomes easy for analysis, comparison, statistical computations, summation of items and detection of errors and omissions,

7.2. Analysis

- It is further transformation of the processed data to look for patterns & relations among data groups
- The computation of certain measures along with searching for r/ships that exist among the data groups
- It involves estimating the values of unknown parameters of the population and testing of hypothesis for drawing inferences
- Analysis can be categorized as:
 - Descriptive Analysis
 - Inferential (Statistical) Analysis

7.2.1 Descriptive analysis

- It is largely the study of distribution of one variable
- Profiles of companies, work groups, persons, etc on any of a multiple of characteristics such as size, composition, efficiency, preference etc. This sort of analysis can be in respect of 1, 2, more than 3 variables (unidimensional, Bivariate, multivariate)
- The calculation of averages, frequency distribution, and percentage distribution is the most common form of summarizing data.

The most common forms of
describing the processed data
are:

- ✓ Tabulation
- ✓ Percentage
- ✓ Measure of central tendency
- ✓ Measure of dispersion
- ✓ Measure of asymmetry

Data transformation

- It is the process of changing original form of data to a form that is more suitable to perform a data analysis that will achieve the research objective.

1) Tabulation

- Refers to the orderly arrangement of data in a table or other summary format.
- It presents responses or the observations on a question-by-question basis & provides the most basic form of information.
- It tells the researcher how frequently each response occurs
- The starting point of analysis requires the counting of responses or observations for each of the categories. E.g. Frequency tables

2) Percentage

- Whether the data are tabulated by computer or by hand, it is useful to have percentages and cumulative percentage.
- Table containing percentage and frequency distribution is easier to interpret.
- Percentages are useful for comparing the trend over time or among categories

3) Measure of central tendency

- It is also known as statistical average. Mean, median and mode are most popular averages.
- *Mean* (arithmetic mean) is the common measure of central tendency
- *Mode* is not commonly used one
- *Median* is commonly used in estimating the average of qualitative phenomenon like estimating intelligence.

4) Measurement of dispersion

- How the value of an item is scattered around the true value of the mean.
- It is a measurement of how far is the value of the variable far from the average value.

Important measures of dispersion are:

- Range:
- Mean deviation: It is the average dispersion of an observation around the mean value. $\Sigma(X_i - \bar{X})/n$
- Variance: It measures the sample variability.

5) Measurement of asymmetry (skew-ness)

- When the distribution of items is happen to be perfectly symmetrical, then we have a normal curve & the distribution is normal. Such curve is perfectly bell shaped curve in which case the value of **Mean = Median = Mode**
- Under this condition the skew-ness is altogether absent. If the curve is distorted (whether on the right or the left side), we have asymmetric distribution which indicates that there is a skew-ness.

7.2.2. Inferential Analysis

- Researchers frequently conduct & seek to determine the r/ship between variables & test statistical significance
- If we have data on two variables we said to have a **bivariate** variable, if the data is more than two variables then the population is known as **multivariate** population
- If for every measure of a variable X, we have corresponding value of variable Y, the resulting pairs of value are called a bivariate population

Continued.....

- In case of bivariate or multivariate population, we often wish to know the relationship between the two or more variables from the data obtained.

E.g. We may like to know, “Whether the number of hours students devote for study is somehow related to their family income, to age, to sex, or to similar other factors.

Continued.....

Two questions should be answered to determine the relationship between variables:

1. Is there exist association or correlation between the two or more variables? If yes, then up to what degree?
- This will be answered by the use of **correlation** technique.

- In case of bivariate population, correlation can be found using
 - Cross tabulation
 - Karl Pearson's coefficient of correlation:
It is simple correlation and commonly used
 - Charles Spearman's coefficient of correlation
- In case of multivariate population correlation can be studied through:
 - Coefficient of multiple correlation
 - Coefficient of partial correlation

2. Is there any cause and effect (causal relationship) between two variables or between one variable on one side and two or more variables on the other side?
- This question can be answered by the use of regression analysis.
 - In regression analysis the researcher tries to estimate or predict the average value of one variable on the basis of the value of other variable.
 - For instance a researcher estimates the average value score on statistics knowing a student's score on a mathematics examination.

- There are different techniques of regression:
 - In case of bivariate population cause and effect relationship can be studied through *simple regression*.
 - In case of multivariate population. causal relationship can be studied through *multiple regression analysis*.

Time series Analysis

- Successive observations of the given phenomenon over a period of time are analyzed through time series analysis. It measures the relationship between variables and time (trend)
- Time series will measure seasonal fluctuation, cyclical irregular fluctuation, and trend.
- The analysis of time series is done to understand the dynamic condition of achieving the short term and long-term goal of business firm for forecasting purpose