Module 6 Part 1 Data Processing
6.1 Descriptive Analysis
6.2 Inferential Analysis

DATA PROCESSING

Data continues to be in raw form, unless and until they are processed and analyzed. Processing is a statistical method by which the collected data is so organized the further analysis and interpretation of data become easy. It is an intermediary stage between the collection of data and their analysis and interpretation.

Processing stages

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.

With regard to points or stages at which editing should be done, one can talk of field editing and central editing.

Field editing consists in the review of the reporting forms by the investigator for completing (translating or rewriting) what the latter has written in abbreviated and/or in illegible form at the time of recording the respondents' responses. This type of editing is necessary in view of the fact that individual writing styles often can be difficult for others to decipher. This sort of editing should be done as soon as possible after the interview, preferably on the very day or on the next day.

Central editing 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 a thorough editing by a single editor in a small study and by a team of editors in case of a large inquiry. Editor(s) may correct the obvious errors such as an entry in the wrong place, entry recorded in months when it should have been recorded in weeks, and the like. In case of inappropriate on missing replies, the editor can sometimes determine the proper answer by reviewing the other information in the schedule. At times, the respondent can be contacted for clarification.

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. They must also possess the characteristic of exhaustiveness (i.e., there must be a class for every data item) and also that of mutual exclusively which means that a specific answer can be placed in one and only one cell in a given

category set. Another rule to be observed is that of unidimensionality by which is meant that every class is defined in terms of only one concept.

Steps in coding

1. Study the answers carefully.

2. Develop a coding frame by listing the answers and by aligning codes to each of them.

3. Prepare a coding manual with the detail of variable names, codes and instructions.

4. If the coding manual has already been prepared before the collection of the data, make the required additions for the open ended and partially coded questions.

Coding rules

1. Give each respondent a code number for identification.

2. Provide code number for each question.

3. All responses including ‘don’t know’, ‘no opinion’. etc is to be coded.

4. Assign additional codes to partially coded questions.

Classification

Classification is the process of reducing large mass of data in to homogeneous groups for meaningful analysis. It converts data from complex to understandable and unintelligible to intelligible forms. It divides data in to different groups or classes according to their similarities and dissimilarities. When the data are classified, they give summary of whole information.

Broadly speaking, there are four types of classification. They are:

(i) Geographical classification, (ii) Chronological classification, (iii) Qualitative classification, and (iv) Quantitative classification.

Classification according to attributes: As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.). Descriptive characteristics refer to qualitative phenomenon which cannot 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.

Classification according to class-intervals: Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomenon 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.

4.Tabulation

Tabulation is the next step to classification. It is an orderly arrangement of data in rows and columns. It is defined as the "measurement of data in columns and rows". Data presented in tabular form is much easier to read and understand than the data presented in the text the main purpose of tabulation is to prepare the data for final analysis. It is a stage between classification of data and final analysis. Objectives of Tabulation

1. To clarify the purpose of enquiry

2. To make the significance of data clear.

3. To express the data in least possible space.

4. To enable comparative study.

5. To eliminate unnecessary data

6. To help in further analysis of the data.

Generally accepted principles of tabulation: Such principles of tabulation, particularly of constructing statistical tables, can be briefly states as follows:

1. Every table should have a clear, concise and adequate title so as to make the table intelligible without reference to the text and this title should always be placed just above the body of the table.

2. Every table should be given a distinct number to facilitate easy reference.

3. The column headings (captions) and the row headings (stubs) of the table should be clear and brief.

4. The units of measurement under each heading or sub-heading must always be indicated.

5. Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the reference symbols used in the table.

6. Source or sources from where the data in the table have been obtained must be indicated just below the table.

7. Usually the columns are separated from one another by lines which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions.

8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines.

9. The columns may be numbered to facilitate reference.

10. Those columns whose data are to be compared should be kept side by side. Similarly, percentages and/or averages must also be kept close to the data.

11. It is generally considered better to approximate figures before tabulation as the same would reduce unnecessary details in the table itself.

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

13. It is important that all column figures be properly aligned. Decimal points and (+) or (–) signs should be in perfect alignment.

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

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

16. Table should be made as logical, clear, accurate and simple as possible. If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient.

17. Total of rows should normally be placed in the extreme right column and that of columns should be placed at the bottom.

18. The arrangement of the categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison. Above all, the table must suit the needs and requirements of an investigation.

Data Entry

Once data collection has been completed and checked, the process of data entry and cleaning starts. During data entry the verbal or numeric data collected using questionnaires, abstraction forms, or observations are entered into a computer, principally as numeric data "codes."

Validity of Data

In general, validity is an indication of how sound your research is. More specifically, validity applies to both the design and the methods of your research. Validity in data collection means that your findings truly represent the phenomenon you are claiming to measure. Data validation means checking the accuracy and quality of source data before using, importing or otherwise processing data. Different types of

validation can be performed depending on destination constraints or objectives. Data validation is a form of data cleansing.

Common Descriptive Techniques

The most common descriptive statistics used in research consist of percentages and frequency tables

(a)Percentages

Percentages are a popular method of displaying distribution. Percentages are the most powerful in making comparisons. In percentages, we simplify the data by reducing all numbers in a range of 10 to 100.

(b) Frequency Tables

One of the most common ways to describe a single variable is with a frequency distribution. Frequency distribution can be depicted in two ways, as table or as a graph. If the frequency distribution is depicted in the form of a table, we call it frequency table.

(c) Contingency Tables

A Contingency table shows the relationship between two variables in tabular form. The term Contingency table was first used by the statistician Karl Pearson in 1904. Contingency tables are especially used in Chi- square test.

Graphs and Diagrams

In research, the data collected may be of complex nature. Diagrams and graphs is one of the methods which simplifies the complexity of quantitative data and make them easily intelligible. They present dry and uninteresting statistical facts in the shape of attracting and appealing pictures. They have a lasting effect on the human mind than the conventional numbers.

Uses of Graphs and Diagrams

1. They help in presenting quantitative facts in simple, clear and effective pictures.

2. They make the whole data readily intelligible.

3. They can be used for comparison purpose.

4. They are useful in analyzing complex economic theories.

5. They save much time in understanding data.

6. Facts can be understood without doing mathematical calculations.

7. They help in lo0cating statistical measures such as median. quartile, mode etc

The following graphs are commonly used to represent data

1. Charts or line graphs

2. Bar charts

3. Circle charts or pie diagram

4. Pictograms

1. Line Graphs

A line graph displays information in a series of data points that each represents an individual measurement or piece of data. The series of points are then connected by a line to show a visual trend in data over a period of time. The line is connected through each piece chronologically.

2.Bar Charts

The bar graph is a common type of graph which consists of parallel bars or rectangles with lengths that are equal to the quantities that occur in a given data set. The bars can be presented vertically or horizontally to show the contrast and record information. Bar graphs are used for plotting discontinuous (discrete) data. Discrete data contains discrete values and are not continuous.

Histogram

A histogram is a graph of frequency distributions. It is a set of vertical bars whose are proportional to the frequencies. While constructing histogram, the variable is always taken on the x- axis and the frequencies on y-axis.

Frequency Polygon

The frequency polygon is a graph of frequency distribution. Here we draw histogram of the data and then join by straight line and mid points of upper horizontal sides of these bars. Join both ends of the frequency polygon with the x- Axis.

Frequency Curves

A continuous frequency distribution can be represented by a smoothed curve known as

Frequency curves.

Ogive or Cumulative Frequency Curve

A frequency distribution can be cumulated in two ways, less than cumulative series and more than cumulative series. Smoothed frequency curves drawn for these two cumulative series are called cumulative frequency curves or ogives.

- Less than ogive curve: In less than ogive curve the upper limit per limit of each class interval is taken on x- axis in increasing order. For each such upper limit on x-axis, the cumulative frequency of all the class intervals from the first class interval to last class interval are taken on the y-axis.

• More than ogive curve: In more than ogive curve the lower limit of each class interval is taken on x- axis in increasing order. For each such lower limit on x- axis the cumulative

frequency of all the class interval from that class interval to the last class interval are taken on y-axis.

3. Circle Charts or Pie Diagram

A pie graph is a circle divided into sections which each display the size of a relative piece of information. Each section of the graph comes together to form a whole. In a pie graph, the length of each sector is proportional to the percentage it represents.

4.Pictograms

A pictogram, also called a pictogram or pictograph, is an ideogram that conveys its meaning through its pictorial resemblance to a physical object. Pictographs are often used in writing and graphic systems in which the characters are to a considerable extent pictorial in appearance. Pictography is a form of writing which uses representational, pictorial drawings. It is a basis of cuneiform and, to some extent, hieroglyphic writing, which also uses drawings as phonetic letters or determinative rhymes.

ANALYSIS OF DATA

Analysis of data is considered to be highly skilled and technical job which should be carried out. Only by the researcher himself or under his close supervision. Analysis of data means critical examination of the data for studying the characteristics of the object under study and for determining the patterns of relationship among the variables relating to it's using both quantitative and qualitative methods.

Purpose of Analysis

Statistical analysis of data saves several major purposes.

1. It summarizes large mass of data in to understandable and meaningful form.

2. It makes descriptions to be exact.

3. It aids the drawing of reliable inferences from observational data.

4. It facilitates identification of the casual factors unde3rlyiong complex phenomena

5. It helps making estimations or generalizations from the results of sample surveys.

6. Inferential analysis is useful for assessing the significance of specific sample results under assumed population conditions.

Steps in Analysis

Different steps in research analysis consist of the following.

1. The first step involves construction of statistical distributions and calculation of simple measures like averages, percentages, etc.

2. The second step is to compare two or more distributions or two or more sub groups within a distribution.

3. Third step is to study the nature of relationships among variables.

4. Next step is to find out the factors which affect the relationship between a set of variables

5. Testing the validity of inferences drawn from sample survey by using parametric tests of significance.

Types of Analysis

Statistical analysis may broadly classified as descriptive analysis and inferential analysis

Descriptive Analysis

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data or the quantitative description itself. In such analysis there are univariate analysis bivariate analysis and multivariate analysis.

• Univariate analysis

• Univariate analysis involves describing the distribution of a single variable, including its central tendency (including the mean, median, and mode) and dispersion (including the range and quartiles of the data-set, and measures of spread such as the variance and standard deviation). The shape of the distribution may also be described via indices such as skewness and kurtosis. Characteristics of a variable's distribution may also be depicted in graphical or tabular format, including histograms and stem-and-leaf display.

• Bivariate analysis

• Bivariate analysis is one of the simplest forms of the quantitative (statistical) analysis. It in volves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them. Common

forms of bivariate analysis involve creating a percentage table or a scatter plot graph and computing a simple correlation coefficient

• Multivariate analysis.

• In multivariate analysis multiple relations between multiple variables are examined simultaneously. Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time.

In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

Usually the following analyses are involved when we make a reference of multivariate analysis:

(a) Multiple regression analysis: This analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.

(b) Multiple discriminant analysis: This analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this

analysis happens to be to predict an entity's possibility of belonging to a particular group based on several predictor variables.

(c) Multivariate analysis of variance (or multi-ANOVA): This analysis is an extension of two way ANOVA, wherein the ratio of among group variance to within group variance is worked out on a set of variables.

The coefficient of variation (CV) is a measure of relative variability. It is the ratio of the standard deviation to the mean (average). For example, the expression "The standard deviation is 15% of the mean" is a CV.

The CV is particularly useful when you want to compare results from two different surveys or tests that have different measures or values.

Formula

The formula for the coefficient of variation is:

Coefficient of Variation = (Standard Deviation / Mean) * 100.

In symbols: CV = (SD/) * 100.

Multiplying the coefficient by 100 is an optional step to get a percentage, as opposed to a decimal.

Factor analysis

Factor analysis is a technique that is used to reduce a large number of variables into fewer numbers of factors. This technique extracts maximum common variance from

all variables and puts them into a common score. As an index of all variables, we can use this score for further analysis. Factor analysis is part of general linear model (GLM) and this method also assumes several assumptions: there is linear relationship, there is no multicollinearity, it includes relevant variables into analysis, and there is true correlation between variables and factors. Several methods are available, but principle component analysis is used most commonly.

Types of factoring:

There are different types of methods used to extract the factor from the data set:

1. Principal component analysis: This is the most common method used by researchers. PCA starts extracting the maximum variance and puts them into the first factor. After that, it removes that variance explained by the first factors and then starts extracting maximum variance for the second factor. This process goes to the last factor.

2. Common factor analysis: The second most preferred method by researchers, it extracts the common variance and puts them into factors. This method does not include the unique variance of all variables.

Canonical analysis

Canonical analysis: This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.

Cluster analysis

Cluster analysis is a multivariate method which aims to classify a sample of subjects (or objects) on the basis of a set of measured variables into a number of different groups such that similar subjects are placed in the same group. Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

Inferential analysis

Inferential analysis is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

Tools and Statistical Methods For Analysis

The tools and technique of statistics can be studied under two divisions of statistics.

(A)Descriptive Statistics

In descriptive statistics we develop certain indices and measures of raw data. They are;

1. Measures of Central Tendency

2. Measures of Dispersion

3. Other measures

1. Measures of Central Tendency.

The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are different types of estimates of central tendency such as mean, median, mode, geometric mean, and harmonic mean.

2. Measures of Dispersion.

Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. It can be used to compare the variability in two statistical series.

3. Measures of correlation

Correlation refers to any of a broad class of statistical relationships involving dependence. When there are two variables, the correlation between them is called simple correlation.

Whenthere are more than two variables and we want to study relation between two of them only, treating the others as constant, the relation is called partial correlation. When there are more than two variables and we want to study relation of one variable with all other variables together, the relation is called multiple correlations.

4. Regression analysis

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

5. Index numbers

An index is a statistical measure of changes in a representative group of individual data points. Index numbers are designed to measure the magnitude of economic changes over time. Because they work in a similar way to percentages they make such changes easier to compare.

6. Time series analysis

A time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. Time series analysis comprises methods for analysing time series data in order to extract meaningful statistics and other characteristics of the data.

Measures of central tendency (averages)

An average is a single significant figure which sums up characteristic of a group of figures. The various measures of central tendency are;

(1) Arithmetic mean

(2) Median

(3) Mode

(4) Geometric mean (5) Harmonic mean

Arithmetic Mean

The Mean or average is probably the most commonly used method of describing central tendency. To compute the mean all you do is add up all the values and divide by the number of value.

Arithmetic mean =

Median

The Median is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample.

Mode

Mode is the value of the item of a series which occurs most frequently. According to Kenny 'the value of the variable which occurs most frequently in a distribution is called a mode". In the case of individual series, the value which occurs more number of times is mode.

Geometric mean

Geometric mean is also useful under certain conditions. It is defined as the nth root of the product of the values of n times in a given series. Symbolically, we can put it thus:

Geometric mean (or G.M.) = $n\sqrt{X1 \times X2 \times X3 \ldots\ldots Xn}$

n= number of items X1,X2 = the various values

Harmonic Mean

The harmonic mean is a type of numerical average. It is calculated by dividing the number of observations by the reciprocal of each number in the series. Thus, the harmonic mean is the reciprocal of the arithmetic mean of the reciprocals.

MEASURES OF DISPERSION

An averages can represent a series only as best as a single figure can, but it certainly cannot reveal the entire story of any phenomenon under study. Specially it fails to give any idea about the scatter of the values of items of a variable in the series around the true value of average. In order to measure this scatter, statistical devices called measures of dispersion are calculated. Important measures of dispersion are (a) range, (b) mean deviation, and (c) standard deviation.

(a) Range is the simplest possible measure of dispersion and is defined as the difference betweenthe values of the extreme items of a series. Thus,

Range = (Highest value of anitem in a series) – (Lowest value of an item in a series)

The utility of range is that it gives an idea of the variability very quickly, but the drawback is that range is affected very greatly by fluctuations of sampling. Its value is never stable, being based on only two values of the variable. As such, range is mostly used as a rough measure of variability and is not considered as an appropriate measure in serious research studies.

(b) Mean deviation is the average of difference of the values of items from some average of the series. Such a difference is technically described as deviation. In calculating mean deviation we ignore the minus sign of deviations while taking their total for obtaining the mean deviation. Mean deviation is, thus, obtained as under:

(c) Standard deviation is most widely used measure of dispersion of a series and is commonly denoted by the symbol 's' (pronounced as sigma). Standard deviation is defined as the square –root of the average of squares of deviations, when such deviations for the values of individual items in a series are obtained from the arithmetic average.

Index Numbers

Index numbers are designed to measure the magnitude of economic changes over time. A statistic which assigns a single number to several individual statistics in order to quantify trends. Index numbers are the indicators of the various trends in an economy. Price index numbers indicate the position of prices whether they are rising or falling and at what rate. Similarly, index numbersregarding agricultural production indicates the trend of change whether it is rising or falling at what rate over a period of time. An index number is an economic data figure reflecting price or quantity compared with a standard or base value. The base usually equals 100 and the index number is usually expressed as 100 times the ratio to the base value. For example, if a commodity costs twice as much in 1970 as it did in 1960, its index number would be 200 relative to 1960. Index numbers are used especially to compare business activity, the cost of living, and employment.

An index number is specialized average. Index numbers may be simple or weighted depending on whether we assign equal importance to every commodities or different importance to different commodities according to the percentage of income spent on them or on the basis of some other criteria. In this chapter, we shall discuss both simple and weighted index numbers.

Simple and weighted index numbers

Simple index numbers are those in the calculation of which all the items are treated as equally important. Here items are not given any weight. Weighted index numbers are those in the calculation of which each item is assigned a particular weight.

Price Index Numbers

Price index numbers measure changes in the price of a commodity for a given period in comparison with another period.

Inferential Analysis

Parameters and Statistics

Parameters are numbers that summarize data for an entire population. Statistics are numbers that summarize data from a sample, i.e. some subset of the entire population.

Testing of hypothesis

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis. Hypothesis testing is used to infer the result of a hypothesis performed on sample data from a larger population.

In hypothesis testing, an analyst tests a statistical sample, with the goal of accepting or rejecting a null hypothesis. The test tells the analyst whether or not his primary

hypothesis is true. If it isn't true, the analyst formulates a new hypothesis to be tested, repeating the process until data reveals a true hypothesis.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

Basic concepts in the context of testing of hypotheses need to be explained.

(a) Null hypothesis and alternative hypothesis : In the context of statistical analysis, we often talk about null hypothesis and alternative hypothesis. If we are to compare method A with method B about its superiority and if we proceed on the assumption that both methods are equally good, then this assumption is termed as the null hypothesis. As against this, we may think that the method A is superior or the method B is inferior, we are then stating what is termed as alternative hypothesis. The null hypothesis is generally symbolized as H0 and the alternative hypothesis as H1.

(b) The level of significance : This is a very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen with great care, thought and reason. In case we take the significance level at 5 per cent, then this implies that H0 will be rejectedwhen the sampling result (i.e., observed evidence) has a less than 0.05 probability of occurring if H0 is true. In other words, the 5 per cent level of significance means that researcher is willing to take as much as a 5 per cent risk of rejecting the null hypothesis when it (H0) happens to be

true. Thus the significance level is the maximum value of the probability of rejecting H0 when it is true and is usually determined in advance before testing the hypothesis.

(c) Type I and Type II errors : In the context of testing of hypotheses, there are basically two types of errors we can make. We may reject H0 when H0 is true and we may accept H0 when in fact H0 is not true. The former is known as Type I error and the latter as Type II error. In other words, Type I error means rejection of hypothesis which should have been accepted and Type II error means accepting the hypothesis which should have been rejected.

Steps in testing hypothesis

1. State the problem

2. Set up a hypothesis

3. Decide the test statistics

4. Select a level of significance

5. Calculate the value of test statistic

6. Obtain the table value

7. Make decision to accept or reject hypothesis.

Test Statistic

The decision to accept or to reject a null hypothesis is made on the basis of a statistic computed from the sample. Such a statistic is called the test statistic. There are different types of test statistics. All these test statistics can be classified into two groups. They are a). Parametric Tests b). Non-Parametric Tests

PARAMETRIC TESTS

The statistical tests based on the assumption that population or population parameter is normally distributed are called parametric tests. The important parametric tests are:-

1.z-test

2.t-test

3.f-test

z-test

z-test is based on the normal probability distribution and is used for judging the significance of several statistical measures, particularly the mean. The relevant test statistic, z, is worked out and compared with its probable value (to be read from table showing area under normal curve) at a specified level of significance for judging the significance of the measure concerned. This is a most frequently used test in research studies. This test is used even when binomial distribution or t- distribution is applicable on the presumption that such a distribution tends to approximate normal distribution as 'n' becomes larger. z-test is generally used for comparing the mean of a sample to some hypothesised mean for the population in case of large sample, or when population variance is known. z-test is also used for judging he significance of difference between means of two independent samples in case of large samples, or when population variance is known. z-test is also used for comparing the sample proportion to a theoretical value of population proportion or for judging the difference in proportions of two independent samples when n happens to be large. Besides, this test may be used for judging the significance of median, mode, coefficient of correlation and several other measures.

t-test

t-test is based on t-distribution and is considered an appropriate test for judging the significance of a sample mean or for judging the significance of difference between the means of two samples in case of small sample(s) when population variance is not known (in which case we use variance of the sample as an estimate of the population variance). In case two samples are related, we use paired t-test (or what is known as difference test) for judging the significance of the mean of difference between the two related samples. It can also be used for judging the significance of the coefficients of simple and partial correlations. The relevant test statistic, t, is calculated from the sample data and then compared with its probable value based on t-distribution (to be read from thetable that gives probable values of t for different levels of significance for different degrees of freedom) at a specified level of significance for concerning degrees of freedom for accepting or rejecting the null hypothesis. It may be noted that t-test applies only in case of small sample(s) when population variance is unknown.

F-test

F-test is based on F-distribution and is used to compare the variance of the two-independent samples. This test is also used in the context of analysis of variance (ANOVA) for judging the significance of more than two sample means at one and the same time. It is also used for judging the significance of multiple correlation coefficients. Test statistic, F, is calculated and compared with its probable value (to be seen in the F-ratio tables for different degrees of freedom for greater and

smaller variances at specified level of significance) for accepting or rejecting the null hypothesis.

ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (abbreviated as ANOVA) is an extremely useful technique concerning researches in the fields of economics, biology, education, psychology, sociology, business/industry and in researches of several other disciplines. This technique is used when multiple sample cases are involved. As stated earlier, the significance of the difference between the means of two samples can be judged through either z-test or the t-test, but the difficulty arises when we happen to examine the significance of the difference amongst more than two sample means at the same time. The ANOVA technique enables us to perform this simultaneous test and as such is considered to be an important tool of analysis in the hands of a researcher. Using this technique, one can draw inferences about whether the samples have been drawn from populations having the same mean.

The ANOVA technique is important in the context of all those situations where we want to compare more than two populations such as in comparing the yield of crop from several varieties of seeds, the gasoline mileage of four automobiles, the

smoking habits of five groups of university students and so on. In such circumstances one generally does not want to consider all possible combinations of two populations at a time for that would require a great number of tests before we would be able to arrive at a decision. This would also consume lot of time and money, and even then certain relationships may be left unidentified (particularly the interaction effects). Therefore, one quite often utilizes the ANOVA technique and through it investigates the differences among the means of all the populations simultaneously.

Chi-square test

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as $\chi 2$ (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used. Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

Interpretation

Interpretation refers to the technique of drawing inference from the collected facts and explaining the significance of those inferences after an analytical and experimental study. It is a search for broader and more abstract means of the research findings. If the interpretation is not done very carefully, misleading conclusions may be drawn. The interpreter must be creative of ideas he should be free from bias and prejudice.

Fundamental principles of interpretation

1. Sound interpretation involves willingness on the part of the interpreter to see what is in the data.

2. Sound interpretation requires that the interpreter knows something more than the mere figures.

3. Sound interpretation demands logical thinking.

4. Clear and simple language is necessary for communicating the interpretation

Need for interpretation (importance of interpretation.)

1. It is through interpretation that the interpreter is able to know the abstract principles lying in his conclusions.

2. On the basis of the principles underlying his findings, a researcher can make various predictions about the various other events which are unrelated to his area of findings.

3. Interpretation leads to the establishment of explaining concepts.

4. A researcher can appreciate only through interpretation, why his findings are and what they are.

5. The interpretation of the findings of exploratory research study usually results in to hypothesis for experimental research.

Steps involved in the technique of interpretation

1. Researcher must give reasonable explanations of the relations he have found. He must be able to see uniformity in diversified research findings so that generalization of findings is possible.

2. If any extraneous information is collected during the study, it must be considered while interpreting the final result of research study.

3. The researcher can consult with those having insight in to the study who can point out the omission and errors in logical arguments.

4. The researcher must consider all relevant factors affecting the problem at the time of interpretation.

5. The conclusions appearing correct at the beginning may prove to be inaccurate later.

%%%%%%%%%%%%