SOCIAL STATISTICS

COURSE AIMS

The main aim of the course is to introduce you to descriptive statistics and give you the understanding of how to present and represent your data and to describe your observations in data form.

This will be achieved by aiming to:

• Introduce you to the nature and types of data

• Outline the importance of data and statistics

• Compare and describe large groups of data in a concise statistical language

• Predict statistical outcome from a number of statistical observations.

DEFINITION OF STATISTICS

Statistics can be defined as the numerical values or indicators computed through mathematical manipulation of the numerical data. This data can be presented in various form such as percentages, index numbers, averages, and medians.

•Encarta Microsoft Dictionary (2008)'a branch of mathematics that deals with analysis and interpretation of numerical data in terms of samples and populations.

•Gupta (1983) "classified facts respecting condition of the people in a state…..especially those facts which can be stated in numbers or in tables or in any other tabular or classified arrangement

Major steps in statistical analysis are:

i. data collection
ii. organizing and summarizing data
iii. analyzing and interpreting the result
iv. using the result to make rational decision

Purposes of statistics

• present a large amount of quantitative information

• in an organised way, go beyond a meaningless record of school test results to a meaningful interpretation

• predict how likely an event will occur

• make inferences from observations

•save us time and energy by condensing large amount of information concisely and conveniently in a table

•The evaluation of the quality of services available to a particular group or organisation,

•Analysing behaviours of groups of people in their environment and special situations,

•Determining the wants or needs of people through statistical sampling.

Role of statistics in social inquiry

Meaning of Social Inquiry

•It means an investigation to determine the facts of a case- a search conducted according to the statistical technique.
• inquiry could mean an investigation to determine the facts of a case.inquiry could mean an investigation to determine the facts of a case.

Roles of Statistics

1. Statistics simplifies complex mass of data and presents them in a comprehensive way that they are at once made easy to comprehend and interpret. Instead of having a large raw data, the data are prepared in percentages and means which can be grasped more easily than a mass of data.
2. Statistics presents data in more comprehensive and definite form: Statistics made conclusion that are stated numerically and more convincing than conclusions stated qualitatively.
3. Statistics interpret conditions that are more presentable: Statistics present conditions in an attractive ways such as pie chart and histogram or bar charts of the phenomenon under investigation.
4. It provides easy way of classifying numerical data: The method of classification in statistics provides the salient features of the variables that is under consideration
5. It provides an easy way of comparing data: Some data may be meaningless unless they are subjected to statistical analysis before they can be compared with similar data at other places.

Statistical research into a given problem could be classified as follows:

• data collection;
• organisation of data;
•data analysis and
• interpretation of facts

Determination of Statistical Unit

•for the interpretation of the results and the presentation of facts

SOME BASIC CONCEPTS

• **Data**- pieces of information that represent the qualitative or quantitative attributes of a variable or set of variables. Data are typically the results of measurements and can be the basis of graphs, images or observations of a set of variables.
• **Variable-** This is any quality that can have a number of values, which may be either discrete or continuous. A variable is a property that can take on different values. Individual in a class may differ in sex, age, intelligence, height etc
• **Quantitative Variables-** This type of variables assumes values that vary in terms of magnitude. Very easy to measure and compare with others e.g. weight, height, age, distance, marks obtained in a test etc.

• **Qualitative Variables-** This type of variable differs in kind. They are only categorized, e.g. gender, nationality, social economic status, academic qualifications, marital status.
• **Independent Variable-** These variables can be manipulated or treated. The effect is reflected on the dependent variable. The value of the dependent variable thus depends on that of the independent variable.
• **Discrete Variable**-This is a variable that can be counted, or for which there is a fixed set of values.
• **Continuous Variable**-This concept is characterized by being related to some numerical scale of measurement e.g. length, height, weight, temperature, volume and time
• **Distribution**-This is the arrangement of a set of numbers classified according to some properties or attributes such as age, height, weight, etc
• **Population**- This consists of the totality of the observations of a particular group.
• **Sample**- This is a subset of a population, a sub-group or sub-aggregate drawn from a population.
• **Parameter**- Any numerical value describing a characteristic of a population is known as parameter. It is a situation when mean (or Average), standard deviation or variance of a population are computed for statistical analysis.
• **Statistic**-This refers to a descriptive measure of a sample, i.e. a numerical value or function computed to describe a sample or population

The Scope of Inquiry

•The Space-a particular scope of the study

•The Time-a specific time lag of the study

■ **Functions of Statistics**

• **Compression**-Generally speaking by the word 'to compress', we mean to reduce or to condense
• **Evaluation**-The two main methods used in condensing data are classification and tabulation method. These help researchers to compare and contrast data collected from different sources
• **Forecasting**-By the word forecasting, we mean to predict, estimate or to project into the future.
• **Estimation**-One of the main objectives of statistics is drawing inference about a population from the analysis for the sample drawn from that population.

The 4 'intellectual foundations' in social inquiry:

• **Regularities**: There are certain discernible uniformities in the social behaviour of human beings that can be expressed in generalisations, as they are capable of explaining and predicting social phenomenon.
•**Verification**: All knowledge must be based on observation and verification- consist of the propositions that may be subjected to empirical investigations.
•**Techniques**: Correct techniques should be adopted for acquiring and interpreting data, use of research tools or methods, which generate valid reliable and comparative data.
•**Quantification**: Data should not only be collected, it should also be measured and quantified so that the conclusions of a researcher may be verified based on quantified evidence

Limitations of Statistics

1. **Neglect the Study of Qualitative Phenomena -**Since statistics is a science and deals with a set of numerical data, it is applicable to the study of only these subjects of enquiry, which can be expressed in terms of quantitative measurements.
2. **Neglect Individuality-** Statistics does not attach any specific importance to the individual items rather; it deals with aggregates of objects.
3. **Lack of Exactitude-** It is well known that mathematical and physical sciences are exact but statistical laws are not as exact but only approximations.
4. **Misuse of Records-**Statistics must be used only by experts otherwise statistical methods are the most dangerous tools in the hands of the inexperienced people. The use of statistical tools by the untrained persons might lead to wrong conclusions.

TYPES OF STATISTICS

■Statistics has three distinct parts namely:

• **Descriptive Statistics-** The event or outcome of events is described without drawing conclusions. It includes the mean, median, mode, standard deviation, range, percentile, kurtosis, correlation coefficient, proportions etc
• **Experimental Statistics -**Relates to the design of experiments to establishing causes and effects of such designs as experimental, Quasi-experiments
• **Inferential Statistics** -This is built on the descriptive statistics (chi-square, t-test, f-test etc) by going a step further to make interpretation with a view to population upon which a decision would be based.

Methods used in Social Statistics

Quantitative Methods
Qualitative methods

PRESENTATION OF DATA TECHNIQUES

Data presentation

Data may be presented in form of tables, graphical presentations or charts.

Tables

A table is an orderly arrangement of data in rows and columns, or possibly in a more complex structure. Tables are very effective medium for the organization and presentation of statistical information.

Uses of Tables

The following are the uses of tables:
1. To present data in a logically arranged and understandable form.
2. To identify peculiar features of the data as well as facilitate comparisons through row and column arrangements 51

3. To show the associations and patterns of relationships among variables
4. To facilitate speedy and easily decision taking since data are presented in understandable form

Characteristics of a Table

a) A table must be simple
b) A table must be easy to understand
c) A table must be numbered if they are more than one
d) A table must have a title
e) The sub headings for row and column must be stated
f) When a table contains secondary data, the source of the data must be shown
g) In a table, the unit of counting should be indicated
h) Items in a table must be arranged alphabetically, chronologically etc.

■Technique of Data Collection

• Literature review: This method involves going through published works by scholars- Secondary source of data collection.

• Survey: The Researcher may conduct special survey or inquiry on the subject-matter-use of interview or administering questionnaires, known as Primary source.

■Two main sources of data collection and the use of a particular source depend on the following considerations:

• the purpose of the inquiry;
• the time required;
• availability of fund;
• accuracy required and
• the nature or status of the researcher

■What is Primary Data?

• It is the statistical data or materials generated directly by the researcher for the purpose of the research

■What is Secondary Data?

• Data collected by investigator from the following printed materials published by another person or organisations are called Secondary data

TYPES OF VARIABLES

i) Independent Variables- It is a variable that influences the behavior of another variable. It represents inputs or causes. It is used to predict the value of another variable. It is thus a variable which we assign value and whose variation does not depend on that of another variable. It is denoted with the symbol X. it is the variable that a researcher can manipulate in a study. For example, income is an independent variable because it causes and influences another variable called consumption.

ii) Dependent variables- It is what is being measured in a research. The dependent variable is sometimes called the outcome variable and it is the variable being predicted. Thus, the value of the dependent variable depends and is determined by the independent variables and other factors: Y = f(X), i.e. Y depends on X

MEASUREMENT OF VARIABLES

Instruments for measuring variables are referred to as measurement scales or levels of measurement and since variables are of different kinds, there are several ways of measuring them. Of all types of measurement, four scales or levels of measurement are popularly used.

Nominal Scale

This is the lowest and simplest scale of measurement. Nominal scales simply sort the objects into categories and assign numbers as labels to identify objects or classes of objects. The assigned numbers have no meaning except as identifiers.

Ordinal Scale

Ordinal scales like the nominal scales sort variables into mutually exclusive categories, assigning numbers to objects and in addition, rank them. The ordinal scale thus categorizes, labels and/or ranks the variables. In the ordinal scale, note that the classes must be put into an order such that each case in one class is considered greater than (or less than) every case in another class.

With expected feedback based on customer satisfaction categorized into

- Very Unsatisfied – 1
- Unsatisfied – 2
- Neutral – 3
- Satisfied – 4
- Very Satisfied – 5

Interval scale

Interval scale contains all the properties of the ordinal scale, that is, an addition of the distance between any pair of adjacent categories is known. That is, it offers a calculation of the difference between variables. The interval scale therefore provides the information on how much one category is more or less than the other.

Ratio scale

This is the highest of the scales discussed so far. Ratio scales have the properties of all the three scales discussed above. In addition, it has an absolute “zero” point. For example, traffic density (measured in vehicles per kilometer) represents a ratio scale. The density of a link is defined as zero when there are no vehicles in a link.

DIAGRAMMATIC PRESENTATION OF DATA

•**Pictogram** -This entails the use of pictures to explain or illustrate statistical data. The pictures can be in form of; picture of men, cars, bottles, pots, houses etc.

•**Bar Chart** -This is a type of diagram in which information given are represented with bars. The length of each bar is associated or corresponds to the value given and the width of each bar is equal to the other.

•**Pie Chart** -Another method of explaining data through diagram (graph, degree, circle etc).

•Histogram is a graph in which class boundaries or class interval is marked on the horizontal axis and the corresponding class frequency on the vertical axis

Range

The range specifies the distance between the highest and the lowest value in a given distributing. It is the difference between the highest value and the lowest value in a data set. Symbolically, the range is summarized as :

Range = H – L

Where H = Highest value

L = Lowest value

Class Interval

Class interval is a set of classes that are used to define the raw data or size of the group chosen. It can be determined by finding the range of the raw data obtained and dividing it by the value of number of classes we desire to have. Using above table of grouped frequency distribution, class interval is regarded as the scores which are: 20-29, 30-39.....70-79.

Class Limit

These are the end numbers of class interval. The lower value for class interval is called lower class limit while the upper value for class interval is called upper class limit. Table can also be used to explain this. For class interval 20-29, 20 is the lower class limit and 29 is the upper class limit. For class interval 30-39, 30 is the lower class limit and 39 is upper class limit etc.

Class Boundaries Class boundaries are easily gotten by subtracting 0.5 from lower class limit or lower value of class interval and adding 0.5 to upper class limit or upper value of class interval. For example, class boundaries for Table 3.2 are 19.5 - 29.5, 29.5 - 39.5 .....69.5 - 79.5

Class Mark

This is the mid-point or value of the class interval. It can be derived by adding lower and upper class limit and dividing by two (2) or adding lower and upper class boundaries and dividing the sum by two (2). Referring to table 3.2 that we have been using, class mark for the first class interval is

20+29/2=49/2=24.5

FREQUENCY DISTRIBUTION

•when dealing with a large form of quantitative data

•Ungrouped Frequency Distribution-a type of frequency distribution in which data are not compressed together in a particular interval.

•Given a set of raw data below, construct a frequency table for it.

2, 3, 2, 1, 4, 4, 1, 1, 3, 2, 5, 3, 5, 1, 5, 1, 4, 2, 3, 0, 4, 1, 0, 3, 5, 2, 5, 0, 4, 1, 4, 0, 5, 2, 3, 0, 4, 4, 1, 2.

•Grouped Frequency Distribution -this grouping may be of equal or unequal interval and each of the group is given as an interval and it is called class interval.

•Class interval is a set of classes that are used to define the raw data or size of the group chosen.

•Class Limit-these are the end numbers of class interval. The lower value for class interval is called lower class limit while the upper value for class interval is called upper class limit.

•Class Boundaries are easily gotten by subtracting 0.5 from lower class limit or lower value of class interval and adding 0.5 to upper class limit or upper value of class interval.

•Class Mark -this is the mid-point or value of the class interval.

•Class Width -this is the size of the class interval and it is obtained by subtracting lower class boundaries from upper class boundary.

NATURE OF MEASUREMENT

It shows that all observations whether it is qualitative or quantitative must be measured to some degree of accuracy.

Errors in measurement

Inaccuracy in the measurement instrument

Changing circumstance of what is measured

The limitations in human judgement

Scales of measurement

It refers to the assignment of numbers to object and events according to logical acceptable rules.

It is essential to know the different kinds of measurement scales, as the number of properties applicable depends upon the measurement scale applied to the objects or events.

MEASURES OF CENTRAL TENDENCY

The Arithmetic Mean

The arithmetic mean (or just mean) is the most important measure of central tendency-all members of the set are used in the calculation of the mean-

$$X = X1 + x2 + x3 + x4 \ldots Xn$$

$$= X1 + x2 + x3 + x4 \ldots Xn/n$$

Arithmetic Mean of Group Data-the values between any class interval are considered as condensed at the mid- point of the class interval or class mark -the class mark of the ith class interval,

The Median-data consisting of "n" members are arranged in order of magnitude, the middle values or member is called THE MEDIAN

The Mode-The mode of a distribution is the value, which occurs most frequently in the distribution.

STANDARD DEVIATION AND OTHER MEASURES OF DISPERSION,

Range-It is the difference between the largest and the smallest values of a data set.

Mean Deviation-It is the mean of the deviations or differences of the scores from either the mean, median or mode- commonly used of the measure of central tendency for computing mean deviation is the median followed by the mean.

Standard deviation is an improvement of mean deviation. Two extra stages are added to the calculation of mean deviation-the most desirable measure of dispersion called standard deviation

Calculation of the Standard Deviation from an Assumed Mean-

The only difference from the mean calculation is that, a guess of an assumed mean is made.