UNIT 1 INTRODUCTION TO STATISTICS

Structure

1.0 INTRODUCTION

The word statistics has different meaning to different persons. Knowledge of statistics is applicable in day to day life in different ways. In daily life it means general calculation of items, in railway statistics means the number of trains operating, number of passenger's freight etc. and so on. Thus statistics is used by people to take decision about the problems on the basis of different type of *quantitative and qualitative* information available to them.

However, in behavioural sciences, the word 'statistics' means something different from the common concern of it. Prime function of statistic is to draw statistical inference about population on the basis of available quantitative information. Overall, statistical methods deal with reduction of data to convenient descriptive terms and drawing some inferences from them. This unit focuses on the above aspects of statistics.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Define the term statistics;
- Explain the status of statistics;
- Describe the nature of statistics;
- State basic concepts used in statistics; and
- Analyse the uses and misuses of statistics.

1.2 MEANING OF STATISTICS

The word statistics has been derived from Latin word 'status' or Italian 'Statista' meaning statesman. Professor Gott Fried Achenwall used it in the 18th century. During early period, these words were used for political state of the region. The word 'Statista' was used to keep the records of census or data related to wealth of a state. Gradually, its meaning and usage extended and thereonwards its nature also changed.

The word statistics is used to convey different meanings in singular and plural sense. Therefore it can be defined in two different ways.

1.2.1 Statistics in Singular Sense

In singular sense, 'Statistics' refers to what is called statistical methods. It deals with the collection of data, their classification, analysis and interpretations of statistical data. Therefore, it is described as a branch of science which deals with classification, tabulation and analysis of numerical facts and make decision as well. Every statistical inquiry should pass through these stages.

1.2.2 Statistics in Plural Sense

'Statistics' used in plural sense means that quantitative information is available called 'data'. For example, information on population or demographic features, enrolment of students in Psychology programmes of IGNOU, and the like. According to Websters "Statistics are the classified facts representing the conditions of the people in a State specifically those facts which can be stated in number or in tables of number or classified arrangement".

Horace Secrist describes statistics in plural sense as follows : " By Statistics we mean aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standard of accuracy , collected in a systematic manner for a pre-determined purpose and placed in relation to each other." Thus Secrist's definition highlights following features of statistics:

i) *Statistics are aggregate of facts:* Single or unrelated items are not considered as statistics.

ii) *Statistics are affected by multiplicity of causes:* In statistics the collected information are greatly influenced by a number of factors and forces working together.

iii) *Statistics are numerical facts:* Only numerical data constitute statistics.

iv) *Statistics are enumerated or estimated with a reasonable standard of accuracy:* While enumerating or estimating data, a reasonable degree of accuracy must be achieved.

v) *Statistics are collected in a systematic manner:* Data should be collected by proper planning by utilising tool/s developed by trained personnel.

vi) *Statistics are collected for a predetermined purpose :* It is necessary to define the objective of enquiry, before collecting the statistics. The objective of enquiry must be specific and well defined.

vii) *Statistics should be comparable:* Only comparable data will have some meaning. For statistical analysis, the data should be comparable with respect to time, place group, etc.

Thus, it may be stated that " All statistics are numerical statements of facts but all numerical statements of facts are not necessarily statistics ".

1.2.3 Definition of Statistics

In this unit emphasis is on the term statistics as a branch of science. It deals with classification, tabulation and analysis of numerical facts. Different statistician defined this aspect of statistics in different ways. For example.

A. L. Bowley gave several definitions of Statistics:

i) "Statistics may be called the science of counting" . This definition emphasises enumeration aspect only.

ii) In another definition he describes it as " Statistics may rightly be called the science of average".

iii) At another place Statistics is defined as, "Statistics is the science of measurement of social organism regarded as a whole in all its manifestations".

All three definitions given by Bowely seem to be inadequate because these do not include all aspects of statistics.

According to **Selligman** "Statistics is the science which deals with the methods of collecting, classifying, presenting , comparing and interpreting numerical data collected to throw some light on any sphere of enquiry".

Croxton and Cowden defined "statistics as the collection , presentation, analysis ,and interpretation of numerical data".

Among all the definitions , the one given by Croxton and Cowden is considered to be most appropriate as it covers all aspects and field of statistics.

These aspects are given below:

Collection of Data : Once the nature of study is decided , it becomes essential to collect information in form of data about the issues of the study. Therefore, the collection of data is the first basic step. Data may be collected either from primary source or secondary or from both the sources depending upon the objective/s of the investigation

Classification and Presentation : Once data are collected , researcher has to arrange them in a format from which they would be able to draw some conclusions. The arrangement of data in groups according to some similarities is known as classification.

Tabulation is the process of presenting the classified data in the form of table. A tabular presentation of data becomes more intelligible and fit for further statistical analysis. Classified and Tabulated data can be presented in diagrams and graphs to facilitate the understanding of various trends as well as the process of comparison of various situations.

Analysis of Data : It is the most important step in any statistical enquiry . Statistical analysis is carried out to process the observed data and transform it in such a manner as to make it suitable for decision making.

Interpretation of Data : After analysing the data, researcher gets information partly or wholly about the population. Explanation of such information is more useful in real life. The quality of interpretation depends more and more on the experience and insight of the researcher.

Self Assessment Questions

1) Complete the following statements

   i) The word statistics has been derived from Latin word .....

   ii) Statistics in plural means .....

   iii) Statistics in singular means .....

   iv) The first step in statistics is .....

   v) The last step in statistics is .....

2) Tick (√) the correct answer

   Statistical data are:

   i) Aggregates of facts

   ii) Unsystematic data

   iii) Single or isolated facts or figure

   iv) None of these

3) Which one of the following statement is true for statistics in singular sense?

   i) Statistics are aggregate of facts.

   ii) Statistics are numerical facts.

   iii) Statistics are collected in a systematic manner.

   iv) Statistics may be called the science of counting.

1.3 TYPES OF STATISTICS

After knowing the concept and definition of statistics, let us know the various types of statistics.

Though various bases have been adopted to classify statistics, following are the two major ways of classifying statistics: (i) on the basis of function and (ii) on the basis of distribution.

1.3.1 On the Basis of Functions

As statistics has some particular procedures to deal with its subject matter or data, three types of statistics have been described.

A) **Descriptive statistics:** The branch which deals with descriptions of obtained data is known as descriptive statistics. On the basis of these descriptions a particular group of population is defined for corresponding characteristics. The descriptive statistics include classification, tabulation measures of central tendency and variability. These measures enable the researchers to know about the tendency of data or the scores, which further enhance the ease in description of the phenomena.

B) **Correlational statistics:** The obtained data are disclosed for their inter correlations in this type of statistics. It includes various types of techniques to compute the correlations among data. Correlational statistics also provide description about sample or population for their further analyses to explore the significance of their differences.

C) **Inferential statistics:** Inferential statistics deals with the drawing of conclusions about large group of individuals (population) on the basis of observations of few participants from them or about the events which are yet to occur on the basis of past events. It provide tools to compute the probabilities of future behaviour of the subjects.

1.3.2 On the Basis of Distribution of Data

Parametric and nonparametric statistics are the two classifications on the basis of distribution of data. Both are also concerned to population or sample. By population we mean the total number of items in a sphere. In general it has infinite number therein but in statistics there is a finite number of a population, like the number of students in a college. According to Kerlinger (1968) "the term population and universe mean all the members of any well-defined class of people, events or objects." In a broad sense, statistical population may have three kinds of properties – (a) containing finite number of items and knowable, (b) having finite number of articles but unknowable, and (c) keeping infinite number of articles.

Sample is known as a part from population which represents that particular population's properties. As much as the sample selection will be unbiased and random, it will be more representing its population. "Sample is a part of a population selected (usually according to some procedure and with some purpose in mind) such that it is considered to be representative of the population as a whole".

Parametric statistics is defined to have an assumption of normal distribution for its population under study. "Parametric statistics refers to those statistical techniques that have been developed on the assumption that the data are of a certain type. In particular the measure should be an interval scale and the scores should be drawn from a normal distribution".

There are certain basic assumptions of parametric statistics. The very first characteristic of parametric statistics is that it moves after confirming its population's property of **normal distribution**. The normal distribution of a population shows its symmetrical spread over the continuum of –3 SD to +3 SD and keeping unimodal shape as its mean, median, and mode coincide. If the samples are from various populations then it is assumed to have same variance ratio among them. The samples are independent in their selection. The chances of occurrence of any event or item out of the total population are equal and any item can be selected in the sample. This reflects the randomized nature of sample which also happens to be a good tool to avoid any experimenter bias.

In view of the above assumptions, parametric statistics seem to be more reliable and authentic as compared to the nonparametric statistics. These statistics are more powerful to establish the statistical significance of effects and differences among variables. It is more appropriate and reliable to use parametric statistics

in case of large samples as it consist of more accuracy of results. The data to be analysed under parametric statistics are usually from interval scale.

However, along with many advantages, some disadvantages have also been noted for the parametric statistics. It is bound to follow the rigid assumption of normal distribution and further it narrows the scope of its usage. In case of small sample, normal distribution cannot be attained and thus parametric statistics cannot be used. Further, computation in parametric statistics is lengthy and complex because of large samples and numerical calculations. T-test, F-test, r-test, are some of the major parametric statistics used for data analysis.

Nonparametric statistics are those statistics which are not based on the assumption of normal distribution of population. Therefore, these are also known as distribution free statistics. They are not bound to be used with interval scale data or normally distributed data. The data with non-continuity are to be tackled with these statistics. In the samples where it is difficult to maintain the assumption of normal distribution, nonparametric statistics are used for analysis. The samples with small number of items are treated with nonparametric statistics because of the absence of normal distribution. It can be used even for nominal data along with the ordinal data. Some of the usual nonparametric statistics include chi-square, Spearman's rank difference method of correlation, Kendall's rank difference method, Mann-Whitney U test, etc.

Self Assessment Questions

1) State true/false for the following statements

   i) Parametric statistics is known as distribution free statistics (T/ F)

   ii) Nonparametric tests assume normality of distribution (T/F)

   iii) T test is an example of parametric test (T/F)

   iv) Nonparametric tests are not bound to be used with interval scale. (T/F)

   v) Parametric tests are bound to be used with either interval or ratio scale. (T/F)

   vi) In case of small sample where normal distribution can not be attained, the use of nonparametric test is more appropriate. (T/F)

2) Define the term sample and population with one example each.

   .....

   .....

   .....

   .....

   .....

   .....

   .....

1.4 SCOPE AND USE OF STATISTICS

Statistical applications have a wide scope. Some of the major ones are given below:

Policy planning: To finalise a policy, it requires some data from previous or expected environment that the policy can be effectively utilised with maximum favourable results. For example, in an organisation the previous sales data are analysed to develop future strategies in the field to obtain maximum benefit in terms of product sale.

Management: Statistics is very useful tool in an organisation to view various aspects of work and well being of the employees as well as keeping an eye on the progress trend of the organisation.

Behavioural and Social Sciences: In social sciences where both types (quantitative and qualitative) of information are used, statistics helps the researchers to alter the information in a comprehensive way to explain and predict the patterns of behaviour/ trend. Where the characteristics of the population being studied are normally distributed, the best and statistically important decision about variables being investigated is possible by using parametric statistics or nonparametric statistics to explain the pattern of activities.

Education: If education is intended to be well dispersed and effective in the interest of the population, the characteristics of students, instructor's contents and infrastructure are very important to understand and again statistics enable these characteristics being analysed in context of needs of the nation. Once the parameters of all components are analysed, areas needing more emphasis become obvious.

Commerce and Accounts: Where money matters are involved, it is essential to take extra care to manage the funds properly enabling efforts in various sectors. The cost and benefit analysis helps to decide putting money and regulating it for maximum benefit at minimum cost.

Industries: Statistics is a basic tool to handle daily matters not only in big organisations but also in small industries. It is required, at each level, to keep data with care and look at them in different perspectives to mitigate the expenditure and enable each employee to have his/ her share in the benefit. Psychologists/ personnel officers dealing with selection and training in industries also use statistical tools to differentiate among employees.

Pure sciences and Mathematics: Statistical tools are also instrumental to have precise measures in pure sciences and to see differences on different occasions in various conditions. Statistics itself is a branch of mathematics which helps them understand differences among properties of various applications in mathematics.

Problem solving: Knowing the useful difference between two or more variables enable the individual to find out the best applicable solution to a problem situation and it is possible because of statistics. During problem solving statistics helps the person analyse his/ her pattern of response and the correct solution thereby minimising the error factor.

Theoretical researches: Theories evolve on the basis of facts obtained from the field. Statistical analyses establish the significance of those facts for a particular paradigm or phenomena. Researchers are engaged in using the statistical measures to decide on the facts and data whether a particular theory can be maintained or challenged. The significance between the facts and factors help them to explore the connectivity among them.

1.5 LIMITATIONS OF STATISTICS

Although Statistics has a very wide application in everyday life as well as in Behavioural Sciences, Physical and Natural Sciences, it has certain limitations also. These limitations are as follow :

Statistics deals with aggregate of facts. It cannot deal with single observation. Thus statistical methods do not give any recognition to an object or a person or an event in isolation. This is a serious limitation of Statistics.

Since Statistics is a science dealing with numerical data, it is more applicable to those phenomenon which can be measured quantitatively. However, the techniques of statistical analysis can be applied to qualitative phenomenon indirectly by expressing them numerically with the help of quantitative standards.

Statistical conclusions are true only on the average . Thus, statistical inferences may not be considered as exact like inferences based on Mathematical laws.

1.6 DISTRUST AND MISUSE OF STATISTICS

Sometimes irresponsible, inexperienced people use statistical tools to fulfill their self motives irrespective of the nature and trend of the data. Because of such various misuses of statistical tools sometimes called an unscrupulous science. There are various misgivings about Statistics . These are as follows :

“Statistics can prove anything”

“Statistics is an unreliable science”

“There are three types of lies , namely, lies, damned lies, and statistics.”

“An ounce of truth will produce tons of Statistics “

Therefore care and precautions should be taken care for the interpretation of statistical data. “ Statistics should not be used as a blind man uses a lamp-post for support instead of illumination”

There are many other fields like, agriculture, space, medicine, geology, technology, etc. where statistics is extensively used to predict the results and find out precision in decision.

Self Assessment Question

1) Write three application of statistics in daily life.

.....

.....

.....

2) List atleast two misuses of statistics.

.....

.....

.....

.....

.....

.....

.....

.....

1.7 LET US SUM UP

In present era people must have some knowledge of statistics. In singular sense, it means statistical methods which include collection, classification, analysis and interpretation of data. In plural sense, it means quantitative information called data. Descriptive, correlational and inferential statistics are three different type of statistics on the basis of their functions. On the other hand, parametric and non parametric are other types of statistics on the basis of the nature of distribution. Statistics has application in almost in all branches of knowledge as well as all sphere of life. Inspite of its wide applicability, it has certain limitations too. Some times inexperienced people misuse statistics to fulfill their own motives.

1.8 UNIT END QUESTIONS

1) What do you mean by statistics? Define its various types with the help of examples of daily life.
2) "Statistical methods are most dangerous tools in the hand of inexpert." Discuss briefly
3) Define following concepts:
   i) Descriptive statistics
   ii) Inferential statistics
   iii) Parametric statistics
   iv) Non parametric statistics
4) Comments on the following statements in two or three lines with reasons:
   i) Statistics in singular sense implies statistical methods.
   ii) Statistics and statistic implies same thing.
   iii) Statistics may rightly be called the science of averages.
   iv) There are lies, damn lies and statistics. Give three examples of misuse of statistics.
5) Write a note on the limitations of statistics.

1.9 GLOSSARY

Statistics in singular sense : In singular sense, it means scientific methods for collection, presentation, analysis and interpretation of data.

Statistics in plural sense : In plural sense it means a set of numerical scores known as statistical data.

Correlational statistics : The statistics which speaks about one or more than one variable's positive or negative magnitude of relationship.

Descriptive statistics : The statistics which describes the tendency or variance of the scores in a distribution.

Inferential statistics : The statistics that enable the researchers to have some conclusions about population or events on the basis of past or observed observations.

Non parametric statistics : The statistics free from the assumptions of normal distribution.

Parametric statistics : The statistics based on assumptions of normal distribution

Statistics : The branch of mathematics that deals with inferring the chances of a particular pattern of population or events on the basis of observed patterns..

1.10 SUGGESTED READINGS

Asthana H.S, and Bhushan, B.(2007) *Statistics for Social Sciences* (with SPSS Applications). Prentice Hall of India

B.L.Aggrawal (2009). *Basic Statistics*. New Age International Publisher, Delhi.

Gupta, S.C.(1990) *Fundamentals of Statistics*. Himalaya Publishing House, Mumbai

What is Data?

Data can be defined as a systematic record of a particular quantity. It is the different values of that quantity represented together in a set. It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis. When arranged in an organized form, can be called information. The source of data (primary data, secondary data) is also an important factor.

Types of Data

Data may be qualitative or quantitative. Once you know the difference between them, you can know how to use them.

- **Qualitative Data:** They represent some characteristics or attributes. They depict descriptions that may be observed but cannot be computed or calculated. For example, data on attributes such as intelligence, honesty, wisdom, cleanliness, and creativity collected using the students of your class a sample would be classified as qualitative. They are more exploratory than conclusive in nature.
- **Quantitative Data:** These can be measured and not simply observed. They can be numerically represented and calculations can be performed on them. For example, data on the number of students playing different sports from your class gives an estimate of how many of the total students play which sport. This information is numerical and can be classified as quantitative.

Source: Wiki Images

Data Collection

Depending on the source, it can classify as primary data or secondary data. Let us take a look at them both.

Primary Data

These are the data that are ***collected for the first time*** by an investigator for a specific purpose. Primary data are 'pure' in the sense that no statistical operations have been performed on them and they are original. An example of primary data is the **Census of India**.

Secondary Data

They are the data that are ***sourced from someplace*** that has originally collected it. This means that this kind of data has already been collected by some researchers or investigators in the past and is available either in published or unpublished form. This information is impure as statistical operations may have been performed on them already. An example is an information available on the **Government of India**, the Department of Finance's website or in other repositories, books, journals, etc.

Learn how to find the Mode of the data and its relationship with the Median and Mean here.

Discrete and Continuous Data

Discrete Data: These are data that can take only certain specific values rather than a range of values. For example, data on the blood group of a certain population or on their genders is termed as discrete data. A usual way to represent this is by using bar charts.

Continuous Data: These are data that can take values between a certain range with the highest and lowest values. The difference between the highest and lowest value is called the range of data. For example, the age of persons can take values even in decimals or so is the case of the height and weights of the students of your school. These are classified as continuous data. Continuous data can be tabulated in what is called a **frequency distribution**. They can be graphically represented using histograms.

Example

Question 1: A dataset is as follows: {3, 8, 7, 4, 6, 1, 5}. Calculate the range of this set.

Answer: Range is a statistical measure that estimates the difference between the highest and lowest values from a dataset. Here, range = 8-1 = 7.

Question 2: What is data?

Answer: Data refers to a systematic record of a specific quantity. It is the diverse values of that quantity together which the sets represent. In other words, it is a set of facts and figures which are useful in a particular purpose like a survey or an analysis. Moreover, when you arrange them in an organized form, they refer to as information. Moreover, the source of data primary or secondary is also an essential factor.

Question 3: What are the two types of data?

Answer: Data can be classified as qualitative and quantitative. Qualitative data represents some characteristics or attributes. Further, they represent descriptions which we may observe

that we cannot compute or calculate. For instance, data on attributes like intelligence, creativity, honesty, cleanliness, wisdom, and more are examples of qualitative data. In other words, they are more exploratory than conclusive.

Further, there is quantitative data which we can measure and not just observe. You can represent them numerically and even perform calculations. For instance, the number of students in basketball will be numerical, thus, it is quantitative data.

Question 4: What is Primary Data?

Answer: Primary data is one which an investigator collects for the first time for a particular purpose. Further, this data is 'pure' in the sense that there haven't been any statistical operations performed on them, plus they are also original. The Census of India is an example of primary data.

Question 5: What s discrete data?

Answer: Discrete data is one which can take into consideration only certain specific values instead of a range of values. For instance, data which is on the blood group of a specific population or on their genders is known as discrete data. Also, bar charts are a common way to represent this data.

DATA

Facts or figures, which are numerical or otherwise, collected with a definite purpose are called data.

Types Of Data

Quantitative Data	Qualitative Data
These represent numerical value.	These represent some characteristics or attributes.
These can be numerically computed.	These depict descriptions that may be observed but cannot be computed.

Primary Data	Secondary Data
Data collected for first time.	Data that is sourced by someone other than the user.

Discrete Data	Continuous Data
These are the data that can take only specific value.	These are the data that can take values from a given range.

Frequency Distribution Table — A list, table or graph that displays the frequency of various outcomes in a sample of data.

Frequency Distribution Table

Ungrouped

It is used for small data set. For eg.

Marks Obtained	Frequency
16	3
17	4
18	8
19	10
20	12
21	6
22	3

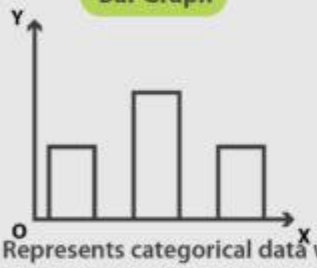
Grouped

It is used for large data set. For eg.

Class Interval	Frequency
0-5	3
5-10	11
10-16	14
15-20	2

Graphical Representation of Frequency Distribution Table

Bar Graph

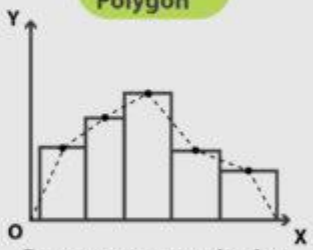


Represents categorical data with rectangular bars whose height is proportional to frequency.

Histogram

Represents continuous data with no gap between bars.

Frequency Polygon



Represents quantitative data using line graph.

Mean for Ungrouped Data

Let the data set be $x_1, x_2, x_3, \ldots, x_n$

$$\text{mean} = \frac{x_1+x_2+x_3+\ldots+x_n}{n}$$

Mean for Grouped Data

(1) Direct Method

$$\text{mean} = \frac{\sum x_i f_i}{\sum f_i}$$

Where

x_i = Corresponding class mark

f_i = Corresponding frequency

(2) Assumed mean method

$$\text{mean} = a + \frac{\sum d_i f_i}{\sum f_i}$$

Where

a = Assumed mean for the given data

d_i = deviation = $x_i - a$

x_i = Corresponding class mark

f_i = Corresponding frequency

(3) Step Deviation method

$$\text{mean} = a + \frac{\sum f_i u_i}{\sum f_i} \times h$$

Where

a = Assumed mean for the given data

$u_i = \frac{x_i - a}{h}$

h = Class width

x_i = Corresponding class mark

f_i = Corresponding frequency

Quantitative Data: Definition, Types, Analysis and Examples

Content Index

Quantitative Data: Definition

Quantitative data is defined as the value of data in the form of counts or numbers where each data-set has an unique numerical value associated with it. This data is any quantifiable information that can be used for mathematical calculations and statistical analysis, such that real-life decisions can be made based on these mathematical derivations. Quantitative data is used to answer questions such as “How many?”, “How often?”, “How much?”. This data can be verified and can also be conveniently evaluated using mathematical techniques.

For example, there are quantities corresponding to various parameters, for instance, “How much did that laptop cost?” is a question which will collect quantitative data. There are values associated with most measuring parameters such as pounds or kilograms for weight, dollars for cost etc.

Quantitative data makes measuring various parameters controllable due to the ease of mathematical derivations they come with. Quantitative data is usually collected for statistical analysis
using surveys, polls or questionnaires sent across to a specific section of a population. The retrieved results can be established across a population.

Types of Quantitative Data with Examples

The most common types of quantitative data are as below:

- **Counter:** Count equated with entities. For example, the number of people who download a particular application from the App Store.
- **Measurement of physical objects:** Calculating measurement of any physical thing. For example, the HR executive carefully measures the size of each cubicle assigned to the newly joined employees.
- **Sensory calculation:** Mechanism to naturally “sense” the measured parameters to create a constant source of information. For example, a digital camera converts electromagnetic information to a string of numerical data.
- **Projection of data:** Future projection of data can be done using algorithms and other mathematical analysis tools. For example, a marketer will predict an increase in the sales after launching a new product with thorough analysis.
- **Quantification of qualitative entities:** Identify numbers to qualitative information. For example, asking respondents of an online survey to share the likelihood of recommendation on a scale of 0-10.

Quantitative Data: Collection Methods

As quantitative data is in the form of numbers, mathematical and statistical analysis of these numbers can lead to establishing some conclusive results.

There are two main Quantitative Data Collection Methods:

Surveys: Traditionally, surveys were conducted using paper-based methods and have gradually evolved into online mediums. Closed-ended questions form a major part of these surveys as they are more effective in

collecting quantitative data. The survey makes include answer options which they think are the most appropriate for a particular question. Surveys are integral in collecting feedback from an audience which is larger than the conventional size. A critical factor about surveys is that the responses collected should be such that they can be generalized to the entire population without significant discrepancies. On the basis of the time involved in completing surveys, they are classified into the following –

- **Longitudinal Studies:** A type of observational research in which the market researcher conducts surveys from a specific time period to another, i.e., over a considerable course of time, is called longitudinal survey. This survey is often implemented for trend analysis or studies where the primary objective is to collect and analyze a pattern in data.
- **Cross-sectional Studies:** A type of observational research in which the market research conducts surveys at a particular time period across the target sample is known as cross-sectional survey. This survey type implements a questionnaire to understand a specific subject from the sample at a definite time period.

Learn more: Cross-sectional vs Longitudinal Survey

To administer a survey to collect quantitative data, the below principles are to be followed.

- **Fundamental Levels of Measurement – Nominal, Ordinal, Interval and Ratio Scales:** There are four measurement scales which are fundamental to creating a multiple-choice question in a survey in collecting quantitative data. They are, nominal, ordinal, interval and ratio measurement scales without the fundamentals of which, no multiple choice questions can be created.
- **Use of Different Question Types:** To collect quantitative data, close-ended questions have to be used in a survey. They can be a mix of multiple question types including multiple-choice questions like semantic differential scale questions, rating scale questions etc. that can help collect data that can be analyzed and made sense of.
- **Survey Distribution and Survey Data Collection:** In the above, we have seen the process of building a survey along with the survey design to collect quantitative data. Survey distribution to collect data is the other important aspect of the survey process. There are different ways of survey distribution. Some of the most commonly used methods are:
  -
    - **Email:** Sending a survey via email is the most commonly used and most effective methods of survey distribution. You can use

the Question Pro email management feature to send out and collect survey responses.

- **Buy respondents:** Another effective way to distribute a survey and collect quantitative data is to use a sample. Since the respondents are knowledgeable and also are open to participating in research studies, the responses are much higher.
- **Embed survey in a website:** Embedding a survey in a website increases a high number of responses as the respondent is already in close proximity to the brand when the survey pops up.
- **Social distribution:** Using social media to distribute the survey aids in collecting higher number of responses from the people that are aware of the brand.
- **QR code:** QuestionPro QR codes store the URL for the survey. You can print/publish this code in magazines, on signs, business cards, or on just about any object/medium.
- **SMS survey:** A quick and time effective way of conducting a survey to collect a high number of responses is the SMS survey.
- **QuestionPro app:** The Question Pro App allows to quickly circulate surveys and the responses can be collected both online and offline.
- **API integration:** You can use the API integration of the QuestionPro platform for potential respondents to take your survey.

One-on-one Interviews: This quantitative data collection method was also traditionally conducted face-to-face but has shifted to telephonic and online platforms. Interviews offer a marketer the opportunity to gather extensive data from the participants. Quantitative interviews are immensely structured and play a key role in collecting information. There are three major sections of these online interviews:

- **Face-to-Face Interviews:** An interviewer can prepare a list of important interview questions in addition to the already asked survey questions. This way, interviewees provide exhaustive details about the topic under discussion. An interviewer can manage to bond with the interviewee on a personal level which will help him/her to collect more details about the topic due to which the responses also improve. Interviewers can also ask for an explanation from the interviewees about unclear answers.
-
- **Online/Telephonic Interviews:** Telephone-based interviews are no more a novelty but these quantitative interviews have also moved to online mediums such as Skype or Zoom. Irrespective of the distance between the interviewer and the interviewee and their corresponding time zones, communication becomes one-click away with online interviews. In case of telephone interviews, the interview is merely a phone call away.
-

- **Computer Assisted Personal Interview:** This is a one-on-one interview technique where the interviewer enters all the collected data directly into a laptop or any other similar device. The processing time is reduced and also the interviewers don't have to carry physical questionnaires and merely enter the answers in the laptop.

All of the above quantitative data collection methods can be achieved by using surveys, questionnaires and polls.

Learn about: Quantitative Research

Quantitative Data: Analysis Methods

Data collection forms a major part of the research process. This data however has to be analyzed to make sense of. There are multiple methods of analyzing quantitative data collected in surveys. They are:

- **Cross-tabulation:** Cross-tabulation is the most widely used quantitative data analysis methods. It is a preferred method since it uses a basic tabular form to draw inferences between different data-sets in the research study. It contains data that is mutually exclusive or have some connection with each other.
- **Trend analysis:** Trend analysis is a statistical analysis method that provides the ability to look at quantitative data that has been collected over a long period of time. This data analysis method helps collect feedback about data changes over time and if aims to understand the change in variables considering one variable remains unchanged.
- **MaxDiff analysis:** The MaxDiff analysis is a quantitative data analysis method that is used to gauge customer preferences for a purchase and what parameters rank higher than the others in this process. In a simplistic form, this method is also called the "best-worst" method. This method is very similar to conjoint analysis but is much easier to implement and can be interchangeably used.
- **Conjoint analysis:** Like in the above method, conjoint analysis is a similar quantitative data analysis method that analyzes parameters behind a purchasing decision. This method possesses the ability to collect and analyze advanced metrics which provide an in-depth insight into purchasing decisions as well as the parameters that rank the most important.
- **TURF analysis:** TURF analysis or Total Unduplicated Reach and Frequency Analysis, is a quantitative data analysis methodology that assesses the total market reach of a product or service or a mix of both. This method is used by organizations to understand the frequency and the avenues at which their messaging reaches customers and

prospective customers which helps them tweak their go-to-market strategies.

- **Gap analysis:** Gap analysis uses a side-by-side matrix to depict quantitative data that helps measure the difference between expected performance and actual performance. This data analysis helps measure gaps in performance and the things that are required to be done to bridge this gap.
- **SWOT analysis:** SWOT analysis, is a quantitative data analysis methods that assigns numerical values to indicate strength, weaknesses, opportunities and threats of an organization or product or service which in turn provides a holistic picture about competition. This method helps to create effective business strategies.
- **Text analysis:** Text analysis is an advanced statistical method where intelligent tools make sense of and quantify or fashion qualitative and open-ended data into easily understandable data. This method is used when the raw survey data is unstructured but has to be brought into a structure that makes sense.

Learn More: MaxDiff Analysis vs Conjoint Analysis

Steps to conduct Quantitative Data Analysis

For Quantitative Data, raw information has to presented in meaningful manner using analysis methods. Quantitative data should be analyzed in order to find evidential data that would help in the research process.

- **Relate measurement scales with variables:** Associate measurement scales such as Nominal, Ordinal, Interval and Ratio with the variables.This step is important to arrange the data in proper order. Data can be entered into an excel sheet to organize it in a specific format.
- **Connect descriptive statistics with data:** Link descriptive statistics to encapsulate available data. It can be difficult to establish a pattern in the raw data. Some widely used descriptive statistics are:

1.
    1. Mean- An average of values for a specific variable
    2. Median- A midpoint of the value scale for a variable
    3. Mode- For a variable, the most common value
    4. Frequency- Number of times a particular value is observed in the scale
    5. Minimum and Maximum Values- Lowest and highest values for a scale
    6. Percentages- Format to express scores and set of values for variables

- **Decide a measurement scale:** It is important to decide the measurement scale to conclude a descriptive statistics for the variable. For instance, a nominal variable score will never have a mean or median and so the descriptive statistics will correspondingly vary. Descriptive statistics suffice in situations where the results are not to be generalized to the population.
- **Select appropriate tables to represent data and analyze collected data:** After deciding on a suitable measurement scale, researchers can use a tabular format to represent data. This data can be analyzed using various techniques such as Cross-tabulation or TURF.

Quantitative Data Examples

Listed below are some examples of quantitative data that can help understand exactly what this pertains:

- I updated my phone **6 times** in a quarter.
- My teenager grew by **3 inches** last year.
- **83 people** downloaded the latest mobile application.
- My aunt lost **18 pounds** last year.
- **150 respondents** were of the opinion that the new product feature will not be successful.
- There will be **30% increase** in revenue with the inclusion of a new product.
- **500 people** attended the seminar.
- **54% people** prefer shopping online instead of going to the mall.
- She has **10 holidays** in this year.
- Product X costs **$1000**.

As you can see in the above 10 examples, there is a numerical value assigned to each parameter and this is known as, quantitative data.

Advantages of Quantitative Data

Some of advantages of quantitative data, are:

- **Conduct in-depth research:** Since quantitative data can be statistically analyzed, it is highly likely that the research will be detailed.
- **Minimum bias:** There are instances in research, where personal bias is involved which leads to incorrect results. Due to the numerical nature of quantitative data, the personal bias is reduced to a great extent.
- **Accurate results:** As the results obtained are objective in nature, they are extremely accurate.

Learn more: Margin of Error

Disadvantages of Quantitative Data

Some of disadvantages of quantitative data, are:

- **Restricted information:** Because quantitative data is not descriptive, it becomes difficult for researchers to make decisions based solely on the collected information.
- **Depends on question types:** Bias in results is dependent on the question types included to collect quantitative data. The researcher's knowledge of questions and the objective of research are exceedingly important while collecting quantitative data.

Differences between Quantitative and Qualitative Data

There are some stark differences between quantitative data and qualitative data. They are:

Quantitative Data	Qualitative Data
Associated with numbers	Associated with details
Implemented when data is numerical	Implemented when data can be segregated into well-defined groups
Collected data can be statistically analyzed	Collected data can just be observed and not evaluated
Examples: Height, Weight, Time, Price, Temperature, etc.	Examples: Scents, Appearance, Beauty, Colors, Flavors, etc.

7-2 Mean, Median, Mode, and Range

UNIT 3RD

PART - I

7-2 Mean, Median, Mode, and Range

Learn to find the mean, median, mode, and range of a data set.

Vocabulary

mean

median

mode

range

outlier

7-2 Mean, Median, Mode, and Range

The **mean** is the sum of the data values divided by the number of data items.

The **median** is the middle value of an odd number of data items arranged in order. For an even number of data items, the median is the average of the two middle values.

Helpful Hint

The mean is sometimes called the average.

The **mode** is the value or values that occur most often. When all the data values occur the same number of times, there is no mode.

The **range** of a set of data is the difference between the greatest and least values. It is used to show the spread of the data in a data set.

Additional Example 1: Finding the Mean, Median, Mode, and Range of Data

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

mean:

$4 + 7 + 8 + 2 + 1 + 2 + 4 + 2 = 30$

8 items

sum

Add the values.

$30 \div 8 = 3.75$

Divide the sum by the number of items.

The mean is 3.75.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

median:

~~1, 2, 2,~~ 2, 4, ~~4, 7, 8~~

Arrange the values in order.

2 + 4 = 6

6 ÷ 2 = 3

There are two middle values, so find the mean of these two values.

The median is 3.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

mode:

1, 2, 2, 2, 4, 4, 7, 8 *The value 2 occurs three times.*

The mode is 2.

Additional Example 1 Continued

Find the mean, median, mode, and range of the data set.

4, 7, 8, 2, 1, 2, 4, 2

range:

1, 2, 2, 2, 4, 4, 7, 8

Subtract the least value from the greatest value.

8 – 1 = 7

The range is 7.

Check It Out: Example 1

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

mean:

$$\underbrace{6 + 4 + 3 + 5 + 2 + 5 + 1 + 8}_{\text{8 items}} = \underbrace{34}_{\text{sum}}$$

Add the values.

$$34 \div 8 = 4.25$$

Divide the sum by the number of items.

The mean is 4.25.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

median:

1, 2, 3, 4, 5, 5, 6, 8

Arrange the values in order.

4 + 5 = 9

9 ÷ 2 = 4.5

There are two middle values, so find the mean of these two values.

The median is 4.5.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

mode:

1, 2, 3, 4, 5, 5, 6, 8 *The value 5 occurs two times.*

The mode is 5.

Check It Out: Example 1 Continued

Find the mean, median, mode, and range of the data set.

6, 4, 3, 5, 2, 5, 1, 8

range:

1, 2, 3, 4, 5, 5, 6, 8

Subtract the least value from the greatest value.

8 – 1 = 7

The range is 7.

Check It Out: Example 2 Continued

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.

mean:

$$\frac{15 + 15 + 15 + 15 + 20 + 20 + 40 + 60 + 60 + 70}{10}$$

$$= \frac{330}{10} = 33$$

The mean is 33. Most of the cheerleaders raised less than $33, so the mean does not describe the data set best.

Check It Out: Example 2 Continued

The line plot shows the number of dollars each of the 10 members of the cheerleading team raised in a week. Which measure of central tendency best describes this data? Justify your answer.

mode:

The greatest number of X's occur above the number 15 on the line plot.

The mode is 15.

The mode focuses on one data value and does not describe the data set.

Additional Example 3 Continued

With the Outlier

55, 88, 89, 90, 94

outlier → 55

mean:	median:	mode:
55+88+89+90+94 = 416 416 ÷ 5 = 83.2 The mean is 83.2.	55, 88, 89, 90, 94 The median is 89.	There is no mode.

Additional Example 3 Continued

Without the Outlier

~~55~~, 88, 89, 90, 94

mean:	median:	mode:
88+89+90+94 = 361 361 ÷ 4 = 90.25 The mean is 90.25.	88, 89, 90, 94 $\frac{89+90}{2}$ = 89.5 The median is 89.5.	There is no mode.

Lesson Quiz: Part I

1. Find the mean, median, mode, and range of the data set. 8, 10, 46, 37, 20, 8, and 11

mean: 20; median: 11; mode: 8; range: 38

Lesson Quiz: Part II

2. Identify the outlier in the data set, and determine how the outlier affects the mean, median, and mode of the data. Then tell which measure of central tendency best describes the data with and without the outlier. Justify your answer.
85, 91, 83, 78, 79, 64, 81, 97

The outlier is 64. Without the outlier the mean is 85, the median is 83, and there is no mode. With the outlier the mean is 82, the median is 82, and there is no mode. Including the outlier decreases the mean by 3 and the median by 1, there is no mode. Because they have the same value and there is no outlier, the median and mean describes the data with the outlier. The median best describes the data without the outlier because it is closer to more of the other data values than the mean.

Geometric Mean and Harmonic Mean

UNIT – 3RD

PART - II

HARMONIC MEAN

Harmonic mean is used to calculate the average of a set of numbers. Here the number of elements will be averaged and divided by the sum of the reciprocals of the elements. The Harmonic mean is always the lowest mean.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} \qquad \ldots . (1)$$

Where,

x = Individual score, n = Sample size (Number of scores)

In other words, Harmonic mean of a number of observations, none of which is zero, is the reciprocal of the arithmetic mean of the reciprocals of the given values. Thus, harmonic mean (H), of n observations x_i, i = 1, 2, …., n is given by:

$$H = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(1/x_i)} \qquad \ldots .(2)$$

HARMONIC MEAN

In case of frequency distribution $x_i \mid f_i$, (i =1,2, ..., n),

$$H = \frac{1}{\frac{1}{N}\left(\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_n}{x_n}\right)}, (N = \sum_{i=1}^{n} f_i) \qquad \ldots.(3)$$

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{n}(f_i/x_i)}, \left[N = \sum_{i=1}^{n} f_i\right] \qquad \ldots.(4)$$

If $x_1, x_2, \ldots\ldots, x_3$ are n observations with weights $w_1, w_2, \ldots\ldots, w_3$, respectively, their weighted harmonic mean is defined as:

$$H = \frac{\sum w_i}{\sum(w_i/x_i)}. \qquad \ldots.(5)$$

Example

Q. A cyclist pedals from his house to his college at a speed of 10 km. p. h. and back from the college to his house at 15 km. p. h. Find the average speed.

Solution: Let the distance from the house to the college be x k. m. In going from house to college, the distance (x kilometre) is covered in $\frac{x}{10}$ hours, while in coming from college to house, the distance is covered in $\frac{x}{10}$ hours. Thus, a total distance of 2x km. is covered in $(\frac{x}{10} + \frac{x}{15})$ hours.

Hence average speed $\frac{Total\ distance\ travelled}{Total\ time\ taken} = \frac{2x}{(\frac{x}{10}+\frac{x}{15})} = 12$ km. p. h

In this case the average speed is given by the harmonic mean of 10 and 15 and not by the arithmetic mean.

HARMONIC MEAN

Example:

a) Milk is sold at the rates of 8, 10, 12 and 15 rupees per litre in four different months. Assuming that equal amounts are spent on milk by a family in the four months, find the average price in rupees per month.

b) An individual purchases three qualities of pencils. The relevant data are given below:

Quality	Price per pencil (Rs.)	Money Spent (Rs.)
A	1.00	50
B	1.50	30
C	2.00	20

Calculate the average price per pencil.

HARMONIC MEAN

Solution:

(a). Since equal amounts of money are spent by the family for each of the four months, the average price of milk per month is given by the harmonic mean of 8, 10, 12 and 15.

∴ Average price of milk per month

$$= Rs.\frac{1}{\frac{1}{4}(\frac{1}{8}+\frac{1}{10}+\frac{1}{12}+\frac{1}{15})} = Rs.\frac{4\times 20}{15+12+10+8} = Rs.\frac{4\times 20}{45} = Rs.10.67$$

(b) Here we are given: Total expenditure = Rs. (50 +30 + 20) = Rs. 100

$$\text{Total number of pencils purchased } = \frac{50}{1}+\frac{30}{1.50}+\frac{20}{2} = 80$$

$$\text{Average price per pencil } = \frac{Total\ expenditure}{Total\ No.\ of\ pencils} = \frac{100}{80} = Rs.1.25.$$

Merits and Demerits of Harmonic Mean

Merits	Demerits
Harmonic mean is rigidly defined, based upon all the observations and is suitable for further mathematical treatment. Like geometric mean, it is not affected much by fluctuation of sampling. It gives greater importance to small items and is useful only when small items have to be given a greater weightage.	Harmonic mean is not easily understood and is difficult to compute.

Geometric Mean

Geometric mean of a set of n observations is the nth root of their product. Thus the geometric mean G, of n observations x_i; i = 1, 2,…., n is given by:

$$G = (x_1.x_2 \ldots x_n)^{1/n} \quad \ldots.(6)$$

The computation is facilitated by use of logarithms. Taking logarithm of both sides,

$$\log G = \frac{1}{n}\left(\log x_1 + \log x_2 + \ldots + \log x_n\right) = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

$$G = \text{Antilog}\left(\frac{1}{n}\sum_{i=1}^{n} \log x_i\right) \quad \ldots.(7)$$

In case of frequency distribution $x_i \mid f_i$, i =1,2, ..., n, Geometric mean, G is:

$$G = \left(x_1^{f_1} . x_2^{f_2} \ldots x_n^{f_n}\right)^{\frac{1}{N}}, \quad \text{where } N = \sum_{i=1}^{n} f_i \quad \ldots.(8)$$

Taking logarithms of both sides, we get

$$\log G = \frac{1}{N}\left(f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n\right) = \frac{1}{N}\sum_{i=1}^{n} f_i \log x_i \quad \ldots.(9)$$

Thus we find that logarithm of geometric mean is the arithmetic mean of the logarithms of the given values.

Geometric Mean

From (4), we get

$$G = \text{Antilog}\left(\frac{1}{N}\sum_{i=1}^{n} f_i \log x\right) \qquad \ldots.(9)$$

In the case of grouped or continuous frequency distribution, x is taken to be the value corresponding to the mid-point of the class intervals.

Uses:

Geometric mean is used:

(i) To find the rate of population growth and the rate of interest.

(ii) In the construction of index numbers.

Step 1 →

Computations for Geometric Mean

X	$\log X$	f	$f \log X$
5	0.6990	10	6.990
15	1.1761	20	23.522
25	1.3979	30	41.937
35	1.5441	50	77.205
45	1.6532	40	66.128
55	1.7404	30	52.212
		$N = 180$	$\Sigma f \log X = 267.994$

Step 2 → $G.M.$ = Antilog $\left[\frac{\Sigma f \log X}{N}\right] = AL\left[\frac{267.994}{180}\right] = AL\,[1.4889] = 30.82$

PRACTICAL STEPS INVOLVED IN THE COMPUTATION OF G.M. IN CASE OF CONTINUOUS SERIES

Step 1 → *Calculate the mid-points of each class and enter these mid-points in the column headed as 'm'.*

Step 2 → *Take the logarithms of each mid-point and enter in the column headed as log m.*

Step 3 → *Multiply these logarithms (log m) with the respective frequencies and enter these products (f log m) in the column headed as f log m and then obtain their total to get* $\Sigma f \log m$.

Step 4 → *Calculate Geometric Mean as follows:*

$$G.M. = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right]$$

ILLUSTRATION 51 [Calculation of *G.M.* in case of Continous Series]

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students	10	20	30	50	40	30

Solution

Step 1 → **Computations for Geometric Mean**

Mark	m	$\log m$	f	$f \log m$
0–10	5	0.6990	10	6.990
10–20	15	1.1761	20	23.522
20–30	25	1.3979	30	41.937
30–40	35	1.5441	50	77.205
40–50	45	1.6532	40	66.128
50–60	55	1.7404	30	52.212
			$N = 180$	$\Sigma f \log m = 267.994$

Step 2 → $G.M. = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right] = \left[AL\,\frac{267.994}{180}\right] = AL\,[1.4889] = 30.82$

ILLUSTRATION 52

The annual rates of growth achieved by a nation for 5 years are 3, 4.5, 5.0, 7.5 and 10 per cent respectively. What is the compound rate of growth for the 5-year period?

ILLUSTRATION

Find the average rate of increase in population which in the first decade has increased by 10%, in the second decade by 20% and in the third decade by 30%.

Solution

Step 1 → **Computations for Geometric Mean**

Decade	*% rise*	*X Population at the end of the decade taking population of the previous decade as 100*	*log X*
1st	10	110	2.0414
2nd	20	120	2.0792
3rd	30	130	2.1139
			$\Sigma \log X = 6.2345$

Step 2 → $G.M. = A.L.\left(\frac{\Sigma \log X}{N}\right) = A.L.\left(\frac{6.2345}{3}\right) = A.L.\ 2.0782 = 119.8$

Step 3 → Average Rate of Increase in population is (119.8 – 100) = 19.8% per decade.

ILLUSTRATION 58

Find the weighted geometric mean from the following data:

Group	*Index Number*	*Weights*
Food	150	40
Fuel & Lighting	250	15
Clothing	350	10
House & Rent	450	20
Misc.	550	15

Solution

Calculation of weighted geometric mean

Group	*Index No.* X	*Weights* W	*Log X*	*W Log X*
Food	150	40	2.1761	87.044
Fuel & Lighting	250	15	2.3979	35.9685
Clothing	350	10	2.5441	25.4410
House Rent	450	20	2.6532	53.0640
Misc.	550	15	2.7404	41.1060
		$\Sigma W = 100$		$\Sigma W \text{ Log } X = 242.6235$

$$G.M._W = A.L.\left[\frac{\Sigma W \log X}{\Sigma W}\right] = AL\left[\frac{242.6235}{100}\right] = AL\,(2.4262) = 266.8$$

Step

COMBINED GEOMETRIC MEAN

If geometric means of two or more groups are given, the geometric mean of the combined group can be ontained, as follows:

$$\text{Combined } G.M. = \text{Antilog}\left[\frac{N_1 \log GM_1 + N_2 \log GM_2 + \ldots N_k \log GM_K}{N_1 + N_2 + \ldots + N_K}\right]$$

where,
GM_1 = Geometric mean of the first group
GM_2 = Geometric mean of the second group
GM_K = Geometric mean of the Kth group.

ILLUSTRATION 57

GM of 100 items is 50 and *GM* of 200 items is 40. Find the combined geometric mean of 300 items.

Solution

$$\text{Combined } G.M. = \text{Antilog}\ \frac{100 \log 50 + 200 \log 40}{300}$$

Merits and Demerits of Geometric Mean

Merits	Demerits
1. It is rigidly defined. **2. It is based upon all the observation.** **3. It is suitable for further mathematical treatment.** **4. It is not affected much by fluctuations of sampling.** **5. It gives comparatively more weight to small items.**	1. Because of its abstract mathematical character, geometric mean is not easy to understand and to calculate for a non-mathematics person. 2. If any one of the observations is zero, geometric mean becomes zero and if any one of the observations is negative, geometric mean becomes imaginary regardless of the magnitude of the other items.