

Business statistics

Unit-1

Statistics: Definition, Origin and Growth, Functions, Applications and Limitations- Classification of data: Types of classifications.

Origin and Growth of Statistics:

The origin of statistics can be traced back to the primitive man, who put notches on trees to keep an account of his belongings. During 5000 BCE, kings used to carry out census of populations and resources of the state. Kings of olden days made their crucial decisions on wars, based on statistics of infantry, and elephantary units of their own and that of their enemies. Later it enhanced its scope in their kingdoms' tax management and administrative domains. Thus, the word 'Statistics' has its root either to Latin word 'Status' or Italian word 'Statista' or German word 'Statistik' each of which means a '**political state**'. The word 'Statistics' was primarily associated with the presentation of facts and figures pertaining to demographic, social and political situations prevailing in a state/government. It is regarded as the "Science of Statecraft". It is the by-product of States administrative activities. It has been the traditional function of Government to keep record of the population, births, deaths, taxes, crop yields and many other type of activities. Its evolution over time formed the basis for most of the science and art disciplines. Statistics is used in the developmental phases of both theoretical and applied areas, encompassing the field of Industry, Agriculture, Medicine, Sports and Business analytics.

In olden days statistics was used for political- war purpose. Later, it was extended to taxation purposes. This is evident from Kautilya's Arthashastra (324 – 300 BCE). Akbar's finance minister Raja Thodarmall collected information regarding agricultural land holdings. During the seventeenth century, statistics entered in vital statistics, which is the basis for the modern day Actuarial Science. Gauss introduced the theory of errors in physical sciences at the end of eighteenth century.

According to a Greek historian, in 1400 B.C. a census of all the lands in the Egypt were taken. Similar reports on the ancient Chinese, Greeks and Romans are also available. People and land were the earliest objects of statistical enquiry.

Statistics came from an Italian Word “Statista” means “Statesman”, German word “Statistik” means Political State.

It was first used by a Professor Gottfried Achenwall (1719-1772), a Professor in Malborough in 1749 to refer to the subject matter as a whole. He defined as a “political science of several countries”.

The word Statistics appeared for the first time in the famous book “Elements of Universal Erudition” by Baran J.F. Von Bielfeld. It is translated by W.Hooper.

Statistics is concerned with scientific method for collecting, organizing, summarizing, presenting, analyzing and interpreting of data. The word statistics is normally referred either as numerical facts or methods.

Statistics is used in two different forms-**singular and plural**. In plural form it refers to the numerical figures obtained by measurement or counting in a systematic manner with a definite purpose such as number of accidents in a busy road of a city in a day, number of people died due to a chronic disease during a month in a state and so on. In its singular form, it refers to statistical theories and methods of collecting, presenting, analyzing and interpreting numerical figures.

Sources of Statistics Origin:

1. Government records
2. Mathematics

1. Government records:

- In ancient Egypt, police prepared registration lists of all the heads of the families.
- In ancient Judea, census of population was taken.
- In Roman- census was recorded about military strength, taxation, births and deaths.

As statistics is used for Government Purpose it is called as “the Science of kings” or “the Science of Statescraft”.

William Petty was the author of “Essays on Political Arthematick”. He regarded Statistics as “Political arthemetic”.

2. Mathematics:

- Statistics is a branch of Applied Mathematics. In 17th century the attention of Mathematicians like Bernoulli, Galileo, Laplace and Karl Gauss has moved towards Statistics. They developed Probability theories like winning and losing in gamble.
- Statisticians were engaged in calculating the risk associate with a particular decision.
- Abraham DeMoivre (1667-1754) –discovered Normal Statistical Theory.
- Jacques Quettlet (1796-1874) – discovered the constancy of great numbers which is the basis for sampling.
- Francis Galton(1822-1911)- regression
- Karl pearson(1857-1936) – Chi-Square test
- Ronald Fisher (1890-1962) – field experimental designs.

These are the real giants in the development of the theory of statistics.

Growth of Statistics:

Though the importance of statistics was strongly felt, its tremendous growth was in the twentieth century. During this period, lot of new theories, applications in various disciplines were introduced. With the contribution of renowned statisticians several theories and methods were introduced, naming a few are Probability Theory, Sampling Theory, Statistical Inference, Design of Experiments, Correlation and Regression Methods, Time Series and Forecasting Techniques.

In early 1900s, statistics and statisticians were not given much importance but over the years due to advancement of technology it had its wider scope and gained attention in all fields of science and management. We also tend to think statistician as a small profession but a steady growth in the last century is impressive. It is pertinent to note that the continued growth of statistics is closely associated with information technology. As a result several new inter- disciplines have emerged. They are Data Mining, Data Warehousing, Geographic Information System, Artificial Intelligence etc. Now-a-days, statistics can be applied in hardcore technological spheres such as Bioinformatics, Signal processing, Telecommunications, Engineering, Medicine, Crimes, Ecology, etc.

Today's business managers need to learn how analytics can help them make better decisions that can generate better business outcomes. They need to have an understanding of the

statistical concepts that can help analyze and simplify the flood of data around them. They should be able to leverage analytical techniques like decision trees, regression analysis, clustering and association to improve business processes.

Factors responsible for development of Statistics in Modern times:

1. Increased demand for statistics:

- In the present century considerable development has taken place in the field of Business and commerce, Governmental activities and Science. Statistics helps in formulating suitable policies.
- Due increase size of the Business- Statistics helps in resolving the complex problems.
- Functions of Government are enlarged (Law and Order)- Statistics helps in proper maintenance of Law and Order.
- Development of Science and technology- Statistics helps in doing research.

2. Decreased cost of Statistics:

- Time and cost of collecting data are the important factors in the use of Statistics. With Calculators, electronic machines e.t.c the cost of analyzing the data has come down. This has led to the use of statistics in solving various problems.
- With the development of statistics the cost of collecting and analyzing the data has come down.
- Many have contributed to the science of statistics.

Inspite of the developments, the list of unsolved statistical problems are long and the statistical research today is more vigorous than ever before.

Definitions:

Statistics is the science of numbers. The English word Statistics originates from the Italian word “Statista” or the Latin word “Status”. Statistics is about creating information about any number-based information by searching and recording it. In other words, statistics is the scientific method of collecting, analyzing, and interpreting numerical data.

Croxton and Cowden, “Statistics may be defined as **the collection, presentation, analysis, and interpretation of numerical data.**”

Edward N. Dubois defines, “Statistics is a body of methods for obtaining and analyzing numerical data in order to make better decisions in an uncertain world.”

According to Bowleg, “Statistics may rightly be called the science of averages.”

Prof. Boddington defines, “The science of estimate and probabilities.”

According to Prof. Horace Secrist, “Statistics is the aggregate of facts affected to a marked extent by the multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a pre-determined purpose and placed in relation to each other.”

American Heritage Dictionary defines statistics as: “The mathematics of the Collection, organization, and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling.”

A.L. Bowley defines, “Statistics are numerical statements of facts in any department of inquiry placed in relation to each other.”

Wallis and Roberts define, “Statistics is a body of methods for making wise decisions in the face of uncertainty.”

According to Prof. Ya Lun Chaou, “ Statistics is a method of decision making in the face of uncertainty on the basis of numerical data and calculated risk.”

According to W.I. King, “The science of statistics is the method of judging collection, natural or social phenomena from the results obtained from the analysis or enumeration or collection of estimates.”

The [Merriam-Webster's](#) Dictionary definition is: “A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.”

Features of Statistics:

1. A numerical expression is required in statistics.
2. Statistics is the sum of information.
3. The search for statistics must be related to a specific field.
4. Statistical information is influenced by multiple factors.
5. Statistical information needs to be collected in a well-organized manner.
6. The statistics should be comparable and homogeneous.
7. There is a need to maintain logical and quantitative accuracy in statistical estimates.

Functions of Statistics:

1. To Present Facts in Definite Form:

We can represent the things in their true form with the help of figures. Without a statistical study, our ideas would be vague and indefinite.

The facts are to be given in a definite form. If the results are given in numbers, then they are more convincing than if the results are expressed on the basis of quality.

The statements like, there is lot of unemployment in India or population is increasing at a faster rate are not in the definite form. The statements should be in definite form like the population in 2004 would be 15% more as compared to 1990.

2. Precision to the Facts:

The statistics are presented in a definite form so they also help in condensing the data into important figures. So statistical methods present meaningful information. In other words statistics helps in simplifying complex data to simple-to make them understandable.

The data may be presented in the form of a graph, diagram or through an average, or coefficients etc. For example, we cannot know the price position from individual prices of all goods, but we can know it, if we get the index of general level of prices.

3. Comparisons:

After simplifying the data, it can be correlated as well as compared. The relationship between the two groups is best represented by certain mathematical quantities like average or coefficients etc. Comparison is one of the main functions of statistics as the absolute figures convey a very less meaning.

4. Formulation and Testing of Hypothesis:

These statistical methods help us in formulating and testing the hypothesis or a new theory. With the help of statistical techniques, we can know the effect of imposing tax on the exports of tea on the consumption of tea in other countries. The other example could be to study whether credit squeeze is effective in checking inflation or not.

5. Forecasting:

Statistics is not only concerned with the above functions, but it also predicts the future course of action of the phenomena. We can make future policies on the basis of estimates made with the help of Statistics. We can predict the demand for goods in 2005 if we know the population in 2004 on the basis of growth rate of population in past. Similarly a businessman can exploit the market situation in a successful manner if he knows about the trends in the market. The statistics help in shaping future policies.

6. Policy Making:

With help of statistics we can frame favourable policies. How much food is required to be imported in 2007? It depends on the food-production in 2007 and the demand for food in 2007. Without knowing these factors we cannot estimate the amount of imports. On the basis of forecast the government forms the policies about food grains, housing etc. But if the forecasting is not correct, then the whole set up will be affected.

7. It Enlarges Knowledge:

Whipple rightly remarks that “Statistics enables one to enlarge his horizon”. So when a person goes through various procedures of statistics, it widens his knowledge pattern. It also widens his thinking and reasoning power. It also helps him to reach to a rational conclusion.

8. To Measure Uncertainty:

Future is uncertain, but statistics help the various authorities in all the phenomenon of the world to make correct estimation by taking and analyzing the various data of the part. So the uncertainty could be decreased. As we have to make a forecast we have also to create trend behaviors of the past, for which we use techniques like regression, interpolation and time series analysis.

9. To simplify mass of figures:

Statistics helps in condensing mass of data into a few significant figures. Statistical methods present meaningful overall information from the mass of data. It is impossible to remember the individual income of the entire population. However remembering the figures of Per capita income is very easy.

Applications of Statistics:

<https://prinsli.com/scope-and-importance-of-statistics/>

[Statistics](#) has such a broad and ever-expanding scope that defining it is not only difficult but also unwise. Statistics are now used in practically every aspect of our life. It has become a recognized subject in its own right, as well as a tool of all disciplines essential to study, research and intelligent judgment. Statistical tools are used in almost every sector, including education, trade, industry or commerce, economics, biology, botany, medicine, physics, chemistry, astronomy, sociology, technology, and psychology.

Statistics has so many applications that it is often said, “Statistics is what statisticians do.” Governments, businesses, and people collect statistical data required in order to perform their activities efficiently and effectively. The following are some areas where statistics is used:

1. Statistics and State: (Importance of Statistics in State):

Since ancient times, ruling kings and chiefs have depended largely on statistics to formulate appropriate military and fiscal policies. Most of the statistics they collected, such as those on the population, military strength, crimes, taxes, and so on, were a by-product of administrative activity. The state's functions have greatly expanded in recent years. The concept of a state has evolved beyond just preserving law and order to one of providing social services.

2. Statistics in Business and Management: (Importance of Statistics in Business and Management):

Statistical methods are almost mainly used in the 20th century to solve commercial difficulties. Statistics are used in almost every aspect of business and industry, including manufacturing, financial analysis, distribution analysis, market research, personnel planning, research and development, and accounting, to name a few.

3. Statistics and Economics: (Importance of Statistics in Economics):

The study of statistics is an important element of economics. In 1890, the famous economist, Prof. Alfred Marshall, stated, "Statistics are the straw out of which I, like every other economist, have to create bricks." Dr Marshall's viewpoint highlights the importance of statistics in economics. Statistical data and statistical analysis techniques have been extremely helpful when dealing with economic issues. Prices, wages, time series analysis, and demand analysis are just a few examples.

4. Statistics and Physical Sciences: (Importance of Statistics in Physical Sciences):

Statistical methods were first developed and applied in the physical sciences, mainly physics, geology, and astronomy, but until recently, these sciences were not as involved in the development of 20th-century statistics as the biological and social sciences.

5. Statistics and Natural Sciences: (Importance of Statistics in Natural Sciences):

In the study of all natural sciences, like biology, medicine, meteorology, zoology, botany, etc., statistical methods have proven to be highly helpful. For example, the doctor had to rely heavily on factual data such as body temperature, pulse rate, BP, and so on when diagnosing the correct disease.

6. Statistics and Research: (Importance of Statistics in Research):

In research, statistics is a must-have. Most of the progress in knowledge has been due to experiments conducted with the help of statistical methods.

7. Statistics and Planning: (Importance of Statistics in Planning):

In the modern era, dubbed “the age of planning,” statistics are essential to planning. Almost everywhere in the world, governments are restoring economic development planning.

8. Statistics and Industry: (Importance of Statistics in Industry):

Statistics are widely employed in industry to control inequality. In production engineering, statistical techniques such as inspection plans, control charts, etc., are used to determine if the product conforms to the standards or not. Other businesses use statistics to help them develop better products and services. Some companies use data from embedded sensors in their goods to provide services such as regular maintenance.

9. Statistics and Mathematics: (Importance of Statistics in Mathematics):

Statistics and mathematics are inextricably linked, and recent advances in statistical methods are the result of a wide range of mathematical applications.

Because statistics is a branch of applied mathematics, it is obvious that it plays an important role in mathematics. Statistics, on the other hand, is more than just a distinct branch of mathematics.

A large portion of math is focused on probability and theory. Statistical approaches help to improve the accuracy of such mathematical theories. Using averages, dispersion, and estimations, you can arrive at conclusions that are closer to the actual answer than simply guessing. Some examples of mathematical statistics are as follows:

- Calculate the average, median, and mode.
- In a group of 20 people, there is a 75% chance that at least two people will share the same birthday, etc.

10. Statistics and Medical Science: (Importance of Statistics in Medical Science):

In medical science, statistical tools for collecting, presenting, and analyzing observed data relating to disease reasons and incidence, as well as the effects of various medications and medicines, are extremely important.

11. Statistics and Psychology & Education: (Importance of Statistics in Psychology & Education):

Statistics has a wide range of applications in education and physiology, including identifying or attempting to determine the reliability and validity of a test, factor analysis, and so on.

12. Statistics and Other Uses: (Importance of Statistics in Other Fields):

Insurance companies, auditors, bankers, brokers, social workers, labor unions, trade unions, chambers of commerce, etc. can all benefit from statistics. Politicians and their supporters can benefit greatly from statistics. They want to see what their chances are and what efforts are required to win the election.

Limitations of Statistics:

1. Qualitative Aspect Ignored:

The statistical methods don't study the nature of phenomenon which cannot be expressed in quantitative terms.

Such phenomena cannot be a part of the study of statistics. These include health, riches, intelligence etc. It needs conversion of qualitative data into quantitative data.

So experiments are being undertaken to measure the reactions of a man through data. Now a days statistics is used in all the aspects of the life as well as universal activities.

2. It does not deal with individual items:

It is clear from the definition given by Prof. Horace Sacrist, "By statistics we mean aggregates of facts.... and placed in relation to each other", that statistics deals with only aggregates of facts or items and it does not recognize any individual item. Thus, individual terms as death of 6 persons

in a accident, 85% results of a class of a school in a particular year, will not amount to statistics as they are not placed in a group of similar items. It does not deal with the individual items, however, important they may be.

3. It does not depict entire story of phenomenon:

When even phenomena happen, that is due to many causes, but all these causes cannot be expressed in terms of data. So we cannot reach at the correct conclusions. Development of a group depends upon many social factors like, parents' economic condition, education, culture, region, administration by government etc. But all these factors cannot be placed in data. So we analyse only that data we find quantitatively and not qualitatively. So results or conclusion are not 100% correct because many aspects are ignored.

4. It is liable to be misused:

As W.I. King points out, "One of the short-comings of statistics is that do not bear on their face the label of their quality." So we can say that we can check the data and procedures of its approaching to conclusions. But these data may have been collected by inexperienced persons or they may have been dishonest or biased. As it is a delicate science and can be easily misused by an unscrupulous person. So data must be used with a caution. Otherwise results may prove to be disastrous.

5. Laws are not exact:

As far as two fundamental laws are concerned with statistics:

(i) Law of inertia of large numbers and

(ii) Law of statistical regularity, are not as good as their science laws.

They are based on probability. So these results will not always be as good as of scientific laws. On the basis of probability or interpolation, we can only estimate the production of paddy in 2008 but cannot make a claim that it would be exactly 100 %. Here only approximations are made.

6. Results are true only on average:

As discussed above, here the results are interpolated for which time series or regression or probability can be used. These are not absolutely true. If average of two sections of students in statistics is same, it does not mean that all the 50 students in section A has got same marks as in B. There may be much variation between the two. So we get average results.

“Statistics largely deals with averages and these averages may be made up of individual items radically different from each other.” —W.L King

7. To Many methods to study problems:

In this subject we use so many methods to find a single result. Variation can be found by quartile deviation, mean deviation or standard deviations and results vary in each case.

“It must not be assumed that the statistics is the only method to use in research, neither should this method of considered the best attack for the problem.” —Croxten and Cowden

8. Statistical results are not always beyond doubt:

“Statistics deals only with measurable aspects of things and therefore, can seldom give the complete solution to problem. They provide a basis for judgement but not the whole judgment.”
—Prof. L.R. Connor

Although we use many laws and formulae in statistics but still the results achieved are not final and conclusive. As they are unable to give complete solution to a problem, the result must be taken and used with much wisdom.

CLASSIFICATION OF DATA:

Meaning of Classification of Data

- It is the process of arranging data into homogeneous (similar) groups according to their common characteristics.
- Raw data cannot be easily understood, and it is not fit for further analysis and interpretation. Arrangement of data helps users in comparison and analysis.
- For example, the population of a town can be grouped according to sex, age, marital status, etc.

Classification of data

The method of arranging data into homogeneous classes according to the common features present in the data is known as classification.

A planned data analysis system makes the fundamental data easy to find and recover. This can be of particular interest for legal discovery, risk management, and compliance. Written methods and sets of guidelines for data classification should determine what levels and measures the company will use to organise data and define the roles of employees within the business regarding input stewardship.

Once a data -classification scheme has been designed, the security standards that stipulate proper approaching practices for each division and the storage criteria that determines the data's lifecycle demands should be discussed.

Objectives of Data Classification

The primary objectives of data classification are:

- To consolidate the volume of data in such a way that similarities and differences can be quickly understood. Figures can consequently be ordered in sections with common traits.
- To aid comparison.
- To point out the important characteristics of the data at a flash.
- To give importance to the prominent data collected while separating the optional elements.
- To allow a statistical method of the materials gathered.

TYPES OF CLASSIFICATION:

There are four types of classification. They are

1. Geographical classification(Area wise e.g.: Cities, districts)
2. Chronological classification (Time)
3. Qualitative classification (Attributes)
4. Quantitative classification (Magnitudes)

(i) Geographical classification

When data are classified on the basis of location or areas, it is called geographical classification

Example: Classification of production of food grains in different states in India.

State wise estimates of Production of Food Grains:

Name of the State	Production of food grains(Thousand Tonnes)
Andhra Pradesh	1093.7
Bihar	12899.0
Haryana	11334.7
Punjab	21148.9
Uttar Pradesh	41828.6
All India	1,92,433.6

- Geographical classifications are listed in alphabetical order for easy reference.
- Items may be listed by size to emphasize the important areas as in ranking the States by Population.

(ii) Chronological classification

Chronological classification means classification on the basis of time, like months, years etc.

Example: Profits of a company from 2001 to 2005.

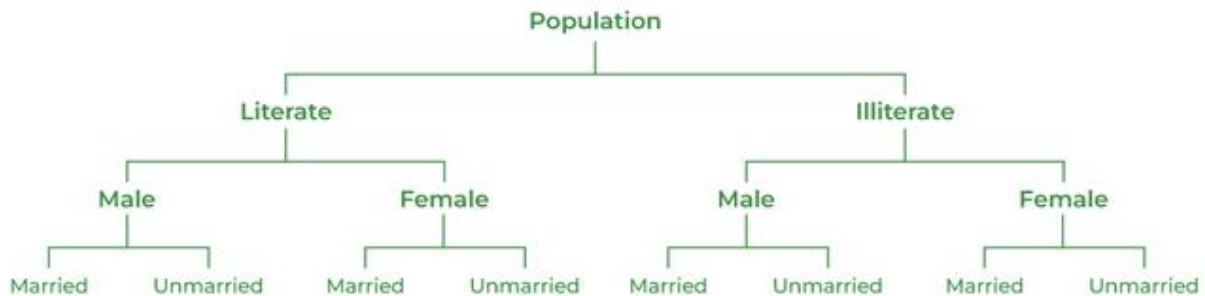
Population of India from 1951 to 2001

Year	Population(in Crores)
1951	36.11
1961	43.92
1971	54.82
1981	68.33
1991	84.63
2001	102.37

Time series are usually listed in chronological order, normally starting with the earliest period. When the major emphasis falls on the most recent events, a reverse time order may be used.

(iii) Qualitative classification

In Qualitative classification, data are classified on the basis of some attributes or quality such as sex, colour of hair, literacy and religion. In this type of classification, the attribute under study cannot be measured. It can only be found out whether it is present or absent in the units of study.



Classification of population on the basis of sex, i.e. into males and females or literacy i.e. into literate and illiterate. The type of classification where only two classes are formed is also called **two fold or Dichotomous Classifications**.

Instead of forming only two classes we further divide the data on the basis of some attributes so as to form several classes, the classification is known as **Manifold Classification**.

(iv) Quantitative classification

Quantitative classification refers to the classification of data according to some characteristics, which can be measured such as height, weight, income, profits etc.

Example: The students of a school may be classified according to the weight as follows.

Weight (in kgs)	No of Students
40-50	50
50-60	200
60-70	300
70-80	100
80-90	30
90-100	20
Total	700

There are **two types of quantitative classification of data**. They are

1. Discrete frequency distribution

2. Continuous frequency distribution

In this type of classification there are **two elements (i) variable (ii) frequency**

Variable

Variable refers to the characteristic that varies in magnitude or quantity. E.g. weight of the students. A variable may be discrete or continuous.

Discrete variable

- A discrete variable can take only certain specific values that are whole numbers (integers). E.g. Number of children in a family or Number of class rooms in a school.
- Varies with finite jumps.
- Manifests every conceivable fractional value.

No. of Children	No. of Families
0	10
1	40
2	80
3	100
4	250
5	150
6	50
Total	680

Continuous variable/ Continuous random Variable:

- A Continuous variable can take any numerical value within a specific interval.
- It is capable of manifesting every conceivable fractional value within the range of possibilities.

Example: the average weight of a particular class student is between 60 and 80 kgs.

Weight passes through all the values between these limits.

Weight(lbs)	No. of persons
100-110	10
110-120	15
120-130	40

130-140	45
140-150	20
150-160	04
Total	134

Frequency

Frequency refers to the number of times each variable gets repeated.

For example there are 50 students having weight of 60 Kgs. Here 50 students is the frequency.

Frequency distribution

Frequency distribution refers to data classified on the basis of some variable that can be measured such as prices, weight, height, wages, no. of units produced and consumed etc.

The following are the two examples of discrete and continuous frequency distribution

The following technical terms are important when a continuous frequency distribution is formed

Class limits: Class limits are the lowest and highest values that can be included in a class. For example take the class 40-50. The lowest value of the class is 40 and the highest value is 50. In this class there can be no value lesser than 40 or more than 50. 40 is the lower class limit and 50 is the upper class limit.

Class interval: The difference between the upper and lower limit of a class is known as class interval of that class. Example in the class 40-50 the class interval is 10 (i.e. 50 minus 40).

Class frequency: The number of observations corresponding to a particular class is known as the frequency of that class

Example:

Income (Rs)	No. of persons
-------------	----------------

1000 - 2000	50
-------------	----

In the above example, 50 is the class frequency. This means that 50 persons earn an income between Rs.1, 000 and Rs.2, 000.

(iv) **Class mid-point:** it is the value lying half way between the lower and upper class limits of a class interval.

