



**NAME:** Azmat Ullah Ibrar

**ROLL NO:** CA634907

**COURSE CODE:** 8614

**COURSE NAME** Educational statistics

**PROGRAM:** B.Ed. 1.5 year

**SEMESTER:** Three

# **ASSINGMENT#01**

## **QUESTION#01**

**What do you understand by statistics? What are the characteristics of statistics? Explain in detail.**

**Ans:**

### **STATISTICS:**

The word “statistics” is derived from the Latin word “Status”, which means a political state. Statistics is a branch of knowledge that deals with facts and figures. The term statistics refers to a set of methods and rules for organizing, summarizing, and interpreting information. Statistics is both a formal science and a practical theory of scientific inquiry, and both aspects are considered in statistics education. Education in statistics has similar concerns as does education in other mathematical sciences, like logic, mathematics, and computer science. At the same time, statistics is concerned with evidence-based reasoning, particularly with the analysis of data.

### **STATISTICS IN EDUCATION:**

Measurement and evaluation are essential part of teaching learning process. In this process we obtained scores and then interpret these scores in order to take decisions. Statistics enables us to study these scores objectively. It makes the teaching learning process more efficient.

The knowledge of statistics helps the teacher in the following way:

1. It helps the teacher to provide the most exact type of description:

When we want to know about the pupil, we administer a test or observe the child. Then from the result we describe about the pupil’s performance or trait. Statistics helps the teacher to give an accurate description of the data.

2. It makes the teacher definite and exact in procedures and thinking:

Sometimes due to lack of technical knowledge the teachers become vague in describing pupil's performance. But statistics enables him to describe the performance by using proper language, and symbol.

3. It enables the teacher to summarize the results in a meaningful and convenient form:

Statistics gives order to the data. It helps the teacher to make the data precise and meaningful and to express it in an understandable and interpretable manner.

4. It enables the teacher to draw general conclusions:

Statistics helps to draw conclusions as well as extracting conclusions. Statistical steps also help to say about how much faith should be placed in any conclusion and about how far we may extend our generalization.

5. It helps the teacher to predict the future performance of the pupils:

Statistics enables the teacher to predict how much of a thing will happen under conditions we know and have measured. For example, the teacher can predict the probable score of a student in the final examination from his entrance test score. But the prediction may be erroneous due to different factors.

6. Statistics enables the teacher to analyze some of the causal factors underlying complex and otherwise be-wildering events: It is a common factor that the behavioral outcome is a resultant of numerous causal factors. The reason why a particular student performs poor in a particular subject are varied and many.

### **CHARACTERISTICS OF STATISTICS:**

Following are the characteristics of statistics.

- **Statistics consists of aggregate facts:** The facts which can be studied in relation to time, place or frequency can be called statistics. In the plural sense, statistics refers to data, but data to be called statistics must consist of aggregate of certain facts. A single isolated and unconnected fact or figure is not statistics because we cannot study it in relation to other facts and figures. Only

aggregate of facts e.g. academic achievement of the students, I.Q. of a group of students, weight of students in a class, profit of a firm etc. are called statistics.

- **Multiple causes affect Statistics:** A phenomena may be affected by so many factors. We cannot study the effects of one factor on the phenomena only by ignoring others. Here we have to go for the effects of all the factors on the phenomenon separately as well as collectively, because effects of the factors can change with change of place, time or situation. To have a true picture we will have to study the effects of all factors on the phenomena separately as well as collectively, because effects of the factors can change with change of place, time or situation. The overall effect is taken and not of one factor only as in other natural sciences. For example, we can say that result of class X in board examination does not depend on any single factor but collectively on standard of teachers, teaching methods, teaching aids, practical's performance of students, standard of question papers, environment of the examination hall, exam supervisory staff and standard of evaluation of answers after the examination.
- **Data should be numerically expressed, enumerated or estimated:** Data to be called statistics should be numerically expressed so that counting or measurement of data can be made possible. It means that the data or the fact must be in quantitative form as achievement scores 60, 50, 85, 78, and 91 out of 100. If it is not in quantitative form, it should be quantified.
- **Statistics are enumerated or estimated according to reasonable standard of accuracy:** For a clear picture of the phenomena under investigation, it should be researched using reasonable standard of accuracy depending upon the nature and purpose of collection of data. Data collection should be free from personal prejudices and biases. Biased and personally prejudiced data leads to inaccurate conclusion. For getting reasonable standard of accuracy the field of enquiry should not be very large. If it is infinite or very large, even enumeration of data is impossible and reasonable standard of accuracy may not be achieved. To achieve it we have to make an estimate according to reasonable standard of accuracy depending upon the nature and purpose of collection of data. e.g. we may measure the height of buildings in metres but we cannot measure the length of small things like bricks in the same unit of metre.
- **Statistics are collected in a Systematic Manner:** In order to have reasonable standard of accuracy statistics/data must be collected in a very systematic manner. The data collected in a haphazard manner will lead to difficulties in the process of analysis, and wrong conclusions. A proper plan should be made and trained investigators should be used to collect data so that they may collect

statistics. If it is not done, in such cases reliability of data gets decreased. So, to get correct results the data must be collected in a precise manner. Any rough and haphazard method of collection will not be desirable for that may lead to improper and wrong conclusion.

- **Statistics for a Pre-determined Purpose:** Before collection of data, investigator/researcher must have a purpose and then should collect data accordingly. Data collected without any purpose is of no use. Suppose we want to know intelligence of a section of people, we must collect data relating to I.O. level and data relating to income, attitude and interest level of that group of people will be of no use. Without having a clear idea about the purpose, we will not be in a position to distinguish between necessary data and unnecessary data or relevant data and irrelevant data.
- **Statistics are Capable of being placed in Relation to each other:** Statistics is a method for the purpose of comparison etc. It must be capable of being compared; otherwise, it will lose much of its significance. Comparison can be made only if the data are homogeneous. Data on memory test can be compared with I.Q. It is with the use of comparison only that we can illustrate changes which may relate to time, place, frequency or any other character, and statistical devices are used for this purpose. It is last but not less important of the characteristics of the statistics. The collection of data is generally done with the motive to compare. If the figures collected are not comparable, in that case, they lose a large part of their significance.

## QUESTION#02

**What do you understand by the term “data”? Write in detail the types of data?**

**Ans:**

### **DATA:**

Data may be defined as information obtained from a survey, an experiment or an investigation. The term “data” refers to the kind of information a researcher obtains to achieve objectives of his research. All research processes start with collection of data, which plays a significant role in the statistical analysis. This term is used in different contexts. Data are individual pieces of factual information recorded and

used for the purpose of analysis. It is the raw information from which statistics are created. Statistics are the results of data analysis - its interpretation and presentation.

Data Types are an important concept in statistics, which needs to be understood, to correctly apply statistical measurements to your data and therefore to correctly conclude certain assumptions about it. Data are the actual pieces of information that you collect through your study. In other words, data can be defined as collection of facts and details like text, figures, observations, symbols, or simply description of things, event or entity gathered with a view of drawing inferences. It is a raw fact which should be processed to get information.

### **TYPES OF DATA:**

Different methods are used to collect data, all of which fall into two categories, i.e. primary data and secondary data. It is a common classification based upon who collected the data.

### **PRIMARY DATA:**

Data that has been generated by the researcher himself/herself, surveys, interviews, experiments, specially designed for understanding and solving the research problem at hand. Primary data is originated by the researcher for the first time for addressing his research problem. It is also known as first hand raw data. The data can be collected using various methods like survey, observations, physical testing, mailed questionnaire, questionnaire filled and sent by enumerators, personal interviews, telephonic interviews, focus groups discussion, case studies, etc. Primary data is often reliable, authentic, and objective in as much as it was collected with the purpose of addressing a particular research problem. It is noteworthy that primary data is not commonly collected because of the high cost of implementation.

### **SECONDARY DATA:**

Using existing data generated by large government Institutions, healthcare facilities etc. as part of organizational record keeping. The data is then extracted from more varied data files. Point towards the second-hand information already collected and recorded by any other person with a purpose not relating to current research problem. It is readily available form of data and saves time and cost of the researcher. But as the data is gathered for the purpose other than the problem under investigation, so the usefulness of the data may be limited in a number of ways like relevance and accuracy. Secondary data is the data that has been collected in the past by someone

else but made available for others to use. They are usually once primary data but become secondary when used by a third party.

### **KEY DIFFERENCES BETWEEN PRIMARY AND SECONDARY DATA:**

- Primary data refers to the data originated by the researcher for the first time. Secondary data is already existing data, collected by other researchers, agencies, and organizations.
- Primary data is real-time data whereas secondary data is one which relates to the past.
- Primary data is collected to address the problem in hand while the purpose behind collection of secondary data is different from the problem in hand.
- Collection of primary data is a laborious process. On the other hand, collection of secondary data is easy and rapid.
- Sources of primary data are survey, observations, physical testing, mailed questionnaire, questionnaire filled and sent by enumerators, personal interviews, telephonic interviews, focus groups discussion, case studies, etc. On the other hand, sources of secondary are censuses data, publications, internal records of the organizations, reports, books, journal articles, websites etc.
- Collection of primary data requires a large amount of resources like time, cost, and human resources. On the other hand, collection of secondary data is expensive and easily available. vii) Primary data is specific to the researcher's needs. He can control the quality of research. On the other hand, secondary data is neither specific to researcher needs nor has he control over the quality of data.
- Primary data is available in the raw form while secondary data has undergone some statistical procedures and is refined from primary data.
- Data collected from primary sources are more reliable and accurate than the secondary sources

### **SOME OTHER TYPES OF DATA:**

#### **QUANTITATIVE DATA & QUALITATIVE DATA:**

#### **QUANTITATIVE DATA:**

- These types of data seem to be the easiest to explain. It tries to find the answers to questions such as

- “How many,
- “How much” and
- “How often”
- It can be expressed as a number, so it can be quantified. In simple words, it can be measured by numerical variables.
- These are easily open for statistical manipulation and can be represented by a wide variety of statistical types of graphs and charts like **line charts, bar graphs, scatter plots**, etc.

### **Examples of quantitative data:**

- Scores of tests and exams e.g. 74, 67, 98, etc.
- The weight of a person.
- The temperature in a room.

### **There are 2 general types of quantitative data:**

- Discrete data
- Continuous data

### **QUALITATIVE DATA:**

- Qualitative data can't be expressed as a number, so it can't be measured. It mainly consists of words, pictures, and symbols, but not numbers.
- It is also known as **Categorical Data** as the information can be sorted by category, not by number.
- These can answer the questions like:
  - “How this has happened”, or
  - “Why this has happened”.

### **Examples of qualitative data:**

- Colors e.g. the color of the sea



- Popular holiday destinations such as Switzerland, New Zealand, South Africa, etc.
- Ethnicity such as American Indian, Asian, etc.

**In general, there are 2 types of qualitative data:**

- Nominal data
- Ordinal data.

### **QUESTION #03**

**What types of characteristics a pictogram should have to successfully convey the meaning? Write down the advantages and drawbacks of using pictograms?**

**Ana:**

#### **PICTOGRAM:**

A pictogram is a graphical symbol that conveys its meaning through its pictorial resemblance to a physical object. A pictogram may include a symbol plus graphic elements such as border, back pattern, or color that is intended to convey specific information. We can also say that a pictogram is a kind of graph that uses pictures instead of bars to represent data under analysis. A pictogram is also called “pictograph”, or simply “picto”. A pictogram or pictograph represents the frequency of data as pictures of symbols. Each picture or symbols may represent one or more units of data.

#### **CHARACTERISTICS OF PICTOGRAM:**

An image that addresses a word or a thought by representation. Picture charts information is the most significant and significant character in science. Pictograms are a visual way of showing factual information. They are otherwise called pictorial unit outlines, pictographs, and pictorial unit bar graphs. A diagram of information should make us center around the real essence of the information, not on different components like eye-getting but rather diverting plan highlights. Try not

to misshape information develop a diagram to uncover the real essence of the information. Practically the entirety of the ink in a diagram ought to be utilized for the information, not for the other plan components. Diagrams that can be far beyond this concise acquaintance may lead you with acceptance. Pictograms form a part of our daily lives. They are used in transport, medication, education, computers etc.

A pictograph uses picture symbols to convey the meaning of statistical information. Pictographs should be used carefully because the graphs may, either accidentally or deliberately, misrepresent the data. A chart giving statistics in pictorial form. For example, using a dollar in increasing sizes to represent the increase in the purchasing power over time.

### **To successfully convey the meaning, a pictogram:**

- ✓ Should be self-explanatory.
- ✓ Should be recognizable by all people.
- ✓ Must represent a general concept.
- ✓ Should be clear concise and interesting.
- ✓ Should be identifiable as a set, through uniform treatment of scale, style and subject.
- ✓ Should be highly visible, easy to reproduce in any scale and in positive or negative form.
- ✓ Should not be dependent upon a border and should work equally well in positive or negative form.
- ✓ Should avoid stylistic fads or a commercial appearance and should imply to wide audience that has a sophisticated, creative culture.
- ✓ Should be attractive when used with their design, elements and typestyles.

### **ADVANTAGES AND DRAWBACKS OF PICTOGRAM:**

#### **DVANTAGES OF PICTOGRAM:**

- ✓ Pictograms can make warnings more eye-catching.
- ✓ They can serve as an “instant reminder” of a hazard or an established message.
- ✓ They may improve warning comprehension for those with visual or literacy difficulties.
- ✓ They have the potential to be interpreted more accurately and more quickly than words.
- ✓ They can be recognized and recalled far better than words.
- ✓ They can improve the legibility of warnings.

- ✓ They may be better when undertaking familiar routine tasks.
- ✓ Express a large amount of information or data in a simple form
- ✓ Since they make the use of symbols, pictographs attract attention i.e, it is an attractive way to represent data
- ✓ Pictographs are easy to read since all the information is available at one glance
- ✓ And since pictographs are universally used, they do not require a lot of explanation.

### **DRAWBACKS OF PICTOGRAM:**

- ✓ Very few pictograms are universally understood.
- ✓ Even well understood pictograms will not be interpreted equally by all groups of peoples and across all cultures, and it takes years for any pictogram to reach maximum effectiveness.
- ✓ They have the potential for interpreting the opposite or often undesired meaning which can create additional confusion.

### **QUESTION#04**

**Define normal curve. Write down the properties of normal curve?**

**Ans:**

One way of presenting out how data are distributed is to plot them in a graph. If the data is evenly distributed, our graph will come across a curve. In statistics this curve is called a normal curve and in social sciences, it is called the bell curve. Normal or bell curved is distribution of data may naturally occur in several possible ways, with a number of possibilities for standard deviation. A standard normal curve has a mean of 0 and standard of 1. The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side.

A normal distribution is determined by two parameters the mean and the variance. A normal distribution with a mean of 0 and a standard deviation of 1 is called a

standard normal distribution. The normal distribution is the most important probability distribution in statistics because many continuous data in nature and psychology displays this bell-shaped curve when compiled and graphed.

For example, if we randomly sampled 100 individuals, we would expect to see a normal distribution frequency curve for many continuous variables, such as IQ, height, weight and blood pressure.

### **PROPERTIES OF NORMAL CURVE:**

- ✓ The mean, mode and median are all equal.
- ✓ The curve is symmetric at the center (i.e. around the mean,  $\mu$ ).
- ✓ Exactly half of the values are to the left of center and exactly half the values are to the right.
- ✓ The total area under the curve is 1.

A bell curve is a common type of distribution for a variable, also known as the normal distribution. The term "bell curve" originates from the fact that the graph used to depict a normal distribution consists of a symmetrical bell-shaped curve. The highest point on the curve, or the top of the bell, represents the most probable event in a series of data (its mean, mode, and median in this case), while all other possible occurrences are symmetrically distributed around the mean, creating a downward-sloping curve on each side of the peak. The width of the bell curve is described by its standard deviation.

The term "bell curve" is used to describe a graphical depiction of a normal probability distribution, whose underlying standard deviations from the mean create the curved bell shape. A standard deviation is a measurement used to quantify the variability of data dispersion, in a set of given values around the mean. The mean, in turn, refers to the average of all data points in the data set or sequence and will be found at the highest point on the bell curve.

### **NUMERICAL MEASURES OF SHAPE:**

One of the fundamental tasks in any statistical analysis is to characterize the location and variability of a data set. Two important measures of shape, skewness and kurtosis, give us a more precise evaluation of the data. Measures of dispersion tell us about the variation of the data set, while skewness tells us about the direction of variation and kurtosis tells us the shape variation.

### **SKEWNESS:**

Skewness tells us about the amount and direction of the variation of the data set. It is a measure of symmetry. A distribution or data set is symmetric if it looks the same to the left and right of the central point. If bulk of data is at the left i.e. the peak is towards left and the right tail is longer, we say that the distribution is skewed right or positively skewed.

On the other hand, if the bulk of data is towards right or, in other words, the peak is towards right and the left tail is longer, we say that the distribution is skewed left or negatively skewed. If the skewness is equal to zero, the data are perfectly symmetrical. But it is quite unlikely in real world.

### **KURTOSIS:**

Kurtosis is a parameter that describes the shape of variation. It is a measurement that tells us how the graph of the set of data is peaked and how high the graph is around the mean. In other words we can say that kurtosis measures the shape of the distribution, .i.e. the fatness of the tails, it focuses on how returns are arranged around the mean. A positive value means that too little data is in the tail and positive value means that too much data is in the tail.

Kurtosis has three types, mesokurtic, platykurtic, and leptokurtic. If the distribution has kurtosis of zero, then the graph is nearly normal. This nearly normal distribution is called mesokurtic. If the distribution has negative kurtosis, it is called platykurtic. An example of platykurtic distribution is a uniform distribution, which has as much data in each tail as it does in the peak. If the distribution has positive kurtosis, it is called leptokurtic. Such distribution has bulk of data in the peak.

### **Bell Curve vs. Non-Normal Distributions:**

The normal probability distribution assumption doesn't always hold true in the financial world, however. It is feasible for stocks and other securities to sometimes display non-normal distributions that fail to resemble a bell curve.

Non-normal distributions have fatter tails than a bell curve (normal probability) distribution. A fatter tail that skews negative signals to investors that there is a greater probability of negative returns.

### **What Are the Limitations of the Bell Curve?**

Although the bell curve is a very useful statistical concept, its applications in finance can be limited because financial phenomena—such as expected stock-market returns—do not fall neatly within a normal distribution. Therefore, relying too heavily on a bell curve when making predictions about these events can lead to

unreliable results. Although most analysts are well aware of this limitation, it is relatively difficult to overcome this shortcoming because it is often unclear which statistical distribution to use as an alternative.

## **QUESTION#05**

**Explain procedure for determining median, with one example each at least, if:**

- 1. The number of scores is even**
- 2. The number of scores is odd.**

**Ans:**

### **MEDIAN:**

The median is the middle number in a sorted, ascending or descending, list of numbers and can be more descriptive of that data set than the average.

The median is sometimes used as opposed to the mean when there are outliers in the sequence that might skew the average of the values.

Median is the middle number in a sorted list of numbers. To determine the median value in a sequence of numbers, the numbers must first be sorted, or arranged, in value order from lowest to highest or highest to lowest. The median can be used to determine an approximate average, or mean, but is not to be confused with the actual mean.

- If there is an odd amount of numbers, the median value is the number that is in the middle, with the same amount of numbers below and above.
- If there is an even amount of numbers in the list, the middle pair must be determined, added together, and divided by two to find the median value.

The median is sometimes used as opposed to the mean when there are outliers in the sequence that might skew the average of the values. The median of a sequence can be less affected by outliers than the mean.

## **Procedure for Determining Median:**

### **1. The number of scores is even**

When there is an even number of scores in the distribution, arrange the scores in order (from lower to higher or from higher to lower). The median will be the average of the middle two score in the list. Consider the set of scores 4, 6, 9, 14 16, 20. The average of the middle two scores 11.5 (i.e.  $9+14/2 = 23/2 = 11.5$ ) is the median of the distribution.

Consider this set of numbers: 1, 2, 2, 4, 5, 7. Since there is an even number of scores, you need to take the average of the middle two scores, calculating their mean. In this case, the mean would be  $2 + 4$  (add the two middle numbers), which equals 6. Then, you take 6 and divide it by 2 (the total number of scores you added together), which equals 3. So, for this example, the median is 3.

### **2. The number of scores is odd.**

When the number of scores is odd, simply arrange the scores in order (from lower to higher or from higher to lower). The median will be the middle score in the list. Consider the set of scores 2, 5, 7, 10, 12. The score “7” lies in the middle of the scores, so it is median.

Consider this set of numbers: 5, 7, 9, 9, 11. Since you have an odd number of scores, the median would be 9. You have five numbers, so you divide 5 by 2 to get 2.5, and round up to 3. The number in the third position is the median.

Median is less affected by outliers and skewed data and is usually preferred measure of central tendency when the distribution is not symmetrical. The median cannot be determined for categorical or nominal data.

## **Merits of Median:**

- ✓ It is rigidly defined.
- ✓ It is easy to understand and calculate.
- ✓ It is not affected by extreme values.
- ✓ Even if the extreme values are not known median can be calculated.
- ✓ It can be located just by inspection in many cases.
- ✓ It can be located graphically.

- ✓ It is not much affected by sampling fluctuations.
- ✓ It can be calculated by data based on ordinal scale.
- ✓ It is suitable for skewed distribution.
- ✓ It is easily located in individual and discrete classes.

**Demerits of Median:**

- ✓ It is not based on all values of the given data.
- ✓ For larger data size the arrangements of the data in the increasing order is somewhat difficult process.
- ✓ It is not capable for further mathematical treatment.
- ✓ It is not sensitive to some change in the data value.
- ✓ It cannot be used for further mathematical processing