1 Write the definition of statistics and also discuss the limitations of statistics and various functions of statistics.
Ans:

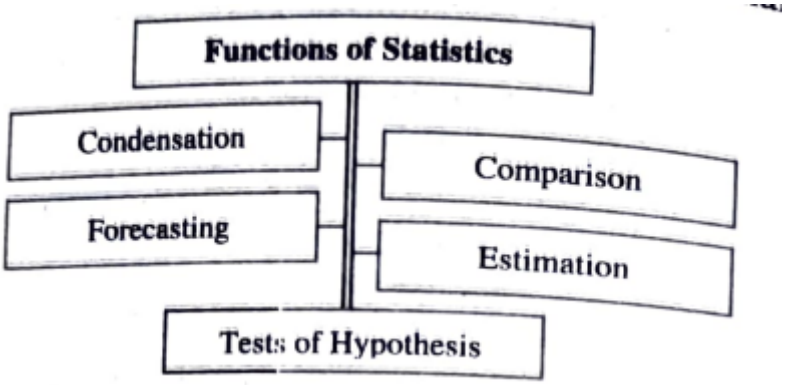
Statistics definition:
Statistical data refers to the collection of quantitative data such as commodity price, production, export, import, births, deaths, etc.

Limitations of statistics:

1. **Statistics does not Deal with Individual Measurements**: The study of individual measurements lies outside the range of statistics as it deals with aggregates of facts. For example, marks obtained by a single student in a class have no statistical significance but the average marks received by the students of the same class are relevant to statistics.
2. **Statistic Deals only with Quantitative Characteristics**: Statistics deals with the quantitative statements of the facts. Various characteristics cannot be expressed quantitatively and hence statistics is unable to analyse these characteristics.
3. **Statistic Results are True only on an Average:** a conclusion which is darwin after statistical study may not be true universally but only under certain conditions
4. **Statistics is only one of the methods of studying a problem:** In all conditions, it is not necessary that statistical tools give the best solution
5. **Statistics can be Misused:** Statistics may be misused when statistical conclusions are based on incomplete information.
6. **Statistics does not Reveal the Entire Story:** Statistics only simplifies and analyses complex data but does not give the real picture of the background about the data.
7. **Statistical Relations do not necessarily bring out the Cause and Effect' Relationship between Phenomena:** Statistics tells the relationships between variables but does not define which one is the cause and which one is the effect.
8. **Statistics is Collected with a given Purpose and cannot be Indiscriminately Applied to any Situation:** The relevance of a

statistical result in one situation does not guarantee its utility in another.

Function of Statis



Condensation: In order to understand the characteristics of a huge amount of data with the help of some observation, 'condensation' is used. For example, if the marks obtained by a particular class of students are mentioned then it serves no purpose. But if the average marks obtained by a group of students are mentioned, it serves a better purpose.

Comparison: The data is condensed with the use of two statistical techniques which are known as 'classification' and 'tabulation'. These methods help in data comparison, collected by various sources. For example, if rice production in two different districts of Uttar Pradesh is known, then it is easier for a comparative study between them. Comparison helps in understanding the data better.

Forecasting: Forecasting refers to the prediction or estimation of the future. The data which is related to the rainfall over the last 10 years in a particular district of Uttar Pradesh helps in predicting rainfall in that region in future.

Tests of Hypothesis: A statistical hypothesis is a statement about some unknown population parameters of a given distribution. The information provided by the sample population is the basis of the statistical hypothesis involved in characterising a population.

2 What is frequency and frequency distribution? What are the reasons for constructing frequency distributions and explain the various types of it.

Ans:

1.4.1. Introduction

Frequency is the number of occurrences of a repeating event per unit time.

Or

Number of times a given quantity (or group of quantities) occurs in a set of data.

1.4.2. Frequency Distribution

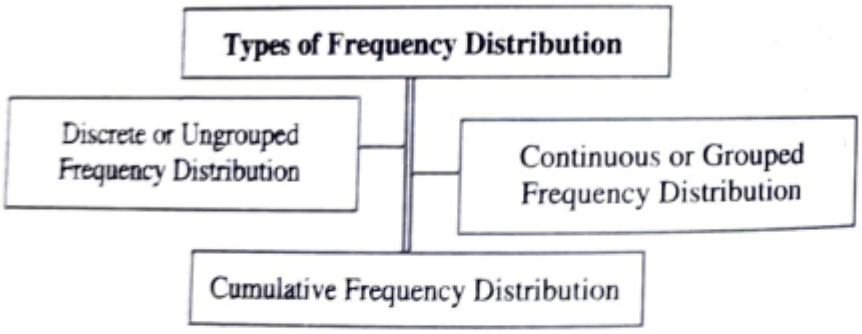
Frequency distribution is a series when a number of observations with similar or closely related values are put in separate bunches or groups, each group being in order of magnitude in a series. It is simply a table in which the data are grouped into classes and the numbers of cases which fall in each class are recorded. It shows the frequency of occurrence of different values of a single phenomenon.

According to Erricker, 'a classification according to the number possessing same value of the variable'.

According to Croxton and Cowden, "Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side".

1.4.4. Types of Frequency Distribution

Basically there are three type of frequency distribution:



1.4.4.1. Discrete or Ungrouped Frequency Distribution

In this form of distribution, the frequency refers to discrete value. Here the data are presented in a way that exact measurement of units is clearly indicated. There is definite difference between the variables of different groups of items. Each class is distinct and separate from the other class. Non-continuity from one class to another class exists. Data such as facts like the number of rooms in a house, the number of companies registered in a country, the number of children in a family, etc.

The process of preparing this type of distribution is very simple. We have just to count the number of times a particular value is repeated, which is called the frequency of that class. In order to facilitate counting prepare a

column for tally marks. In another column, place all possible values of variable from the lowest to the highest. Then put a bar (Vertical line) opposite the particular value to which it relates. To facilitate counting, blocks of five bars are prepared and some space is left in between each block. We finally count the number of bars and get frequency.

Preparation of Simple Frequency Table from Discrete Data

Simple frequency table from discrete data is framed by the following steps:

Step 1: A table of two columns is framed. In the first column, the variables are placed in ascending order.

Step 2: Repetition number of each variable is placed opposite to each variable in the second column as shown in the frequency table below:

Example 1: In a survey of number of coconuts per tree in a field having 30 coconut trees gives the following results.

28, 22, 26, 27, 24, 28, 29, 26, 24, 22, 21, 20, 23, 25, 27,
27, 26, 25, 28, 23, 23, 25, 26, 28, 22, 25, 27, 29, 22, 29

Prepare a discrete frequency distribution table from the above details.

Solution:

Variable	Tally marks	Frequency (f)
20	|	1
21	|	1
22	||||	4
23	|||	3
24	||	2
25	||||	4
26	||||	4
27	||||	4
28	||||	4
29	|||	3
30		0
Total		**N = 30**

Example 2:

1.4.4.2. Continuous or Grouped Frequency Distribution

Continuous series is one where measurements are only approximations and are expressed in class intervals, i.e., within certain limits.

According to **Boddington**, "the variable which can take any intermediate value between the smallest and longest value in the distribution".

In a continuous frequency distribution the class intervals theoretically continue from the beginning of the frequency distribution to the end

without break. The continuous frequency distribution can always be distinguished from the discrete frequency distribution in that it will contain two limits upper limits and lower limits of each class interval while the discrete frequency distribution will possess only one list of classification of values. **For example,** Wage distribution of 100 employees.

Weekly Wages (₹)	Number of Employees
50-100	4
100-150	12
150-200	22
200-250	33
250-300	16
300-350	8
350-400	5
Total	**100**

3 What is Geometric mean? Write it's properties, limitations and Applications. Describe the method of calculating geometric mean for individual, discrete, continuous series.

Ans:

The geometric mean is a measure of central tendency that is used to describe the average value of a set of numbers by multiplying them together and taking the nth root of the result, where n is the number of numbers in the set. It is often used to describe the average growth rate of an investment or a portfolio of investments over a specified period of time.

Properties of the Geometric Mean:

1. It is always less than or equal to the arithmetic mean.
2. The geometric mean is only defined for positive numbers.
3. It is a measure of central tendency that is less sensitive to outliers than the arithmetic mean.
4. It is always less than the harmonic mean for sets of two or more numbers.

Limitations of the Geometric Mean:

1. The geometric mean is not defined for negative numbers or for sets of numbers that include zero.
2. The geometric mean can be influenced by outliers and extreme values.

Applications of the Geometric Mean:

1. The geometric mean is commonly used in finance to describe the average growth rate of an investment or a portfolio of investments over a specified period of time.
2. It is also used in biology to describe the average growth rate of populations.
3. The geometric mean is used in engineering and physics to describe the average of logarithmically distributed data.

Method of calculating Geometric Mean for individual numbers:

The formula for calculating the geometric mean of individual numbers is given by:

Geometric Mean = (x1 * x2 * ... * xn)^(1/n)

Where x1, x2, ..., xn are the individual numbers in the set and n is the number of numbers in the set.

Method of calculating Geometric Mean for discrete series:

The formula for calculating the geometric mean of a discrete series is given by:

Geometric Mean = (f1 * f2 * ... * fn)^(1/n)

Where fi is the frequency of the ith value in the set and n is the number of values in the set.

Method of calculating Geometric Mean for continuous series:

The formula for calculating the geometric mean of a continuous series is given by:

Geometric Mean = exp(sum(ln(x))/n)

Where x is the value in the set, exp is the exponential function and ln is the natural logarithm function, and n is the number of values in the set.

4 Discuss the types of measures of central tendency. What are the requisites for ideal measures of central tendency? Also list the various application of it.

Ans:

Measures of central tendency, also known as measures of location, are used to describe the central or typical value of a set of data. The most common measures of central tendency are the mean, median, and mode.

Types of Measures of Central Tendency:

1. Mean: The mean is the sum of all values in a set of data divided by the number of values in the set. It is also known as the arithmetic average.
2. Median: The median is the middle value in a set of data when the values are ordered from least to greatest. It is a robust measure of central tendency, meaning that it is not greatly affected by outliers or extreme values.

3. Mode: The mode is the value that occurs most frequently in a set of data. A set of data may have one mode, more than one mode, or no mode.

Requisites for Ideal Measures of Central Tendency:

1. Unbiased: The ideal measure of central tendency should not be biased towards any specific value or group of values in the data set.
2. Sensitivity to Outliers: The ideal measure of central tendency should not be greatly affected by outliers or extreme values in the data set.
3. Easy to Compute: The ideal measure of central tendency should be easy to calculate and understand.

Applications of Measures of Central Tendency:

1. Descriptive Statistics: Measures of central tendency are commonly used in descriptive statistics to describe the central or typical value of a set of data.
2. Inferential Statistics: Measures of central tendency are also used in inferential statistics to make predictions about a population based on a sample of data.
3. Business and Economics: Measures of central tendency are commonly used in business and economics to describe the central or typical value of a set of data, such as sales data, economic data, or financial data.
4. Health and Medical Research: Measures of central tendency are commonly used in health and medical research to describe the central or typical value of a set of data, such as patient data, clinical trial data, or population health data.
5. Environmental Science: Measures of central tendency are also used in environmental science to describe the central or typical value of a set of data, such as atmospheric data, water quality data, or wildlife population data.

5 What is range? Write it's advantages and disadvantages. Also discuss the co-efficient of range and describe the method of calculating range.
Ans:

Range is a measure of dispersion that describes the difference between the largest and smallest values in a set of data. It provides a simple and easy-to-calculate measure of how spread out the data is.

Advantages of Range:

1. Simple to calculate: Range is easy to calculate and requires only the knowledge of the largest and smallest values in the data set.
2. Easy to interpret: Range provides a straightforward measure of dispersion that is easy to understand and interpret.

Disadvantages of Range:

1. Sensitivity to outliers: Range can be greatly affected by outliers or extreme values in the data set, making it an unreliable measure of dispersion for data sets with a large number of outliers.
2. Not informative about the distribution of data: Range provides no information about the distribution of the data within the range, making it an insufficient measure of dispersion for data sets with complex distributions.

Coefficient of Range:

The coefficient of range is a measure of dispersion that is calculated by dividing the range by the mean or median of the data set. It provides a relative measure of dispersion that is useful for comparing the dispersion of different data sets.

Method of Calculating Range:

To calculate the range, simply subtract the smallest value in the data set from the largest value in the data set. The formula for range is:

Range = Maximum Value - Minimum Value

Where Maximum Value is the largest value in the data set and Minimum Value is the smallest value in the data set.

6 What is quartile? Define the quartile deviation and co-efficient of quartile deviation. Compute Quartile Deviation and co-efficient of quartile Deviation for the following data.
24,14,21,100,104,68,34,106,100,72,16,21,14,21,72, 100,106,34,100
Ans:

Quartile is a measure of dispersion that divides a set of data into four equal parts or quarters. Quartiles are commonly used to describe the distribution of a data set and to identify the range of values that contain a certain percentage of the data.

There are three types of quartiles:

1. The first quartile, also known as Q1, represents the 25th percentile of the data and is the value that separates the lowest 25% of the data from the rest.
2. The second quartile, also known as Q2 or the median, represents the 50th percentile of the data and is the middle value in the data set.
3. The third quartile, also known as Q3, represents the 75th percentile of the data and is the value that separates the highest 25% of the data from the rest.

Quartile Deviation:

Quartile deviation, also known as interquartile range (IQR), is a measure of dispersion that is calculated as the difference between the third quartile (Q3) and the first quartile (Q1). It provides a measure of the spread of the middle 50% of the data.

Coefficient of Quartile Deviation:

The coefficient of quartile deviation, also known as the coefficient of IQR, is a measure of dispersion that is calculated by dividing the IQR by the median (Q2) of the data set. It provides a relative measure of dispersion that is useful for comparing the dispersion of different data sets.

Computing Quartile Deviation and Coefficient of Quartile Deviation:

To compute the quartile deviation and the coefficient of quartile deviation, follow these steps:

1. Order the data set from smallest to largest.
2. Compute Q1, Q2, and Q3. To do this, you can use the following formulas:

- Q1 = 25th percentile = (N + 1) * 0.25, where N is the number of values in the data set.
- Q2 = 50th percentile = (N + 1) * 0.5, where N is the number of values in the data set.
- Q3 = 75th percentile = (N + 1) * 0.75, where N is the number of values in the data set.

3. Compute the IQR by subtracting Q1 from Q3. IQR = Q3 - Q1
4. Compute the coefficient of quartile deviation by dividing the IQR by the median (Q2). Coefficient of Quartile Deviation = IQR / Q2

For the given data set:

24, 14, 21, 100, 104, 68, 34, 106, 100, 72, 16, 21, 14, 21, 72, 100, 106, 34, 100

Number of values in the data set (N) = 19

Order the data set:

1. 14, 14, 16, 21, 21, 21, 24, 34, 34, 68, 72, 72, 100, 100, 100, 104, 106, 106
2. Compute Q1, Q2, and Q3:

- Q1 = (N + 1) * 0.25 = (19 + 1) * 0.25 = 5.25 = 21
- Q2 = (N + 1) * 0.5 = (19 + 1) * 0.5 = 10
- Q3 = (N + 1) * 0.75 = (19 + 1) * 0.75 = 14.75 = 100

7 Write a short note on Tabulation.
Ans
Tabulation is a method of organizing data into a table or chart for easier understanding and analysis. Tabulation is used to summarize large and complex sets of data, making it easier to see patterns, relationships, and trends in the data. In tabulation, data is organized into columns and rows, with each row representing an individual observation and each column representing a variable or category.

8 Write a note on weighted Average.
Ans
Weighted average, also known as weighted mean, is a statistical measure that calculates the average value of a set of numbers by assigning different weights to each number based on its relative importance. Unlike the traditional average, which assigns equal weights to each number in the data set, the weighted average gives more emphasis to certain values based on their relative importance.

The formula for calculating the weighted average is:

Weighted average = (sum of (weight of each value * value)) / (sum of weights)

9 Write a short note on standard deviations.
Ans
Standard deviation is a measure of the variability or dispersion of a set of data around the mean. It is calculated as the square root of the variance, which is the average of the squared deviations of each value in the data set from the mean.

The standard deviation gives a sense of how spread out the values in the data set are. If the standard deviation is small, it means that the values are close to the mean, while a large standard deviation indicates that the values are spread out over a wider range. The standard deviation can be used to compare the dispersion of two or more data sets, to determine the level of risk or uncertainty in a data set, and to make inferences about a population based on a sample.

The standard deviation is often represented by the symbol "σ" (sigma) and is calculated using the following formula:

σ = √(Σ(x - μ)^2 / n)

where x is each value in the data set, μ is the mean of the data set, and n is the number of values in the data set.

Standard deviation is a widely used measure of variability in various fields, including statistics, finance, and engineering. It is a useful tool for summarizing and analyzing data, and for making inferences about a population based on a sample. However, it also has its limitations, such as being sensitive to outliers and being affected by the sample size.

10 Give a note on "Distrust of Statistics".

Ans

Distrust of statistics refers to the skepticism or skepticism towards the use and interpretation of statistical data and results. This distrust can arise from several factors, including:

1. Misinterpretation or misuse of statistics: Statistics can be misinterpreted or misused by individuals who do not have a deep understanding of statistical methods, leading to incorrect conclusions and decision making.
2. Bias and manipulation: Statistics can be manipulated to support a particular agenda or bias, leading to a distrust in the validity of the data and results.
3. Data quality: The quality of data collected and used in statistical analysis can affect the validity and reliability of results, leading to distrust if the data is inaccurate or incomplete.
4. Subjectivity in data analysis: The choice of statistical methods and interpretation of results can be subjective, leading to distrust if the results are perceived to be influenced by personal or political bias.
5. Conflicting results: Different statistical methods or data sources can produce conflicting results, leading to distrust in the accuracy and reliability of statistical data.

Distrust of statistics can have serious implications, such as hindering decision making and policy making based on accurate data, and can lead to public mistrust in the scientific and academic communities. To overcome distrust in statistics, it is important to ensure the transparency and accuracy of data collection and analysis, and to promote critical thinking and education about the use of statistics.

11 Short note on Harmonic mean

Ans

Harmonic mean is a measure of central tendency that is used to calculate the average value of a set of numbers. Unlike the arithmetic mean, which is calculated by summing the values and dividing by the number of values, the harmonic mean is calculated by taking the reciprocal of each value, summing the reciprocals, and then taking the reciprocal of the sum.

The formula for calculating the harmonic mean of a set of numbers is:

Harmonic mean = n / (Σ(1/x))

12 Write a short note on Arithmatic mean.

Ans

Arithmetic mean, also known as the average, is a measure of central tendency that is used to describe the average value of a set of numbers. It is calculated by summing all the values in a data set and dividing the sum by the number of values.

The formula for calculating the arithmetic mean of a set of numbers is:

Arithmetic mean = (Σx) / n