Question: what the effect will occure in the result of mean, median, mode, varience, and standred deviation if we add or subtract a number from the data?

Answer: mean is not effected by change of origion(addition or subtraction of any no) and change of scale(multiplication or dividion of any no) but variance is effected by change of origion and scale.with change of origion it will remain unchanged i-e v(X+Y)=v(X)+v(Y) or v(X-Y)=v(X)+v(Y) and when one number is multiplied or divided then it will become double.i-e v(5X)=25v(X) or v(1/5x)=1/25v(X)

Question: what is practical importances of median and quartile or in which cases these comodities are used?

Answer: Median is one of the measure of central location. It is good to use the median when a frequency distribution involves "open end" classes like those of income and prices. In a highly skewed distribution, median is an appropriate average to use as it is not affected by extreme values. It can be located when the values are not capable of quantitative measurement. While quartiles are used when the same nature of data is to be dealt with but they are used to divide the data into four equal parts.

Question: What is the important of relation between Arithmetic, Geometric and harmonic?

Answer: Relation between arithmetic mean, geometric mean and harmonic mean is given below: Arithmetic Mean > Geometric Mean >Harmonic Mean I.e. for a data arithmetic mean is greater than geometric mean and harmonic mean. And geometric mean is greater than harmonic mean.

Question: state what is Grouped and Row data?

Answer: Grouped data The data presented in the form of frequency distribution is also known as grouped data. Raw data Data that have not been processed in any manner. It often refers to uncompressed text that is not stored in any priority format. It may also refer to recently captured data that may have been placed into a database structure, but not yet processed.

Question: how will decide number of classes and class interval for the given data?

Answer: There are no hard and fast rules for deciding on the number of classes which actually depends on the size of data. Statistical experience tells us that no less than 5 and no more than 20 classes are generally used. Use of too many classes will defeat the purpose of condensation and too few will result in too much loss of information. Deciding on the number of classes does not depend on the value of range. In the given example no. of classes was chosen 8. It is chosen with respect the size of data. It is not decided after seeing the value of range which is 1.38 in this example. To find class interval ‘h’ we should first find the range and divide it by number of classes

Question: define theMean Deviation.

Answer: The mean deviation is used to characterize the dispersion among the measures in a given population. To calculate the mean deviation of a set of scores it is first necessary to compute their average (mean or median) and then specify the distance between each score and that mean without regard to whether the score is above or below(negative and positive) the mean. The mean deviation is defined as the mean of these absolute values.

Question: What is meant by variability?

Answer: Variability is the spread or dispersion in a set of data. Consider the following sets of data. 9, 9, 9, 9, 9, 9, 9, 9, 9, 9 10, 6, 2, 8, 4, 14, 16, 12 13, 10, 7, 6, 21, 3, 7, 5 All these three sets of data have same mean (9) but they are different in variability. First set of values has no dispersion and there is greater variability is third data set as compared to second set of data as its values are more spread away as compared to the values of second set of data.

Question: What is EQUALLY LIKELY EVENTS?

Answer: The two events are said to be equally likely if they have the same chance of occurring. For example,

in our coin-tossing experiment, the two events, heads and tails, are equally likely. Both have the same chances of occurring. There is 50% chance for occurring both events.

Question: what is the difference between dependent and independent event?

Answer: Independent and Dependent Events By independent we mean that the first event does not affect the probability of the second event. Coin tosses are independent. They cannot affect each other's probabilities; the probability of each toss is independent of a previous toss and will always be 1/2. Separate drawings from a deck of cards are independent events if you put the cards back. An example of a dependent event, one in which the probability of the second event is affected by the first, is drawing a card from a deck but not returning it. By not returning the card, you've decreased the number of cards in the deck by 1, and you've decreased the number of whatever kind of card you drew. If you draw an ace of spades, there are 1 fewer aces and 1 fewer spades. This affects our simple probability: (number of favorable outcomes)/ (total number of outcomes. This type of probability is formulated as follows: If A and B are not independent, then the probability of A and B is P(A and B) = P(A) × P(B|A) where P(B|A) is the conditional probability of B given A. Example If someone draws a card at random from a deck and then, without replacing the first card, draws a second card, what is the probability that both cards will be aces? Solution Event A is that the first card is an ace. Since 4 of the 52 cards are aces, P(A) = 4/52 = 1/13. Given that the first card is an ace, what is the probability that the second card will be an ace as well? Of the 51 remaining cards, 3 are aces. Therefore, p(B|A) = 3/51 = 1/17, and the probability of A and B is 1/13 × 1/17 = 1/221.

Question: explain the Conditional Probability.

Answer: Conditional Probability In many situations, once more information becomes available; we are able to revise our estimates for the probability of further outcomes or events happening. For example, suppose you go out for lunch at the same place and time every Friday and you are served lunch within 15 minutes with probability 0.9. However, given that you notice that the restaurant is exceptionally busy, the probability of being served lunch within 15 minutes may reduce to 0.7. This is the conditional probability of being served lunch within 15 minutes given that the restaurant is exceptionally busy. The usual notation for "event A occurs given that event B has occurred" is "A | B" (A given B). The symbol | is a vertical line and does not imply division. P(A | B) denotes the probability that event A will occur given that event B has occurred already. A rule that can be used to determine a conditional probability from unconditional probabilities is: P(A|B) = P(AÇB)/P(B) Where: P(A | B) = the (conditional) probability that event A will occur given that event B has occurred already P(AÇB) = the (unconditional) probability that event A and event B both occur P(B) = the (unconditional) probability that event B occurs

Question: what is the difference between mutually exclusive events & exshastive events with the help of an example?

Answer: If only one of two or more events can occcur, the evetns are called mutually exclusive events. For example, in our coin-tossing experiment, the two evetns, heads and tails, are mutually exclusive: if one occurs, the other cannot occur. When a set of events for an experiment includes every possilbe outcome, the set is said to be collectively exhaustive. Thus, heads and tails are a collectivly exhaustive set of evetns for our coin-tossing experiment. One requirement we place on probability numbers is that the sum of the probalities for a collectively exhaustive set of mutuallly exclusive events be equal to 1.

Question: explain the concept of "Central limit theorem".

Answer: The central limit theorem states that given a distribution with a mean μ and variance s2, the sampling distribution of the mean approaches a normal distribution with a mean (μ) and a variance s2/n as ‘n’, the sample size, increases. The amazing thing about the central limit theorem is that no matter what the shape of the original distribution, the sampling distribution of the mean approaches a normal distribution. Furthermore, for most distributions, a normal distribution is approached very quickly as ‘n’ increases.

Question: explain "P -value"? with examples.

Answer: Each statistical test has an associated null hypothesis, the p-value is the probability that your sample

could have been drawn from the population(s) being tested (or that a more improbable sample could be drawn) given the assumption that the null hypothesis is true. A p-value of .05, for example, indicates that you would have only a 5% chance of drawing the sample being tested if the null hypothesis was actually true.

Question: what is the difference between average and Central tedency?

Answer: A single value used to represent the distribution is called average. Most commonly used averages are Mean, Median and Mode. And measures of dispersion are used to measure how the data are dispersed about the average. For example, it is quite possible that two or more sets of data may have the same average (mean, median and mode) but their individual observations may differ considerably from the average. Thus the value of central tendency does not describe data. So we therefore need some additional information concerning with how the how the data are dispersed about the average. There are several measures of dispersion, the most common being the range, quartile deviation, mean deviation and standard deviation.

Question: what is the coefficient of variation?

Answer: Co-efficient of variation is used to compare the variability and to check the consistency of two or more series. It is most commonly used relative measure of dispersion. Symbolically, the coefficient of variation, denoted by C.V., is given by C.V = [Standard deviation / Arithmetic mean] × 100 It is used as a criterion of consistent performance; the smaller the coefficient of variation, the more consistent is the performance It is also used as the criterion of variability; the larger the coefficient of variation, the more variability in the data.

Question: define mean deviation from median.

Answer: Mean deviation from median deviation is defined as the average of the deviations of the values from median; the deviations are taken without considering algebraic signs. The median deviation of a set of n values X1, X2, Xn, denoted by M.D., is given by M.D = summation | X – median | /n Where | X – median | indicate the absolute deviations of the observations from the median of a sample

---

Question: which line is best fitted line, how we can juge?

Answer: According to the principal of least squares, the best-fitting line to a set of points is the one for which the sum of the squares of the vertical distances between the points and the line is minimum

---

Question: What are Covariance & Correlation?

Answer: The Covariance of two r.v.'s X and Y is a numerical measure of the extent to which their values tend to increase or decrease together. The correlation is used to describe the degree to which one variable is linearly related to another. Often, correlation is used in conjunction with regression to measure how well the regression line explains the variation of the dependent variable; Y. Correlation can also be used by itself, however, to measure the degree of association between two variables. Statisticians have developed measure for describing the correlation between two variables: the coefficient of correlation.

Question: describe the method of least squares in details.

Answer: The Method of least squares (LS): The method of least squares (LS) consists of determining the values for the unknown parameters that will minimize the sum of squares of errors (or residuals) where errors are defined as the difference between observed values and the corresponding values predicted or estimated by the fitted model equation. The observed Yi may be expressed in a linear form of the population parameter as Yi= a + ßXi+ei, Or in terms of sample data Yi= a +b Xi+ei,

Where a and b are least square estimates of parameters a and ß, ei called residual is the deviation of the observed Yi from the estimate provided by ?i = a + bXi, According to the method of least square we determine those values of a and b which will minimize the sum of square of residuals. Summation [Yi – (a+bx)]2

Question: explain the steps of a hypothesis testing ?

Answer: General procedure for testing hypotheses: The procedure for testing hypotheses about a population parameter involves following six steps. State your problem and formulate an appropriate hypothesis Hº with an alternative hypothesis H1, which is to be accepted when Hº is rejected. Decide upon significant level, a of the test, which is the probability of rejecting the null hypothesis if it is true. Choose an appropriate test- statistic, determine and sketch the sampling distribution of the test-statistic, assuming Hº is true. Determine the rejection or critical region in such a way that the probability of rejecting the null hypothesis Hº , if it is true, is equal to the significance level, a. The location of the critical region depends upon the form of H1. The significance level will separate the acceptance region from the rejection region. Compute the value of the test- statistic from the sample data in order to decide whether to accept or reject the null hypothesis Hº. Formulate the decision rule as below: a) Reject the null hypothesis Hº if the computed value of the test- statistic falls in the rejection rejoin and conclude that H1 is true. b) Accept the null hypothesis Hº otherwise

Question: What is the different between percantage and Angle pie chart and why we draw pie chart in 360 angle?

Answer: The difference between and percantage and Angle Pie chart is that in percent we represent thing on a scale of 100, that is out of 100. For example, we say that I got 85 % marks in an examination. It means that we have divided the numbers that we secured by the total marks and then multiplied it by 100 in order to get % marks. In Angle Pie Chart, we are representing things on a scale of 360 that is out of 360. Why? Pie chart is circular and a circle has 360 degrees, that is the central angle is of 360 degrees. So if you want to make a pie chart for the marks that you got in Matric. You will divide the maks that you got in Physics by the total marks in English and then multiply it with 360. You will carry out this process with all the remaining subjects in order to get the exact angle or sector.

Question: What is the different between percantage and Angle pie chart and why we draw pie chart in 360 angle?

Answer: The difference between and percantage and Angle Pie chart is that in percent we represent thing on a scale of 100, that is out of 100. For example, we say that I got 85 % marks in an examination. It means that we have divided the numbers that we secured by the total marks and then multiplied it by 100 in order to get % marks. In Angle Pie Chart, we are representing things on a scale of 360 that is out of 360. Why? Pie chart is circular and a circle has 360 degrees, that is the central angle is of 360 degrees. So if you want to make a pie chart for the marks that you got in Matric. You will divide the maks that you got in Physics by the total marks in English and then multiply it with 360. You will carry out this process with all the remaining subjects in order to get the exact angle or sector.

Question: How can we define Cumalative destribution?

Answer: The cumulative distribution is as follows Classes f c.f 5-10 2 2 10-15 4 6 15-20 3 9 20-25 1 10 Total 10

Question: What is meant by Transformation?

Answer: If we change one variable into another variable, this is called transformation. For example, If we have values of variable X, then we can find the values of other variables using transformations like Y = X + 3 or Z = 2X - 5

Question: what is difference b/w Geometric mean and Harmonic means?

Answer: The Geometric Mean is used primarily to average data for which the ratio of successive terms remain approximately constant. This occurs with data as rates of change, ratios, economic index numbers, and population sizes over successive time periods and the like. On the other hand, Harmonic Mean is most frequently used in average speeds of various distances covered. Where the distances remain constant, and also in finding the average cost of commodity, such as mutual funds, when several different purchases are made by investing the same amount of money each time.

Question: what is a logistic system and what is a inventry?

Answer: Logistics system: The total flow of products from the acquisition of raw materials to the delivery of finished goods to users, including the related flow of information that controls and records the movement of those products. Inventory System: Procedures that govern how supplies are received, stored, handled, and issued is called inventory system.

Question: Define the standard normal distribution.

Answer: The Standard normal distribution: A normal distribution whose mean is zero and whose standard deviation is 1 is known as the standard normal distribution. This distribution has a very important role in computing areas under the normal curve. The reason is that the mathematical equation of the normal distribution is so complicated that it is not possible to find areas under the normal curve by ordinary integration. Areas under the normal curve have to be found by the more advanced method of numerical integration. The point to be noted is that areas under the normal curve have been computed for that particular normal distribution whose mean is zero and whose standard deviation is equal to 1, i.e. the standard normal distribution.

Question: Explain the concept of Continuity Correction.

Answer: Continuity Correction: In testing of hypothesis, we use a continuity correction of ±1/2 whenever we consider the normal approximation to the binomially distributed random variable X. Because the normal distribution can take all real numbers (is continuous) but the binomial distribution can only take integer values (is discrete) so there for in using normal curve areas to approximate binomial probabilities, a discrete value of the binomial variable is to be replaced by an interval before the z values are computed. Accordingly a discrete value x becomes the interval from x-0.5 to x+0.5 and this sort of adjustment is called continuity correction. Thus, the discrete value 5, adjusted means 4.5 to 5.5.

Question: Define confidence interval.

Answer: Confidence interval: A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. In interval estimation of an unknown population parameter we find an interval for which we have very high confidence (probability) that it contains the unknown parameter. This level of confidence is dented by (1 – a). It is always very high probability usually 95%, 98%, 99% etc. and the chance that our interval does not contain unknown parameter is called level of significance and it is denoted by a.

Question: Component bar chart e. What is range?

Answer: Component bar charts: When you want to draw a bar chart to illustrate your data, it is often the case that the totals of the figures can be broken down into parts or components. You start by drawing a simple bar chart with the total figures. The columns or bars are then divided into the component parts. Remember to put a key on the diagram. Range: It is the difference between the largest and the smallest value of the data. Suppose largest value is denoted by Xm and smallest value is denoted by

Xo then, range is given by, Range = Xm - Xo

Question: Explain " Weighted Mean" .

Answer: Weighted Mean Weighted Mean is used when one is concerned with averaging average values, which is best explained with an example: If I have a class of 30 students for whom the mean score on a test is 75, and another class of 50 students for whom the mean score is 80; then the sum of all the scores is (30)(75)+(50)(80), hence the overall mean is ((30)(75)=(50)(80))/(30+50) = 78.12. This is obtained by "weighting" the means by 30 and 50, respectively, and dividing by the sum of the weights. The result is called the weighted mean of the means.

Question: Describe the properties of the discrete probability distribution.

Answer: Properties of a Discrete Probability Distribution: First property of probability distribution is that $0 < P(X) < 1$ It means that the value of probability always lies between 0 and 1. i.e. the value of probability can neither be negative nor it can exceed 1. Sigma represents the standard deviation of the probability distribution.

Question: when to use Geometric mean .

Answer: The geometric mean is a measure of central tendency it uses multiplication rather than addition to summarize data values. The geometric mean is a useful summary when we expect that changes in the data occur in percentages. For example adjustments in salary are often a percentage amount. Geometric means are often useful summaries for highly skewed data. They are also natural for summarizing ratios. Don't use a geometric mean, though, if you have any negative or zero values in your data

---

Question: Explain point estimator and method of least squares.

Answer: Point estimator: A point estimator is a formula or expression producing a single value estimate of the population parameter. Method of least squares: A method of determining the curve that best describes the relationship between expected and observed sets of data by minimizing the sums of the squares of deviation between observed and expected values.

Question: Explaine clearly the concept of "statistical inference" and also the concept of estimation.

Answer: Statistical inference is inference about a population from a random sample drawn from it. point estimation interval estimation hypothesis testing (or statistical significance testing) prediction Estimation is the procedure by which we obtain an estimate of the unknown population parameter using sample data. For example we may estimate the mean and the variance of population by computing the mean and the variance of a sample drawn from the population

Question: In which condition we use multiple bar chart and component bar chart?

Answer: Multiple bar chart: A multiple bar chart shows two or more characteristics corresponding to the values of a common variable. Component bar chart. While the component bar chart is effective technique in which each bar is divided into two or more sections.

Question: Statistic and statistics.

Answer: Statistic and statistics: The word statistics is used as the plural of the word statistic, which means some numerical quantity calculated from sample observations. You can say that in the plural sense it refers to a collection of numerical facts and in singular sense, denotes the science of basing decision on numerical data.

Question: What is sampling & how many types of samling?

Answer: Sampling: It is a statistical technique which is used in almost every filed to collect information about the population. Types: 1. Probability sampling 2. Non - probability sampling

---

Question: What are the limitations of statistics .

Answer: Limitations: Statistics deal with the aggregate of observations of the same kind but it has nothing to do with what is happening to a particular individual or object of the aggregate .The statistical laws are valid in the long run but there is no guarantee that a certain law will hold in all cases. Statistical results might be misleading due to incorrect collection, processing and interpreting the data.

Question: What is meant by "attribute"?

Answer: Attributes: Qualitative characteristics of variable is also called attributes e.g.poverty,intelligence etc.

Question: What is proper definition of DATA?

Answer: DATA: Collection of facts & figure for specific purpose.

Question: Explain what is Ratio scale?

Answer: RATIO SCALE: Values on a ratio scale have the features of an interval scale but also have an absolute zero point, allowing ratio statements. Absolute or true zero point: corresponds to the absence of the thing being measured.

Question: What is the interval scale?

Answer: INTERVAL SCALE: A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale.

Question: Explain the measurment scals.

Answer: NOMINAL SCALE: The classification or grouping of the observations into mutually exclusive qualitative categories or classes is said to constitute a nominal scale. A set of data is said to be nominal if the values / observations belonging to it can be assigned a code in the form of a number where the numbers are simply labels. You can count but not order or measure nominal data For example; students are classified as male and female. Number 1 and 2 may also be used to identify these two categories. Similarly, rainfall may be classified as heavy moderate and light. We may use number 1, 2 and 3 to denote the three classes of rainfall. The numbers when they are used only to identify the categories of the given scale carry no numerical significance and there is no particular order for the grouping. ORDINAL SCALE: A measurement scale for which the relative values of data are defined solely in terms of being lesser, equal-to or greater as compared with other data on the ordinal scale. It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurements. For example, the performance of students (or players) is rated as excellent, good fair or poor, etc. Number 1, 2, 3, 4 etc. are also used to indicate ranks RATIO SCALE: It is a special kind of an interval scale where the sale of measurement has a true zero point as its origin. Such a scale will have a zero point which is meaningful in the sense that it indicates complete absence of the property which the scale measures. The ratio scale is used to measure weight, volume, distance, money, etc.

Question: What is meant by "cumulative or systematic errors"?

Answer: Cumulative or Systematic Errors: An error is said to be biased when the observed value is consistently and constantly higher or lower than the true value. Biased errors arise from the personal limitations of the observer, the imperfection in the instruments used or some other conditions which control the measurements. These errors are not revealed by repeating the measurements. They are cumulative in nature, that is, the greater the number of measurements, the greater would be the magnitude of error. They are thus more troublesome. These errors are also called cumulative or systematic errors.

Question: What are the points which should be kept in view while constructing a graph?

Answer: The following points should be remembered while constructing the graph. Use clear titles and indicate when and how the data were collected (i.e. the theme of the graphs and the source of data). Ensure that the scales are clear, understandable and represent the data accurately. When possible, use symbols for extra data. Always keep in mind the reason why a graph is being used (i.e. to highlight some information or data in a striking and unambiguous way) and anything that facilitates this objective is desirable.

Question: What is variable? Explain its types.

Answer: Variable: A characteristic that can vary or differ is called a variable, such as age, location, or education level. It is a term used in statistics to describe the factors that are to be studied. A variable can be classified in to qualitative and quantitative according to the form of characteristics of interest. Qualitative variable: A variable is called a quantitative variable when a characteristic can be expressed numerically such as age, weight, income or number of children. Quantitative variable: A variable is called a quantitative variable when a characteristic can not be expressed numerically such as education, sex, eye-color, quality, intelligence, poverty, satisfaction, etc A quantitative variable can be classified in to two categories, discrete variable and continuous variable. Discrete variable: A discrete variable can assume values by counting process, such as the number of persons in a family, the number of rooms in a house, the number of deaths in an accident, the income of an individual, etc. It can take on only a discrete set of integers or whole numbers. Continuous variable: A continuous variable can assume values by measuring process such as such as the age of a person, the height of a plant, the weight of a commodity, the temperature at a place, etc. It can take on any value-fractional or integral—within a given interval.

Question: Explain Sampled population and Target population.

Answer: Sampled population - The population from which the sample is taken. Target population - The population about which inferences are made. I explain u with the help of the following example Suppose we want to know the opinions of college students in the province of Punjab with regard to the present examination system. Then our Population will consist of the total number of students in all the colleges of Punjab. Suppose we conduct a survey only on five colleges through out the province due to shortage of resources. In such a case, the target population consists of the students of all the colleges in Punjab while on the other hand, the sampled populations consists of students of five colleges. The students of these five colleges are the representative of the students of all the colleges; the result would be applicable to all the colleges.

Question: Differentiate between inferential and descriptive statistics.

Answer: Descriptive Statistics: It is that branch of statistics which deals with concepts and methods concerned with summarization & description of the important numerical data. Inferential Statistics: It deals with procedures for making inferences about the characteristics of the larger group of data or the whole called the population, from the knowledge derived from only the part of data.

Question: define infrentional statistics.

Answer: Inferential Statistics uses sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data (population). It is further divided in two main areas: 1 Estimation 2 Testing of Hypothesis

Question: What is observation give details?

Answer: OBSERVATION: In statistics, an observation often means any sort of numerical recording of information, whether it is a physical measurement such as height or weight; a classification such as heads or tails, or an answer to a question such as yes or no.

Question: What is random error?

Answer: RANDOM ERROR: An error is said to be unbiased or random error when the deviations, i.e. the excesses and defects, from the true value tend to occur equally often. Unbiased errors and revealed when measurements are repeated and they tend to cancel out in the long run. These errors are therefore compensating and are also known as accidental errors.

Question: What is the relation between Probability and Statistics?

Answer: Probability and statistics are fundamentally interrelated. Probability is often called the vehicle of statistics. The area of inferential statistics in which we are mainly concerned with drawing inferences from experiments or situations involving an element of uncertainty, leans heavily upon probability theory.

Question: What is Random sampling and its is usefullness.

Answer: Random sampling : A process for obtaining a sample from a population that requires that every individual in the population has the same chance of being selected for the sample.It is widely used in various areas such as industry, business etc.

Question: Explain the frames or sample frames and what are the uses of them in Statistics ?

Answer: sampling frame or simply frame is a complete list or a map that contains all the sampling units of population, for example, a complete list of the names of all the students in the Virtual University. A list of all households in a city, a map of a village showing all fields, etc. The requirements of a good frame are 1 It does not contain inaccurate sampling units. 2 It is complete 3 It is free of errors 4 It is as up-to-date as possible at the time of use. It helps us to select a good sample from the population.

Question: What is error of measurement and give detail with real life examples?

Answer: ERRORS OF MEASUREMENT: Experience has shown that a continuous variable can neverbe measured with perfect fineness because of certain habits and practices, methods of easurements, instruments used, etc. the measurements are thus always recorded correct to the nearest units and hence are of limited accuracy. The actual or true values are, however, assumed to exist. Example: If a student's weight is recorded as 60 kg (correct to the nearest kilogram), his true weight in fact lies between 59.5 kg and 60.5 kg, whereas a weight recorded as 60.00 kg means the true weight is known to lie between 59.995 and 60.005 kg. Thus there is a difference, however small it may be between the measured value and the true value.

Question: What are the advantages of sampling in daily life?

Answer: Sampling is procedure of taking sample from population. It is not a place or an area which can be empty or full. We are all familiar with the idea of sampling in our everyday life. A cook takes a bit of

the cooked food to see whether it has been properly cooked. Customers, by observation, sample the quality of fruits and vegetables they intend to buy. A food inspector takes a sample of the food items like milk, oil, flour, etc.to find out whether they are pure our not. Medical doctors receive samples of various medicines to try them on a sample of patients to determine their effectiveness in curing the disease

---

Question: What is Quantitative Analysis ?

Answer: Quantitative Analysis: It means the analysis of those variables which can be expressed numerically such as age, income or number of children.

Question: what are population parameters and sample statistic?

Answer: Population Parameters or simply Parameters are numerical values that describe the characteristics of a whole population. Commonly represented by Greek letters. Sample Statistic or simple Statistic are numerical values describing the characteristics of a sample. Commonly represented by Roman letters. Note that the term statistic refers to a sample quantity and the term parameter refers to a population quantity.

Question: What is statistical error, in what way it differs from a mistake?

Answer: Statistical error: A continuous variable can never be measured with perfect fineness because of certain habits and practices, methods of measurements, instruments used, etc. the measurements are thus always recorded correct to the nearest units and hence are of limited accuracy. In statistics the error does not mean mistake which is a chance of in accuracy because the actual or true values are, however, assumed to exist.

Question: What is the difference between a nominal and an ordinal scale?

Answer: ORDINAL SCALE It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurements. For example, the performance of students (or players) is rated as excellent, good fair or poor, etc. Number 1, 2, 3, 4 etc. are also used to indicate ranks

---

Question: What is Finite population and Infinite population?

Answer: Finite Population: The population is Finite when it contains countable number of units. Examples: 1.Population of all licensed cars. 2.Population of all students in college. 3.Population of all houses in a country. Infinite Population: The population is Infinite when it contains uncountable number of units. Examples: 1.Population of all points in line. 2.Population of pressures at various points in the atmosphere.

Question: Which is better QUOTA SAMPLING or RANDOM SAMPLING?

Answer: Both Random & Quota sampling has their advantages & disadvantages. Both are used by organizations for their surveys. 1.The main advantage of Random sampling is that it provides a valid estimate of sampling error, But it is impossible to assess objectively the error in quota sampling. 2.When quota sampling is cheap (and fast) it is usually done poorly. When it is done better, it is not all much cheaper really then efficient probability (Random) sampling. Random sampling is widely used in various areas such as industry, agriculture, business etc.

Question: Define grouped data, ungrouped data and frequency.

Answer: Grouped data - Data available in class intervals as summarized by a frequency distribution. Individual values of the original data are not available. Or Data that are presented in the form of frequency distribution are called grouped data. We often group the data of a sample into intervals to produce a better over all picture of the unknown population, but in doing so we lose the identity of individual observations in the sample. Ungrouped data - Ungrouped data is that in which raw data is not grouped. Example: 2, 3, 9, 0, 4, 4, 1, 5, 4, 8, 5, 3, 6, 6, 0, 2, 2, 7, 6, 4, 8, 4, 3, 3, 1, 0, 8, 7, 5, 1, 3, 4, 7, 2, 4, 7, 5, 2, 6, 3, 1, 7, 5, 4, 6, 4, 2, 5, 3, 4, Definition of frequency: Number of observations in each class or group is called is the frequency of that class. It means "how frequently something happens?"

Question: In which situation we use Pie chart simple, Bar chart and multiple bar chart ?

Answer: Pie Chart consists of a circle divided into sectors whose areas are proportional to the various parts into which the whole quantity is divided. It is an effective way of showing percentage parts when the whole quantity is taken as 100. It is also used when the basic categories are not quantifiable. For example as with expenditure, classified into food, clothing, fuel and light etc. Simple Bar Diagram is used when the data consist of a single component and do not involve much variation. Multiple Bar Diagram is used represent two or more related sets of data. It is a diagram which supplies more than one information at the same time.

Question: Brifly decribe the primary data and secondary data.

Answer: Primary Data: Primary data are data collected by the investigators for the purposes of the study. This allows the opportunity to improve precision and to minimize measurement bias through the use of precise definitions, systematic procedures, trained observers, and blinding during data collection. Such data are usually expensive to acquire compared to secondary data. Secondary Data: Secondary data are data collected for purposes other than that of the study, such as patient clinical records, and are used frequently for case-control studies. Because the investigator has no control over definitions, collection procedures, observers (clinicians) or other opportunities for measurement bias reduction, the opportunity for bias is large. The advantages of secondary data are that these data are usually considerably less expensive and much more readily available than are primary data. The severe disadvantage is the opportunity for the presence of large amounts of measurement bias.

Question: define Frequency Polygon .

Answer: A frequency polygon is obtained by plotting the class frequencies against the mid-points of the classes, and connecting the points so obtained by straight line segments. In order to construct the frequency polygon, the mid-points of the classes are taken along the X-axis and the frequencies along the Y-axis.

Question: Explian how you allocate frequency.And also explain bivarite table?

Answer: Steps in Frequency Distribution: Following are the basic rules to construct frequency distribution: 1. Decide the number of classes into which the data are to be grouped & it depends upon the size of data. 2. Determine the RANGE (difference between the smallest &largest values in data) of data. 3. Decide where to locate the class limit (numbers typically use to identify the classes). 4. Determine the reaming class limits by adding the class interval repeatedly. 5. Distribute the data into classes by using tally marks and sum it in frequency column. Finally, total the frequency column to see that all data have been accounted for. Bivarite Table: In bivarite frequency table we have two variables & their respective frequencies.

Question: Define interval in simple words.

Answer: Class interval is defined as the length of class which is equal to the difference between the upper boundary and the lower boundary of the class. Class interval is usually denoted by h. Or Class interval is obtained by finding the difference between either two successive upper class limits or

lower class limits. Note that the lower class limits should not be subtracted from its upper limit to get the class interval. Suppose we have the following frequency distribution: Class Limits Class boundaries f 5 – 9 4.5 – 9.5 2 10 – 14 9.5 – 14.5 5 15 – 19 14.5 – 19.5 9 20 – 24 19.5 – 24.5 12 25 – 29 24.5 – 29.5 3 Here class interval is 5

Question: What is the procedure of Tally?

Answer: TALLY MARKS: These are used to show that how many times a value appears in a data. This is a method of showing frequency of particular class. Example: If the no 4appears one time in the data then in column of tally we use l.But if it appears four times then we use llll.

Question: What are definitions of Bivariate & Univariate?

Answer: BIVARITE TABLE: In bivarite frequency table we have two variables & their frequency. Example: Medium of schooling and sex of the students of a particular college. UNIVARITE TABLE: In unviarite frequency table we have one variable & its frequency. Example: Medium of schooling of the students of a particular college.

Question: Is there any ASYMMETRICAL DISTRIBUTIONS and also What can we say about Raito Charts or Semi-Logarithmic Graphs?

Answer: Asymmetrical Distribution: The Moderately Skewed Distribution is also known as Asymmetrical Distribution. Frequency distribution or curve is said to be “Skewed” when it departs from symmetry. In this the frequencies tend to pile up at one end or the other end of the distribution or curve .This is the most common type. Ratio Charts or Semi-Logarithmic Graph: In ordinary types of graph, the scales used are called the natural scales or the arithmetic scales. These graphs can only be used to compare absolute changes. More often than not we are interested in studying relative changes or ratio. In practice , the difficulty of looking up logarithms can be dispensed with using another type of graph paper ,called Semi-logarithmic paper or ratio paper. “Graphs obtained by plotting the values on Semi-logarithmic paper or ratio paper and joining the successive points by means of straight line segments are called Semi-logarithmic Graph or Ratio Graph”.

Question: defineclass frequency and class bounderies .

Answer: Class frequency: The no of observations falling in a particular class is called class frequency or simply frequency. Or The numbers in each class are referred to as frequencies. Class Boundaries: Class boundaries are the precise numbers which separates one class from another. A class boundary located midway between the upper limit of the class and the lower limit of the next higher class.

Question: Define frequency curves and decomulatative frequency .

Answer: Frequency curves reveal the general pattern or shape of the distribution. When the frequencies are cumulated from the highest value to the lowest value, it is called a "more than" type cumulative frequency or decumulative frequency. It is used the to answer the questions like How many students have weights more than 100 pounds?

Question: Define types of Frquency Curves?

Answer: Types of Frequency Curves: The frequency distribution occurring in practice, usually belong to one of the following four types. You will study about them in your next lecture. 1.The Symmetrical Distribution. 2.Moderately Skewed Distribution. 3.Extremely Skewed or J-shaped Distribution 4.U-Shaped Distribution

Question: HOw frequency distribution is formed from raw data?

Answer: The frequency distribution of an ungrouped data is formed in the following steps. Step - 1 Identify the smallest and the largest measurements in the data set. Step - 2 Find the range which is defined as the difference between the largest value and the smallest value. Step - 3 Decide on the number of classes into which the data are to be grouped. (By classes, we mean small sub-intervals of the total interval) There are no hard and fast rules for this purpose. The decision will depend on the size of the data. When the data are sufficiently large, the number of classes is usually taken between 10 and 20. Step - 4 Divide the range by the chosen number of classes in order to obtain the approximate value of the class interval i.e. the width of our classes. Class interval is usually denoted by h. Step - 5 Decide the lower class limit of the lowest class. Where should we start from? The answer is that we should start constructing our classes from a number equal to or slightly less than the smallest value in the data. Step - 6 Determine the lower class limits of the successive classes by adding h successively. Step - 7 Determine the upper class limit of every class. The upper class limit of the highest class should cover the largest value in the data. It should be noted that the upper class limits will also have a difference of h between them. Step - 8 After forming the classes, distribute the data into the appropriate classes and find the frequency of each class.

Question: Give details to find the INTERVALS ?

Answer: Class interval: It is the length of class and is equal to the difference between the upper class boundary & lower class boundary. A uniform class interval, usually denoted by h or c.Determination of the class interval width The class interval width is determined by the using following formula. Class interval h = RANGE/No. of classes For Example: If the range of data is 61& its no. of classes is 10. Then its class interval = 61/10 = 6.1i.e. 6

Question: Explain the hypergeometric experiment.

Answer: There are many experiments in which the condition of independence is violated and the probability of success does not remain constant for all trials. Such experiments are called Hypergeometric experiments. In other words, a Hypergeometric experiment has the following properties: PROPERTIES OF HYPERGEOMETRIC EXPERIMENT: i) The outcomes of each trial may be classified into one of two categories, success and failure. ii) The probability of success changes on each trial. iii) The successive trials are not independent. iv) The experiment is repeated a fixed number of times. The number of success, X in a Hypergeometric experiment is called a Hypergeometric random variable and its probability distribution is called the Hypergeometric distribution. Consider the example of a bag which contains 4 red balls and 6 black balls. If we draw 4 balls from the bag one by one without replacing the drawn balls into the bag. Let X be the number of red balls contained in the sample, then, it is a hypergeometric experiment because, (i) The result of each draw may be classified as either red (success) or black (failure). (ii) The probability of success changes on each draw. (iii) Successive draws are dependent as the selection is made without replacement. (iv) The drawing is repeated a fixed number of times (n = 4)

Question: What are the concepts of sampling with replacement and sampling without replacement.

Answer: In sampling with replacement, the units are replaced back before the next unit is selected. In this sampling procedure, number of units in population remains same for all selections. Let ‘N’ be the population size and ‘n’ be the sample size then number of possible samples that can be drawn with replacement are Nn. In sampling without replacement, the units are not replaced back before the next unit is selected. In this sampling procedure, number of units in population is reduced after each unit. Let ‘N’ be the population size and ‘n’ be the sample size then number of possible samples that can be drawn with replacement are NCn .

Question: How we calculate the boundries?

Answer: CLASS BOUNDARIES The true class limits of a class are known as its class boundaries.It should be noted that the difference between the upper class boundary and the lower class boundary of any class

is equal to the class interval.

Question: What is Raw data ?And What is central tendency of variable data?

Answer: Raw data: In statistics, a listing of values that has not yet been treated, arranged, or interpreted Measures of Center Tendency: In this context, the first thing to note is that in any data-based study, our data is always going to be variable. Plotting data in a frequency distribution shows the general shape of the distribution and gives a general sense of how the numbers are bunched. Several statistics can be used to represent the "center" of the distribution. These statistics are commonly referred to as measures of central tendency. Such as Mean, Median & Mode. They remain unchanged by rearrangement of the observations in a different order.

Question: What is role of statics in the real life?

Answer: Role of Statistics: Statistics is perhaps a subject that is used by everybody. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. Statistics play its role in the real life which can be identified as follows.A businessman, an industrial and a research worker all employ tatistical methods in their work. Banks, Insurance companies and Government all have their statistics departments. A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision. A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles. A social scientist uses statistical methods in various areas of socio-economic life a nation. It is sometimes said that “a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there”.

Question: Who was the founder of statics and probality?

Answer: The word statistics ultimately derives from the New Latin term statisticum collegium ("council of state") and the Italian word statista ("statesman" or "politician"). The German Statistik, first introduced by Gottfried Achenwall (1749), originally designated the analysis of data about the state, signifying the "science of state" (then called political arithmetic in English). It acquired the meaning of the collection and classification of data generally in the early 19th century. It was introduced into English by Sir John Sinclair. The mathematical methods of statistics emerged from probability theory, which can be dated to the correspondence of Pierre de Fermat and Blaise Pascal (1654). Christiaan Huygens (1657) gave the earliest known scientific treatment of the subject. Jakob Bernoulli's Ars Conjectandi (posthumous, 1713) and Abraham de Moivre's Doctrine of Chances (1718) treated the subject as a branch of mathematics.[1] In the modern era, the work of Kolmogorov has been instrumental in formulating the fundamental model of Probability Theory, which is used throughout statistics.

Question: What is the defination of Mid Range ?

Answer: MID-RANGE: If there are n observations with X0 and Xm as their smallest and largest observations respectively, then their mid-range is defined as Mid range= X0+ Xm / 2 It is obvious that if we add the smallest value with the largest, and divide by 2, we will get a value which is more or less in the middle of the data-set

---

Question: What is difference b/w Discrete and continous variable?

Answer: Discrete Variable: a variable with a limited number of values (e.g., gender (male/female), college class (freshman/junior/senior).Continuous Variable: a variable that can take on many different values, in theory, any value between the lowest and highest points on the measurement scale. Difference: A discrete variable represents count data such as the number of persons in a family, the number of rooms in a house, the number of deaths in an accident, the income of an individual, etc

Question: explain the agvantages and stem an leaf display

Answer: A frequency table has the disadvantage that the identity of individual observations is lost in grouping process. To overcome this difficulty Stem and leaf technique offers a quick and novel way for simultaneously sorting and displaying data sets where each number in the data set is divided into two parts, a Stem and a Leaf. A stem is the leading digit(s) of each number and is used in sorting, while a leaf is the rest of the number or the trailing digit(s). A vertical line separates the leaf (or leaves) from the stem. It is technique that simultaneously ranks orders quantitative data and provides insight about the shape of the distribution.

Question: Why we have a need to take coefficient of standard deviation and variance?

Answer: Co-efficient of variation: The standard deviation is expressed in absolute terms and is given in the same unit of measurement as the variable itself. The dispersion of two or more sets of data can not be compared unless we have relative measure of variation which is known as Co-efficient of variation. It expresses the standard deviation as percentage of arithmetic mean. It is used to compare sets of data or distributions which are expressed in difficult units of measurement, e.g. one may be in hours & the other may be in kilograms or rupees.It is also used as criteria for consistent performance, the smaller the co -efficient of variation; the more consistent is the performance.

Question: what is a grouping error and tally?

Answer: Grouping error refers to the error that is introduced by the assumption that all the values falling in a class are equal to the mid-point of the class interval. In reality, it is highly improbable to have a class for which all the values lying in that class are equal to the mid-point of that class. This is why the mean that we calculate from a frequency distribution does not give exactly the same answer as what we would get by computing the mean of our raw data. This grouping error arises in the computation of many descriptive measures such as the geometric mean, harmonic mean, mean deviation and standard deviation. But, experience has shown that in the calculation of the arithmetic mean, this error is usually small and never serious. Only a slight difference occurs between the true answer that we would get from the raw data, and the answer that we get from the data that has been grouped in the form of a frequency distribution. We use Tally marks for the convenience for making the frequency distribution as it is used to record each & every value that falls in the particular class & after adding them we write it in the digit in the frequency column.

Question: what is value of central tendecy? and why weapply it ?and how many types of centeral tendency

Answer: Central Tendency means the tendency of the data to gather around some central value and the value around which all the observations tend to gather is called measure of central tendency. Measures of central tendency of central tendency are generally known as Averages. The most common types of averages are: i) The arithmetic mean ii) Geometric Mean iii) Harmonic Mean iv) Median v) Mode

---

Question: How we find mediam from the data?

Answer: In order to find Median, we following the steps: i) Arrange the values in increasing order. ii) Count the number of values. iii) a. If the no. of values is odd then Median is (n+1)/2 th value. b. If the no. of values is even then Median is the average of n/2 th and [(n/2) +1] th observations.

Question: what is meant by imperical relation ?

Answer: Empirical Relationship: This is a concept which is not based on a mathematical formula; rather, it is based on observation. In fact, the word ‘empirical’ implies ‘based on observation’. This is a rule of thumb that applies to data sets with frequency distributions that are mound-shaped and symmetric. According to this empirical rule: a) Approximately 68% of the measurements will fall within 1 standard deviation of the mean, i.e. within the interval (`X – S,`X + S) b) Approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval (`X – 2S,`X

+ 2S). c) Approximately 100% (practically all) of the measurements will fall within 3 standard deviations of the mean, i.e. within the interval (`X – 3S,`X + 3S).

Question: What is the relation between these two Moments & Moment Ratios?

Answer: Moments: A moment designates the power to which deviations are raised before averaging them. Moment ratio: These are certain ratios in which both numerators and the denominators are moments.

Question: what is difference between arbitrary form and dispersion?

Answer: Arbitrary form: We find the moment form any value other than the mean that is called the moments about the arbitrary form. Dispersion: By which we mean the extent the observation in a sample or population are spread out. And the second moment about the mean is exactly the same thing as the variance, the positive square root of which is the standard deviation, the most important measure of dispersion?

Question: what is the conditinal and un conditinal probability?

Answer: In many situations, once more information becomes available, we are able to revise our estimates for the probability of further outcomes or events happening. For example, suppose you go out for lunch at the same place and time every Friday and you are served lunch within 15 minutes with probability 0.9. However, given that you notice that the restaurant is exceptionally busy, the probability of being served lunch within 15 minutes may reduce to 0.7. This is the conditional probability of being served lunch within 15 minutes given that the restaurant is exceptionally busy

---

Question: explain What is Moment ratios?

Answer: Moment ratios are certain ratios in which both the numerator and the denominator are moments. The most common of these moment-ratios are denoted by b1 and b2 and defined by the relations: i) b1 = (m3)2 / (m2)3 ii) b2 = m4 / (m2)2 These are independent of origin and units of measurement, i.e. they are pure numbers. b1 is used to measure the Skewness of distribution, and b2 is used to measure the kurtosis of the distribution.

Question: Why the significance level is consider 0.05?

Answer: By a = 5%, we mean that there are about 5 chances in 100 of incorrectly rejecting a true null hypothesis. That is, we want to make the significance level as small as possible in order to protect the null hypothesis and to prevent, as far as possible, the investigator from inadvertently making false claims.

Question: what is scatter diagram . what is its work ,and what is its advantages ?

Answer: A scatter plot is a useful summary of a set of bivariate data (two variables), usually drawn before working out a linear correlation coefficient or fitting a regression line. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model. Each unit contributes one point to the scatter plot, on which points are plotted but not joined. The resulting pattern indicates the type and strength of the relationship between the two variables

---

Question: What is the specific definition of "MOments"?

Answer: A moment represents the power to which deviations are raised before averaging them. The first four

moments play important rule in describing the characteristics of the frequency distribution We use Moments to describe the Frequency Distribution. For example, the First moment about x = 0 is the arithmetic mean, the Second moment about the mean is the Variance & its positive square root is the standard deviation. The Third moment is a measure of Skewness while the Fourth central moment is used to measure Kurtosis.

Question: explain the term "Degrees of freedom of t distribution"?

Answer: Where the term degree of freedom represents the number of independent random variables that express the t-distribution. Recall that in estimating the population's variance, we used (n-1) rather than n, in the denominator. The factor (n-1) is called "degrees of freedom.

Question: What is the difference between p(type 1 error) and p(type2 error)?

Answer: Type I error: On the basis of sample information, we may reject the null hypothesis H0, when it is, in fact true. This type of error is called the type I error. Type II error: On the basis of sample information we may accept the null hypothesis H0, when it is actually false. This type of error is called the type II error.

Question: What are Mutually Exclusive Events? Define it with an example and also explain the classic definition of probability.

Answer: Mutually Exclusive Events: Two events are mutually exclusive (or disjoint) if it is impossible for them to occur together. Formally, two events A and B are mutually exclusive if and only if AnB = F Examples: 1. Experiment: Rolling a die once Sample space S = {1, 2, 3, 4, 5, 6} Events A = 'observe an odd number' = {1, 3, 5} B = 'observe an even number' = {2, 4, 6} AnB=F= the empty set, so A and B are mutually exclusive. Classical Definition: If a random experiment can produce n mutually exclusive and equally likely outcomes, and if m out to these outcomes are considered favorable to the occurrence of a certain event A, then the probability of the event A, denoted by P(A), is defined as the ratio m/n. P(A)= m/n.

Question: what is subjective approach and objective approach to probability in statistics?

Answer: There are two approaches to interpret the probability which are as follows Subjective Approach: As its name suggests, the subjective or personalistic probability is a measure of the strength of a person's belief regarding the occurrence of an event. This definition may be applied to those real world situations where neither an equally likely nor the relative frequency approach is possible. Objective Approach: Objective probability relates to those situations where everyone will arrive at the same conclusion. There are three different definitions in objective approach. 1) Classical Definition 2) Relative Frequency Definition 3) Axiomatic Definition

---

Question: write down the LAW OF COMPLEMENTATION and ADDITION LAW.

Answer: LAW OF COMPLEMENTATION: If A is the complement of an event A relative to the sample space S, then P (A) = 1 – P (A) Complementary probabilities are very useful when we want to solve questions of the type 'What is the probability that, in tossing two fair dice, at least one even number will appear?' ADDITION LAW If A and B are any two events defined in a sample space S, then P (AÈB) = P (A) + P (B) – P (AÇB)

Question: Difference Between Papulation Data and Sample Data ? 2.Why we measure skewness ? 3.Why we measure Kurtosis ?

Answer: . Population data consists of all individual members or objects whether finite or infinite, relevant to some characteristics of interest. And as sample is only a part of population so sample data consists of

some of the observation. 2. Skew ness is a lack of symmetry around some central value i.e. mean, median, mode. It is to find that whether the curve of distribution is positively skewed or negatively skewed. 3. We measure kurtosis to find the shape of the distribution. It is used to find the degree of peak ness or flatness of a unimodal curve.

Question: Define Multiplication theorem of probability for independent events. what is marginal probability.

Answer: Multiplication theorem of probability for independent events is as follows: P(A Ç B) = P(A) P(B) Here A and B are independents events. P(A) and P(B) are called marginal probabilities whereas, P(A Ç B) is called joint probability of A and B.

Question: what is bivariate probability. how define it.

Answer: distribution of two or more random variables which are observed simultaneously when an experiment is performed, is called their JOINT distribution. It is customary to call the distribution of a single random variable as univariate. Likewise, a distribution involving two, r.v.'s simultaneously is referred to as bivariate. Bivariate Probability Function: Let X and Y be two discrete r.v.'s defined on the same sample space S, X taking the values x1, x2, …, xm and Y taking the values y1, y2, …, yn. Then the probability that X takes on the value xi and, at the same time, Y takes on the value yj , denoted by f(xi, yj) or pij, is defined to be the joint probability function or simply the joint distribution of X and Y. Thus the joint probability function, also called the bivariate probability function f(x, y) is a function whose value at the point (xi, yj) is given by: f(xi, yj) = P(X = xi and Y = yj), i = 1, 2, …, m. j = 1, 2, …, n.

Question: what is hypergeometric and its properties?

Answer: There are many experiments in which the condition of independence is violated and the probability of success does not remain constant for all trials. Such experiments are called Hypergeometric experiments. In other words, a Hypergeometric experiment has the following properties: PROPERTIES OF HYPERGEOMETRIC EXPERIMENT: i) The outcomes of each trial may be classified into one of two categories, success and failure. ii) The probability of success changes on each trial. iii) The successive trials are not independent. iv) The experiment is repeated a fixed number of times. The number of success, X in a Hypergeometric experiment is called a Hypergeometric random variable and its probability distribution is called the Hypergeometric distribution. Consider the example of a bag which contains 4 red balls and 6 black balls. If we draw 4 balls from the bag one by one without replacing the drawn balls into the bag. Let X be the number of red balls contained in the sample, then, it is a hypergeometric experiment because, (i) The result of each draw may be classified as either red (success) or black (failure). (ii) The probability of success changes on each draw. (iii) Successive draws are dependent as the selection is made without replacement. (iv) The drawing is repeated a fixed number of times (n = 4)

Question: how we define the standard normal distribution?

Answer: A normal distribution whose mean is zero and whose standard deviation is 1 is known as the standard normal distribution. A normal probability distribution depends upon the values of parameters μ and s 2 and the various values for these parameters will result in an unlimited number of different normal distributions. Every normally distributed r.v. x with mean = μ and variance = s 2 is therefore conveniently transformed into a new normal r.v. z with zero mean and unit variance by using the following expression. Z = x - μ / s Then the p.d.f of z, denoted by Ø (z) becomes Ø (z) = 1/ (2?) ½* e –Z2/2 -8 < x < 8 A normal probability distribution of Z with a mean of zero and a standard deviation of one is called Standardized normal distribution and is denoted by N (0, 1). As you know the density function of normal distribution N (μ, s) = 1/s*(2?) 1/2 * e -1/2[(x-μ)/ s] 2 -8 < x < 8 Now if we put μ = 0 and s = 1 we get N (0, 1) = 1/ (2?) ½* e –x2/2

Question: define sampling with replacement and sampling without replacement.

Answer: In sampling with replacement, the units are replaced back before the next unit is selected. In this sampling procedure, number of units in population remains same for all selections. Let 'N' be the population size and 'n' be the sample size then number of possible samples that can be drawn with replacement are Nn. In sampling without replacement, the units are not replaced back before the next unit is selected. In this sampling procedure, number of units in population is reduced after each unit. Let 'N' be the population size and 'n' be the sample size then number of possible samples that can be drawn with replacement are NCn .

Question: explain Point Estimator and what does it mean by a good point estimator.

Answer: Point Estimator: A single value calculated from the sample that is likely to be close in value to the unknown parameter. It is to be noted that a point estimate will not, in general, be equal to the population parameter as the random sample used is one of the many possible samples which could be chosen from the population. Good Point Estimator: A point estimator is considered a good estimator if it satisfies various criteria. Four of these criteria are: (i) Unbiasedness (ii) Consistency (iii) Efficiency (iv) Sufficiency

---

Question: what is one Tailed and two Tailed

Answer: ONE-TAILED AND TWO-TAILED TESTS: A test, for which the entire rejection region lies in only one of the two tails – either in the right tail or in the left tail – of the sampling distribution of the test-statistic, is called a one-tailed test or one-sided test. If, on the other hand, the rejection region is divided equally between the two tails of the sampling distribution of the test-statistic, the test is referred to as a two-tailed test or two-sided test.

Question: What are the application of the and in which conditions for the use of following tests? F-test chi square test z-test and t-test are not fulfilling need

Answer: (i) F-test is used to compare the variances of two populations. (ii) Chi-square test is used to test a specific value of population variance. (iii) Z-test is used to test the mean of a population or equality of two population means when population variance is known or sample size is greater than 30. (iv) t-test is used to test the mean of a population or equality of two population means when population variance is unknown or sample size is less than 30.

Question: what is Confidence Interval?

Answer: Confidence Interval: When we can not get information about the whole population because of our limited resources and time we then resort to estimate the population parameters using the part of the population called sample. The information we obtain from sample is called an estimate. An estimate could be of two types (i) point estimate and (ii) interval estimate. In interval estimation we find an interval and we expect with certain degree of confidence (usually 95%) our parameter lies in this interval. Obviously using sample data we can never be 100% confident that our parameter lies in the given interval as sample is the small part of the population but we try to be close to 100% like 95% and 99%. Usually we find 95% confidence interval but when we are extra careful and need more accurate information then we find 99% confidence interval but it could be expensive for we have to select a very large sample.

Question: what is the difference between f-distribution , chi-square distribution t-distribution?

Answer: These distributions have their own applications and these are used in separate situations. (i) f-distribution is used to test the equality of two populations variances. It is also used to test the equality

of population means when we have more than two populations. (ii) t-distribution is used to test the mean of a population and equality of two population means in case of small sample size. (iii) Chi-square distribution is used to test the variance of a population. It is also used to test the association of attributes.

Question: Explain the concept of goodness of fit test.

Answer: goodness of fit test is a hypothesis test that is concerned with the distribution which may be the uniform, binomial, poisson, Normal or any other distribution. This is a kind of hypothesis test for problems where we do not know the probability distribution of the random variable under consideration, say X, and we wish to test the hypothesis that X follows a particular distribution. In this test procedure, the rage of all possible values of the random variable assumed to follow a particular distribution is divided into k mutually exclusive classes and the probabilities pi are calculated for each of the classes, using the estimates of the parameters of the probability distribution specified in H0. The npi represents the expected number of observations that fall in that class. The differences between observed and expected number of observations can arise form sampling error or from H0 being false. Small differences are generally attributed to sampling error, large differences which are considered to arise form H0 being false, are unlikely if the hypothesized distribution gives the satisfactory fit to the sample data.

Question: what is Point Estimation.

Answer: Point estimation of a population parameter provides, as an estimate, a single value calculated from the sample that is likely to be close in value to the unknown parameter. For example the value of the statistic (Xbar) computed from a sample of size n, is a point estimate of the population parameter (u)

Question: what is the difference between bar chart and scatter diagram ?

Answer: Scatter plot is used to represent the type & strength of the relationship between two variables. Whereas bar chart is used to illustrate the major features of the distribution of the data in a convenient form Scatter Plot: A scatter plot is a useful summary of a set of bivariate data (two variables), usually drawn before working out a linear correlation coefficient or fitting a regression line. It gives a good visual picture of the relationship between the two variables, and aids the interpretation of the correlation coefficient or regression model. Each unit contributes one point to the scatter plot, on which points are plotted but not joined. The resulting pattern indicates the type and strength of the relationship between the two variables. Bar Chart: A bar chart is a way of summarizing a set of categorical data. It is often used in exploratory data analysis to illustrate the major features of the distribution of the data in a convenient form. It displays the data using a number of rectangles, of the same width, each of which represents a particular category. The length (and hence area) of each rectangle is proportional to the number of cases in the category it represents, for example, age group, religious affiliation. Bar charts are used to summaries nominal or ordinal data.

Question: explain difference set, complement set, disjoint set.

Answer: Difference set is defined as: The difference of set A &B is the set of elements which belongs to A but not B e.g A{1,2,3,4,5,6,7,} & B{1,3,5}hence A-B{2,4,6,7}is the required difference set. Complement 0f set Bis the set of elements which do not belongs to B e.g. if U={1,2,3,4,5,……100}and B={2,4,6,8,} then the complement of B isB' ={1,3,5,7,9,10,11……100} Disjoint set two sets A and B are said to be disjoint set if AnB=?. For example A={2,4,6,8,}and B={1,3,5,9}hence A and B are said to be disjoint set.

Question: What do mean by order? how we Differeniate between permutation and combination?

Answer: Order: Placement of objects is known as order. Permutations: When our purpose is to arrange the objects with respect to order out of" n" then we use permutations. Combinations: When we select our objects out of "n" with out considering order then we apply combination

Question: what is quartile.

Answer: Quartile: The values which divide the distribution into four equal parts are called quartiles. Quartiles divide the data into four equal-sized and non-overlapping parts. One fourth of the data lies below the Q1 (first quartile). Half of the data lies below Q2 (second quartile) similarly, three quarters of the data lies below Q3 (third quartile) Note: Q2 (second quartile) is also known as median. Use of quartiles: In order to describe a data set without listing all the data, we have measures of location such as the mean and median, measures of spread such as the range and standard deviation. Quartiles are also used to describe the data in combination with other measures. For example they are used in five number summary of the data. The five number summary, i.e., the minimum, Q1, Q2 (median), Q3, and maximum, give a good indication of where data lie.

Question: Explain the binomial Distribution and fitting a binomial distribution to real data in detail.

Answer: Binomial distribution: The binomial distribution is a very important discrete probability distribution. Binomial probability distribution - A probability distribution showing the probability of x successes in n trials of a binomial experiment. Binomial probability function - The function used to compute probabilities in a binomial experiment. The binomial distribution is a very important discrete probability distribution Binomial Distribution and fitting a binomial distribution to real data in detail is given in handouts of your lecture number 28. And it is very comprehensive. Read it care fully and then ask from where you didn't understand.

Question: Explain the method of Maximum Likelihood in Point Estimation?

Answer: The Method of Maximum Likelihood (ML): "To consider every possible value that the parameter might have, and for each value, compute the probability that the given sample would have occurred if that were the true value of the parameter. That value of the parameter for which the probability of a given sample is greatest, is chosen as an estimate." An estimate obtained by this method is called the maximum likelihood estimate (MLE).

Question: what is purpose of calculation the index number ?

Answer: Index number: An index number is a statistical measure of average change in a variable or group of variables with respect to time or space. It is a device that measures the changes occurring in data from time to time or from place to place.

Question: what is one and two tail test?

Answer: One-tailed test: A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of the sampling distribution is called One-tailed test. Or The entire rejection region lies in only one of the two tails, either in the right tail or in the left tail, of the sampling distribution of the test-statistic, is called a one-tailed test or one-sided test. Two-Tailed Test: The test of a given statistical hypothesis in which a value of the statistic that is either sufficiently small or sufficiently large will lead to rejection of the hypothesis tested. The statistical tables for Z and for t provide critical values for both one and two tailed tests. That is, they provide the critical values that cut off an entire alpha region at one or the other end of the sampling distribution as well as the critical values that cut off the 1/2 alpha regions at both ends of the sampling distribution. Or In two tailed test rejection region is divided equally between the two tails of the sampling distributions of the test statistic.

Question: Define the trial and the outcome.what is meant by sample distribution. Define the following term ƒæ Null Hypothesis ƒæ Alternative hypothesis.

Answer: Trial: A single performance of an experiment is called a trial. Outcome: The result obtained from an

experiment or a trial is called an outcome. Null hypothesis: The hypothesis tentatively assumed true in the hypothesis testing procedure. It is denoted by H0. Alternative hypothesis: The hypothesis concluded to be true if the null hypothesis is rejected. It is denoted by H1. Sampling distribution: A probability distribution consisting of all possible values of a sample statistic is called sampling distribution. Or Sampling distribution is defined as probability distribution of sample statistic such as a mean, a standard deviation, and a proportion etc, computed from all possible samples of the same size, which might be selected with or without replacement from a population.

Question: what is the main difference between Discrete & Continuious Variables?

Answer: Variable is a characteristic under study that assumes different values for different elements. For example, Height of students in a class, No. of class rooms in a college Discrete variable Continuous variable 1. Values belonging to it are distinct and separate 2. Value are obtained by counting 3. Examples include § The number of defective light bulbs in a box § The number of children in a family § The number of patients in a doctor's clinic 1. Values belonging to it may take on any value within a finite or infinite interval 2. Values are obtained by measuring 3. Examples include § The amount of sugar in an orange § The time required to run a mile § Temperature Variable is a characteristic under study that assumes different values for different elements. For example, Height of students in a class, No. of class rooms in a college Discrete variable Continuous variable 1. Values belonging to it are distinct and separate 2. Value are obtained by counting 3. Examples include § The number of defective light bulbs in a box § The number of children in a family § The number of patients in a doctor's clinic 1. Values belonging to it may take on any value within a finite or infinite interval 2. Values are obtained by measuring 3. Examples include § The amount of sugar in an orange § The time required to run a mile § Temperature

---

Question: Define these terms Discriptive statistics,Infrencial statistics.data,types of measuring scales and measurement of errors in simple form thanks.

Answer: Descriptive Statistics uses graphical and numerical techniques to summarize and display the information contained in a data set. Inferential Statistics uses sample data to make decisions or predictions about a larger population of data. Data: these are the results of observation. Examples, · Statements given to a police officer or physician or psychologist during an interview are data. · So are the correct and incorrect answers given by a student on a final examination. · Almost any athletic event produces data. · The time required by a runner to complete a marathon, · The number of errors committed by a baseball team in nine innings of play. And, of course, data are obtained in the course of scientific inquiry: · the positions of artifacts and fossils in an archaeological site, · The number of interactions between two members of an animal colony during a period of observation, · The spectral composition of light emitted by a star. Scales of Measurement: Nominal Scales When measuring using a nominal scale, one simply names or categorizes responses. The essential point about nominal scales is that they do not imply any ordering. Nominal scales embody the lowest level of measurement. It is used for identifying individuals, groups or regions. Ordinal Scales Where nominal scales don't allow comparisons in degree, this is possible with ordinal scales. Say you think it is better to live in Karachi than in Lahore but you don't know by how much. Interval Scales Interval sales are numerical scales in which intervals have the same interpretation throughout. As an example, consider the Fahrenheit scale of temperature. The difference between 30 degrees and 40 degrees represents the same temperature difference as the difference between 80 degrees and 90 degrees. This is because each 10 degree interval has the same physical meaning (in terms of the kinetic energy of molecules). Interval scales do not have a true zero point even if one of the scaled values happens to carry the name "zero". The Fahrenheit scale illustrates the issue. Zero degrees Fahrenheit does not represent the complete absence of temperature (the absence of any molecular kinetic energy). In reality, the label "zero" is applied to its temperature for quite accidental reasons connected to the history of temperature measurement. Ratio scales If data have a natural zero then they can be assigned to a ratio scale, which is the scale that has the most information attached to it because you can make both interval and ratio comparisons.

Question: Define statistical Quality control.

Answer: Statistical Quality Control (SQC) Statistical Quality Control is the process of inspecting enough product from given lots to probabilistically ensure a specified quality level. The purpose of statistical

quality control is to ensure, in a cost efficient manner, that the product shipped to customers meets their specifications. Inspecting every product is costly and inefficient, but the consequences of shipping non conforming product can be significant in terms of customer dissatisfaction.

Question: Explaine the "Distributive law" and "complementation law"by using them in an example.

Answer: Distributive laws A Ç (B È C) = (A Ç B) È (A Ç C) and A È (B Ç C) = (A È B) Ç (A È C) Complementation laws A È A= S, A ÇA = f, (`A ) = A, `S = f `f = S S={0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15} A= {0, 1, 2, 3, 4, 5} B= {4, 5, 6, 7, 8, 9, 10} C= {9, 10, 11, 12, 13, 14, 15} Distributive law: AUB = {0, 1, 2, 3, 4, 5} U {4, 5, 6, 7, 8, 9, 10} AUB = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10} AUC = {0, 1, 2, 3, 4, 5} U {9, 10, 11, 12, 13, 14, 15} AUC = {0, 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 14, 15} BUC = {4, 5, 6, 7, 8, 9, 10} U {9, 10, 11, 12, 13, 14, 15} BUC = {4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15} A Ç B = {0, 1, 2, 3, 4, 5} Ç {4, 5, 6, 7, 8, 9, 10} A Ç B = { 4, 5} A Ç C = {0, 1, 2, 3, 4, 5} Ç {9, 10, 11, 12, 13, 14, 15} A Ç C = f B Ç C = {4, 5, 6, 7, 8, 9, 10} Ç {9, 10, 11, 12, 13, 14, 15} B Ç C = {9, 10} A Ç (B È C) = {0, 1, 2, 3, 4, 5} Ç {4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15} A Ç (B È C) = {4, 5} (A Ç B) È (A Ç C) = {4, 5} È f (A Ç B) È (A Ç C) = {4, 5} We can see that (A Ç B) È (A Ç C) = (A Ç B) È (A Ç C) Similarly, second equation can be proved. Law of Complement A = Ac = S – A A = {0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15} – {0, 1, 2, 3, 4, 5} A = {6, 7, 8, 9, 10, 11, 12, 13, 14, 15}

Question: Distinguish between standard deviation and variance.

Answer: Variance : Sample variance is a measure of the spread of or dispersion within a set of sample data. Standard Deviation: It is calculated by taking the square root of the variance and is symbolized by s.d, or s. The more widely the values are spread out, the larger the standard deviation. For example, say we have two separate lists of exam results from a class of 30 students; one ranges from 31% to 98%, the other from 82% to 93%, then the standard deviation would be larger for the results of the first exam.There is no such difference in them as S.D is the square root of variance.

Question: can we use Shapperd's correctness in raw data?

Answer: We use the Shepard's Correction in case of the frequency distribution. Because in the calculation of moments from a grouped frequency distribution, an error is introduced by the assumption that the frequencies associated with a class are located at the midpoint of the class interval. The important point to note here is that these corrections are not applicable to highly skewed distributions and distributions having unequal class-intervals.

Question: Explain co-efficient of correlation.

Answer: Correlation is a statistical technique which can show whether and how strongly pairs of variables are related. For example, height and weight are related. The correlation is used to describe the degree to which one variable is linearly related to another. Often, correlation is used in conjunction with regression to measure how well the regression line explains the variation of the dependent variable; Y. Correlation can also be used by itself, however, to measure the degree of association between two variables. Statisticians have developed measure for describing the correlation between two variables: the coefficient of correlation. . The correlation coefficient for n pairs of observations usually denoted by r, is

---

Question: explain the theoram of chebychev's .

Answer: Chebychev's inequality: A useful rule that illustrates the relationship between dispersion and standard deviation is given by Chebychev's thorem.Chebychev's Rule applies to any data set, regardless of the shape of the frequency distribution of the data. Chebychev's theorem: For any number k greater than 1, at least 1 – 1/k2 of the data-values fall within k standard deviations of the mean, i.e., within the interval (`X – kS,`X + kS) This means that: a) At least 1-1/22 = 3/4 will fall within 2 standard deviations of the mean, i.e. within the interval (`X – 2S,`X + 2S). B) At least 1-

1/32=8/9 of the data-values will fall within 3 standard deviations of the mean, i.e. within the interval (`X – 3S,`X + 3S) Because of the fact that Chebychev's theorem requires k to be greater than 1, therefore no useful information is provided by this theorem on the fraction of measurements that fall within 1 standard deviation of the mean, i.e. within the interval (X–S,`X+S).

Question: What is the difference between Probability Distribution and discrete Probability Distribution. Define E(X), the expected value of a random variable ?

Answer: Probability distribution is an arrangement of the possible values of random variable along with its corresponding probabilities. In Probability distribution, there are two type of distributions 1-Discrete probability distribution 2-Contineous probability distribution Following are the conditions of discrete probability distribution. (i) There must be a finite probability (0 to 1) against every possible value of the random variable i.e. 0

< (X) E as, defined is by denoted X of expectation mathematical then (x)="1" P Sum that such (xn) P(x1),P(x2),.....P probabilities respective with values,x1,x2,x3,.....xn the assume variable random discrete a If 1.i.e., to equal be must all sum The (ii) (x)

Question: What is the relationship between discrete and continuous random variables.

Answer: A random variable 'X' is defined to be discrete random variable, if it can assume finite or countably infinite number of values. Examples: The number of defective bulbs in a lot. The number of road accidents on motor-way per day. A random variable 'X' is defined to be a continuous random variable if it can assume unlimited values within a given range of possible values. Examples: The temperature of a room The amount of rain fall

Question: What is difference between independent and independence variable.

Answer: Two events A and B in the same sample space S, are defined to be independent if the probability that one event occurs, is not affected by whether the other event has or has not occurred. Two events A and B in the same sample space S, are defined to be Dependent if the probability that one event occurs, is affected by whether the other event has or has not occurred.

Question: What is difference between independent and independence variable.

Answer: Two events A and B in the same sample space S, are defined to be independent if the probability that one event occurs, is not affected by whether the other event has or has not occurred. Two events A and B in the same sample space S, are defined to be Dependent if the probability that one event occurs, is affected by whether the other event has or has not occurred.

Question: Explain the Condititional Probability with the help of example.

Answer: In conditional probability we are dealing with two events .One event is that for which we have to find the probability and about 2nd event we have some priori information.To illustrate the concept of conditional probability let us consider an example. Let a die is rolled. S={1,2,3,4,5,6} A is the event of getting a" 5" & a prior information is given that on a particular throw of a die ,the outcome is an odd number (event B) .Hence B={1,3,5}now the probability of getting a "5" in this reduce sample space is 1/3 which is known as conditional probability of event "A". Note. Priori means already known information before starting the experiment

---

Question: what is loaded die?

Answer: We can say that, a biased unfair die is a loaded die

Question: Explain Nominal and ordinal levels of measurement and also tell me what is EPAmileage rating.

Answer: Nominal Scales When measuring using a nominal scale, one simply names or categorizes responses. The essential point about nominal scales is that they do not imply any ordering. Nominal scales embody the lowest level of measurement. It is used for identifying individuals, groups or regions. Ordinal Scales Where nominal scales don't allow comparisons in degree, this is possible with ordinal scales. Say you think it is better to live in Karachi than in Lahore but you don't know by how much. EPA means Environmental Protection Agency US government agency for the protection of the environment which ranks the most fuel-efficient vehicle.

Question: Explain bivarite.

Answer: Bivariate Data Before we looked at one measurement on an observation (or individual), say X is height. Now we're interested in more than one measurement per observation (individual), say X is height and Y is weight. Let's say we have n individuals we're taking the measurements on. Then our data would be as follows (X1,Y1), (X2, Y2).....(Xn, Yn)

Question: What is permutation and combination .

Answer: Permutations: When our purpose is to arrange the objects with respect to order out of" n" then we use permutations. Combinations: When we select our objects out of "n" with out considering order then we apply combination. For the examples, solve the questions in the book.

Question: What is meant by " biased coin and fair coin

Answer: Suppose you have a coin which is not in balanced form, for example if any side of coin is damage then it will not give the fair result. In this case coin will called BIASED. Proper coin will be called FAIR coin.

Question: Define Quartile Deviation.

Answer: Quartile Deviation: The quartile deviation is defined as half of the difference between the third and first quartiles. The quartile deviation has featured that the range “Median + Q.D.” contains approximately 50% of the data. Quartile deviation is regarded as that measure of dispersion which is associated with the median. The quartile deviation is superior to range as it is not affected by extremely large or small observations. It is not as widely used as other measures of dispersion. It is, however, used in situation where extreme observations are thought to be unrepresentative. The quartile deviation should always be employed to indicate dispersion when the median has been adopted as the most appropriate average. The quartile deviation is also an absolute measure of dispersion

---

Question: what is the meaning of Dispersion of data-set? What is the purpose of calculating Mean Deviation? What is the The coefficient of variation?

Answer: Dispersion: The data values in a sample are not all the same. This variation between values is called dispersion. When the dispersion is large, the values are widely scattered; when it is small they are tightly clustered. The width of diagrams such as dot plots, box plots, stem and leaf plots is greater for samples with more dispersion and vice versa. There are several measures of dispersion, the most common being the standard deviation. These measures indicate to what degree the individual observations of a data set are dispersed or 'spread out' around their mean. Mean Deviation: As quartile deviation measures the dispersion of the data-set around the median. But the problem is that the sum of the deviations of the values from the mean is zero (No matter what the amount of dispersion in a data-set is, this quantity will always be zero, and hence it cannot be used to measure

the dispersion in the data-set.) By ignoring the sign of the deviations we will achieve a NON-ZERO sum, and averaging these absolute differences, again, we obtain a non-zero quantity which can be used as a measure of dispersion. This quantity is known as the MEAN DEVIATION. As the absolute deviations of the observations from their mean are being averaged, therefore the complete name of this measure is Mean Absolute Deviation but generally, it is simply called "Mean Deviation. The coefficient of variation measures the spread of a set of data as a proportion of its mean. It is often expressed as a percentage. It is the ratio of the sample standard deviation to the sample mean:

Question: what is the main difference between five number summary and box and whisker plot?

Answer: There is no such difference between the box & whisker plot & five number summary because the Five number summary is another name for the visual representations of the box-and-whisker plot. Box and Whisker Plot: A box & whisker graph is used to display a set of data so that you can easily see where most of the numbers are. A box-and-whisker plot can be useful for handling many data values. They allow people to explore data and to draw informal conclusions when two or more variables are present. It shows only certain statistics rather than all the data. Five-number summary is another name for the visual representations of the box-and-whisker plot. The five-number summary consists of the median, the quartiles, and the smallest and greatest values in the distribution. Immediate visuals of a box-and-whisker plot are the center, the spread, and the overall range of distribution. Five Number Summary: Five-number summary is another name for the visual representations of the box-and-whisker plot. The five-number summary consists of the median, the quartiles, and the smallest and greatest values in the distribution. Immediate visuals of a box-and-whisker plot are the center, the spread, and the overall range of distribution.

Question: What is meant by population?

Answer: A population is any entire collection of people, animals, plants or things from which we may collect data. It is the entire group we are interested in, which we wish to describe or draw conclusions about.

Question: What is meant by discrete?

Answer: Discrete Variable: A variable that is made up of distinct and separate units or categories and is, most of the times, counted only in whole numbers. For example the number of computers in a lab. The number of students in a class.

Question: What is the basic difference b/w Discrete and Continous Frequency Distribution?

Answer: Discrete and Continuous Data Data or variables that are continuous are those that can take any value within a range. An obvious example is height. The height of human beings can take any value between biologically determined limits. So if someone is 1.63 metres tall and someone else is 1.64 metres tall, then it is possible for someone to have an intermediate height between these two values, e.g. 1.633 On the other hand, discrete data can only take certain values, e.g. the total number of goals scored by a football team. It could have scored 42 goals, or 43 goals, or 44 goals....., but definitely not 42 ½ or 42 ¼ , etc. . Discrete Frequency Distribution: The frequency distribution for the discrete variable is called discrete frequency distribution. Suppose the numbers of children in 20 families are as follows: 2, 3, 0, 4, 4, 1, 5, 4, 8, 5, 3, 6, 6, 0, 2, 2, 7, 6, 4, 8 We arrange the above data in frequency distribution as follows: Discrete Frequency Distribution Number of children Tally frequency 0 || 2 1 | 1 2 ||| 3 3 || 2 4 |||| 4 5 || 2 6 ||| 3 7 | 1 8 || 2 Total 20 Note that we have used two tally marks for 0 as it is repeated two times and one tally mark for 1 as it is repeated once and three tally marks for 2 as it is repeated three times and so on. Continuous Frequency Distribution: The frequency distribution for the continuous variable is called continuous frequency distribution. EX. Given the following set of measurements for a particular sample: 2.5 5.9 3.2 1.4 7.0 4.3 8.9 0.7 4.2 9.9 3.4 4.6 5.0 6.4 1.1 9.2 7.7 0.9 4.0 2.3 5.6 2.2 3.1 4.7 5.5 6.6 1.9 3.9 6.1 5.2 8.2 3.3 2.2 5.8 4.1 3.8 1.2 6.8 9.5 0.8 We arrange the above data in frequency distribution as follows: Classes f 0.0 - 1.0 3 1.0 - 2.0 4 2.0 - 3.0 4 3.0 - 4.0 7 4.0 - 5.0 6 5.0 - 6.0 5 6.0 - 7.0 5 7.0 - 8.0 1 8.0 - 9.0 2 9.0 - 10.0 3 Totals 40

Question: define the Distribution function.

Answer: Distribution Function: The distribution function D(x) (also called the cumulative density function (CDF) or probability distribution function), describes the probability that a variate /variable X takes on a value less than or equal to a number x. The distribution function is sometimes also denoted by F(x) .The function F(x) gives the probability of the event that X takes a value less than or equal to a specified value x.

Question: what is discrete probability distribution explain it in simple words.

Answer: Discrete probability distribution: It gives the probability of every possible value of a discrete random variable. Properties: f(xi) greater than or equal to 0 i.e. the probability cannot be negative. Sum of the total probability is one i.e. ?(xi)=1. These two are the fundamental properties. Example: All questions based on the binomial, hyper & Poisson distributions are the examples of discrete probability distributions.

Question: explain the fitting of bionmial distribution. and also what is the benifit of it.

Answer: Fitting of the binomial distribution involves two steps: 1. Estimating the values of the two parameters (n,p) that completely determine a binomial distribution. 2. Calculate the probabilities as well as the expected frequencies for x-values ranging from 0 to n. Application/benefit. The binomial distribution is useful for describing distributions of binomial events, such as the number of males and females in a random sample of companies, or the number of defective components in samples of n units taken from a production process and etc.

Question: What is the difference b/w normal and poissio distribution ?

Answer: Poisson distribution. The Poisson distribution is referred to as the distribution of rare events. Examples of Poisson distributed variables are number of accidents per person, number of sweepstakes won per person, or the number of catastrophic defects found in a production process. While: Normal Distribution. The normal distribution (the "bell-shaped curve" which is symmetrical about the mean) is a theoretical function commonly used in inferential statistics as an approximation to sampling distributions. In general, the normal distribution provides a good model for a random variable, when: There is a strong tendency for the variable to take a central value; Positive and negative deviations from this central value are equally likely; The frequency of deviations falls off rapidly as the deviations become larger.

Question: What is the difference between continuous confidence interval and confidence interval.

Answer: A confidence interval is a range of values that has a specified probability of containing the parameter being estimated. The 95% and 99% confidence intervals which have .95 and .99 probabilities of containing the parameter respectively are most commonly used. As we know that the confidence interval formula is based on z and z is a continuous SNV, so there is no difference between continuous confidence interval and confidence interval

---

Question: The poisson distribution resembles the binomial distribution why, give some basic reson?

Answer: As we know that when “n” is sufficiently large tends to infinity and “p” the probability of success is tends to zero and the “the product “np”remains constant, then we can derive the poisson distribution from binomial under certain conditions we can say that both distributions resemble.

Question: Mean deviation.

Answer: Mean Deviation: Given a set of numbers, and its mean, we can find the difference between each of the numbers and the mean. If we take the mean of these differences, the result is called the mean deviation of the numbers.

Question: What is coefficient of skewness?

Answer: The 'Coefficient of Skewness' shows the tendency of the data set values to 'bunch' at one end of its distribution, with the values at the other end being relatively dispersed. The mode is the measure indicating the value where most bunching happens. Skewness is measured by working out the extent to which the mode departs from the mean. If the mode is towards the lower values in the data set, then the skewness is said to be positive; if it occurs towards the higher values, the skewness is negative. Sk = mean - mode/ S.D. An alternative formula for the coefficient of skewness is often used. (This is based on the knowledge that the difference between the mean and the mode is generally about three times the difference between the mean and the median): Sk = 3 (mean - median) / S.D.

Question: What is FIVE-NUMBER SUMMARY?

Answer: Five number summary consists of five numbers Minimum value Q1 Median Q2 Maximum value These numbers give a good identification of center and spread of data. The five number summary is the best shorts description for most distribution.

Question: what is joint probability. Please discribe about the joint probabaility.

Answer: Joint probability: The probability of two events both occurring is called joint probability. Where JOINT PROBABILITY DISTRIBUTION The joint probability distribution of two discrete random variables X and Y is a function whose domain is the set of ordered pairs (x, y) , where x and y are possible values for X and Y, respectively, and whose range is the set of probability values corresponding to the ordered pairs in its domain. This is denoted by pX,Y(x, y) and is defined as pX,Y(x, y) = P (X = x and Y = y) or JOINT DISTRIBUTIONS: The distribution of two or more random variables which are observed simultaneously when an experiment is performed is called their joint distribution.

Question: Explain the "Covariance & Correlation.

Answer: The Covariance of two r.v.'s X and Y is a numerical measure of the extent to which their values tend to increase or decrease together. The correlation is used to describe the degree to which one variable is linearly related to another. Often, correlation is used in conjunction with regression to measure how well the regression line explains the variation of the dependent variable; Y. Correlation can also be used by itself, however, to measure the degree of association between two variables. Statisticians have developed measure for describing the correlation between two variables: the coefficient of correlation.

Question: Explain in simple words poisson process.

Answer: A Poisson Process is a type of counting process. Typically we see Poisson Process models applied to counting the number of times an event occurs during a specified time period. Consider X as a Random Variable for a situation where we are interested in the number of occurrences of an event during a specified period of time. For X to be a Poisson Random Variable the following must be true: 1. The number of occurrences in one interval of time is unaffected by the number of occurrences in any other non-overlapping time interval. 2. The expected (or average) number of occurrences over any time period is proportional to the size of this interval. 3. Events cannot occur simultaneously. Examples of Poisson Process may include: The number of cars that pass through a certain point on a road during a given period of time. The number of spelling mistakes a secretary makes while typing a single page. The number of phone calls at a call center per minute.

Question: what is a logistic system and what is a inventry.

Answer: Logistics system: The total flow of products from the acquisition of raw materials to the delivery of

finished goods to users, including the related flow of information that controls and records the movement of those products. Inventory System: Procedures that govern how supplies are received, stored, handled, and issued is called inventory system.

Question: Explain pearson's coefficient of correlation.

Answer: Pearson's Product Moment Correlation Coefficient: Pearson's product moment correlation coefficient, usually denoted by r, is one example of a correlation coefficient. It is a measure of the linear association between two variables that have been measured on interval or ratio scales, such as the relationship between height in inches and weight in pounds. However, it can be misleadingly small when there is a relationship between the variables but it is a non-linear one. There are procedures, based on r, for making inferences about the population correlation coefficient. However, these make the implicit assumption that the two variables are jointly normally distributed. When this assumption is not justified, a non-parametric measure such as the Spearman Rank Correlation Coefficient might be more appropriate

---

Question: How define the standard normal distribution.

Answer: Standard normal probability distribution: A normal distribution with a mean of zero and a standard deviation of one is called standard normal probability distribution. The Standard normal distribution: A normal distribution whose mean is zero and whose standard deviation is 1 is known as the standard normal distribution. This distribution has a very important role in computing areas under the normal curve. The reason is that the mathematical equation of the normal distribution is so complicated that it is not possible to find areas under the normal curve by ordinary integration. Areas under the normal curve have to be found by the more advanced method of numerical integration. The point to be noted is that areas under the normal curve have been computed for that particular normal distribution whose mean is zero and whose standard deviation is equal to 1, i.e. the standard normal distribution.

Question: What is meant by stsndard error?

Answer: Standard Error: Standard error is the standard deviation of the values of a given function of the data (parameter), over all possible samples of the same size. Standard error is the standard deviation of sampling distribution of sample statistic (sampling distribution of means or sampling distribution of proportions). You will study about it in details in next lectures.

Question: what is the use of hypothesis testing in pratical bussiness ?

Answer: Setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms. In each hypothesis testing problem, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted H0, against the alternative hypothesis, denoted H1. These two competing claims / hypotheses are not however treated on an equal basis, special consideration is given to the null hypothesis. We have two common situations: 1. The experiment has been carried out in an attempt to disprove or reject a particular hypothesis, the null hypothesis, thus we give that one priority so it cannot be rejected unless the evidence against it is sufficiently strong. For example, H0: there is no difference in taste between coke and diet coke against H1: there is a difference. 2. If one of the two hypotheses is 'simpler' we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one. For example, it is 'simpler' to claim that there is no difference in flavor between coke and diet coke than it is to say that there is a difference.

Question: explain the concept of degrees of freedom.

Answer: In testing of hypotheses you can simply denote it by alpha (the type one error) it is a fixed probability

of wrongly rejecting the null hypothesis H0, if it is in fact true more practically statisticians use the terms "degrees of freedom" to describe the number of values in the final calculation of a statistic that are free to vary.

Question: Explain the concept of type (1)error and type(2) error .

Answer: Type-I error: In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H0 is wrongly rejected. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H0: there is no difference between the two drugs on average. A type I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them. Type-II error: In a hypothesis test, a type II error occurs when the null hypothesis H0, is not rejected when it is in fact false. For example, in a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug; that is H0: there is no difference between the two drugs on average. A type II error would occur if it was concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different ones.

Question: Explain the concept of Continuity Correction.

Answer: Continuity Correction: In testing of hypothesis, we use a continuity correction of ±1/2 whenever we consider the normal approximation to the binomially distributed random variable X. Because the normal distribution can take all real numbers (is continuous) but the binomial distribution can only take integer values (is discrete) so there for in using normal curve areas to approximate binomial probabilities, a discrete value of the binomial variable is to be replaced by an interval before the z values are computed. Accordingly a discrete value x becomes the interval from x-0.5 to x+0.5 and this sort of adjustment is called continuity correction. Thus, the discrete value 5, adjusted means 4.5 to 5.5.

Question: Explain student's t-statistic .

Answer: Student's t Distribution. The student's t distribution is symmetric about zero, and its general shape is similar to that of the standard normal distribution. It is most commonly used in testing hypothesis about the mean of a particular population. The student's t distribution is defined as (for n = 1, 2,…..).It is a family of curves depending on a single parameter new (the degrees of freedom). As degrees of freedom goes to infinity, the t distribution converges to the standard normal distribution.

Question: what is confidence interval.

Answer: Confidence interval: A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. In interval estimation of an unknown population parameter we find an interval for which we have very high confidence (probability) that it contains the unknown parameter. This level of confidence is dented by (1 – a). It is always very high probability usually 95%, 98%, 99% etc. and the chance that our interval does not contain unknown parameter is called level of significance and it is denoted by a.

Question: what is construction of rendamized design.

Answer: In randomized design treatments are applied randomly there for conclusions drawn are supported by statistical tests. In this way inference are applicable in wide range. The random process also minimizes both systematic and random errors. Selection of an experimental design: The selection of an experimental design of the randomize nature necessities knowledge of variability of the material under test. The selection of the no of replicates to be used is determined by the minimum size of the treatment differences, which the experiment wishes to, detect. Before considering the selection of the design one might investigate the way of controlling the variability of the experimental material or

area. Snedecor describes four ways in which variability of treatment means can be reduced. I. Selection of more homogeneous material II. Stratification of experimental into homogeneous sub groups III. Increasing the no of observations or replications IV. Measurement of one or more related characteristic in order to use regression techniques. The selection of an experimental is dependent upon the nature of the resources of variation in the experimental areas. There are three basic principles of Experimental Design. Randomization Replication Local Control Randomization: Because it is generally extremely difficult for experimenters to eliminate bias using only their expert judgment, the use of randomization in experiments is common practice. In a randomized experimental design, objects or individuals are randomly assigned (by chance) to an experimental group. Using randomization is the most reliable method of creating homogeneous treatment groups, without involving any potential biases or judgments. Completely Randomized Design In a completely randomized design, objects or subjects are assigned to groups completely at random. One standard method for assigning subjects to treatment groups is to label each subject, then use a table of random numbers to select from the labeled subjects. This may also be accomplished using a computer. In MINITAB, the "SAMPLE" command will select a random sample of a specified size from a list of objects or numbers. Randomized Block Design: If an experimenter is aware of specific differences among groups of subjects or objects within an experimental group, he or she may prefer a randomized block design to a completely randomized design. In a block design, experimental subjects are first divided into homogeneous blocks before they are randomly assigned to a treatment group. If, for instance, an experimenter had reason to believe that age might be a significant factor in the effect of a given medication, he might choose to first divide the experimental subjects into age groups, such as under 30 years old, 30-60 years old and over 60 years old. Then, within each age level, individuals would be assigned to treatment groups using a completely randomized design. In a block design, both control and randomization are considered. Replication: The second principle of experimental design is replication, which is a repetition of the basic experiment. In other words, it is a complete run of all the treatments to be tested in the experiment. In all experiments some variations are introduced because of the fact that the experimental units such as individuals or plot of land in agricultural experiments cannot be physically identical. This type of variation can be removed by using the number of experimental units. We there fore perform the experiment more then once. Replication reduces variability in experimental results, increasing their significance and the confidence level with which a researcher can draw conclusions about an experimental factor. Local Control: It has been observed the all extraneous sources of variation are not removed by randomization and replication. As we need to choose a design in such a manner that all the extraneous sources of variation are brought under control. For this purpose, we make use of local control, a term referring to the amount of balancing, blocking and grouping of the experimental units. Balancing means that the treatments should be assign to the experimental units in such a way that result is a balance arrangement of treatments. Blocking means that the experimental units should be collected together to form a homogeneous group. The purpose of local control is to increase the efficiency of the experimental design by decreasing the experimental error

---

Question: Define quantitative and qualitative variables.

Answer: Quantitative Variable: It is a variable that can be measured numerically. e.g. heights, yield, age, weight. Data collected on such a variable are called quantitative data It is of two types: Continuous variables: A variable that can assume any numerical value. 1.34, 2.45 Discrete variables: A variable whose values can be countable. 12, 17, 80 Suppose, you and I both measure the heights of a group of people. We should get the same value for each person. Height is a quantitative variable that we can measure. This is true for any variable where we can say "how much?" or "how many?" Qualitative Variable: It is a variable that cannot assume a numerical value but can be classified into nonnumeric categories e.g. gender, hair color, health status. Data collected on such a variable is called a qualitative data. Suppose, you and I both judge the honesty of a group of people. We may well get different answers for each person. Honesty is a qualitative variable that we judge rather than measure. This is true for any variable which we qualify only as "more" or "less" You and I categorize the gender of a group of people. Although we should get the same value for each person .... Gender is only a qualitative variable that we categorize rather than measure This is true for any variable we judge as "type A", "type B" etc..

Question: Explain strata and homogeneous .

Answer: Strata: There may often be factors which divide up the population into sub-populations (groups) These sub-populations (groups) are called strata. Example Suppose a farmer wishes to work out the average milk yield of each cow type in his herd which has four different types of cows. He could divide up his herd into the four sub-groups (strata) and take samples from these. Homogeneous: The data in which the values are very close to one another, we say these data are homogeneous. Values within strata are homogeneous

---

Question: Error of measurement by deferent examples.

Answer: Errors of measurement: Experience has shown that a continuous variable can never be measured with perfect fineness because of certain habits and practices, methods of measurements, instruments used, etc. the measurements are thus always recorded correct to the nearest units and hence are of limited accuracy. The actual or true values are, however, assumed to exist. For example, if a student's weight is recorded as 60 kg (correct to the nearest kilogram), his true weight in fact lies between 59.5 kg and 60.5 kg, whereas a weight recorded as 60.00 kg means the true weight is known to lie between 59.995 and 60.005 kg. Thus there is a difference, however small it may be between the measured value and the true value. This sort of departure from the true value is technically known as the error of measurement. In other words, if the observed value and the true value of a variable are denoted by x and x + e respectively, then the difference (x + e) – x, i.e. e is the error. This error involves the unit of measurement of x and is therefore called an absolute error. An absolute error divided by the true value is called the relative error.

Question: Explain the " Emprical Realtion.

Answer: Empirical Relation: In a single-peaked frequency distribution, the values of mean, median & mode coincide if the frequency distribution is absolutely symmetrical. But if these values differ, the frequency distribution is said to be skewed. Experience has shown that in unimodel curve of moderate skewness, the median is usually between mean& mode and between them the following approximate relation holds good. MEAN-MODE = 3(MEAN – MEDIAN) OR MODE = 3MEDIAN – 2MEAN The empirical relation does not hold in case of J-shaped or extremely skewed distribution.

Question: What is coefficient of dispersion .

Answer: Coefficient of Dispersion: It is the relative measure of the range. It is defined as: Coefficient of Dispersion = (Xm – X0) / (Xm + X0) Where, Xm stands for the largest value and X0 stands for the smallest one.

Question: explaine the concept of "MARGINAL PROBABILITY"and "joint probability".

Answer: Conditional probability is the probability of some event A, assuming event B. Conditional probability is written P(A|B), and is read "the probability of A, given B". Joint probability is the probability of two events in conjunction. That is, it is the probability of both events together. The joint probability of A and B is written as P(AÇB) or P(A,B) or P(AB). Marginal probability is the probability of one event, ignoring any information about the other event. Marginal probability is obtained by summing (or integrating, more generally) the joint probability over the ignored event. The marginal probability of A is written P(A), and the marginal probability of B is written P(B).

Question: what is difference between raw data and grouped data,please explain it with some example.

Answer: Raw data Data that have not been processed in any manner is called raw data. It often refers to uncompressed text that is not stored in any priority format. It may also refer to recently captured data that may have been placed into a database structure, but not yet processed. Grouped data The data presented in the form of frequency distribution is also known as grouped data.

Question: Explain relative measure of dispersion

Answer: Relative measure of dispersion is one that is expressed in the form of a ratio, co-efficient of percentage and is independent of the units of measurement. A relative measure of dispersion is useful for comparison of data of different nature. A measure of central tendency together with a measure of dispersion gives an adequate description of data. We will be discussing four measures of dispersion i.e. the range, the quartile deviation, the mean deviation, and the standard deviation.

WHAT IS MOMENTS

Answer: Moments are the arithmetic means of the powers to which the deviations are raised. Thus the mean of the first power of the deviations from mean is the first moment about the mean; the mean of the second power of the deviations from mean is the second moment about the mean and so on. First four moments about mean are defined as: m1 = (X – Xbar)/n m2 = (X – Xbar)2/n m3 = (X – Xbar)3/n m4 = (X – Xbar)4/n

---

Question: Explain about the Measure of dispersion ?how we find the co-efficient of variation from the data.

Answer: Co-efficient of variation is used to compare the variability and to check the consistency of two or more series. It is most commonly used relative measure of dispersion. Symbolically, the coefficient of variation, denoted by C.V., is given by C.V = [Standard deviation / Arithmetic mean] × 100 It is used as a criterion of consistent performance; the smaller the coefficient of variation, the more consistent is the performance It is also used as the criterion of variability; the larger the coefficient of variation, the more variability in the data. Dispersion is the spread or variability in a set of data. Consider the following sets of data. 9, 9, 9, 9, 9, 9, 9, 9, 9, 9 10, 6, 2, 8, 4, 14, 16, 12 13, 10, 7, 6, 21, 3, 7, 5 All these three sets of data have same mean (9) but they are different in dispersion. First set of values has no dispersion and there is greater dispersion is third data set as compared to second set of data as its values are more spread away as compared to the values of second set of data. Measure of dispersion is used to measure the dispersion of value from the average. There are several measures of dispersion, the most common being the range, quartile deviation, mean deviation and standard deviation.

Question: Explain the easy way of the pearson;s coeffcient of skewness?

Answer: Pearson’s coefficient of Skewness is defined as: Pearson’s Coefficient of Skewness = (mean – mode) / standard deviation OR Pearson’s Coefficient of Skewness = 3(Mean – median) / standard deviation Example: We calculate the Pearson’s Coefficient of Skewness for the following data Data: 20, 6, 15, 22, 13, 9, 2, 20, 10 First we calculate mean, median, mode and standard deviation of this data. Mean = 13 Median = 13 Mode = 20 Standard Deviation = 6.87 Now we calculate Pearson's Coefficient of Skewness Pearson’s Coefficient of Skewness = (13 – 20) / 6.87 = –1.02 OR Pearson’s Coefficient of Skewness = 3(13 – 13) / 6.87 = 0

Question: what are the main and detailable concept of dispersion

Answer: Dispersion means the extent to which the data/values are spread out from the average. Example: There are many situations in which two different data having the same average e.g. Data 1:5, 5,5,5,5 having mean=5 Data 2:1, 5,6,6,7 having mean=5 Hence in such a situation we, need a measure which tell us how dispersed the data are. The measure used for this purpose is called measure of dispersion.

Question: if X1=20.7% and X2=14.56% then which is more variant? Explain relative measure of dispersion?

Answer: Suppose we have two distributions having coefficient of variations: CV(X1) =20.7% CV(X2)

=14.56% Than the first distribution has more variation as compare to second as: CV(X1)> CV(X2) Relative measure of dispersion is one that is expressed in the form of a ratio, co-efficient of percentage and is independent of the units of measurement. A relative measure of dispersion is useful for comparison of data of different nature. A measure of central tendency together with a measure of dispersion gives an adequate description of data. We will be discussing FOUR measures of dispersion i.e. the range, the quartile deviation, the mean deviation, and the standard deviation.

Question: what is difference beween correlation and regression.

Answer: Correlation: Correlation is a measure of the strength or the degree of relationship between two random variables. Or Interdependence of two variables is called correlation. Regression: Dependence of one variable on the other variable is called regression. Or Estimation or prediction of one variable on the basis of other variable is called regression.

Question: MADAM! WHAT IS CHEBYCHEVS THEOREM?PLEASE GIVE ME DETAIL.

Answer: Knowing the mean and standard deviation of a sample or a population gives us a good idea of where most of the data values are because of the following two rules: Chebychev's Rule: The proportion of observations within k standard deviations of the mean, where k > 1, is at least1 – 1 / k2, i.e., at least 75%, 89%, and 94% of the data are within 2, 3, and 4 standard deviations of the mean, respectively. Empirical Rule: If data follow a bell-shaped curve, then approximately 68%, 95%, and 99.7% of the data are within 1, 2, and 3 standard deviations of the mean, respectively. Chebysehev's theorem Chebysehev's theorem allows you to understand how the value of a standard deviation can be applied to any data set. Theorem: The fraction of any data set lying within k standard deviations of the mean is at least 1 – 1/k2 where k = a number greater than 1. This theorem applies to all data sets, which include a sample or a population.

Question: Explain pie chart.

Answer: Pie Chart: A pie chart is a way of summarizing a set of categorical data. It is a circle which is divided into segments. Each segment represents a particular category. The area of each segment is proportional to the number of cases in that category. To construct the Pie chart we divide the circle into different sectors. The proportion that each component part or category bears to the whole quantity will be the corresponding proportion 360 degree. Angles are calculated by the formula Angle: component part /whole quantity x 360 degree We can construct pie chart with the help of Excel. In this we write the angles in the columns & with the help of chart wizard we draw the pie chart.

Question: write a note on the concept of dot- plot'

Answer: Dot Plot: A dot plot is a way of summarizing data, often used in exploratory data analysis to illustrate the major features of the distribution of the data in a convenient form. For nominal or ordinal data, a dot plot is similar to a bar chart, with the bars replaced by a series of dots. Each dot represents a fixed number of individuals. For continuous data, the dot plot is similar to a histogram, with the rectangles replaced by dots. A dot plot can also help detect any unusual observations (outliers), or any gaps in the data set. In the Example: The numerical value of each measurement in the data set is located on the horizontal scale by a dot. When data values repeat, the dots are placed above one another

---

Question: Waht is grouping error?

Answer: Grouping Error: Grouping error refers to the error that all the values falling in a class are equal to the mid-point of the class interval. It is highly improbable to have a class for which all the values lying in that class are equal to the mid-point of that class. There is no such formula to correct this. This error is usually small and never serious. Only a slight difference occurs between the true answer that we would get from the raw data, and the answer that we get from the data that has been grouped in the form of a frequency distribution

Question: What is difference b/w attribute frequency and relative frequency and why?

Answer: Frequency: The number of measurements in an interval of a frequency distribution is called frequency. The number of observations falling in a particular class is referred to as the class frequency or simply frequency and is denoted by f. Attribute frequency is that in which we find the frequency of some attribute in data. For example we have data containing no of male staff in an office. And we have to find the no of males who do smoking. So frequency of those males is called attribute frequency. The relative frequency of a class is the frequency of the class divided by the total number of frequencies of the class and is generally expresses as a percentage. Example: The weights of 100 persons were given as under: Relative frequency table Weight No. of persons (f) Relative frequency 60 – 62 5 5/90 = 0.056 63 – 65 8 8/90 = 0.089 66 – 68 42 42/90 = 0.467 69 – 71 27 27/90 = 0.3 72 – 74 8 8/90 = 0.08 Total 90

Question: What is the sampling error?

Answer: SAMPLING ERROR: The difference between the estimate derived from the sample (i.e. the statistic) and the true population value (i.e. the parameter) is technically called the sampling error. it is determine because sample is only a part of population & cannot represent the population, no matter how carefully the sample is selected BAR graphs & PIE charts are used to represent data graphically. Such visual representation of statistical data is in the most general terms known as" Graphical Representation". Statistical data can be studied with this method without going through figures, presented in the forms of table

---

Question: what is the purpose of mid range

Answer: MID-RANGE: If there are n observations with x0 and xm as their smallest and largest observations respectively, then their mid-range is defined as Mid range=X0+Xm/2 It is obvious that if we add the smallest value with the largest, and divide by 2, we will get a value which is more or less in the middle of the data-set.

Question: Distinguish between continues data table and discreat data table.

Answer: Continuous Data: A set of data is said to be continuous if the values / observations belonging to it may take on any value within a finite or infinite interval. You can count, order and measure continuous data. For example height, weight, temperature, the amount of sugar in an orange, the time required to run a mile. Discrete Data: A set of data is said to be discrete if the values / observations belonging to it are distinct and separate, i.e. they can be counted (1,2,3,....). Examples might include the number of kittens in a litter; the number of patients in a doctors surgery; the number of flaws in one meter of cloth; gender (male, female); blood group (O, A, B, AB). Difference: A discrete data represents count data such as the number of persons in a family, the number of rooms in a house, the number of deaths in an accident, the income of an individual, etc. Where as the continuous data represents measurement data such as the age of a person, the height of a plant, the weight of a commodity, the temperature at a place, etc.

Question: WHAT IS THE USE OF CUMULATIVE FREQUENCY

Answer: Cumulative frequency is used to determine the number of observations that lie above (or below) a particular value in a data set. The cumulative frequency is calculated using a frequency distribution table, which can be constructed from stem and leaf plots or directly from the data. We need to calculate cumulative frequency for finding median, mode and to draw cumulative frequency curves.

Question: Difference b/w H.M and G.M.

Answer: The Harmonic mean, H, of a set of n values is defined as the reciprocal of the arithmetic mean of the reciprocals of the values. i.e. Harmonic mean is an appropriate type to be used in averaging certain kinds of ratios or rates of change. Uses of Harmonic Mean The Harmonic mean is a measure of

central tendency. In general the it is used when averaging ratio values, as in the problem of determining average trip speed when you travel 30 miles an hour for the first half and 90 miles an hour for the second. (Answer- 45 mph). Geometric Mean

---

Question: what is the importance of deviation in our real time life?

Answer: Finding of deviation is important in real life because it helps in data analysis and compiling it easily.

Question: what are the formulas we are using in quartiles and deciles and their interpretaion and their application in real life.

Answer: Quartile: Quartiles are values that divide a sample of data into four groups containing (as far as possible) equal numbers of observations. A data set has three quartiles. References to quartiles often relate to just the outer two, the upper and the lower quartiles; the second quartile being equal to the median. The lower quartile is the data value a quarter ways up through the ordered data set; the upper quartile is the data value a quarter ways down through the ordered data set. Q1 = l+ h/f (n/4 + c) Percentile: Percentiles are values that divide a sample of data into one hundred groups containing (as far as possible) equal numbers of observations. For example, 30% of the data values lie below the 30th percentile. P2 = l+ h/f (2n/100 + c) Uses: In certain situations, we may be interested in describing the relative quantitative location of a particular measurement within a data set. Quartiles provide us with an easy way of achieving this. Out of these various quartiles, one of the most frequently used is percentile ranking If oil company 'A' reports that its yearly sales are at the 90th percentile of all companies in the industry, the implication is that 90% of all oil companies have yearly sales less than company A's, and only 10% have yearly sales exceeding company A's: It is evident from the above example that the concept of percentile ranking is quite a useful concept, but it should be kept in mind that percentile rankings are of practical value only for large data sets.

Question: How a frriquecy distributation is formed form a data?

Answer: There is no formula of frequency. Basic steps of frequency distribution are as follows Steps in Frequency Distribution: Following are the basic rules to construct frequency distribution: Decide the number of classes into which the data are to be grouped & it depends upon the size of data. Determine the RANGE (difference between the smallest &largest values in data) of data. Decide where to locate the class limit (numbers typically use to identify the classes). Determine the reaming class limits by adding the class interval repeatedly. Distribute the data into classes by using tally marks and sum it in frequency column. Finally, total the frequency column to see that all data have been accounted for.

Question: Define R& D percentage in the dot plot.

Answer: R & D Percentage stands for Percentage of Revenues Spent on Research and Development Dot Plot show almost all of the R & D percentage are b/w 6% and 7%. It means that it is clear from the figure of dot plot that it is clearly showing all the values which falls between 6% & 7%.

Question: Explaine weighted Arithmetic Mean.

Answer: Weighted Arithmetic Mean: The value of each observation is multiplied by the number of times it occurs. The sum of these products is divided by the total number of observations to give the weighted mean. The multipliers or a set of numbers which express more or less adequately the relative importance of various observations in a set of data are technically called the weights. The formula of weighted mean is as follows. Weighted Mean = ? xw/?w

---

Question: How median is calculated from the data?

Answer: Median The median is the value halfway through the ordered data set, below and above which there lies an equal number of data values.The median is the 0.5 quantile. Example: With an odd number of data values, for example 21, we have: Data 96 48 27 72 39 70 7 68 99 36 95 4 6 13 34 74 65 42 28 54 69 Ordered Data 4 6 7 13 27 28 34 36 39 42 48 54 65 68 69 70 72 74 95 96 99 Median 48, leaving ten values below and ten values above With an even number of data values, for example 20, we have: Data 57 55 85 24 33 49 94 2 8 51 71 30 91 6 47 50 65 43 41 7 Ordered Data 2 6 7 8 24 30 33 41 43 47 49 50 51 55 57 65 71 85 91 94 Median Halfway between the two 'middle' data points - in this case halfway between 47 and 49, and so the median is 48 In Case of frequency distribution first we find the median group with n / 2.Then we apply the following formula to find the median of data. Median = l +h/f(n/2 – C) Where l = lower boundary of median group h = class interval f = frequency of median group C = Cumulative frequency of the class above the median group

---

Question: What is Histogram? commulative? Percentole? Symectric and curtile? Quartile?Why we use these? class marks? decile?

Answer: Histogram: A histogram consists of a set of adjacent rectangles whose bases are marked off by class boundaries along the X-axis, and whose heights are proportional to the frequencies associated with the respective classes. Cumulative frequency: Any particular cumulative frequency meant that we were counting the number of observations starting from the very first value of X and going up to that particular value of X against which that particular cumulative frequency was falling. Quartile: The values which divide the distribution into four equal parts are called quartiles. in other words the quartiles Q1,Q2,Q3are the values at or below which lie respectively, the lowest 25,50,75perceent of data. Deciles & Percentiles: The deciles and the percentiles given the division of the total area into 10 and 100 equal parts respectively. Uses: In certain situations, we may be interested in describing the relative quantitative location of a particular measurement within a data set. Quantiles provide us with an easy way of achieving this. Out of these various quantiles, one of the most frequently used is percentile ranking Class Mark: Class mark of a class or class interval is that point which divides the class into two equal parts. It is also known as Mid-point. Symmetrical Distributions: A frequency distribution or curve is said to be symmetrical if values equidistant from a central maximum have the same frequencies, i.e. the curve can be folded along the central maximum in such a way that the two halves of the curve coincide. Please tell that in which lesson you listen or read these words Curtile stream value. I think it is quartile. Stream value means the values appear in data.

Question: What is the concept of cumulative frequency.

Answer: Cumulative frequency is used to determine the number of observations that lie above (or below) a particular value in a data set. The cumulative frequency is calculated using a frequency distribution table Cumulative frequency table Age (X) No. of Students f Cumulative Frequency cf 13 6 6 14 61 67 15 270 337 16 491 828 17 153 981 18 15 996 19 4 1000 Total 1000 The cumulative frequency is determined by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations. Random Number Problem: You can select any of the number randomly because Random Number Tables are constructed according to certain mathematical principles so that each digit has the same chance of selection.

Question: why we construct chart in Statistics?

Answer: Charts are used to illustrate quantitative relationships between the variables.

Question: Quartile deviation is a pure number, describe it.

Answer: The Quartile Deviation is a pure number and is used for comparing the variation in two or more sets of data. In this the pure number means dimensionless number.which is explained below. Dimensionless Number: A dimensionless number is a quantity which describes a certain physical system and which is a pure number without any physical units. Such a number is typically defined as a product or ratio of quantities which do have units, in such a way that all units cancel

Question: what is Relative Frequency

Answer: About tally bars: In construction of frequency distribution, when the number of observations is large, tally bars help us to avoid counting the numbers in the data again and again. If we use frequency numbers directly, we have to read whole data many times for determining the frequency of every class. So, tally bars save our time and energy. The numbers in each class are referred to as frequencies. Example: Suppose the numbers of children in 20 families are as follows: 2, 3, 0, 4, 4, 1, 5, 4, 8, 5, 3, 6, 6, 0, 2, 2, 7, 6, 4, 8 We arrange these values in frequency distribution. Number of children Tally frequency 0 || 2 1 | 1 2 ||| 3 3 || 2 4 |||| 4 5 || 2 6 ||| 3 7 | 1 8 || 2 Total 20 Note that we have used two tally marks for 0, as it is repeated two times and one tally mark for 1 as it is repeated once and three tally marks for 2 as it is repeated three times and so on. The relative frequency of a class is the frequency of the class divided by the total number of frequencies of the class and is generally expresses as a percentage. Example: The weights of 100 persons were given as under: Relative frequency table Weight No. of persons (f) Relative frequency 60 – 62 5 5/90 = 0.056 63 – 65 8 8/90 = 0.089 66 – 68 42 42/90 = 0.467 69 – 71 27 27/90 = 0.3 72 – 74 8 8/90 = 0.08 Total 90

Question: Define MEAN DEVIATION.

Answer: Mean Deviation: As quartile deviation measures the dispersion of the data-set around the median. But the problem is that the sum of the deviations of the values from the mean is zero(No matter what the amount of dispersion in a data-set is, this quantity will always be zero, and hence it cannot be used to measure the dispersion in the data-set.) By ignoring the sign of the deviations we will achieve a NON-ZERO sum, and averaging these absolute differences, again, we obtain a non-zero quantity which can be used as a measure of dispersion. This quantity is known as the MEAN DEVIATION. As the absolute deviations of the observations from their mean are being averaged, therefore the complete name of this measure is Mean Absolute Deviation but generally, it is simply called “Mean Deviation”.

Question: what is positively and negatively skewed?also explain about the what is CUMULATIVE FREQUENCY DISTRIBUTION.

Answer: A frequency distribution or curve is said to be skewed when it departs from symmetry. If the right tail is longer the distribution is positively skewed and if the left tail of the distribution is longer, the distribution is said to be negatively skewed. Cumulative frequency distribution: A cumulative frequency distribution is a plot of the number of observations falling in or below an interval. The graph shown here is a cumulative frequency distribution of the scores on a statistics test.

Question: How can we make the class boundries?

Answer: To find the class boundary of first class, firstly we find the difference between the upper class limit of first class (group) and lower class limit of second class; secondly we divide that difference by two. Then we subtract that resulting value in each lower class limit of each class and add in upper class limit of each class in such a way we can make the class boundaries. For example: For the data given below, we can make class boundaries easily. Data: Class Limits Class Boundaries 3.5-4.4 3.45-4.45 4.5-5.4 4.45-5.45 5.5-6.4 5.45-6.45 Firstly, we find the difference between 4.4(upper class limit of first class) and 4.5(lower class limit of second class). 4.5-4.4=0.1 Secondly, we divide the difference by 2. 0.1/2=0.05 Finally we subtract this resulting value from 3.5 and we get 3.45. And then we add this value in 4.4 and we get 4.45 and so on.

Question: why we use dot plot and what is the main purpose of this even we have so many other ways to plot the data

Answer: DOT PLOT: A dot plot is a way of summarizing data, often used in exploratory data analysis to illustrate the major features of the distribution of the data in a convenient form. For nominal or ordinal data, a dot plot is similar to a bar chart, with the bars replaced by a series of dots. Each dot represents a fixed number of individuals. For continuous data, the dot plot is similar to a histogram,

with the rectangles replaced by dots. A dot plot can also help detect any unusual observations (outliers), or any gaps in the data set. The horizontal axis of a dot plot contains a scale for the quantitative variable that we want to represent. The numerical value of each measurement in the data set is located on the horizontal scale by a dot. When data values repeat, the dots are placed above one another, forming a pile at that particular numerical location.

Question: what is difference between Quartiles and percentiles.

Answer: Quartile: The values which divide the distribution into four equal parts are called quartiles. Quartiles divide the data into four equal-sized and non-overlapping parts. One fourth of the data lies below the Q1 (first quartile). Half of the data lies below Q2 (second quartile) similarly, three quarters of the data lies below Q3 (third quartile) Note: Q2 (second quartile) is also known as median. Percentiles: Percentiles are values that divide a sample of data into one hundred groups containing (as far as possible) equal numbers of observations.

Question: What is the relation b/w Arthimetic, Geometric and Harmonic Mean explain mid-range as well.

Answer: Relation between arithmetic mean, geometric mean and harmonic mean is given below: Arithmetic Mean > Geometric Mean >Harmonic Mean I.e. for a data arithmetic mean is greater than geometric mean and harmonic mean. And geometric mean is greater than harmonic mean. Mid Range: Mid range is the arithmetic mean of the smallest and largest value.

Question: what is empirical relation and why we use it.

Answer: Empirical Relation: In a single-peaked frequency distribution, the values of mean, median & mode coincide if the frequency distribution is absolutely symmetrical. But if these values differ, the frequency distribution is said to be skewed. Experience has shown that in unimodel curve of moderate skew ness, the median is usually between mean& mode and between them the following approximate relation holds good. MEAN-MODE = 3(MEAN – MEDIAN) OR MODE = 3MEDIAN – 2MEAN The empirical relation does not hold in case of J-shaped or extremely skewed distribution.

Question: How we select the number of classes for a given data set?

Answer: The number of classes actually depends on the size of data. When the data are sufficiently large, the number of classes should lie between 10 and 25. In the ranges provided by you the no of classes can be 5 , 10 or more because it depends upon the no of values in the data.As there is no hard & fast rule to determine the no of classes. H.A.Sturges has proposed an empirical rule for determining the number of classes into which a set of observation should be grouped. The rule is K=1+3.3 log N Where K denotes the number of classes & N is equal the total number of observation. For example if there are 100observations, then we have K=1+3.3(2.000) =7.6i.e=8 classes This rule is rarely used in practice.

Question: What are the usefulness of frequency poygon?

Answer: Frequency polygon: A graph that consists of line segments connecting the points formed by the intersection of the class midpoint and the class frequency. A frequency polygon can be used for comparing two or more data sets. It also gives roughly position of the mode, some idea of skewness and kurtosis of the curve.

Question: Define frequency and how a frequency distribution is formed?

Answer: Frequency: It is a record of how often each value (or set of values) of the variable in question occurs. It may be enhanced by the addition of percentages that fall into each category Steps in Frequency

Distribution: Following are the basic rules to construct frequency distribution: 1. Decide the number of classes into which the data are to be grouped & it depends upon the size of data. 2. Determine the RANGE (difference between the smallest &largest values in data) of data. 3. Decide where to locate the class limit (numbers typically use to identify the classes). 4. Determine the reaming class limits by adding the class interval repeatedly. 5. Distribute the data into classes by using tally marks and sum it in frequency column. Finally, total the frequency column to see that all data have been accounted for.

Question: Define Qualitative and Quantitative variable.

Answer: Qualitative Variable: A variable based on categorical data. if the characteristic is non-numerical such as education, sex, eye-color, quality, intelligence, poverty, satisfaction, etc. the variable is referred to as a qualitative variable Quantitative Variable: A variable based on quantitative data. A variable is called a quantitative variable when a characteristic can be expressed numerically such as age, weight, income or number of children

---

Question: How we guess which chart is best for the data?

Answer: WE use different charts to represents different types of data. As we use pie chart & bar graph to represent the single variable and multiple bar & component bar chart to represent the two or more variables of the data. The component bar chart should be used when we have available to us information regarding totals and their components. BUT multiple bar chart should be used when we have the data, which do not add up to give us the totality of some one thing. e.g. imports & exports

Question: what is the difference between the line chart and simple bar cha?

Answer: We use bar chart as well as line chart to best represent the data. In Histogram we use Class Boundaries along X-axis where as in frequency polygon we use Mid point along X-axis HISTOGRAM: A histogram consists of a set of adjacent rectangles whose bases are marked off by class boundaries along the X-axis, and whose heights are proportional to the frequencies associated with the respective classes. FREQUENCY POLYGON: A frequency polygon is obtained by plotting the class frequencies against the mid-points of the classes, and connecting the points so obtained by straight line segments

---

Question: What is the importance of class boundaries?

Answer: Class Boundaries: The true class limits of a class are known as its class boundaries It should be noted that the difference between the upper class boundary and the lower class boundary of any class is equal to the class interval. The problem with class intervals is the space between the intervals. To solve this problem, class boundaries are used. Class boundaries remove space between intervals by dividing it in half. One half is added to the upper limit of one interval and the other half is subtracted from the lower limit of the next interval. By subtracting the class interval from upper class boundary of first class we can find the lower class boundary of first class.

Question: Waht is the difference between Component bar chat and Multiple bar chart.

Answer: We use the Component bar chart & Multiple bar chart according to the requirement of presentation of data. multiple bar & component bar chart to represent the two or more variables of the data. The component bar chart should be used when we have available to us information regarding totals and their components. But multiple bar charts should be used when we have the data, which do not add up to give us the totality of some one thing. e.g. imports & exports. Component bar charts: They are used to represent the cumulation of the various components of data & percentage. Multiple bar charts:

Question: what is the telly and its adventages?

Answer: TALLY MARKS: These are used to show that how many times a value appears in a data. This is a method of showing frequency of particular class. We use Tally marks for the convenience for making the frequency distribution as it used to record each & every value fall in the particular class & after adding them we write it in the digit in the frequency column.

Question: Explain the difference between interval and ratio scale and the concept of zero point in these scales.

Answer: Interval Scale: A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale. Temperature measured on either the Celsius or the Fahrenheit scale is an outstanding example of interval scale because the same difference exists between 20o C (68o F) and 30o C (86o F) as between 5o C (41o F) and 15o C (59o F). It cannot be said that a temperature of 40 degrees is twice as hot as a temperature of 20 degree, i.e. the ratio 40/20 has no meaning. The arithmetic operation of addition, subtraction, etc. is meaningful. Intervals between adjacent scale values are equal with respect the attribute being measured. E.g., the difference between 8 and 9 is the same as the difference between 76 and 77. Ratio Scale: It is a special kind of an interval scale where the sale of measurement has a true zero point as its origin. The ratio scale is used to measure weight, volume, distance, money, etc. There is a rationale zero point for the scale. Ratios are equivalent, e.g., the ratio of 2 to 1 is the same as the ratio of 8 to 4. The, key to differentiating interval and ratio scale is that the zero point is meaningful for ratio scale.

Question: what is the difference between the line chart and simple bar cha?

Answer: We use bar chart as well as line chart to best represent the data. In Histogram we use Class Boundaries along X-axis where as in frequency polygon we use Mid point along X-axis HISTOGRAM: A histogram consists of a set of adjacent rectangles whose bases are marked off by class boundaries along the X-axis, and whose heights are proportional to the frequencies associated with the respective classes. FREQUENCY POLYGON: A frequency polygon is obtained by plotting the class frequencies against the mid-points of the classes, and connecting the points so obtained by straight line segments

---

Question: What is the importance of class boundaries?

Answer: Class Boundaries: The true class limits of a class are known as its class boundaries It should be noted that the difference between the upper class boundary and the lower class boundary of any class is equal to the class interval. The problem with class intervals is the space between the intervals. To solve this problem, class boundaries are used. Class boundaries remove space between intervals by dividing it in half. One half is added to the upper limit of one interval and the other half is subtracted from the lower limit of the next interval. By subtracting the class interval from upper class boundary of first class we can find the lower class boundary of first class.

Question: Waht is the difference between Component bar chat and Multiple bar chart.

Answer: We use the Component bar chart & Multiple bar chart according to the requirement of presentation of data. multiple bar & component bar chart to represent the two or more variables of the data. The component bar chart should be used when we have available to us information regarding totals and their components. But multiple bar charts should be used when we have the data, which do not add up to give us the totality of some one thing. e.g. imports & exports. Component bar charts: They are used to represent the cumulation of the various components of data & percentage. Multiple bar charts:

Question: what is the telly and its adventages?

Answer: TALLY MARKS: These are used to show that how many times a value appears in a data. This is a

method of showing frequency of particular class. We use Tally marks for the convenience for making the frequency distribution as it used to record each & every value fall in the particular class & after adding them we write it in the digit in the frequency column.

Question: Explain the difference between interval and ratio scale and the concept of zero point in these scales.

Answer: Interval Scale: A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale. Temperature measured on either the Celsius or the Fahrenheit scale is an outstanding example of interval scale because the same difference exists between 20o C (68o F) and 30o C (86o F) as between 5o C (41o F) and 15o C (59o F). It cannot be said that a temperature of 40 degrees is twice as hot as a temperature of 20 degree, i.e. the ratio 40/20 has no meaning. The arithmetic operation of addition, subtraction, etc. is meaningful. Intervals between adjacent scale values are equal with respect the attribute being measured. E.g., the difference between 8 and 9 is the same as the difference between 76 and 77. Ratio Scale: It is a special kind of an interval scale where the sale of measurement has a true zero point as its origin. The ratio scale is used to measure weight, volume, distance, money, etc. There is a rationale zero point for the scale. Ratios are equivalent, e.g., the ratio of 2 to 1 is the same as the ratio of 8 to 4. The, key to differentiating interval and ratio scale is that the zero point is meaningful for ratio scale.

Question: What is the differnce between statistics and statistic?

Answer: Statistics: Statistics is a branch of mathematics (i.e. Statistics is a subject) which involves the collection, organization, interpretation, and presentation of data (information). The goal is to make some sort of inference about the data that you have collected (i.e., more than half of the class spent one hour in doing a math homework) Statistic: It is a numerical quantity computed from the sample

---

Question: Define absolute error.

Answer: The difference between the measured value of a quantity and its actual value , given by This difference is called an absolute error. We use it because a continuous variable can never be measured with perfect fineness because of certain habits and practices, methods of measurements, instruments used, etc. the measurements are thus always recorded correct to the nearest units and hence are of limited accuracy. The actual or true values are, however, assumed to exist. For example, if a student's weight is recorded as 60 kg (correct to the nearest kilogram), his true weight in fact lies between 59.5 kg and 60.5 kg, whereas a weight recorded as 60.00 kg means the true weight is known to lie between 59.995 and 60.005 kg. Thus there is a difference, however small it may be between the measured value and the true value. This sort of departure from the true value is technically known as the error of measurement. In other words, if the observed value and the true value of a variable are denoted by x and x + e respectively, then the difference (x + e) – x, i.e. e is the error. This error involves the unit of measurement of x and is therefore called an absolute error.

Question: What is cumulative frequency distributation?

Answer: It is the tabular presentation of the number of data items whose numerical values is less than a given value. To obtain cumulative frequency distribution, we add the frequencies of our frequency table column-wise, we obtain the column of cumulative frequencies. The Cumulative frequency of the last class is the sum of all frequencies in the distribution.

Question: what is Strata and Stratum?

Answer: A sample selected from a population which has been divided into a number of non-overlapping groups or sub populations called strata, such that part of the sample is drawn at random from each stratum.

Question: Are quantitative and qualitative variables were same like Discrete and continuous variables?

Answer: Quantitative Variable: It is a variable that can be measured numerically. e.g. heights, yield, age, weight. Data collected on such a variable are called quantitative data It is of two types: (i) Discrete Random Variable A discrete random variable is one that can take only a discrete set of integers or whole numbers. For discrete variables values are obtained by counting process. For example, if we toss three dice together, and let X denote the number of heads, then the random variable X consists of the values 0, 1, 2, and 3. Obviously, in this example, X is a discrete random variable A discrete random variable represents count data such as the number of persons in a family, the number of rooms in a house, the number of deaths in an accident, the income of an individual, etc. (ii) Continuous Random Variable A variable is called a continuous variable if it can take on any value-fractional or integral—within a given interval, i.e. its domain is an interval with all possible values without gaps. For continuous variables values are obtained by measuring process. A continuous variable represents measurement data such as the age of a person, the height of a plant, the weight of a commodity, the temperature at a place, etc. Suppose, you and I both measure the heights of a group of people. We should get the same value for each person. Height is a quantitative variable that we can measure. This is true for any variable where we can say "how much?" or "how many?" Qualitative Variable: It is a variable that cannot assume a numerical value but can be classified into nonnumeric Categories e.g. gender, hair color, health status. Data collected on such a variable is called a qualitative data or non-numeric data. . Suppose, you and I both judge the honesty of a group of people. We may well get different answers for each person. Honesty is a qualitative variable that we judge rather than measure. This is true for any variable which we qualify only as "more" or "less" You and I categorize the gender of a group of people. Although we should get the same value for each person. Gender is only a qualitative variable that we categorize rather than measure. This is true for any variable we judge as "type A", "type B" etc.

Question: Explain the use of word STATISTICS in singula & plural sense.

Answer: Latin words status, meaning a political state is believed to be the origin of the word “statistics” Statistics: Today the word statistics is used in three different meaning. Firstly, it is used in the sense of data for example price statistics, death statistics etc Secondly, it is used as the plural of the word “statistic” meaning the information obtained from the sample data. Thirdly, it means the science of collecting, presenting, analyzing, and interpreting the numerical facts obtained as a result of a survey.

Question: State about the types of statistics ?

Answer: Statistics as a subject is divided into descriptive and inferential statistics. Descriptive Statistics uses graphical and numerical techniques to summarize and display the information contained in a data set. Inferential Statistics uses sample data to make decisions or predictions about a larger population of data.

Question: What is bais and how it is differnt from random error?

Answer: A systematic error which deprive our resluts from there representativeness. Biase id different from random error in the sence that random error balance out in the long run while biase is cumulative (addition of error) and does not become balance out in long the run.

Question: What are the basic techniques to better understand statistics because i am a new student in this field and i have heard that it is very tough subject.

Answer: The objective of this course is to provide knowledge of basic concepts that will inculcate in the students an attitude of statistical and probabilistic thinking, and will enable them not only to apply statistical techniques to real-world problems but also to use Statistics as a tool for data-based research.You are provided the statistics material in the form of CD's and Handouts and this is the better source to understand the Statistics. Listen and study this material and also do practice on it and if you have any problem put your queries on MDB.

Question: write a brief note on statistics.

Answer: The word "Statistics" which comes from the Latin words status, meaning a political state, originally meant information useful to the state, for example, information about the sizes of population and armed forces. But this word has now acquired different meanings. • In the first place, the word statistics refers to "numerical facts systematically arranged". In this sense, the word statistics is always used in plural. We have, for instance, statistics of prices, statistics of road accidents, statistics of crimes, statistics of births, statistics of educational institutions, etc. In all these examples, the word statistics denotes a set of numerical data in the respective fields. This is the meaning the man in the street gives to the word Statistics and most people usually use the word data instead. • In the second place, the word statistics is defined as a discipline that includes procedures and techniques used to collect process and analyze numerical data to make inferences and to research decisions in the face of uncertainty. It should of course be borne in mind that uncertainty does not imply ignorance but it refers to the incompleteness and the instability of data available. In this sense, the word statistics is used in the singular. As it embodies more of less all stages of the general process of learning, sometimes called scientific method, statistics is characterized as a science. Thus the word statistics used in the plural refers to a set of numerical information and in the singular, denotes the science of basing decision on numerical data. It should be noted that statistics as a subject is mathematical in character. • Thirdly, the word statistics are numerical quantities calculated from sample observations; a single quantity that has been so collected is called a statistic. The mean of a sample for instance is a statistic. The word statistics is plural when used in this sense. Formal Definition of Statistics: Statistics is a branch of mathematics which involves the collection, organization, interpretation, and presentation of data (information). The goal is to make some sort of inference about the data that you have collected (i.e., more than half of the class spent one hour in doing a math homework).

Question: Define sample and Probability.

Answer: Sample: A sample is a group of units selected from a larger group (the population). By studying the sample it is hoped to draw valid conclusions about the larger group Probability: A probability provides a quantitative description of the likely occurrence of a particular event. Probability is conventionally expressed on a scale from 0 to 1; a rare event has a probability close to 0, a very common event has a probability close to 1

Question: what is startified random sampling?

Answer: Stratified Sampling: A stratified sample is obtained by taking samples from each stratum or sub-group of a population. When we sample a population with several strata, we generally require that the proportion of each stratum in the sample should be the same as in the population. Stratified sampling techniques are generally used when the population is heterogeneous, or dissimilar, where certain homogeneous, or similar, sub-populations can be isolated (strata). Simple random sampling is most appropriate when the entire population from which the sample is taken is homogeneous. Some reasons for using stratified sampling over simple random sampling are: the cost per observation in the survey may be reduced; estimates of the population parameters may be wanted for each sub-population; Increased accuracy at given cost.

Question: What is Simple Random Sampling?

Answer: Simple Random Sampling: Simple random sampling is the basic sampling technique where we select

a group of subjects (a sample) for study from a larger group (a population). Each individual is chosen entirely by chance and each member of the population has an equal chance of being included in the sample. Every possible sample of a given size has the same chance of selection; i.e. each member of the population is equally likely to be chosen at any stage in the sampling process

---

Question: what are the basic statistical techniques.

Answer: summarization, graphical representation, averages, dispersion, regression, correlation, index number are basic statistical techniques.

Question: Statistics can help in computer give an example .

Answer: Statistics is a science of facts and figures. This subject is equally important as other subjects. Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. Statistics is Information Science and Information Science is Statistics. It is an applicable science as its tools are applied to all sciences including humanities and social sciences.

Question: Explain the measuring scales with examples.

Answer: Measurement Scales: By measurement scale, we usually mean the assigning of number to observations or objects and scaling is a process of measuring. The four scales of measurements are briefly mentioned below: NOMINAL SCALE: The classification or grouping of the observations into mutually exclusive qualitative categories or classes is said to constitute a nominal scale. Example: Students are classified as male and female. Number 1 and 2 may also be used to identify these two categories.. ORDINAL OR RANKING SCALE: It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurements. For example, the performance of students (or players) is rated as excellent, good fair or poor, etc. Number 1, 2, 3, 4 etc. are also used to indicate ranks. The only relation that holds between any pair of categories is that of “greater than” (or more preferred). INTERVAL SCALE: A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale. Example: Temperature measured on either the Celsius or the Fahrenheit scale is an outstanding example of interval scale because the same difference exists between 20o C (68o F) and 30o C (86o F) as between 5o C (41o F) and 15o C (59o F). It cannot be said that a temperature of 40 degrees is twice as hot as a temperature of 20 degree, i.e. the ratio 40/20 has no meaning. The arithmetic operation of addition, subtraction, etc. is meaningful. RATIO SCALE: It is a special kind of an interval scale where the sale of measurement has a true zero point as its origin. The ratio scale is used to measure weight, volume, distance, money, etc. The, key to differentiating interval and ratio scale is that the zero point is meaningful for ratio scale

---

Question: Is internet the cheapest mode for statistical calculation? What kind of graphs do we need for this subject ? What is the Traditional method of writing down the statistical data/information? Are there any programs(Comuter Softwares) of Statistics for appropriate manipulation of the data? What kind of scale is most commonly used in our country?

Answer: Is internet the cheapest mode for statistical calculation? Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. data sent over the Internet consists of discrete packets that can follow different channels in a sequence over time and rejoin at the final destination, in a process known as packet switching. For that reason, important information was able to flow around damaged or destroyed cables and telephone switching equipment. What kind of graphs do we need for this subject? Graphs exhibit a relationship, often functional, between two sets of numbers as a set of points having coordinates determined by the relationship. Graphs are used to illustrate quantitative

relationships. Details of all graphs are given in next few lectures. What is the Traditional method of writing down the statistical data/information? Factual information, especially information organized for analysis or used to reason or make decisions is called data. The most important part of statistical work is collection of data. There are two types methods of collecting data. Primary and Secondary data: When people think of market research, they tend to think of collecting data directly from customers, prospects, or other stake holders (this is called primary data collection). However, secondary data can also provide a rich source of information. Secondary data are data that already exist in industry-specific reports, previous research on the topic of interest, or data from an organization's own data base. Qualitative sources of secondary data include magazine and newspaper articles and annual reports of industry participants. Data can be representing by graphs, charts and tables. You will study in details about these. Are there any programs(Comuter Softwares) of Statistics for appropriate manipulation of the data? We use SPSS software for statistics. But the use of this software is not included in your course. What kind of scale is most commonly used in our country? All scales are equally important and useful. We use scales for two main reasons. The scale determines the amount of information contained in the data. The scale indicates the data summarization and statistical analyses that are most appropriate.

Question: Define Hypothetical population and non random sampling.

Answer: Hypothetical population: A population is not necessarily real; it may be hypothetical or imaginary. For example, outcomes of an experiment, that is carried out infinitely, make a hypothetical population. It consists of all conceivable ways in which an event can occur, e.g. all possible throws of a die. Such a population does not exist in an actual manner but is only to be thought of. Non-random Sampling: 'Nonrandom sampling' implies that kind of sampling in which the population units are drawn into the sample by using one's personal judgment. In this sampling personal judgment (of an every person) decide that which sampling unit (of population) should be selected for the sample.

Question: What is the difference between quantitative and qualitative?

Answer: Quantitative variable: A variable is called a quantitative variable when a characteristic can be expressed numerically such as age, weight, income or number of children. QUALITATIVE variable: A variable if the characteristic is non-numerical such as education, sex, eye-colour, quality, intelligence, poverty, satisfaction, etc. the variable is referred to as a qualitative variable. A qualitative characteristic is also called an attribute.

Question: What's the role of statistics in today's life? Also, what's the difference between statistics and the probability? Both are same subject?

Answer: Statistics is a science of facts and figures. This subject is equally important as other subjects. Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. Statistics is Information Science and Information Science is Statistics. It is an applicable science as its tools are applied to all sciences including humanities and social sciences. Statistics is divided into two main areas: Descriptive Statistics; all the charts, tables and graphs are examples of descriptive statistics. Probability theory is not needed for this part. Inferential Statistics; all the statistical tests in inferential statistics are based on probability theory. We can safely say that probability theory is the backbone of Inferential Statistics. The Importance of Statistics H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on

unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. Engineers gather data on the quality and reliability of manufactured products. Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. Doctors must understand the origin and trustworthiness of the data that appear in medical journals if they are to offer their patients the most effective treatment Doctors can determine the likelihood that you will develop cancer or a have a heart attack. Political scientists use statistics to determine how citizens feel about current issues and their likelihood to vote for a particular candidate. Statistics is a science of facts and figures. This subject is equally important as other subjects. Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. Statistics is Information Science and Information Science is Statistics. It is an applicable science as its tools are applied to all sciences including humanities and social sciences. Statistics is divided into two main areas: Descriptive Statistics; all the charts, tables and graphs are examples of descriptive statistics. Probability theory is not needed for this part. Inferential Statistics; all the statistical tests in inferential statistics are based on probability theory. We can safely say that probability theory is the backbone of Inferential Statistics. The Importance of Statistics H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. Engineers gather data on the quality and reliability of manufactured products. Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. Doctors must understand the origin and trustworthiness of the data that appear in medical journals if they are to offer their patients the most effective treatment Doctors can determine the likelihood that you will develop cancer or a have a heart attack. Political scientists use statistics to determine how citizens feel about current issues and their likelihood to vote for a particular candidate.

Question: What is the roal of Statistics and Probability in BS(Commerce).

Answer: The Importance of Statistics H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. · Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. · Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision. A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments. A social scientist uses statistical methods in various areas of socio-economic life of a nation. It is sometimes said that “a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there.

Question: what is the difference between inferential and descriptive statistics?

Answer: Descriptive Statistics: Methods of organizing, summarizing, and presenting of data in an informative

way. Its grounds are measure of central tendency and measure of dispersion. Inferential statistics: The methods used to find out something about a population, based on a sample.

Question: What is the role of Statistics and Probability concerned to Information Technology and what is meant Probability?

Answer: The Importance of Statistics H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. Engineers gather data on the quality and reliability of manufactured products. Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. Doctors must understand the origin and trustworthiness of the data that appear in medical journals if they are to offer their patients the most effective treatment Doctors can determine the likelihood that you will develop cancer or a have a heart attack. Political scientists use statistics to determine how citizens feel about current issues and their likelihood to vote for a particular candidate. IMPORTANCE OF STATISTICS IN VARIOUS FIELDS As stated earlier, Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision. A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles. A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments. A social scientist uses statistical methods in various areas of socio-economic life of a nation. It is sometimes said that “a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there”. Probability is the numerical measure of uncertainty. It tell us how likely it is to happen a certain event. In any experiment there are certain possible outcomes; the set of all possible outcomes is called the sample space of the experiment. To each element of the sample space (i.e., to each possible outcome) is assigned a probability measure between 0 and 1 inclusive (0 is sometimes described as corresponding to impossibility, 1 to certainty). Furthermore, the sum of the probability measures in the sample space must be 1. P(A) = The Number Of Ways Event A Can Occur The Total Number Of Possible Outcomes 1. The classical definition of probability The probability P(A) of an event A is equal to the number of possible simple events (outcomes) favorable to A divided by the total number of possible simple events of the experiment, i.e., P(A) =m/n Where m = number of the simple events favorable to A. 2. The relative frequency definition of probability The probability of an event A can be approximated by the proportion of times that A occurs when the experiment is repeated a very large number of times. 3. Axiomatic definition of probability With each random event A in a field of events S, there is associated a non-negative number P(A),called its probability. The probability of the certain event E is 1, i.e., P(E)=1. (Addition axiom) If the event A1, A2, ..., An are mutually exclusive events then P(A1+ A2+ ...+An)= P(A1)+P(A2)+ ...+P(An)

Question: explain the Porpotional Allocation and Hypothetical population.

Answer: Proportional Allocation means when the total sample size is distributed among different strata in proportion to the sizes of strata. Hypothetical means Conditional. it is used with the Population & A hypothetical population is defined as the aggregate of all the conceivable ways in which a specified event can happen.

Question: what is Sampled and Target Population?

Answer: Target population The target population is the total population for which the information is required. For example, if you were to conduct a survey about the most popular types of cars in Lahore, then the target population would be every car in Lahore. Sampled Population: Sampled population is that from which a sample is chosen.

Question: Explain the multi-stage samplig.

Answer: Multistage sampling is a complex form of cluster sampling. Using all the sample elements in all the selected clusters may be prohibitively expensive or not necessary. Under these circumstances, multistage cluster sampling becomes useful. Instead of using all the elements contained in the selected clusters, the researcher randomly selects elements from each cluster. Constructing the clusters is the first stage. Deciding what elements within the cluster to use, is the second stage. The technique is used frequently when a complete list of all members of the population does not exist.

Question: How we can plan right QUESTIONNAIRES for the intended audince.

Answer: COLLECTION THROUGH QUESTIONNAIRES: This method is considered as the standard method for routine business and administrative inquiries. Questions should be few, brief, very simple, and easy for all respondents to answer, clearly worded and not offensive to certain respondents.

Question: Define random sampling.

Answer: In random sampling, all items have some chance of selection that can be calculated. Five common random sampling techniques are: simple random sampling, systematic sampling, stratified sampling, cluster sampling, and multi-stage sampling

---

Question: sampling frame and sampling units.

Answer: The list or map that identifies every unit within the target population is the sampling frame. Such a map or list is needed so that every individual member of the population can be identified unambiguously. The individual members of the target population whose characteristics are to be measured are the sampling units.

Question: What is Ogive and polygon.

Answer: In statistics, an ogive is the curve of a cumulative distribution function. polygon and ogive are same.

Question: What is simple random and stratified sampling.

Answer: Simple random sampling: With simple random sampling, each item in a population has an equal chance of inclusion in the sample. Stratified sampling: In stratified sampling, the population is divided into groups called strata. A sample is then drawn from within these strata. Some examples of strata commonly used by the ABS are States, Age and Sex. Other strata may be religion, academic ability or marital status.

Question: Define cluster sampling.

Answer: Cluster sampling divides the population into groups, or clusters. A number of clusters are selected

randomly to represent the population, and then all units within selected clusters are included in the sample.

Question: ACCEPTANCE AND REJECTION REGIONS.

Answer: ACCEPTANCE AND REJECTION REGIONS: All possible values which a test-statistic may assume can be divided into two mutually exclusive groups: One group consisting of values which appear to be consistent with the null hypothesis (i.e. values which appear to support the null hypothesis), and the other having values which lead to the rejection of the null hypothesis The first group is called the acceptance region and the second set of values is known as the rejection region for a test. The rejection region is also called the critical region.

Question: Null & Alternative hypothesis:

Answer: Null hypothesis: The hypothesis tentatively assumed true in the hypothesis testing procedure. or A null hypothesis, generally denoted by the symbol H0,is any hypothesis which is to be tested for possible rejection or nullification under the assumption that it is true. Alternative hypothesis: The hypothesis concluded to be true if the null hypothesis is rejected. It is denoted by H1.

Question: What is stem and leaf method?

Answer: Stem and leaf display is the method of summarizing data in such a way that no information in the data is lost.

Question: Relation between A.M,G.M and H.M.

Answer: Relation between arithmetic mean, geometric mean and harmonic mean is given below: Arithmetic Mean > Geometric Mean >Harmonic Mean I.e. for a data arithmetic mean is greater than geometric mean and harmonic mean. And geometric mean is greater than harmonic mean.

Question: Quartiles & their Uses.

Answer: Quartile: The values which divide the distribution into four equal parts are called quartiles. Quartiles divide the data into four equal-sized and non-overlapping parts. One fourth of the data lies below the Q1 (first quartile). Half of the data lies below Q2 (second quartile) similarly, three quarters of the data lies below Q3 (third quartile) Q2 (second quartile) is also known as median. Use of quartiles: In order to describe a data set without listing all the data, we have measures of location such as the mean and median, measures of spread such as the range and standard deviation. Quartiles are also used to describe the data in combination with other measures. For example they are used in five number summary of the data. The five number summary, i.e., the minimum, Q1, Q2 (median), Q3, and maximum, give a good indication of where data lie. The five number summary is sometimes represented graphically as a (box-and-)whisker plot.

Question: What is Discrete Random Variable?

Answer: Discrete Variable: A variable that is made up of distinct and separate units or categories and is, most of the times, counted only in whole numbers. Or Discrete Random Variable: A discrete random variable is one that can take only a discrete set of integers or whole numbers. For discrete variables values are obtained by counting process. For example, if we toss three dice together, and let X denote the number of heads, then the random variable X consists of the values 0, 1, 2, and 3. Obviously, in this example, X is a discrete random variable A discrete random variable represents count data such as the number of persons in a family, the number of rooms in a house, the number of deaths in an accident, the income of an individual, etc.

Question: What is Continuous Random Variable?

Answer: Continuous Random Variable: A variable is called a continuous variable if it can take on any value-fractional or integral—within a given interval, i.e. its domain is an interval with all possible values without gaps. For continuous variables values are obtained by measuring process. A continuous variable represents measurement data such as the age of a person, the height of a plant, the weight of a commodity, the temperature at a place, etc.

Question: Explain coefficient of skewness & its formula.

Answer: The 'Coefficient of Skewness' shows the tendency of the data set values to 'bunch' at one end of its distribution, with the values at the other end being relatively dispersed. The mode is the measure indicating the value where most bunching happens. Skewness is measured by working out the extent to which the mode departs from the mean. If the mode is towards the lower values in the data set, then the skewness is said to be positive; if it occurs towards the higher values, the skewness is negative. Sk = mean - mode/ S.D. An alternative formula for the coefficient of skewness is often used. (This is based on the knowledge that the difference between the mean and the mode is generally about three times the difference between the mean and the median): Sk = 3 (mean - median) / S.D.

Question: What is Co-efficient of variation?

Answer: Co-efficient of variation: It is used to compare the variability and to check the consistency of two or more series. It is most commonly used relative measure of dispersion. Symbolically, the coefficient of variation, denoted by C.V., is given by C.V = [Standard deviation / Arithmetic mean] × 100 It is used as a criterion of consistent performance; the smaller the coefficient of variation, the more consistent is the performance It is also used as the criterion of variability; the larger the coefficient of variation, the more variability in the data.

Question: What is empirical rule?

Answer: According to the empirical rule: a) Approximately 68% of the measurements will fall within 1 standard deviation of the mean, i.e. within the interval (`X – S,`X + S) b) Approximately 95% of the measurements will fall within 2 standard deviations of the mean, i.e. within the interval (`X – 2S,`X + 2S). c) Approximately 100% (practically all) of the measurements will fall within 3 standard deviations of the mean, i.e. within the interval (`X – 3S,`X + 3S).

Question: Explain Kurtosis.

Answer: KURTOSIS: The term kurtosis was introduced by Karl Pearson. This word literally means 'the amount of hump', and is used to represent the degree of PEAKEDNESS or flatness of a unimodal frequency curve. When the values of a variable are closely bunched round the mode in such a way that the peak of the curve becomes relatively high, we say that the curve is LEPTOKURTIC. On the other hand, if the curve is flat-topped, we say that the curve is PLATYKURTIC: The normal curve is a curve which is neither very peaked nor very flat, and hence it is taken as a basis for comparison. The normal curve itself is called MESOKURTIC. Kurtosis measures the shape of a distribution,how values are ranged around the mean. A normal distribution has a kurtosis of 3.

Question: What is meant by positive & negative skewness.

Answer: Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. A skewed curve describes a population whose values are not equally distributed about the mean. In a positive skewness there are a small number of very large values; this means that when the curve is drawn there is a long tail after the peak. Put statistically, the mode is lowest, then the median, then the mean

is highest, effectively dragged upwards by the few high results. In a negative skewness the reverse occurs; there are a small number of small values. The tail appears before the peak, and the mean is the smallest, followed by the median and the mode.

Question: What is Measure of Central Tendency?

Answer: Measure of Central tendency: The tendency of the observation to cluster in the central part of the data set is called Central tendency and the summary value is called measure of central tendency. The measures of central tendency are generally known as Averages.The most common are the arithmetic mean (simple average), the median, and the mode.

Question: What is Dispersion?

Answer: Dispersion: The data values in a sample are not all the same. This variation between values is called dispersion. When the dispersion is large, the values are widely scattered; when it is small they are tightly clustered. The width of diagrams such as dot plots, box plots, stem and leaf plots is greater for samples with more dispersion and vice versa. There are several measures of dispersion, the most common being the range, quartile deviation, mean deviation and standard deviation. In many ways, measures of central tendency are less useful in statistical analysis than measures of dispersion of values around the central tendency.

Question: What is Average?

Answer: A single value used to represent the distribution is called average. Most commonly used averages are Mean, Median and Mode.

Question: Uses of Geometric Mean & Harmonic Mean.

Answer: Uses of Harmonic Mean: The Harmonic mean is a measure of central tendency. In general the it is used when averaging ratio values, as in the problem of determining average trip speed when you travel 30 miles an hour for the first half and 90 miles an hour for the second. (Answer- 45 mph). Uses of geometric mean: The geometric mean is a measure of central tendency it uses multiplication rather than addition to summarize data values. The geometric mean is a useful summary when we expect that changes in the data occur in percentages. For example adjustments in salary are often a percentage amount. Geometric means are often useful summaries for highly skewed data. They are also natural for summarizing ratios. Don't use a geometric mean, though, if you have any negative or zero values in your data

---

Question: Explain Primary and Secondary data.

Answer: Primary and Secondary data: When people think of market research, they tend to think of collecting data directly from customers, prospects, or other stake holders (this is called primary data collection). However, secondary data can also provide a rich source of information. Secondary data are data that already exist in industry-specific reports, previous research on the topic of interest, or data from an organization’s own data base. Qualitative sources of secondary data include magazine and newspaper articles and annual reports of industry participants.

Question: What is meant by Frequency?

Answer: Frequency: The number of measurements in an interval of a frequency distribution is called frequency. The number of observations falling in a particular class is referred to as the class

frequency or simply frequency and is denoted by f.

Question: Difference between Grouped and Ungrouped data.

Answer: Ungrouped data: Collection of data in a survey results in a massive volume of statistical data, which are in the form of individual measurements or counts. It is called raw data or ungrouped data. It is difficult to learn anything by examining the ungrouped data. Grouped data: One useful way to organize data is to divide them into similar categories or classes and then count the number of observations that fall into each category. This method produces a grouped data or frequency distribution. Grouped data and frequency distribution are same.

Question: What is the importance of Statistics ?

Answer: The Importance of Statistics: H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. Engineers gather data on the quality and reliability of manufactured products. Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. Doctors must understand the origin and trustworthiness of the data that appear in medical journals if they are to offer their patients the most effective treatment Doctors can determine the likelihood that you will develop cancer or a have a heart attack. Political scientists use statistics to determine how citizens feel about current issues and their likelihood to vote for a particular candidate. IMPORTANCE OF STATISTICS IN VARIOUS FIELDS: As stated earlier, Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. • A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision. • A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles. • A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments. • A social scientist uses statistical methods in various areas of socio-economic life of a nation. It is sometimes said that “a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there”.

Question: How to determine the number of Classes.

Answer: There are no hard and fast rules for deciding on the number of classes which actually depends on the size of data. Statistical experience tells us that no less than 5 and no more than 20 classes are generally used. Use of too many classes will defeat the purpose of condensation and too few will result in too much loss of information. Deciding on the number of classes does not depend on the value of range. To find class interval ‘h’ we should first find the range and divide it by number of classes.

Question: What is Absolute Error?

Answer: The difference between the measured value of a quantity X0 and its actual value X, given by Absolute error= X0-X This difference is called an absolute error.

Question: What are Class Boundries?

Answer: The true class limits of a class are known as its class boundaries. It should be noted that the difference between the upper class boundary and the lower class boundary of any class is equal to the class interval. The problem with class intervals is the space between the intervals. To solve this problem, class boundaries are used. Class boundaries remove space between intervals by dividing it in half. One half is added to the upper limit of one interval and the other half is subtracted from the lower limit of the next interval. By subtracting the class interval from upper class boundary of first class we can find the lower class boundary of first class.

Question: Why we do tabulation?

Answer: A table is a systematic arrangement of data into vertical columns and horizontal rows. The process of arranging data into rows and columns is called tabulation. We need to arrange the data in tabular form in order to comprehend it and getting more information easily.

Question: What is meant by Sampling Frame and Target Population?

Answer: Sampling Frame: Sampling frame (also called survey frame) is the actual set of units from which a sample has been drawn. It consists of all the N sampling units in the population. A list of sampling units from which a sample can be drawn is called sampling frame. E.g. A complete list of the names of students in the virtual university. Sampled and target population is also falls in the category of finite population. Target population: The target population is the total population for which the information is required. For example, if you were to conduct a survey about the most popular types of cars in Lahore, then the target population would be every car in Lahore. Sampled Population: Sampled population is that from which a sample is chosen.

Question: Difference between Skewed and symmetrical distribution.

Answer: A frequency distribution or curve is said to be skewed when it departs from symmetry. If the right tail is longer the distribution is positively skewed and if the left tail of the distribution is longer, the distribution is said to be negatively skewed. A frequency distribution or curve is said to be symmetrical if values equidistant from a central maximum have the same frequencies.

Question: Define Measurement Scales.

Answer: By measurement scale, we usually mean the assigning of number to observations or objects and scaling is a process of measuring. The four scales of measurements are briefly mentioned below: Nominal Scale: The classification or grouping of the observations into mutually exclusive qualitative categories or classes is said to constitute a nominal scale. Example: Students are classified as male and female. Number 1 and 2 may also be used to identify these two categories.. Ordinal or Ranking Scale: It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurements. For example, the performance of students (or players) is rated as excellent, good fair or poor, etc. Number 1, 2, 3, 4 etc. are also used to indicate ranks. The only relation that holds between any pair of categories is that of “greater than” (or more preferred). Interval Scale: A measurement scale possessing a constant interval size (distance) but not a true zero point, is called an interval scale. Example: Temperature measured on either the Celsius or the Fahrenheit scale is an outstanding example of interval scale because the same difference exists between 20o C (68o F) and 30o C (86o F) as between 5o C (41o F) and 15o C (59o F). It cannot be said that a temperature of 40 degrees is twice as hot as a temperature of 20 degree, i.e. the ratio 40/20 has no meaning. The arithmetic operation of addition, subtraction, etc. is meaningful. Ratio Scale: It is a special kind of an interval scale where the sale of measurement has a true zero point as its origin. The ratio scale is used to measure weight, volume, distance, money, etc. The, key to differentiating interval and ratio scale is that the zero point is meaningful for ratio scale

---

Question: Explain the concept of Cumulative Frequency.

Answer: Cumulative frequency is used to determine the number of observations that lie above (or below) a particular value in a data set. The cumulative frequency is calculated using a frequency distribution table. The cumulative frequency is determined by adding each frequency from a frequency distribution table to the sum of its predecessors. The last value will always be equal to the total for all observations.

Question: Why statistics is less implemented as compare to other sciences?

Answer: Statistics is a science of facts and figures. This subject is equally important as other subjects. Statistics is a discipline that has finds application in the most diverse fields of activity. It is perhaps a subject that should be used by everybody. Statistical techniques being powerful tools for analyzing numerical data are used in almost every branch of learning. In all areas, statistical techniques are being increasingly used, and are developing very rapidly. Statistics is Information Science and Information Science is Statistics. It is an applicable science as its tools are applied to all sciences including humanities and social sciences. The Importance of Statistics H.G. Wells anticipated that statistical thinking (numerical literacy) would one day be as necessary for efficient citizenship as the ability to read and write. · Statistics allows a trained person to see the significance of data, the relationship between seemingly unrelated phenomena, and predict what may happen in the future or determine what may have happened in the past. The study and collection of data are important in the work of many professions, so that training in the science of statistics is valuable preparation for a variety of careers. Each month, for example, Government statistical offices release the latest numerical information on unemployment and inflation. Economists and financial advisors as well as policy makers in government and business study these data to make informed decisions. Market research data that reveal consumer tastes influence business decisions. Farmers study data from field trials of new crop varieties. Engineers gather data on the quality and reliability of manufactured products. · Insurance agencies use actuary tables to determine the likelihood that you will have a car accident and will adjust your premiums accordingly. · Doctors must understand the origin and trustworthiness of the data that appear in medical journals if they are to offer their patients the most effective treatment Doctors can determine the likelihood that you will develop cancer or a have a heart attack. · Political scientists use statistics to determine how citizens feel about current issues and their likelihood to vote for a particular candidate. A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision. A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles. A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments. A social scientist uses statistical methods in various areas of socio-economic life of a nation. It is sometimes said that “a social scientist without an adequate understanding of statistics, is often like the blind man groping in a dark room for a black cat that is not there .

Question: Explain me how to use Tally bars.

Answer: About tally bars: In construction of frequency distribution, when the number of observations is large, tally bars help us to avoid counting the numbers in the data again and again. If we use frequency numbers directly, we have to read whole data many times for determining the frequency of every class. So, tally bars save our time and energy. The numbers in each class are referred to as frequencies. Example: Suppose the numbers of children in 20 families are as follows: 2, 3, 0, 4, 4, 1, 5, 4, 8, 5, 3, 6, 6, 0, 2, 2, 7, 6, 4, 8 We arrange these values in frequency distribution. Number of children Tally frequency 0 || 2 1 | 1 2 ||| 3 3 || 2 4 |||| 4 5 || 2 6 ||| 3 7 | 1 8 || 2 Total 20 Note that we have used two tally marks for 0, as it is repeated two times and one tally mark for 1 as it is repeated once and three tally marks for 2 as it is repeated three times and so on.

Question: What is meadian

Answer: abc

Question: Explain the Term hypothesis.

Answer: Dear Student, The term Hypothesis is also called Statistical Hypothesis and it is defined as: "An assumption or statement about the value of unknown population parameter which may or may not be true is called Statistical hypothesis." It is of two types: 1. Null Hypothesis 2. Alternative Hypothesis Null Hypothesis: Any hypothesis which is to be tested for possible rejection under the assumption that it is true is called Null Hypothesis. It is generally denoted by Ho. The hypothesis is usually assigned a numerical value. For example, suppose we think that the average height of students in all colleges is 62 inches. This statement is taken as null hypothesis and is written symbolically as Ho: $\mu = 62$. Alternative Hypothesis: "Any other hypothesis which we accept when the null hypothesis is rejected is called Alternative hypothesis" It is generally denoted by H1 or HA. A null hypothesis is thus tested against an alternative hypothesis H1. For example, if our null hypothesis is Ho: $\mu = 62$, then our alternative hypothesis my be H1: $\mu \neq 62$ or H1: $\mu > 62$ or H1: $\mu < 62$.

Question: What is correletion coefficient?

Answer: Correlation Coefficient: A correlation coefficient is a number between -1 and 1 which measures the degree to which two variables are linearly related. If there is perfect linear relationship with positive slope between the two variables, we have a correlation coefficient of 1; if there is positive correlation, whenever one variable has a high (low) value, so does the other. If there is a perfect linear relationship with negative slope between the two variables, we have a correlation coefficient of -1; if there is negative correlation, whenever one variable has a high (low) value; the other has a low (high) value. A correlation coefficient of 0 means that there is no linear relationship between the variables.

Question: what is ment by percentile coefficient of kurtosis?

Answer: Kurtosis: Karl Pearson introduced the term Kurtosis for the degree of peakedness or flatness of a unimodal frequency curve. Percentile Co-efficient of Kurtosis is another measure of kurtosis which is not widely used. it is given by K=Q.D/ P90 –P10 Where Q.D is the semi inter quartile range & P's are the percentiles. It has been shown that K for a normal distribution is .263 and it lies between 0 and 0.50.

Question: Explain Conditional Probability, Marginal Probability and Joint Probability.

Answer: Conditional probability is the probability of some event A, assuming event B. Conditional probability is written P(A|B), and is read "the probability of A, given B". Joint probability is the probability of two events in conjunction. That is, it is the probability of both events together. The joint probability of A and B is written as P(AÇB) or P(A,B) or P(AB). Marginal probability is the probability of one event, ignoring any information about the other event. Marginal probability is obtained by summing (or integrating, more generally) the joint probability over the ignored event. The marginal probability of A is written P(A), and the marginal probability of B is written P(B).

Question: what is Random Variable?

Answer: Random Variable: A random variable is a rule that assigns a value to each possible outcome of an experiment. For example, if an experiment involves measuring the height of people, then each person who could be a subject of the experiment has associated value, his or her height. A random variable may be discrete (the possible outcomes are finite, as in tossing a coin) or continuous (the values can take any possible value along a range, as in height measurements).

Question: Explain the Concept of "Continous Random Variable"

Answer: Continuous random variable: A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

Question: Explain the concept of inferential statisticts.

Answer: Inferential statistics: In Inferential Statistics we try to get an idea about population parameters using sample data because it is not possible, in many situations, for us to study the whole of population. We therefore resort ourselves to the sample estimates. In drawing conclusion, the decision maker makes use of probability theory

Question: What is continuity correction?

Answer: Continuity Correction Factor A value of .5 that is added to and/or subtracted from a value of a Binomial random variable X when the continuous normal probability distribution is used to approximate the discrete binomial probability distribution

---

Question: what is hypereomatric distribution.

Answer: Hypergeometric Distribution: In probability theory and statistics, the hypergeometric distribution is a discrete probability distribution that describes the number of successes in a sequence of n draws from a finite population without replacement.

Question: What is probablity density function and what is it's significance.

Answer: Dear Student, Probability density function (pdf) is a mathematical expression or formula which gives probabilities for a range of values of a continuous random variable. It is denoted by f(x). It has certain very important properties which we have sent you by email. Probability density functions are of great significance in Statistics. In fact all the conclusions that are made in Inferencial Statistics are due to using appropriate probability density function. Most important probability distributions which are used in Inferencial Statistics are normal distribution, t-distribution, F distribution and chi-square distribution.

Question: What is random variable and how the fdp is related to it?

Answer: RANDOM VARIABLE: Such a numerical quantity whose value is determined by the outcome of a random experiment is called a random variable. For example, no. of children in a family, daily income of a medical store etc. It is of two types (i) Discrete random variable (ii) Continuous random variable Probability density function (pdf) is the expression or formula which gives us the probability for given range of values of the continuous random variable.

Question: What is the concept of normal distribution.

Answer: Gaussian (Normal) Distribution The Normal or Gaussian distribution plays a central role in statistics and has been found to be a very good model for many continuous distributions that occur in real situations. The function is symmetric about the mean, it gains its maximum value at the mean, the minimum value is at plus and minus infinity. The distribution is often referred to as "bell shaped".

Question: What is ORDINAL or RANKING SCALE.

Answer: Where nominal scales don't allow comparisons in degree, this is possible with ordinal scales. Say you think it is better to live in Karachi than in Lahore but you don't know by how much. Example: 1-People or objects with a higher scale value have more of some attribute. 2-The intervals between adjacent scale values are indeterminate. 3-Scale assignment is by the property of "greater than," "equal to," or "less than."

Question: What is the descriptive and inferential Statistics.

Answer: Descriptive Statistics uses graphical and numerical techniques to summarize and display the information contained in a data set. Inferential Statistics uses sample data to make decisions or predictions about a larger population of data.

Question: Differentiate between Quantitative and Qualitative Variable.

Answer: Quantitative Variable: It is a variable that can be measured numerically. e.g. heights, yield, age, weight. Data collected on such a variable are called quantitative data It is of two types: Continuous variables: A variable that can assume any numerical value. 1.34, 2.45 Discrete variables: A variable whose values can be countable. 12, 17, 80 Suppose, you and I both measure the heights of a group of people. We should get the same value for each person. Height is a quantitative variable that we can measure. This is true for any variable where we can say "how much?" or "how many?" Qualitative Variable: It is a variable that cannot assume a numerical value but can be classified into nonnumeric categories e.g. gender, hair color, health status. Data collected on such a variable is called a qualitative data. Suppose, you and I both judge the honesty of a group of people. We may well get different answers for each person. Honesty is a qualitative variable that we judge rather than measure. This is true for any variable which we qualify only as "more" or "less" You and I categorize the gender of a group of people. Although we should get the same value for each person .... Gender is only a qualitative variable that we categorize rather than measure This is true for any variable we judge as "type A", "type B" etc..

Question: DISCUSS STATUS,STATISTICS AND STATISTIC.

Answer: Latin words status, meaning a political state is believed to be the origin of the word “statistics” Statistics: Today the word statistics is used in three different meanings. Firstly, it is used in the sense of data for example price statistics, death statistics etc Secondly, it is used as the plural of the word “statistic” meaning the information obtained from the sample data. Thirdly, it means the science of collecting, presenting, analyzing, and interpreting the numerical facts obtained as a result of a survey.

Question: Define the error of instrument in respect of the ratio measurment?

Answer: Error of Instrument arises when we are measuring any quantity because of the fault in the measuring instrument. For ratio scale we can use the following example. If a student’s weight is recorded as 60 kg (correct to the nearest kilogram), his true weight in fact lies between 59.5 kg and 60.5 kg, whereas a weight recorded as 60.00 kg means the true weight is known to lie between 59.995 and 60.005 kg. Thus there is a difference, however small it may be between the measured value and the true value. This sort of departure from the true value is technically known as the error of measurement.

Question: Define Variable,Discrete Variable and continuous Variable.

Answer: Variable is a characteristic under study that assumes different values for different elements. For example, Height of students in a class, No. of rooms in a house Discrete Variable: A DISCRETE variable is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, 4,..... Discrete variables are usually (but not necessarily) counts. If a variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a box of ten. Continuous Variable: A CONTINUOUS variable is one which takes an infinite number of possible values. Continuous variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

Question: What is the difference between the confidence interval and confidence limits?

Answer: Confidence Interval**:**

A confidence interval gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

If independent samples are taken repeatedly from the same population, and a confidence interval calculated for each sample, then a certain percentage (confidence level) of the intervals will include the unknown population parameter. Confidence intervals are usually calculated so that this percentage is 95%, but we can produce 90%, 99%, 99.9% (or whatever) confidence intervals for the unknown parameter.

The width of the confidence interval gives us some idea about how uncertain we are about the unknown parameter (see precision). A very wide interval may indicate that more data should be collected before anything very definite can be said about the parameter.

Confidence intervals are more informative than the simple results of hypothesis tests (where we decide "reject H0" or "don't reject H0") since they provide a range of plausible values for the unknown parameter

Confidence Limits

Confidence limits are the lower and upper boundaries / values of a confidence interval, that is, the values which define the range of a confidence interval.

The upper and lower bounds of a 95% confidence interval are the 95% confidence limits. These limits may be taken for other confidence levels, for example, 90%, 99%, 99.9%.

If the interval is 26% then α is .74

Question: What is the difference between type-I error and type -II error ?.

Answer: Type-I error:

In a hypothesis test, a type I error occurs when the null hypothesis is rejected when it is in fact true; that is, H0 is wrongly rejected. For example, suppose that an accused is, in fact, innocent (i-e Ho is true) but the finding of the judge is guilty. The judge has rejected the true null hypothesis and is so doing ,has made a type-I error.

Type-II error:

In a hypothesis test, a type II error occurs when the null hypothesis H0, is not rejected when it is in fact false. For example if the accused is, in fact, guilty (i-e Ho is false) and the finding of the judge is innocent, the judge has accepted the false null hypothesis and by accepting the false null hypothesis he has committed a type –II error.

Question: What do you understand by hypotheses.?

Answer: Dear student:

Hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved, for example, claiming that a new drug is better than the current drug for treatment of the same symptoms.

In each hypothesis testing problem, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted H0, against the alternative hypothesis, denoted H1. These two competing claims / hypotheses are not however treated on an equal basis, special consideration is given to the null hypothesis. We have two common situations:

1. The experiment has been carried out in an attempt to disprove or reject a particular hypothesis, the null hypothesis, thus we give that one priority so it cannot be rejected unless the evidence against it is sufficiently strong. For example, H0: there is no difference in taste between coke and diet coke against H1: there is a difference.

2. If one of the two hypotheses is 'simpler' we give it priority so that a more 'complicated' theory is not adopted unless there is sufficient evidence against the simpler one. For example, it is 'simpler' to claim that there is no difference in flavor between coke and diet coke than it is to say that there is a difference.

In simplest way you can say that hypotheses is an assumption which may or may not be true.

Question: What is the difference between the Poisson distribution and the normal distribution?

Answer: Poisson distribution. The Poisson distribution is referred to as the distribution of rare events. Examples of Poisson distributed variables are number of accidents per person, number of sweepstakes won per person, or the number of catastrophic defects found in a production process. While: Normal Distribution. The normal distribution (the "bell-shaped curve" which is symmetrical about the mean) is a theoretical function commonly used in inferential statistics as an approximation to sampling distributions. In general, the normal distribution provides a good model for a random variable, when: There is a strong tendency for the variable to take a central value; Positive and negative deviations from this central value are equally likely; The frequency of deviations falls off rapidly as the deviations become larger.

Question: What is meant by Loaded die?.

Answer: A biased die is known as Loaded die.

Question: What is the difference between Probability distribution and sampling distribution?.

Answer: The probability distribution of any statistic (such as the mean, the standard deviation, the proportion of successes in a sample, etc.) is known as its sampling distribution.

Question: What is the difference between these two limits when we are dealing with continuous random variable: 0<X< . ≤5 x ≤ 0 and>

Answer: In case ofcontinuous random variable there is no differecce both are discribing the same thing either we mention the equal sign or not that is ,the random variable ranging from 0 to 5.

Question: What is meant by standard deviation?.

Answer: Standard deviation tells how tightly a set of values is clustered around the average of those same values.

Question: what is meant by marginal probability function?.

Answer: The individual probability function of the random variables,from the joint probability function,is known as marginal probability function.

Question: In which distributions we used empirical rule & chebychev rule?.

Answer: Empirical rule is applicable to the mound- shape, symmetrical and unimodle (bell shaped)distributions while chebychev apply to any distribution regardless of the shape of the frequency distribution of the data.

Question: What is the difference between frequency and frequency distribution.?

Answer: Frequency:

The number of observations falling in a particular class is known as class frequency or simply frequency.

Frequency distribution.

When we arrange the frequencies in a form of table then it is known as Frequency distribution.

Question: What is the difference between permutation and combination.

Answer: Permutations:

When our purpose is to arrange the objects with respect to order out of" n" then we use permutations.

Combinations:

When we select our objects out of "n" with out considering order then we apply combination.

Question: What is the difference between cumulative frequency distribution and Cumulative Frequency Polygon?.

Answer: There is no difference between cumulative frequency distribution & Cumulative Frequency Polygon,because the graph of cummulative frequency distrbution is known as Cumulative Frequency Polygon/ogive.

Question: What is the relation between these two Moments & Moment Ratios . ?

Answer: Moments: A moment designates the power to which deviations are raised before averaging them. Moment ratio: These are certain ratios in which both numerators and the denominators are moments.

Question: What is meant by mid-rang and mid-quartile range and what is the difference between these two ranges.?

Answer: MID-RANGE: If there are n observations with x0 and xm as their smallest and largest observations respectively, then their mid-range is defined as Mid range=X0+Xm/2. It is obvious that if we add the smallest value with the largest, and divide by 2, we will get a value which is more or less in the middle of the data-set. MID-QUARTILE RANGE: If x1, x2… xn are n observations with Q1andQ3 as their first and third quartiles respectively, then their mid-quartile range is defined as Mid Quartile range= Q1+Q3 /2. Difference: They both used as measures of central tendency because they both provide us with more or less the middle value of data. The difference is that the mid-quartile range is an attempt to address the problem of the range being heavily dependent on extreme scores. An mid-quartile range represents the middle 50% of the scores in the distribution.

Question: What is Mean, Median & Mode?

Answer:

Mean:

The arithmetic mean is the statistician's term for what the layman knows as the average. The arithmetic mean or simply the mean is a value obtained by dividing the sum of all the observations by their number.

THE MEDIAN:

The median is the middle value of the series when the variable values are placed in order of magnitude.

THE MODE:

The mode is defined as that value which occurs most frequently in a set of data i.e. it indicates the most common result.

The median indicates the middle position while the mode provides information about the most frequent value in the distribution or the set of data.

Both median & mode are different methods of calculating the average value of data and they have their advantages & disadvantages .They are used by the statisticians according to their requirement.

Question: What is meant by Dispersion?

Answer: Dispersion means the extent to which the data/values are spread out from the average.

Example:

There are many situations in which two different data having the same average e.g.

Data 1:5, 5,5,5,5 having mean=5

Data 2:1, 5,6,6,7 having mean=5

Hence in such a situation we, need a measure which tell us how dispersed the data are. The measure used for this purpose is called measure of dispersion.

Question: What is meant by Statistics ? What are its Branches ,restrictions & uses?

Answer:

MEANING OF STATISTICS:

The word "Statistics" which comes from the Latin words **status**, meaning political state, originally meant information useful to the state.

In the first place, the word statistics refers to "numerical facts systematically arranged

In the second place, the word statistics is defined as a discipline that includes procedures and techniques used to collect, process and analyze numerical data to make inferences and to research decisions in the face of uncertainty.

Thirdly, the word statistics are numerical quantities calculated from sample observations

Formal Definition of Statistics:

Statistics is a branch of mathematics which involves the collection, organization, interpretation, and presentation of data (information). The goal is to make some sort of inference about the data that you have collected (i.e., more than half of the class spent one hour in doing a math homework).

Descriptive Statistics:

It is that branch of statistics which deals with concepts and methods concerned with summarization & description of the important numerical data.

Inferential Statistics:

It deals with procedures for making inferences about the characteristics of the larger group of data or the whole called the population, from the knowledge derived from only the part of data.

Restrictions:

- It only deals with behavior of aggregates or large groups of data. It has

nothing to do with what is happening to a particular individual or object of the aggregate

- It deals with those characteristics of things which can be numerically described
- Statistical laws are valid on the average or in the long run. There is no guarantee that a certain law will hold in all cases.
- Statistical results might be misleading or incorrect if sufficient care in collecting, processing and interpreting the data is not exercised

Uses:

· A modern administrator whether in public or private sector leans on statistical data to provide a factual basis for decision.

· A politician uses statistics advantageously to lend support and credence to his arguments while elucidating the problems he handles.

· A businessman, an industrial and a research worker all employ statistical methods in their work. Banks, Insurance companies and Government all have their statistics departments.

· A social scientist uses statistical methods in various areas of socio-economic life a nation

---

Question: What is Frequency? What are the steps for making frequency distribution?

Answer: **Frequency:**

It is a record of how often each value (or set of values) of the variable in question occurs. It may be enhanced by the addition of percentages that fall into each category

Steps in Frequency Distribution:

Following are the basic rules to construct frequency distribution:

1. Decide the number of classes into which the data are to be grouped & it depends upon the size of data.
2. Determine the RANGE (difference between the smallest &largest values in data) data.
3. Decide where to locate the class limit (numbers typically use to identify the classes).
4. Determine the reaming class limits by adding the class interval repeatedly.
5. Distribute the data into classes by using tally marks and sum it in frequency column. Finally, total the frequency column to see that all data have been accounted for.

Question: What is Box & Whisker Plot?

Answer: **Box and Whisker Plot (or Box plot):**

A box and whisker plot is a way of summarizing a set of data measured on an interval scale. It is often used in exploratory data analysis. It is a type of graph which is used to show the shape of the distribution, its central value, and variability. The picture produced consists of the most extreme values in the data

set (maximum and minimum values), the lower and upper quartiles, and the median.
A box plot (as it is often called) is especially helpful for indicating whether a distribution is skewed and whether there are any unusual observations (outliers) in the data set.
Box and whisker plots are also very useful when large numbers of observations are involved and when two or more data sets are being compared.

Question: What is Skewness?

Answer: Skewness is defined as asymmetry in the distribution of the sample data values. Values on one side of the distribution tend to be further from the 'middle' than values on the other side.
For skewed data, the usual measures of location will give different values, for example, mode<median<mean would indicate positive (or right) skewness.
Positive (or right) skewness is more common than negative (or left) skewness.
If there is evidence of skewness in the data, we can apply transformations, for example, taking logarithms of positive skew data.

Question: What is population?

Answer: A population is any entire collection of people, animals, plants or things from which we may collect data. It is the entire group we are interested in, which we wish to describe or draw conclusions about.

In order to make any generalizations about a population, a sample, that is meant to be representative of the population, is often studied. For each population there are many possible samples. A sample statistic gives information about a corresponding population parameter. For example, the sample mean for a set of data would give information about the overall population mean.
It is important that the investigator carefully and completely defines the population before collecting the sample, including a description of the members to be included.
Example:
The population for a study of infant health might be all children born in the Pakistan in the 1980's. The sample might be all babies born on 7th May in any of the years

---

Question: What is a Sample?

Answer: A sample is a group of units selected from a larger group (the population). By studying the sample it is hoped to draw valid conclusions about the larger group.

A sample is generally selected for study because the population is too large to study in its entirety. The sample should be representative of the general population. This is often best achieved by random sampling. Also, before collecting the sample, it is important that the researcher carefully and completely defines the population, including a description of the members to be included.
Example:
The population for a study of infant health might be all children born in the Pakistan in the 1980's. The sample might be all babies born on 7th May in any of

the years.

Question: What is Statistic?

Answer: A statistic is a quantity that is calculated from a sample of data. It is used to give information about unknown values in the corresponding population. For example, the average of the data in a sample is used to give information about the overall average in the population from which that sample was drawn.

It is possible to draw more than one sample from the same population and the value of a statistic will in general vary from sample to sample. For example, the average value in a sample is a statistic. The average values in more than one sample, drawn from the same population, will not necessarily be equal.
Statistics are often assigned Roman letters (e.g. m and s), whereas the equivalent unknown values in the population (parameters) are assigned Greek letters (e.g. μ)

Question: What are the different ways of representing the frequency distribution graphically?

Answer: There are three ways of graphical representation of frequency distribution.

HISTOGRAM:

A histogram consists of a set of adjacent rectangles whose bases are marked off by class boundaries along the X-axis, and whose heights are proportional to the frequencies associated with the respective classes.

FREQUENCY POLYGON:

A frequency polygon is obtained by plotting the class frequencies against the mid-points of the classes, and connecting the points so obtained by straight line segments.

FREQUENCY CURVE:

When the frequency polygon constructed over class intervals made sufficiently small for a large number observation, is smoothed, it approaches a continuous curve, such a curve is called Frequency Curve.

Types of Frequency Curves:

The frequency distribution occurring in practice, usually belong to one of the following four types. You will study about them in your next lecture.

1. The Symmetrical Distribution.
2. Moderately Skewed Distribution.
3. Extremely Skewed or J-shaped Distribution
4. U-Shaped Distribution

Question: What is meant by 5-Number Summary?

Answer:

5-Number Summary:

A 5-number summary is especially useful when we have so many data that it is sufficient to present a summary of the data rather than the whole data set. It consists of 5 values: the most extreme values in the data set (maximum and minimum values), the lower and upper quartiles, and the median.
A 5-number summary can be represented in a diagram known as a box and whisker plot. In cases where we have more than one data set to analyze, a 5-number summary is constructed for each, with corresponding multiple box and whisker plots.

Pie Chart : Pie Chart consists of a circle which is divided into two or more mars in accordance with the number

	of distinct classes that we have in our data.
Statistical Inference :	Statistical Inference is an estimate or prediction or some other generalization about a Population based on information contained in a sample.
Statistics :	Statistics is that science which enables to draw conclustions about various phenomena one the basis of real data collected on sample basis.
Sample :	Sample is that part of the Population from which information is collected.
What is meant by order?. :	Order: Arrangement of objects in ascending or descending way is known as order.
Population :	The collection of all individuals,items or data under consideration in statistical study is called Population.
Nominal Scale :	The classification or grouping of observatoins into mutually exclusive qualitative categories is said to constitute a nominal scale e.g students are classified as male and female.
Ordinal Scale :	It includes the characteristic of a nominal scale and in addition has the property of ordering or ranking of measurments e.g the performance of students can be rated as excellent,good or poor.
Interval Scale :	A measurment scale possessing a constant interval size but not true zero point is called an Interval Scale.
Ratio Scale :	It is a special kind of an interval scale in which the scale of measurment has a true zero point as its origin.
Median :	Median of a set of values arranged in ascending or descending order of magnitude is defined as middle value if the number of values is odd and mean of two middle values if the number of values is even.Median is a value at or below which 50% of data lie.
Average :	A single value which intended to represent a distribution or a set of data as a whole is called an average. It is more or less a central value around which the observations tend to cluster so it is called measure of central tendency. Since measure of central tendency indicate the location of the distribution on X axis so it is also called measure of location.

Mean Deviation :	The mean deviation is defined as the arithmetic mean of the deviations measured either from the mean or from the median, all deviations being counted as positive.
Chebshev's Theorm :	Chebshev's Theorem states that "For any number K greater than one at least 1-1/k2 of the data values fall with in K standard deviations of the mean i.e. within the interval.
Moments :	Moments are the arithmetic means of the powers to which the deviations are raised.
Kurtosis :	kurtosis is the degree of peakness of a distribution usually taken relative to a normal distribution.
Correlation :	Correlation is a measure of the strength or the degree of relationship between two random variables. OR Interdependence of two variables is called correlation.
Venn Diagram :	A diagram that is understood to represent sets by circular regions, parts of circular regions or their complements with respect to a rectangle representing the space S is called a Venn diagram. The Venn diagrams are used to represent sets and subsets in a pictorial way and to verify the relationship among sets and subsets.
Mutually Exclusive Event :	Two events are said to be mutually exclusive events if and only if they can not both occur together at the same time. OR Two events are said to be mutually exclusive events if the occurrence of one event discard the occurrence of other event.
Independent events :	Two events A and B in the same sample space S, are defined to be independent (or statistically independent) if the probability that one event occurs, is not affected by whether the other event has or has not occured.

Random variable :	A numerical quantity whose value is determined by the outcome of a random experiment is called a random variable.
Distribution Function :	The function which gives the probability of the event that X takes a value less than or equal TO a specified value x is called a distribution function and is also called the cumulative distribution function.
Cumulative	The function which gives the probability of the event that X takes a value less than or equal TO a

Distribution Function :	specified value x is called a cumulative distribution function and is also called the distribution function.
Sampling Frame :	A sampling frame is a complete list of all the elements in the population.
Sampling Error :	The sampling error is the difference between the the sample statistic and the population parameter.

Probability Samples :	Probability samples are those in which following the sampling plan each unit in the poplation has a known probability of being included in the sample.
Non probability samples :	Non probability samples are those in which the sample elements are the arbitrarily selected by the sampler because in this judgment the elements thus chosen will most effectively represent the Population.
Frequency Polygon :	A frequency polygon is obtained by plotting the class frequencies against the mid-points of the classes, and connecting the points so obtained by straight line segments.
Variable :	A measurable quantity which can vary from one individual or object to another is called a variable.
Constant :	A quantity which can assume only one value is called a constant
Event. :	the possible outcomes of an experiment is known as event.
Data. :	A well defined collection of objects is known as data.
Mode :	The mode is a value which occures most frequently in a set of data i.e. it indicates the most common result

Box and Whisker plot : A Box and Whisker plot provides a graphical representation of data through its five number summary.

The five number summary : A five number summary consists of X0, Q1, median, Q3, and Xm. It enables us to find the shape of the distribution without drawing a graph.

EXHAUSTIVE EVENTS : Two or more than two mutually exclusive events are said to be exhaustive events when their union constitute the entire sample space

Equally likely events : Two events A and B are said to be equally likely when one event is as likely to occur as other

Probability : Probability is defined as the ratio of favorable cases over equally likely cases.

Table : Table is a systematic arrangement of data into vertical columns and horizontal rows.

Tabulation : The process of arranging data into rows and columns is called tabulation.

Classification : The process of arranging data in classes or categories according to some common characteristics present in the data is called classification.

Class Mark or Mid Point : The class mark or mid point is that value which divides a class into two equall parts.

Mid Poin or Class Mark : The mid point or class mark is that value which divides a class into two equal parts.

Measure of location : A single value which intended to represent a distribution or a set of data as a whole is called an average. It is more or less a central value around which the observations tend to cluster so it is called measure of central tendency. Since measure of central tendency indicate the location of the distribution on X axis so it is also called measure of location.

The Semi-interquartile Range : The quartile deviation or the Semi-interquartile Range is defined as half of the difference between the first and third quartiles.

The coefficient The coefficient of variation expresses the standard deviation as the percentage of the arithmetic

of variation : mean.

Disjoint Set : Two sets A and B are said to be disjoint Sets if they have no elements in common.

DISTRIBUTION FUNCTION: : The distribution function of a random variable X, denoted by F(x), is defined by F(x) = P(X < x). The function F(x) gives the probability of the event that X takes a value LESS THAN OR EQUAL TO a specified value x. The distribution function is abbreviated to d.f. and is also called the cumulative distribution function (cdf) as it is the cumulative probability function of the random variable X from the smallest value up to a specific value x.

Experimental design: : An experimental design is a set of rules or a plan to collect the data relevant to the problem under investigation in such a way as to provide the basis for valid and objective inferences about the stated problem. The plan usually consists of collection of the treatments, specification of experimental layout, allocation of treatments.

Experimental Unit: : An experimental unit is the basic unit to which the experiment is performed. It is the basic unit to which the treatment is applied and in which the variable under investigation is measured and analyzed.

Randomized Deigns; : These designs are those in which treatments are applied to experimental units randomly and conclusions are supported by the statistical results.

Basic Randomized Designs: : Randomization Replication Local Control

---

Randomization: : It is a random process of assigning treatments to the experimental unit. The random process implies that every possible allocation of treatments has the same probability.

Replication: : The second principle of an experimental design is replication which is the repetition of the basic experiment. It is a complete run of all the treatments to be tested in the experiment.

Local Control: : It is used to bring all extraneous sources of variations under control. For this purpose we use Local Control, a term referring to the amount of balancing, blocking and grouping of the experimental units.

Complete Randomized Designs : In this design treatments are applied to the experimental units completely at random, that is randomization is done without any restrictions. Design is completely flexible, any number of treatments and any number of units per treatments can be applied.

ANOVA :	Analysis of variance is defined as the procedure by means of which the total variability of the set of data measured by total sum of square is partitioned into components that measure different sources of variations. The procedure thus permits the decomposition of the total SS into to the component SS which are corresponding to the real and suspected sources of variations.
Randomized complete block Design (RCB): :	Randomized complete block Design (RCB) is a design in which • Experimental material is divided into groups or blocks in such a manner that experimental units within a particular block are relatively homogeneous. • Each block contains complete set of treatments i.e. it constitutes a replication of treatments. • Treatments are assigned at random to the experimental units with in each block which means the randomization is restricted with blocks.
Latin Square Design: :	LS design is an arrangement of k treatments in a k*k square, where the treatments are grouped in blocks in two directions, the direction being orthogonal to each other and to the treatments, and where the treatments appear once and only once in each in each direction. It should be noted that in Latin square design, the number of rows, the number of columns and number of treatments must be equal
Critical Value :	The value that separates the critical region from the acceptance region, is called the critical value(s).
Level of significance :	Level of significance of a test is the probability used as a standard for rejecting null hypothesis Ho when Ho is assumed to be true. The level of significance acts as a basis for determining the critical region of the test.
statistics 2 :	Statistics is a science of facts and figures.
Deciles :	Deciles are those nine quantities that divide the distribution into ten equall parts.
Percentiles :	Percentiles are those ninety nine quantities that divide the distribution into hundred equall parts
Arithmetic Mean :	Arithmetic Mean is a value obtained by dividing the sum of the observations by their numbers.
Geometric Mean :	The Geometric Mean G, of a set of n positive values is defined as the positive nth root of their product.
Absolute Measure of Dispersion :	An absolute measure of dispersion is one that measures the dispersion in terms of the same units, or in the square of units as the units of the data.

Dispersion :	The variability that exists between data set.
Relative Measure of Dispersion :	A Relative Measure of Dispersion is one that measures the dispersion in terms of a ratio, coefficient or percentage and is independent of the units of measurement.
Range :	The range is defined as the difference between the maximum and minimum values of a data set.
Quartile Deviation :	The quartile deviation is defined as half of the difference between the first and third quartiles.
Set :	A set is any well defined collection or list of distinct objects.
standard error of estimate :	The degree of scatter of the observed values about the regression line measured by what is called standard deviation of regression or standard error of estimate.
Class of Sets :	A set of sets is called a class.
Primary Data :	The data published or used by an organization which originally collected them are called primary data thus the primary data are thr first hand information collected, complied, and published by an organization for a certain purpose.
Secondary Data :	The data published or used by an organization other than the one which origninally collected them are known as secondary data.
Harmonic Mean :	Harmonic mean is defined as the reciprocal of the arithmetic mean of the reciprocals of the values.
Quartiles :	Quartiles are those three quantities that divide the distribution into four equal parts.
Quantiles :	Collectively the quartiles, the deciles,percentiles and other values obtained by equall sub-division of the data are called quantiles.
Index Number :	An Index Number is a statistical measure which shows changes in a variable or group of related variables with respect to time, geographic location or other characteristics such as income, profession etc.

Standard Deviation :	Standard Deviation is defined as the positive square root of the mean of the squared deviations of the values from their mean.
Variance :	Variance is defined as the square of the standard deviation.
Regression :	Dependence of one variable on the other variable is called regression. OR Estimation or prediction of one variable on the basis of other variable is called regression.
Random Experiment :	An experiment which produces different results even though it is repeated a large number of times under essentially similar conditions is called a random experiment.
Sub Set :	A set that consists of some elements of an other set is called a subset of that set.
Non-Sampling Error :	Such errors which are not attributable to sampling but arise in the process of data collection even if a complete count is carried out.
Skewness :	Skewness is the lack of symmetry in a distribution around some central value (mean,median or mode).It is thus the degree of a symmetry.
Permutation :	an arrangement of all or some of a set of objects in a definite order is called permutation.
Universal Set :	All sets are subsets of one particular set called universal set.
Sample Space :	The set or collection of all possible outcomes of an experiment is called the sample space.
Conditional Probability :	The probability of the occurence of an event A when it is known that some other event B has already occured is called the conditional probability.
Degrees of freedom :	Degrees of freedom can be defined as the number of observations in the sample minus the number of population parameters that are estimated from the sample data (from those observations
P value :	The p-value is a property of the data, and it indicates “how improbable” the obtained result really is.
Test Statistic :	A statistic (i.e. a function of sample data not containing any parameter), which provides a basis for

testing a null hypothesis, is called a test statistics.

Addition law : A probability law used to compute the probability of a union of two events, denoted A and B. It is P(AÈB)=P(A)+P(B)-P(AÇB). For mutually exclusive events, because P(AÇB)=0, it reduces to P(AÈB)=P(A)+P(B).

Alternative hypothesis : The hypothesis concluded to be true if the null hypothesis is rejected.

ANOVA table : A table used to summarize the analysis of variance computations and results. It contains columns showing the source of variation, the sum of squares, the degrees of freedom, the mean square, and the F values.

Bayes' theorem : A method used to compute posterior probabilities.

Binomial probability distribution : A probability distribution showing the probability of x successes in n trials of a binomial experiment.

Binomial probability function : The function used to compute probabilities in a binomial experiment.

Blocking : The process of using the same or similar experimental units for all treatments. The purpose of blocking is to remove a source of variation from the error term and hence provide a more powerful test for a difference in population or treatment means.

Box plot : A graphical summary of data. A box, drawn from the first to the third quartiles, shows the location of the middle 50% of the data. Dashed lines, called whiskers, extending from the ends of the box show the location of data values greater than the third quartile and data values less than the first quartile. The locations of any outliers are also noted.

Central limit theorem : A theorem that enables one to use the normal probability distribution to approximate the sampling distribution of the sample mean and sample proportion whenever the sample size is large.

Consistency : A property of a point estimator that is present whenever larger sample sizes tend to provide point estimates closer to the population parameter

Histogram :	A graphical presentation of a frequency distribution, relative frequency distribution, or percent frequency distribution of quantitative data constructed by placing the class intervals on the horizontal axis and the frequencies on the vertical axis.
Null hypothesis :	The hypothesis tentatively assumed true in the hypothesis testing procedure. or A null hypothesis, generally denoted by the symbol H0, is any hypothesis which is to be tested for possible rejection or nullification under the assumption that it is true.
Normal probability distribution :	A continuous probability distribution. Its probability density function is bell shaped and determined by its mean m and standard deviation s.
Observation :	The set of measurements obtained for a single element.
Ogive :	A graph of a cumulative distribution.
One-tailed test :	A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of the sampling distribution. or The entire rejection region lies in only one of the two tails, either in the right tail or in the left tai, of the sampling distribution of the test-statistic, is called a one-tailed test or one-sided test.
Parameter :	numerical characteristic of a population, such as a population mean, a population standard deviation, a population proportion, and so on.
Point estimate :	A single numerical value used as an estimate of a population parameter.
Point estimator :	The sample statistic that provides the point estimate of the population parameter.
Poisson probability distribution :	A probability distribution showing the probability of x occurrences of an event over a specified interval of time or space.
Poisson probability function :	The function used to compute Poisson probabilities.
Population parameter :	A numerical value used as a summary measure for a population of data (e.g., the population mean, the population variance, and the population standard deviation).

Posterior probabilities :	Revised probabilities of events based on additional information.
Power curve :	A graph of the probability of rejecting H0 for all possible values of the population parameter not satisfying the null hypothesis. The power curve provides the probability of correctly rejecting the null hypothesis.
Power :	The probability of correctly rejecting H0 when it is false.
Probability density function :	A function used to compute probabilities for a continuous random variable. The area under the graph of a probability density function over an interval represents probability.
Probability function :	A function, denoted by f(x), that provides the probability that x assumes a particular value for a discrete random variable.
Qualitative data :	Data that are labels or names used to identify an attribute of each element. Qualitative data may be nonnumeric or numeric.
Qualitative variable :	A variable with qualitative data.
Quantitative data :	Data that indicate how much or how many of something. Quantitative data are always numeric.
t Distribution :	A family of probability distributions that can be used to develop interval estimates of a population mean whenever the population standard deviation is unknown and the population has a normal or near-normal probability distribution.
Target population :	The population about which inferences are made.
Treatment :	Different levels of a factor.
Tree diagram :	A graphical representation helpful in identifying the sample points of an experiment involving multiple steps.

Two-tailed test :	A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in either tail of the sampling distribution.
Type I error :	The error of rejecting H0 when it is true.
Type II error - :	The error of accepting H0 when it is false.
Unbiasedness :	A property of a point estimator when the expected value of the point estimator is equal to the population parameter it estimates.
Union of events A and B :	The event containing all sample points that are in A, in B, or in both. The union is denoted AÈB.
Types of Experimental Designs: :	Systematic Designs Randomized design
Systematic Designs: :	These designs are those in which treatments are applied to the experimental units by some systematic manner that is choice of the experimenter
Acceptance and rejection region: :	All possible values which a test-statistic may assume can be divided into two mutually exclusive groups: One group consisting of values which appear to be consistent with the null hypothesis (i.e. values which appear to support the null hypothesis), and the other having values which lead to the rejection of the null hypothesis. The first group is called the acceptance region and the second set of values is known as the rejection region for a test
Type I error: :	When we perform a hypothesis test, we derive evidence from the sample in the form of a test statistics. There is a possibility that sample may lead us to make a wrong decision. We may reject the hypothesis when it is in fact true. This type of error is called an error of first kind or type I-error. The probability of committing a type I error is denoted by α. Thus α is the probability of rejecting null hypothesis Ho when Ho true.
Type II error: :	When we perform a hypothesis test, we derive evidence from the sample in the form of a test statistics. There is a possibility that sample may lead us to make a wrong decision. We may accept the hypothesis when it is in fact false. This type of error is called an error of second kind or a Type II error. The probability of committing a type II error is denoted by β. Thus β is the probability of accepting null hypothesis Ho when Ho false.
Class midpoint	The point in each class that is halfway between the lower and upper class limits.

:

Complement of event A : The event consisting of all sample points that are not in A.

Dependent variable : The variable that is being predicted or explained. It is denoted by y.

Descriptive statistics : Tabular, graphical, and numerical methods used to summarize data.

Dot plot : A simple graphical summary of data with each observation represented by a dot placed above a horizontal axis that shows the range of values for the observations.

Discrete random variable : A random variable that may assume either a finite number of values or an infinite sequence of values.

Empirical rule : A rule that states the percentages of items that are within one, two, and three standard deviations from the mean for mound-shaped, or bell-shaped, distributions.

Quick Find:

Experiment : A process that generates well-defined outcomes.

Binomial experiment : A probability experiment having the following four properties: consists of n identical trials, two outcomes (success and failure) are possible on each trial, probability of success does not change from trial to trail, and the trials are independent.

Factorial experiment : An experimental design that allows statistical conclusions about two or more factors.

Five-number summary : An exploratory data analysis technique that uses the following five numbers to summarize the data set: smallest value, first quartile, median, third quartile, and largest value.

Frame : A list of the sampling units for a study. The sample is drawn by selecting units from the frame.

Quick Find:

Frequency distribution : A tabular summary of data showing the number (or frequency) of items in each of several non-overlapping classes.

Grouped data : Data available in class intervals as summarized by a frequency distribution. Individual values of the original data are not available.

Independent variable : The variable that is doing the predicting or explaining. It is denoted by x.

Intersection of A and B : The event containing all sample points that are in both A and B. The intersection is denoted AÇB.

Joint probability : The probability of two events both occurring; that is, the probability of the intersection of two events.

Quick Find:

Judgment sampling : A nonprobabilistic method of sampling whereby element selection is based on the judgment of the person doing the study.

Interquartile range (IQR) : A measure of variability, defined to be the difference between the third and first quartiles.

Least squares method : The method used to develop the estimated regression equation. It minimizes the sum of squared residuals (the deviations between the observed values of the dependent variable, yi, and the estimated values of the dependent variable, yi)

Regression equation : The equation that describes how the mean or expected value of the dependent variable is related to the independent variable.

Rejection region :	The range of values that will lead to the rejection of a null hypothesis.
Replication :	The number of times each experimental condition is repeated in an experiment.
Residual :	The difference between the observed value of the dependent variable and the value predicted using the estimated regression equation.
Sample point :	An element of the sample space. A sample point represents an experimental outcome.
Sample statistic :	A numerical value used as a summary measure for a sample (e.g., the sample mean, the sample variance, and the sample standard deviation). The value of the sample statistic is used to estimate the value of the population parameter.
Sampled population :	The population from which the sample is taken.
Sampling unit :	The units selected for sampling. A sampling unit may include several elements.
Sampling with replacement :	Once an element has been included in the sample, it is returned to the population. A previously selected element can be selected again and therefore may appear in the sample more than once.
Sampling without replacement :	Once an element has been included in the sample, it is removed from the population and cannot be selected a second time.
Scatter diagram :	A graph of bivariate data in which the independent variable is on the horizontal axis and the dependent variable is on the vertical axis.
Simple linear regression :	Regression analysis involving one independent variable and one dependent variable in which the relationship between the variables is approximated by a straight line.
Simple random sampling :	Finite population: a sample selected such that each possible sample of size n has the same probability of being selected. Infinite population: a sample selected such that each element comes from the same population and the elements are selected independently.
Standard error :	The standard deviation of a point estimator.

Stem-and-leaf display :	An exploratory data analysis technique that simultaneously rank orders quantitative data and provides insight about the shape of the distribution.
Stratified random sampling :	A probability sampling method in which the population is first divided into strata and a simple random sample is then taken from each stratum.
Hypergeometric probability function :	The function used to compute the probability of x successes in n trials when the trials are dependent.
Multiplication law :	A probability law used to compute the probability of an intersection of two events, denoted A and B. It is P(AÇB)=P(A)P(B|A) or P(AÇB)= P(B)P(A|B). For independent events it reduces to P(AÇB)=P(A)P(B).
Goodness of fit test :	A statistical test conducted to determine whether to reject a hypothesized probability distribution for a population.
Sampling distribution :	A probability distribution consisting of all possible values of a sample statistic.