

STA301

Questions Answers

Prepared by SONU ILYAS MUGHAL

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 conclusions 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 observations into mutually exclusive qualitative categories are 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 measurements e.g. the performance of students can be rated as excellent, good or poor.

Interval Scale:

A measurement 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 measurement 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'sTheorm:

Chebyshev's Theorem states that "For any number K greater than one at least 1-1/k2 of the data values fall within 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 weakness 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 cannot 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 another 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 occurred.

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.

Distribution Function:

less than or equal TO a 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 population 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 occurs 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 Mx. 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 Two events A and B are said to be equally likely when one event is as likely

events: 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 equal parts.

Mid-Point or Class Mark:

The midpoint 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 of variation:

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

Randomized complete block Design (RCB) is a design in which • Experimental material is divided into groups or blocks in such a manner

block

Design

(RCB):

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 within 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

equal parts.

Percentiles

:

Percentiles are those ninety-nine quantities that divide the distribution into hundred equal 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 the 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 originally 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 equal 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 another 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 another 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 spaces.

Conditional

Probability

:

The probability of the occurrence of an event A when it is known that some other event B has already occurred 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 curves
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 the event containing all sample points that are in A, in B, or in both. The

.

events A

and B:

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.

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.

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.

.

:

Judgment

sampling:

A no probabilistic 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, yet, and the estimated values of

the dependent variable, yet)

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 are 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.

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 cumulative frequency distribution is known as Cumulative Frequency Polygon/ogive.

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 are 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 Mean, Median & Mode?

Answer: **Mean:**

The arithmetic mean is the statistician's term for what the layman knows as the
average. The arithmetic means 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 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 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 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 the difference between these two limits when we are dealing with continuous random variable: $0<5$ and $0 \leq x \leq 5$.

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

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 is equally likely; The frequency of deviations falls off rapidly as the deviations become larger.

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 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" without considering order then we apply combination.

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: 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: 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: 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 correlation 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 meant by percentile coefficient of kurtosis?

Answer: Kurtosis: Karl Pearson introduced the term Kurtosis for the degree of
pawedness or flatness of a unimodal frequency curve. Percentile Coefficient
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 "Continuous 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 statistics.

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 hypergeometric 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 probability density function and what is its 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 Inferential Statistics are due to using appropriate probability density function. Most important probability distributions which are used in Inferential Statistics are normal distribution, t-distribution, F distribution and chi-square distribution.

Question: What is random variable and how the fop 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: 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 every person) decide that which sampling unit (of population) should be selected for the sample.

Question: what are the main and derailable 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 between 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 are 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: What is population?

Answer: A population is consisting of the totality of the observation with which we re
concerned.

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.

Question: In which distributions we used empirical rule & Chebyshev rule?

Answer: Empirical rule is applicable to the mound- shape, symmetrical and

unimodal (bell shaped) distributions while Chebyshev apply to any distribution regardless of the shape of the frequency distribution of the data.

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.

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

Answer: MID-RANGE: If there are n observations with x0 and am as them 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… an 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. A mid-quartile range represents the middle 50% of the scores in the distribution.

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: 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 measurement?

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: 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 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 may be H1: $\mu \neq 62$ or H1: $\mu > 62$ or H1: $\mu < 62$.

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.

Question: WHAT ARE 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 – Bar)/n m2 = (X – Bar)2/n m3 = (X – Bar)3/n m4 = (X – Bar)4/n

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 Conditional 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 Permillages 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 bivariate.

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).....(XNA, Yn)

Question: What is median

Answer: ABC

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 lie 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 summaries 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: 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

Generated by Foxit PDF Creator © Foxit Software
http://www.foxitsoftware.com For evaluation only.

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: 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: 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: Explain the use of word STATISTICS in singular & 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 bias and how it is different from random error?

Answer: A systematic error which deprive our results from there representativeness. Biased id different from random error in the sense that

random error balance out in the long run while biased is cumulative (addition of error) and does not become balance out in long the run.

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 (Bar) computed from a sample of size n, is a point estimate of the population parameter (u)

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 formats. It may also refer to recently captured data that may have been placed into a database structure, but not yet processed.

Question: define the Mean 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 are 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 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: 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 Average?

Answer: A single value used to represent the distribution is called average. Most

commonly used averages are Mean, Median and Mode.

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: why we construct chart in Statistics?

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

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: How we calculate the 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.

Question: what is value of central tendency? and why we apply it? and how many

types of central 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 is 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 media 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 the value. b. If the no. of values is even then
Median is the average of n/2 tie and [(n/2) +1] the observations.

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 conditional and un conditional 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 is 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 considered 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 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: 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: 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: 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 is Nan. 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 NC.

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 is 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) Zest
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 the difference between f-distribution, chi-square distribution 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.