PROBABILITY AND STATISTICS

ASSIGNMENT SET-I

1. Define vector space and vector subspace.

*** vector space :** A vector space is a non-empty set V of objects called vectors on which are defined two operations called addition and multiplication by scalars (real numbers) subject to the axioms (or rules) listed below.

• The axioms held for all vectors $\bar{u}, \bar{v}, \bar{w}$ in V and for all scalar c and d,

1. The sum of u and v denoted by $\bar{u}+\bar{v}$ is in V i.e $\bar{u},\bar{v} \in \bar{V}$ (closure property).
2. $\bar{u}+\bar{v} = \bar{v}+\bar{u}$ (commutative property).
3. $(\bar{u}+\bar{v})+(\bar{w}) = \bar{u}+(\bar{v}+\bar{w})$ (Associative property)
4. There is a zero vector O in V such that $\bar{u}+0 = 0+\bar{u} = \bar{u}$
5. For each $\bar{u}$ in V there is a vector $-\bar{u} \in V \ni \bar{u}+(-\bar{u}) = 0$
6. The scalar multiplication $c \in F, \bar{u} \in \bar{V} \Rightarrow c\bar{u} \in V$.
7. $c(\bar{u}+v) = c\bar{u}+c\bar{v}$ (u, v are vectors). $c \in F, \bar{u} \in \bar{V}$ (c-scalar).
8. $(c+d)u = c\bar{u}+d\bar{u}$ (u-vector). $c,d \in F, \bar{u} \in \bar{V}$.
9. $c(d\bar{u}) = (cd)\bar{u}$ $\quad c,d \in F, \bar{u} \in \bar{V}$
10. $1\bar{u} = \bar{u}$

*** Vector subspace :** A subset of a vector space V is called a subspace of V if W is itself a vector space under the addition and scalar multiplication defined on V.

• The properties that should be satisfied are :

(i) The zero vector of V is in H i.e $\bar{0} \in H$.

(ii) If $\bar{u}, \bar{v} \in H$ then $\bar{u}+\bar{v} \in H$.

(iii) $\bar{u} \in H$ and c is any scalar then $c\bar{u} \in H$.

2. State and prove the addition theorem of probability.

If A and B are two events then the probability of occurance of A or B is given by

$P(A)$ or $P(B) = P(A \cup B) = P(A)+P(B)-P(A \cap B)$.

when A and B are mutually exclusive events then

$P(A \text{ or } B) = P(A) + P(B)$.

Proof: Since events are sets,

From set theory, we have

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$.

Dividing eqn with $n(S)$; S - sample space.

$n(A \cup B) / n(S) = n(A)/n(S) + n(B)/n(S) - n(A \cap B) / n(S)$

Then by the definition of probability

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

3. Explain sampling Techniques.

The sampling techniques are used to examine the selected sample from the population is known as sampling Techniques. Sampling Technique is practical and its scope is vast. The whole data is analyzed with better supervision. It require less time and less cost. It gives reliable data.

The sampling techniques are as follows:

- Random sampling
- Stratified sampling (Random sampling).
- Systematic Random sampling

⊛ Random Sampling: In this sampling, each item in the population has an equal and likely possibility of getting selected in the sample. "Method of chance selection".

⊛ Stratified Random sampling: In this method, the population is divided into subgroups to obtain a simple random sample from each group and complete the sampling process. The small groups are also called strata.

⊛ Systematic Random sampling: In this sampling method, the items are chosen from the destination population by choosing the random selecting point and picking other methods after fixed sample period.

4. Write about Testing hypothesis

A Hypothesis is a statement about the population parameter. Hypothesis testing is a procedure that helps us to ascertain the likelihood of hypothesized parameter being correct by making use of sample statistics. The two hypothesis in a statistical test are normally referred to as

a. Null hypothesis: It is which, is tested to be actually tested for acceptance or rejection is termed as Null hypothesis. According to R.A. Fisher, "Null hypothesis is the hypothesis which is tested for rejection under the assumption that it is true".

$H_0 : \mu = \bar{x}$; μ = population mean, $\bar{x}$ = sample mean.

b. Alternative Hypothesis: Any hypothesis which is compliment-ary to the null hypothesis is called an alternative hypothesis. Rejection of H_0 leads to the acceptance of alternative hypo-thesis which is denoted by H_1.

$H_1 : \mu \neq \bar{x}$ (two tailed test).

$H_1 : \mu > \bar{x}$ or $H_1 : \mu < \bar{x}$ (right tailed and left tailed tests).

5. Differentiate correlation of regression analysis.

CORRELATION.	REGRESSION.
• It determines the interconnection or a co-relationship between the variables.	• It explains how an independe-nt variable is numerically associated with the dependent variable.
• In correlation, both the indepen-dent and dependent values have no difference.	• In regression, both the dependent and independent variables are different.
• The main objective of correlation is to find out ~~of~~ a quantitative	• The main purpose is to calculate the values of a random variable

or numerical value expressing the association between the values.	based on the values of a fixed variable.
• It stipulates the degree to which both variables can move together.	• It specifies the effect of the change in the unit in the known variable (p) on the evaluated variable (q).
• It helps to constitute the connection between the two variables.	• It helps in estimating a variable's value based on another given value.

ASSIGNMENT SET - II.

1. Write about linear Transformations and linearly independent sets.

Linear Transformations: A transformation or a function T from R^n in R^n is said to be linear if

1. $T(\bar{u}+\bar{v}) = T(\bar{u})+T(\bar{v})$ where $\bar{u},\bar{v} \in R^n$
2. $T(c\bar{u}) = cT(\bar{u}) \forall \bar{u} \in R^n$ & $c \in R$.

Let $T: V \to W$ be a linear transformation then the kernel (T) is the set of all $\bar{u}$ in V such that a $T(\bar{u}) = \bar{0}$ (the zero vector in W).

The range of T is set of all vectors $\bar{w} \ni T(\bar{u}) = \bar{w}$

Kernel of $T = \{c, d\}$.

Domain $= V = \{a, b, c, d, e\}$.

co-domain $= W = \{a, b, o, e, f\}$.

Range $= \{a, b, o, e\}$.

Linearly independent set: An indexed set of vectors $\{\bar{v}_1 \bar{v}_2 \ldots \bar{v}_n\}$ is a vector space V is said to be linearly independent if the vector equation

$c_1\bar{v}_1 + c_2\bar{v}_2 + \ldots + c_n\bar{v}_n = \bar{0}$ has only the trivial solution

$c_1 = 0, \; c_2 = 0, \ldots \; c_n = 0.$

2. Define conditional probability and state and prove the multiplication theorem of probability.

conditional probability: A and B are two events in a sample space, then conditional probability of A/B is defined as the probability of A after the occurance of B and is given by

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \; P(B) \neq 0.$$

similarly

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, \; P(A) \neq 0.$$

Multiplication theorem of probability:

> If A and B are dependent events, then the probability of both events occuring simultaneously is given by:

$P(A \cap B) = P(B) \cdot P(A/B)$.

> If A and B are independent events, then the probability of both events occuring simultaneously is given by:

$P(A \cap B) = P(A) \cdot P(B)$.

proof: W.K.T, the conditional probability of event A given that B has occurred is denoted by $P(A/B)$ and is given by:

$P(A/B) = \dfrac{P(A \cap B)}{P(B)}$; where, $P(B) \neq 0$.

$P(A \cap B) = P(B) \times P(A/B) \longrightarrow (1)$

$P(B/A) = \dfrac{P(B \cap A)}{P(A)}$; where, $P(A) \neq 0$.

$P(B \cap A) = P(A) \times P(B/A)$

since, $P(A \cap B) = P(B \cap A)$

$P(A \cap B) = P(A) \times P(B/A) \longrightarrow (2)$

From (1) and (2), we get:

$P(A \cap B) = P(B) \times P(A/B) = P(A) \times P(B/A)$ where,

$P(A) \neq 0$, $P(B) \neq 0$.

For indpt. events A and B, $P(B/A) = P(B)$, then eq(2) can be modified into, $P(A \cap B) = P(B) \times P(A)$

3. Describe Discrete and continuous distributions.

Discrete distributions: Discrete probability distribution is a type of probability distribution that defines the probability of occurences of discrete random variables. It is expressed in tabular form.

example: If a coin is tossed twice then the statistical experiment can give four possible outcomes i.e: HH, TT, HT and

TH respectively. Assume that, 'x' is a random variable that denotes the number of tails (ie: 0, 1 and 2) in this experiment since, values are different, they said to be discrete random variables.

No. of tails	probability
0	0.25
1	0.50
2	0.25

The above table depicts the discrete random variable associated with its probability of occurance.

Continuous distributions: continuous distribution is another type of probability distribution that defines the probability of occurences of continuous random variables. It can not be expressed in the tabular form. Insted, it makes use of an equation known as probability density function inorder to define continuous probability distribution.

<u>example</u>: Let $f(x) = a$ be the probability density function where $0 < a < -\infty$. If the probability of a random variable x is calculated in such a way that $P(x \leq a)$ then it implies that the area of the curve is bounded under the curve from a to $-\infty$. This can be as shown in the below graph.

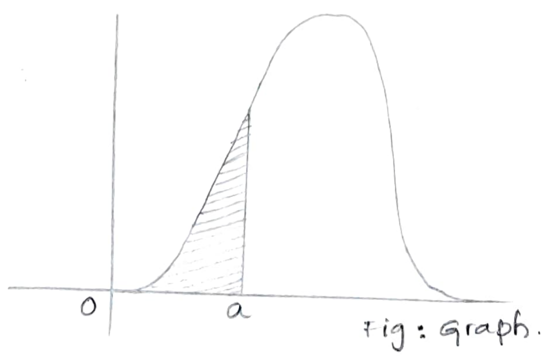

Fig: Graph.

4. Explain point estimate, interval estimate, and confidence level.

<u>point estimate</u>: A point estimator is a single number that is used to estimate an unknown population parameter. It

is useful if we have an idea of the error that might be involved.

Eg: sample mean ($\bar{x}$) is an estimator when population μ is unknown.

interval estimate: An interval estimate is a range of values used in making estimation of a population parameter. It has the advantage of showing the error in two ways -

1. by the extent of its range, and.
2. by the probability of true population parameter lying within the range.

confidence levels: The probability that we associate with an interval estimate is called the confidence level. It indicates how confidently we can say that the interval estimate will include the population parameter. The high the probability the more is the confidence.

5. What is ANOVA.

Analysis of variance is the separation of variances ascribble to one group of causes from the variance ascribble to another group. It is a collection of powerful statistical models and their associated techniques which help us to examine variance (σ^2) between a number (more than two) of samples.

ANOVA is a powerful statistical tool which separates the assignable & chance variations in the data.

The main purpose of ANOVA is to test the homogenity of several means.

This technique is also helpful in comparing the estimates due to various factors.