Applied Statistical Methods

Dr. Abhishek Das

Call: 9864103203

Mail: abhishek.das@vit.ac.in

Office: SJT 711A

Module:1	**Introduction to Statistics:**	**7 hours**
Introduction to Statistics and data analysis-Measures of central tendency, Measures of dispersion, Skewness and Kurtosis.		
Module:2	**Correlation and regression:**	**5 hours**
Correlation and Regression–Rank Correlation-Partial and Multiple Correlation Regression, Multiple Regression.		
Module:3	**Random Variables**	**5 hours**
Introduction to discrete random variables – Binomial – Poisson – Geometric, continuous random variables-Normal, Student's T, expectation of random variables, mean and variance.		
Module:4	**Testing of hypothesis I:**	**5 hours**
Introduction-Types of errors, Critical region, procedure of testing hypothesis- tests of hypotheses-Z- test for Single Proportion, Difference of Proportion, Single mean and difference of means.		

Module:5	Testing of hypothesis II:	6 hours
Small Sample Tests - Student t-test, F-test, Chi-Square test for independence of Attributes, Analysis of Variance-Principles of experimental design, Completely randomized design, Randomized block design, Latin Square design- Problems.		

Mode of Evaluation

Digital Assignments-1
Quiz-2
Continuous Assessment Test-2
Final Assessment Test-1

Few definitions:

To understand the following definitions, we will use the following study:
As a quality control expert you want to know the proportion of good computer chips produced by a manufacturing unit.

Populations:

A population is a collection of all distinct individuals or objects or items under study. The number of entities in a population, called the population size.

Example: collection of all computer chips produced by the manufacturing unit.

Variable:

Once a population is fixed, we then want to study a characteristic of the individuals (or objects or items) in the population. This individual characteristic is called a variable (value of which can vary from individual to individual, or from one object to another within the population).

Example: quality of a computer chip.

Most of the variables can be classified in two broad categories:

(i) Categorical (or qualitative) variable: If it assigns a categorical (non-numerical) value to each individual or object. Examples: gender, nationality, etc.

(ii) Quantitative variable: If it assigns a numerical value to each individual or object. Examples: age, income, height, etc.

Example: quality of a computer chip is a categorical variable (since chip is classified as either good or bad).

Parameter:

A summary value of the variable for the population is called a parameter. In a statistical study we are interested in a parameter since often it gives a fairly good idea about the population.

Example: the percentage of good computer chips in the population is a parameter of interest.

Note: A variable is a characteristic of an individual or object, and a parameter is a characteristic of the population (with respect to a variable).

Sample:

A sample is a part of a population and the sample size is denoted by *n*. A sample should be a representative of the population.

Basic Steps in a Statistical Study

For any statistical study, there are some basic steps to be followed once we draw a sample. These are:

- Step 1: Gather first-hand information from the sample and this is called the raw data.
- Step 2: Tabular representation of the raw data, i.e., represent the raw data in a table.
- Step 3: Pictorial representation of the data, i.e., draw diagrams with the organized data in a table.
- Step 4: Numerically summarize the data, i.e., describe the entire data set with some key numbers.
- Step 5: Analyze the data using mathematical formulae.
- Step 6: Draw the final inference or conclusion about the population under study.

Representation of the data

Given a raw data set, we can rearrange it in two different ways.

- **Frequency distribution or Discrete frequency distribution:**

  Using the frequency of the variable we can arrange it. This representation of the data is known as frequency distribution.

- **Grouped frequency distribution or Continuous frequency distribution:**

  Again we can arrange it for the class intervals. For this situation, it is called as Grouped frequency distribution of the variable.

Frequency distribution of ungrouped data:

Given below are marks obtained by 20 students in Math out of 25.

21, 23, 19, 17, 12, 15, 15, 17, 17, 19, 23, 23, 21, 23, 25, 25, 21, 19, 19, 19

Marks Obtained	Tally Marks	Frequency
12	|	1
15	||	2
17	|||	3
19	𝍸	5
21	|||	3
23	||||	4
25	||	2

Frequency distribution of grouped data:

The presentation of the above data can be expressed into groups. These groups are called classes or the **class interval**.

Each class interval is bounded by two figures called the **class limits**.

Marks	Number of Students (Frequency)
0 - 10	0
10 - 20	11
20 - 30	9

Note: The lower value of a class interval is called lower limit and upper value of that class interval is called the upper limit. Thus, each class interval has lower and upper limits.

For Example:

In the class interval 10 - 20, 10 is the lower limit and 20 is the upper limit.

Exclusive form of data:

The previous table is expressed in the exclusive form.

The class intervals are 0 - 10, 10 - 20, 20 - 30.

In this, we include lower limit but exclude upper limit.

So, 10 - 20 means values from 10 and more but less than 20.

20 - 30 would mean values from 20 and more but less than 30.

Data in the inclusive form:

Marks obtained by 20 students in Math test are given below.

23, 0, 14, 10, 15, 3, 8, 16, 18, 20, 1, 3, 20, 23, 24, 15, 24, 22, 14, 13

Let us represent this data in the inclusive form.

Marks	Number of Students (Frequency)
0 - 10	6
11 - 20	9
21 - 30	5

How to make it continuous?

If "*d*" is the gap between the upper limit of any class and the lower limit of the succeeding class, the class boundaries for any class are then given by:

$$\textbf{Upper class boundary} = \textbf{Upper class limit} + \frac{d}{2}$$

$$\textbf{Lower class boundary} = \textbf{Lower class limit} - \frac{d}{2}$$

Marks	Number of Students (Frequency)
-0.5 – 10.5	6
10.5 – 20.5	9
20.5 – 30.5	5