

Module 2

INTRODUCTION TO PROBABILITY AND STATISTICS

Introduction to Probability & Statistics

- **Statistics**

A branch of science that deals with collection, processing and analysis of data.

Past world war II, a lot of emphasis was laid on quality improvement in industrial products. Developed countries made extraordinary advances on this.

Much of this success is attributed to use of statistical methods and statistical thinking among managers.

Basic Statistical Problem

- Suppose one is interested in information on some numerical characteristic of a big population.
- Due to factors such as cost/time/size of the population, it is not possible to take observations on each and every unit of the population (would be the best thing to do if possible, and in that case no statistics would be required)
- We take observations on a selected sample of units in population and try to make inferences (by developing various statistical methods) about the big population based on this sample.

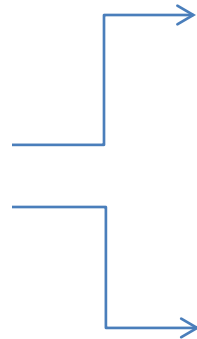
Example

- Consider the population of all college going students in a India. Suppose one is interested in estimating the mean height of these students (or say one is interested in testing a claim that mean height of these students is at least 5 feet and 6 inches).
- Due to various reasons it may not be possible to measure heights of all college going students in the country. Thus we take a sample of students from the population and measure their heights.
- Based on this sample we try to estimate the mean height of population of all college going students in the country (or test the claim that the mean height of the population is at least 5 feet and 6 inches).

Issues Involved

- How to take sample so that it is a good representative of the population under study?
- For example it may be inappropriate to take whole sample from just one state (say, Haryana, where students are known to be taller)
- Which estimate (statistical procedure) to use?

Statistics



Descriptive Statistics
(deals with collection,
processing and organization
of data/sample)

Inferential Statistics
(deals with drawing inference
about the population based
on sample collected)

Probability Theory

- Data (sample) generated from the population has randomness yielding variability (e.g. data sets collected by different persons may yield different estimates of the population mean).
- To arrive at good statistical procedures it is important to study this variability.
- Probability theory is useful in studying the randomness/variability.
- Probability Theory is also helpful in developing good (optimal) statistical inferential procedures and in comparing different statistical inferential procedures developed for the same situation.

Example (Use of Probability Theory in developing Statistical Inferential Procedure):

- A new manufacturing process is to be installed.
- In the long run company can afford only up to 5% defectives.
- Question: Whether the new manufacturing process should be installed?
- Sample/Data: 100 items are produced using the new manufacturing process out of which 10 are found to be defective

New manufacturing process is acceptable (i.e., it produces at most 5% defectives)

⇒ the probability of obtaining 10 or more defectives in random sample of 100 is at most

$$\sum_{j=10}^{100} \binom{100}{j} \left(\frac{5}{100}\right)^j \left(1 - \frac{5}{100}\right)^{100-j} = 0.0282 \quad (\text{Binomial Probability})$$

⇒ If the process is good then probability of obtaining given data is very low (0.0282)

⇒ the new process does not seem to be acceptable

In the above example

- Population: All items ever produced by new manufacturing process.
- Sample/Data: 100 items taken for study
- Statistical Inference drawn: The manufacturing process does not seem to be acceptable. Clearly in arriving at this inference we have used probability theory.

Abstract of Next Module

- We saw that probability theory plays a vital role in Statistics.
- In next module we will study the concept of probability as a measure of uncertainty (or certainty).

Thank you for your patience

