

## **ENGINEERING STATISTIC**

### **1<sup>st</sup> LEC TURE: INTRODUCTION OF STATISTICA**

**Statistics:-** Statistics is the area of science that deals with collection, organization, analysis, and interpretation of data.

**A variable:-** is a characteristic, or attribute under study that can assume different values like, length, weight.

**Data:-** are the values (measurement or observations) that the variables can assume.

**Random variables:-** are variables whose values are determined by chance.

- A collection of data values forms a data set.
- Each value in the data set is called a data value or a datum .

Depending on how data are used, the body of knowledge called statistics is sometimes divided into two main areas:

- 1-Descriptive Statistics.*
- 2-Inferential Statistics.*

**Descriptive Statistics :** consists of organization-summarization, and-presentation of data by using tables, graphs and summary measures.

**Inferential Statistics:** deals with making decision , inferences, predictions, and forecasts about populations based on results obtained from samples. It uses probability (the chance of on event occurring) to achieve inferences.

**Population:** consists of all elements (human or otherwise) that are being studied.

**Sample:** is a group of elements selected from a population.

**Variables:** divided into

a -Qualitative: variables that can't be measured numerically but can classified into different categories according to some characteristics or attribute such as color, gender of person.

b -Quantitative : variables that can be measured numerically . It can divided into :

a- discrete: assume values that can be counted such as number of houses, cars accidents.

b- continuous: assume all values between any two specific values such as length, time, Wight. They are obtained by measuring. Data must be measured and rounded due to the limits of the measuring device.

-Data of continuous variables are written in boundaries. boundaries are given in one additional decimal place and always end with the digit 5.

**Measurement Scales:** how variables are categorized, counted, or measured.

**1- Nominal level of measurement:** classifies data into mutually exclusive (non-overlapping), exhausting categories in which no order or ranking can be imposed on the data. Exp., (Democratic, Republican, Independent), Gender (male, female)

**2- Ordinal level of measurement:** classifies data into categories that can be ranked, however, precise differences between the ranks do not exist.

A, B, C, D, E, F a large variation exists among the individuals in each class.  
Grads (A, B, C, D,...) rating (poor, good, excellent)

**3- Interval level of measurement:** ranks data and precise differences between units of measure do exist however, there is no meaningful zero. 72F°, 73F°, 0F° (0F° does not mean no heat).

**4- Ratio level of measurement:** possesses all the characteristics of interval measurement, and there exists a true zero, true ratio exists when the same variable is measured on two different members of the population. (2:1)  
Ex.: Height, weight, Time, Salary, Age

**Data Collection:** a variety of ways. One of the most common methods is through the use of surveys.

#### **1- Telephone surveys**

- less costly.
- people may be more candid in their opinions.
- not all people have a chance of being surveyed.

#### **2- Mailed questionnaire surveys.**

- cover a wider geographic area, less expensive to conduct.
- a low number of responses and inappropriate answers to questions.
- difficulty reading or understanding the questions.

#### **3- Personal interview surveys.**

- obtaining in-depth responses.
- need training in asking questions and recording responses.
- more costly.
- interviewer may be biased in selections of respondents.

#### **4- Other ways of data collection**

- surveying records, direct observation of situations.

#### **Methods of Sampling:**

Obtaining samples that are unbiased, give each subject in the pop. an equally likely chance of being selected.

#### **1- Random Sampling:**

chance method or random numbers. Generating random numbers with a computer or calculator.

**2-Systematic Sampling:** each element has an equal probability of selection, but combinations of elements have different probabilities.

Population size  $N$ , desired sample size  $n$ , sampling interval  $k=N/n$ .

Randomly select a number  $j$  between 1 and  $k$ , sample element  $j$  and then every  $k^{\text{th}}$  element thereafter,  $j+k$ ,  $j+2k$ , etc.

**$N = 100$**

**want  $n = 20$**

**$N/n = 5$**

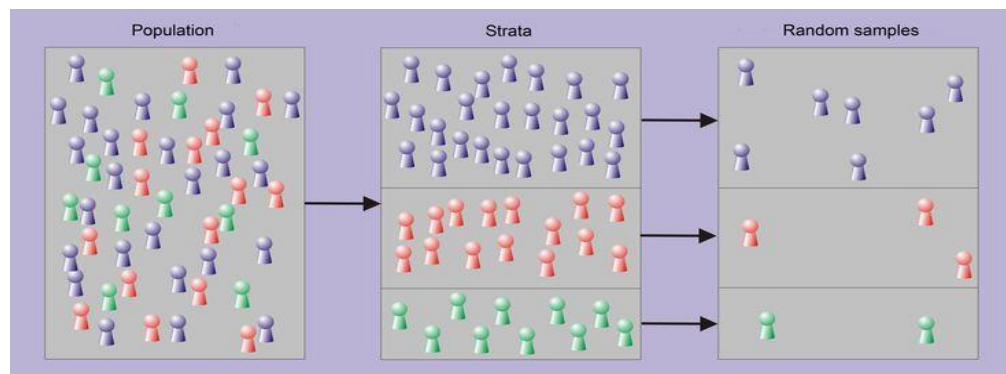
**$K = 5$**

**select a random number from 1-5:  
chose 4**

**start with #4 and take every 5th unit**

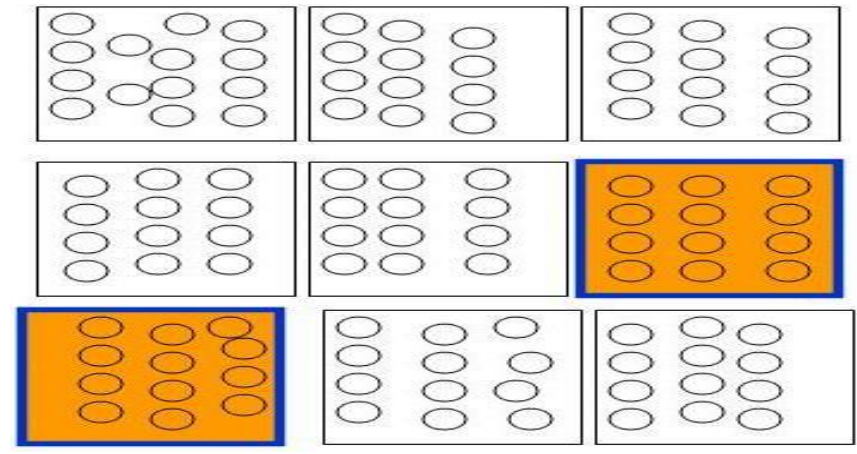
1	26	51	76
2	27	52	77
3	28	53	78
4	29	54	79
5	30	55	80
6	31	56	81
7	32	57	82
8	33	58	83
9	34	59	84
10	35	60	85
11	36	61	86
12	37	62	87
13	38	63	88
14	39	64	89
15	40	65	90
16	41	66	91
17	42	67	92
18	43	68	93
19	44	69	94
20	45	70	95
21	46	71	96
22	47	72	97
23	48	73	98
24	49	74	99
25	50	75	100

**3- Stratified Sampling:** divided population into groups (strata) according to some specified characteristics such as age, grade level or income and subsamples are randomly selected from each strata.



#### 4- Cluster Sampling:

The population is divided into groups or clusters of elements usually geographic or organizational. Some of the groups are randomly chosen. This method is useful when it is difficult or costly to develop a complete list of the population members or when the population elements are widely dispersed geographically.



### **Statistical Studies:-**

1-Observational Study: the researcher merely observes what is happening or what has happened in the past and tries to draw conclusions based on these observations.

2-Experimental study: the researcher manipulates one of variables and tries to determine how the manipulation influences other variables.

**Independent variable:** in an experimental study is the one that is being manipulated by the researcher. (Explanatory variable). The resultant variable is called dependent variable or the outcome variable.

