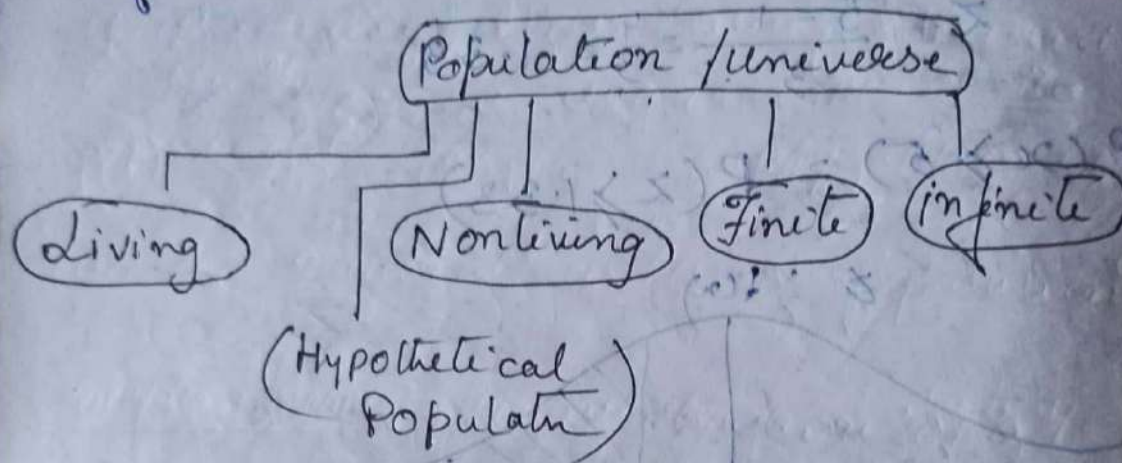


① Population : Population or universe in statistics means aggregate of anything; living or non living; finite or infinite



Hypothetical Population - If a popu. can only be imagined without having any existence in reality. It is called Hypothetical population

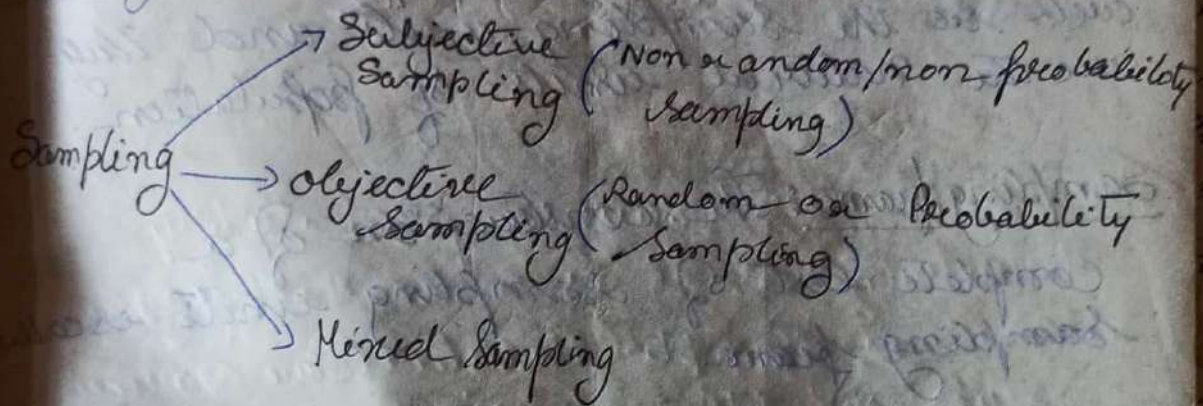
For example:- The results population of heads in case of tossing of an unbiased coins for infinite no. of times

Sample:- Sample refers to a part of population which is selected to study some characteristics of the population. Eg - Out of 5000 students in a college if we select 200 out of it to



study about avg talent of the students of the college, then 200 students will constitute the sample.

Sampling:- Sampling is a process of collecting sample from a population. Sampling could be of 3 types as represented in the following arrow diagram.



Population Size: Total number of units in a statistical population is called population size. It is usually denoted by 'N'.

Sample Size: Total number of units in a sample is called sample size. It is denoted by 'n'.

Sample Survey: It refers to the enumeration of the units included in the sample only.



Sampling units: To draw samples from a population, it is necessary to divide the entire population into some smaller units. These units are called sampling units. Sampling units may be the natural units or cluster of natural units. For eg- if we are interested in studying the avg income of the household in a locality then households will be the sampling units and they are the natural units of population.

Sampling frame: The construction of a complete list of sampling units is called sampling frame.

Parameter: statistical constant: The term parameter is used to denote the characteristic of the statistical population. In practice, parameters are unknown & they are estimated by the statistics. Parameters are also called statistical constant. For eg- population Mean ( $\mu$ ), population variance ( $\sigma^2$ ), population proportion ( $P$ ) etc.

Statistic: The term statistic is used to denote the characteristics of the sample. The formal definition of statistic is concerned, statistic is defined as



the for of sample values.

For eg - If  $\Theta = (x_1, x_2, x_3, \dots, x_n)$

where  $x_1, x_2, \dots, x_n$  are sample values, then  $\Theta$  is called a statistic.

Sample mean ( $\bar{x}$ ), sample variance ( $s^2$ ) & sample proportion ( $p$ ) etc are some of the examples of statistic. Statistics are variables. They are known phenomenon & their values are used for estimating the values of the parameter.

H/W

\* Difference bet<sup>n</sup> parameter & statistic

11. Estimator & Estimate: In the theory of sampling, statistic is called the estimator of a parameter and the value of the statistic is called the estimate.

Sampling Distribution of a Statistic

Sampling distribution of statistic is a probability distribution which shows the diff values taken by a statistic along with their respective probabilities. For example - small 't' represents of statistic and  $P(t)$  represents the probabilities of various values taken by the statistic 't'.



then the sampling distribution of the statistic  $t$  can be shown as follows

Sampling Distribution of the Statistic  $t$

$$t : t_1 \quad t_2 \quad t_3 \quad \dots \quad t_n$$

$$P(t) : P(t_1) \quad P(t_2) \quad P(t_3) \quad \dots \quad P(t_n)$$

Parameter doesn't have a probability distribution as it is a constant figure whereas as sample statistics have various values as diff sample have diff statistics.  $\therefore$  it has a P. distribution.

Note

① Population Mean ( $\mu$ )  $= \frac{1}{N} \sum_{i=1}^N X_i$

② Sample Mean ( $\bar{x}$ )  $= \frac{1}{n} \sum_{i=1}^n X_i$

③ Population Variance ( $\sigma^2$ )  $= \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2$

④ Sample Variance ( $s^2$ )  $= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2$

⑤ Population Mean Sq ( $S^2$ )  $= \frac{1}{N-1} \sum_{i=1}^N (X_i - \mu)^2$



Sample Mean  $\bar{x} = \frac{\sum x_i}{n}$

$$\text{Sample Mean sq } (s'^2) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

## Estimation of the Parameters

Point Estimation

Interval Estimation

### Point

Estimation: It is a method of ~~from~~ estimating the value of the Populn parameter. In this method, we get a specific value of the estimator of the parameter. For eg, if  $\theta$  denotes a parameter &  $\hat{\theta}$  denotes the estimator of the parameter, then the value of  $\hat{\theta}$  will provide a point estimate of the parameter.

Now

### Points Estimation

method of point estimation

Method of Moments

Least sq method

Maximum Likelihood method

Generalised method of Moment.



Subjective Sampling:- Purposive sampling / Judgement Sampling

when sample is derived from population without following any probabilistic / scientific sampling.

It is also called Non-Random Sampling

Random Sampling - or probability sampling where each & every unit of the popu has an equal and independent chance of being included in the sample.

Simple Random Sampling -

It is a type of random sampling where the samples are selected from the population by following a probabilistic rule. The probability of selecting a sample in simple random sampling for each & every unit of popu is same equal to  $\frac{1}{N}$



# Simple Random Sampling

Simple Random Sampling  
(with replacement)  
SRSWR

Simple Random  
Sampling without  
replacement

SRSWOR

Distinguish bet SRSWR & SRSWOR -

SRSWR & SRSWOR are 2 diff types of Simple Random sampling. There exist some imp points of differences betw these 2 types of SRS. These differences are:-

① SRSWR

① In case of SRSWR, a sample unit is selected from the population and then it is again returned back to the original population before selecting the 2nd sample.

② The population size in case of SRSWR does not diminish as the procedure of sample

SRSWOR

① In case of SRSWOR on the other hand the sample unit after selecting from the population is not returned back to the original population.

② In case of SRSWOR the population size diminished with every single draw of a sample from popu.



Selection continues,

③ In case of SRSWR a ~~sample unit~~ unit may appear ~~no~~ unit in the popu may appear more than once in the course of sampling

4) SRSWR provides an inefficient estimate of a parameter, because the variance of the estimator in case of SRSWR is high.

5) In case of SRSWR, sample of size  $n$ , from population of size  $N$  can be drawn in  $N^n$  ways.

③ In case of SRSWOR unit of the popu can appear only once in the sample.

4) But in case of SRSWOR, the estimate of parameter is more efficient because the variance of the estimator in case of SRSWOR is low.

5) In SRSWOR, sample of size  $n$ , from a popu of size  $N$  can be drawn in  $N^n$  ways.



## Methods of Sampling

The various methods of sampling also called (sampling design) can be grouped into 2 broad categories:-

- ① Random Sampling - / Probability Sampling / Proportional Sampling
- ② Non Random Sampling -

① Random Sampling - Random Sampling refers to that method of sample selection in which every item of the population has an equal chance of being selected in the sample. In practice the members of the sample are drawn one-by-one.

There are 2 methods of drawing simple random sampling:-

- ① SRSWR, ② SRSSWR

A) SRSWR - SRS is said to be with replacement when the members of the sample are drawn from the population one by one & after each drawing the selected member is returned to the population. Thus if 'N' is the size of the population and 'n' is the size of sample. Then we can have ' $N^n$ ' no. of samples. So that at each drawing any member of the population has the same probability of  $\frac{1}{N}$  of being selected. Here the population remains the same both in size & composition throughout the sampling procedure & any of the population members may appear



more than one in the sample.

b) SRSWOR: The simple random sampling is said to be 'Without Replacement' when the samples are drawn from the population one by one & the member selected at any drawing is not returned to the population. Thus in case of SRSWOR the member selected once is no longer offered for selection. Thus, at each drawing each of the remaining members of the population gets the same chance of being included in the sample. In SRSWOR, the total no. of possible samples are  $N_c_n$ , so the probability that a particular sample is selected is given by  $\frac{1}{N_c_n}$ .

### Advantage of Random Sampling

- ① This technique is simple to understand & easy to operate.
- ② The sampling technique eliminates the effect of any personal biasness & samples are selected entirely on chance.
- ③ This technique gives a better representation of the population compared to judgement sampling provided the sample size is sufficiently large, it is one.
- ④ It is inexpensive & requires much less time & energy.



### Disadvantages:

- ① It requires very high level of skill & experience for its use.
- ② The cost involved are generally high as compared to non random sample.
- ③ The use of SRS ~~needs~~ necessitate the numbering of the whole population. In case of very large population, the numbering process is extremely costly & time consuming & in many cases it is often impossible.
- ④ In case of hetero heterogeneous population, the SRS will fail to depict the true characteristic of the population.

### There are Methods of Drawing Random Sampling

- ① Lottery Method - Under this method all the items of the population are numbered, the numbers are written on a card & the pack of cards are ~~well~~ shuffled. A blind-fold selection is made from these cards. In order to get accurate results it is necessary that the card should be same in size, shape, thickness & in all other respects. However this will be not practical if the ~~eq~~ parent universe is large.
- ② Table of Random Numbers: This is most widely used method of obtaining a random sample. J. J. Tippett, Kendall & Stuart, Rand Corporation & many others have constructed Table of Random Sampling numbers. Each of



these tables numbers consist of large number of Random digit in Rows & columns.

II/ Non-Random Sampling:- Non-Random is a process of sampling without the use of Randomization. A non-random sample is selected on the basis of someone's experience about the population or accordance to the convenience of the user. That is <sup>why</sup> it is also called as purposeful sampling. In this method the criterion for selection is laid down first & items in accordance with it.

### Advantages

- ① This sampling technique can give quick result.
- ② This ~~the~~ technique is very flexible in the sense that even at the time of conducting the survey the investigator may replace a unit of sample by any other convenient member of the population.
- ③ This technique is economical.

### Disadvantages

- ① In this sampling technique, the bias & prejudice means of the investigator has enough scope to influence the selection.
- ② It lacks degree of accuracy

|   |   |   |
|---|---|---|
| R | S |   |
| 5 | 7 | 1 |



## Stratified Sampling

When the population from which sample is drawn is ~~hetero~~ heterogeneous in nature, then the technique of SRS fails to collect a truly representative sample. In such cases stratified sampling is frequently considered as a more efficient procedure for ensuring accurate representation. Under stratified sampling, the population is divided into different non overlapping sub groups called 'strata' or classes & then samples are drawn from each strata or class. For achieving maximum efficiency, the strata should be so formed that the units within a stratum are as homogeneous as possible & betw the strata are as heterogeneous as possible. These samples are then combined to form a single sample.

There are 2 main objectives of stratification

- ① To increase the accuracy of the overall population estimates
- ② To ensure that subdivisions of the populations are adequately represented in the sample.

## Advantages

- ① This method ensures better representation of the characteristic of population in the sample.



② This method is most effective in dealing with highly skewed population such as income data, retail sales.

③ Stratification also being about a game plan in precision of the estimate obtained. This technique is helpful to avoid the heterogeneity of population.

### Disadvantages

- ① If the basis of stratification is not properly decided, the results may be misleading.
- ② Selective judgement used in dividing the universe into strata may affect the accuracy of the result.

### ③ Principles of Sample Survey

① Validity -

② Optimization -

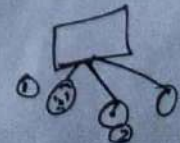
③ Statistical Regularity -

} Xerox

} X

100000

④ Inertia of Large numbers:- According to this law, other things being equal - The larger the size of the sample, the more accurate the results are likely to be. Thus in other words, if the sample size is large, it is more likely to depict the true nature of the population & be a perfect representative of the population.



111



## Sampling Bias:-

Sampling Bias is inherent in the method of sampling & arises because the information collected relates only to a part in actual practice, certain inaccuracy arise out of complex reaction bet<sup>co</sup> an enumerators & informants or in processing the collected data which effect the conclusions derived therefrom. This kind of error is known as bias. Bias arises solely due to human factor & its difficult to detect & detect. It occurs even in census where sampling error is completely absent. Bias may arise in several way-

- ① Bias may arise in selecting the sample itself, if the proper process of selection is not strictly followed
- ② Bias may arise because of substitution of a selected unit by another
- ③ It may arise in the process of estimation in parameters
- ④ It may arise due to negligence or carelessness on the part of investigators in asking &



## Systematic Sampling

Under this method, all units of the universe are arranged in a systematic manner ~~on~~ based on some basis viz numerical, alphabetical or geographical etc. This method <sup>can be</sup> ~~is~~ used when a complete list of whole universe is available. Under the method of systematic sampling every  $k^{\text{th}}$  item from a properly arranged universe (is selected for the sample). The sampling interval (i.e.  $k$ ) is decided on the basis of the total number of units in the sample & the size of the universe. For example, out of 2000 student, a sample of 100 student is to be taken. Then sampling interval is 20 i.e.  $(\frac{2000}{100})$ .

The 1st sampling sample unit will be taken from first 20 students, the subsequent sample units will be selected by using the sampling interval in the serial order.

$$\text{Sampling interval} = \frac{\text{Size of Universe}}{\text{Sample Size.}}$$

### Adv

- ① It helps in saving time & money.
- ② Results are reliable & accurate.

### Demerit

- (i) If the basis of popu is arranged in a wrong manner, the result will be misleading.
- (i) Systematic sampling are not generally random samples



## Multi Stage Sampling

Under this method sampling is taken out in several stages. In the 1st stage, various homogeneous groups are selected from the population depending on the nature of enquiry. These are termed as primary sampling units. In the 2nd stage elementary sampling units are selected from each of these groups. Further stages may be added if required.

### Merits

- ① It saves time & money.
- ② At every stage the size of sample goes on decreasing.

### Demerit

- ① If wrong sample is taken at 1st stage, the sample at other stages are liable to be inaccurate.
- ② A multi-stage sample is less accurate.
- ③ This method is not suitable for the cases where the scope of investigation is limited.

## \* Sampling Error

The <sup>drawn</sup> conclusions about the population by studying a small part of it, may be inaccurate. It may be affected by some sort of error. However, such error disappears in case of a census study. Thus the diff. bet<sup>n</sup> sample & census result is called a sampling error.

///



## Quota Sampling

Under this method the whole population is divided into various groups based on some diff characteristics such as sex, occupation, age etc. Quota of persons are fixed for each group for each investigator.

But within the given quota, the selection of sample units depends upon the personal judgement of the investigator.

### Merit

- ① Requires less time & money.
- ② If the interviewers are trained, the results obtained under this method is quite satisfactory.

### Demerit

- ① It is effected by the bias & judgement of the interviewers.
- ② This method is not very popular. It is rarely used in practical work.