

## B.Sc. STATISTICS

(For the candidates admitted from 2017 – 2018 onwards)

Core Course - V

SEMESTER – V

P. Code:

### SAMPLING TECHNIQUES

#### Unit – I:

Concept of sampling and population: Need for sampling – Design, Organization and execution of sample survey – Principal steps in sample surveys – preparation of questionnaire and schedules – Pilot survey – Sampling and Non-sampling Errors – Limitations of sampling.

#### Unit – II:

Sampling from finite population – Simple Random Sampling with and without replacement – Unbiased estimate of mean and Variance – finite population correction – Estimation of standard error from a sample – Determinations of sample size – Simple Random Sampling for attributes.

#### Unit – III:

Stratified Random Sampling: Concept of stratifying factor - Unbiased estimate of the mean and variance of the estimated mean – Proportional and optimum allocation – Relative precision of stratified random sampling and simple random sampling

#### Unit – IV:

Ratio and regression estimators (based on simple random sampling only) – concept of Auxiliary variate – Ratio estimators – Bias of Ratio estimates – Variance of the ratio estimates – comparison of Ratio estimator with mean per unit.

Regression Estimators: Linear regression estimate, Regression estimate with pre assigned 'b' and regression estimates computed from sample.

#### Unit – V:

Systematic sampling: Estimation of the mean and variance of the estimated mean – comparison of simple, stratified and systematic sampling – circular systematic sampling.

NSSO and its functions – Other agencies undertaking sample surveys.

#### Reference Books

1. W.G.Cochran (1985) Sampling Techniques, Wiley Eastern Ltd, New Delhi.
2. S.C. Gupta and V.K.Kapoor (2007), Fundamentals of Applied Statistics, Sultan Chand & Sons, New Delhi.
3. Goon. A.M., Gupta. M.K, B.Das Gupta. Fundamentals of Statistics.

# SAMPLING TECHNIQUES

(1)

## UNIT - I

### POPULATION :

In Statistics, Population is an aggregate of Objects under study. The objects may be animate or inanimate.

Ex : No. of students in a college  
No. of stars in the sky.

### FINITE POPULATION :

A Population consists of finite no. of elements is called as finite Population.

Ex : No. of students in the college.

### POPULATION SIZE :

The total no. of elements in the Population is known as Population Size. It is usually denoted by  $N$ .

### INFINITE POPULATION :

A population having infinite no. of objects is known as infinite population.

Ex : No. of stars in the sky.

### HYPOTHETICAL POPULATION :

A Population does not consist concrete objects is called as Hypothetical Population

Ex : The Population of throws of a coins thrown infinite no. of times is called Hypothetical Population.



## Methods of collecting information on population

There are two methods of collecting information on population. They are

1. Census Method
2. Sample Method.

### 1. Census Method [Complete Enumeration]

The object of a census or complete enumeration is to collect information for each and every unit of the population.

Ex: Assume that there are 20,000 students in the Periyar University. If you want to study the average expenditure of these students, we must consider the expenditure of all these 20,000 students.

#### Merits of Census Method:

- (i) The data are collected from each and every item of the population.
- (ii) The results are more accurate and reliable.
- (iii) Intensive study is possible.
- (iv) The data collected may be used for various surveys, analysis etc.,

#### Demerits of Census Method:

- (i) It requires more cost, labour, time, energy etc.,
- (ii) It requires a large of enumerators.
- (iii) If the population is too large, it may not

be possible to use census method.

(iv) If testing is destructive, this method is impracticable.

### Sample Method:

In the sample method only a part of the population studied and conclusions are drawn on that basis for the entire population.

Ex: A Doctor examine a few drop of blood and draws conclusion about the blood constitution of the whole body.

### Merits of Sample Method:

- (i) The sample method involves less cost, labour, time, energy etc.
- (ii) It gives more detailed information.
- (iii) If the population is too large, if testing is destructive sample method is the only method to be used.
- (iv) More reliable results can be obtained because highly skilled and trained persons can be employed for scientific processing and analysis of sample data.

### Demerits of Sample Method:

- (i) If the sample is not representative, the results may be false and misleading.
- (ii) If the sample size is not appropriate it may lead to wrong results.



- (iii) There may be personal bias with recording to the choice of technique of drawing of sampling units.
- (iv) If the information is required for every population unit then a sample method is not applicable.

### Sample :

A finite subset of a statistical population is called a sample.

Ex: From the whole bag of sugar, if we take a small quantity of sugar for testing its quality, it is called as sample.

### Sample Size :

The number of individuals in a sample is called the sample size. It is usually denoted by 'n'

### Characteristic of a good sample :

- (i) **Representativeness** : The sample must be a representative part of the population.
- (ii) **Adequate sample size** : The size (no. of units in the sample) of the sample should be adequate.
- (iii) **Homogeneity** : The sample units and the population units must be similar in nature.
- (iv) **Optimization** : The sample should provide optimum results, both in terms of cost and efficiency.

## SAMPLING :

Sampling is the process of choosing a sample from a population so that some inference about the population can be made by studying the sample.

## PARAMETERS :

The statistical constants of the population are known as parameters.

Ex: Population mean  $\mu$ , Population Variance  $\sigma^2$

## STATISTIC :

Statistical measures computed from the sample observations are known as statistic

Ex: Sample mean  $\bar{x}$ , Sample Variance  $s^2$

## Need for Sampling :

For any statistical investigation complete enumeration of the population is rather impracticable under the following situations census method is practically impossible.

- (i) If the population is infinite, complete enumeration is not possible and we need sampling.  
For ex, if we want to have an idea of the avg per capital (monthly) income of the people in India, we will have to enumerate all the earning individuals in the country, which is rather a very difficult task.
- (ii) If the testing units are destructive (e) if the units are destroyed in the course of inspection the 100% inspection is not at all desirable



(6)

Ex - Inspection of crackers, explosive materials,

(ii) The sample method involves lesser cost, labour, energy, time etc., than Census method.

### PRINCIPLES OF SAMPLE SURVEY:

The theory of Sampling is based on the following important Principles.

1. Law of Statistical regularity
2. Law of inertia of large numbers
3. Principle of validity.
4. Principle of optimisation.

### LAW OF STATISTICAL REGULARITY

The law of statistical regularity states that a moderately large number of items chosen at random from a very large group are almost sure on the average to possess the characteristic of the large group.

This law is based on two assumptions

1. The selection of sample must be random
2. The no. of items in the sample should be large.

### LAW OF INERTIA OF LARGE NUMBERS

This law is based on a corollary of the law of statistical regularity. It states that "other things being equal as sample size increases the results tends to be more reliable and accurate". This law states that large groups of data show higher

degree of stability as compared to smaller one.

PRINCIPLE OF VALIDITY :

A Sampling design is turn as valid , If it enables us to obtain valid test and estimates about the Parameters.

PRINCIPLE OF OPTIMISATION :

The principle aims at

- 1. Obtaining desired level of efficiency at minimum cost.
- 2. Obtaining maximum Possible efficiency for a given cost.

Advantages of Sampling Over complete Census Method (or) Essential aspects of a sample survey :

1. Less time :

It requires less time. The sampling results can be obtained more quickly.

2. Less costs :

The total cost of the sample survey is smaller than that of a complete census.

3. Greater Accuracy of Results :

The results of a sample survey are usually much more reliable than those obtained from a complete census due to the following reasons.

- a) It is always possible to determine the extent of the sampling error and
- b) The non-sampling errors :



Effective control of non-sampling errors can be made. More sophisticated statistical techniques can be employed to obtain relatively more reliable results.

#### 4. Greater Scope :

Sample Survey has generally greater scope as compared with complete census. The complete enumeration is impracticable if the survey requires a highly trained personnel and more sophisticated equipment for the collection and analysis of the data. Since Sample Survey saves time and money, it is possible to have a thorough and extensive enquiry because a more detailed information can be obtained from a small group of respondents.

5. If the population is too large, we have to use only the sample method.
6. If testing is destructive we have to use only the sample method. Ex - Crackers.
7. If the population is hypothetical, sampling method is the only scientific method of estimating the population.

#### Limitations of Sampling :

1. Proper care should be taken in the planning and execution of the sample survey.
2. Sampling theory requires the services of trained

and qualified personnel and sophisticated equipment for its planning, execution and analysis. In the absence of these, the results of the sample survey are not reliable.

3. If the information is required about each and every unit of the population sampling method cannot be applied.

### Statistical Survey or enquiry :

A statistical survey is a way of investigation carried out to collect relevant information and to analyse the data using statistical method.

The principal steps in a sample survey (or) organization or execution in a sample survey :

The main steps involved in the planning and execution of a sample survey may be grouped under the following heads

#### 1. Objectives of the survey

The first step is to define in clear and concrete terms, the objectives of the survey

#### 2. Defining the population to be sampled

The population, (i.e.) the aggregate of objects (animate or in-animate) from which sample is chosen should be defined in clear and unambiguous terms. For ex - in sampling of forms clear-cut rules must be framed to define a form regarding shape, size etc.,



### 3. The frame and sampling units

The population must be capable of division into what are called sampling units for purposes of sample selection.

The sampling units must cover the entire population and they must be distinct, unambiguous and non-overlapping.

Ex - In socio-economic survey for selecting people in a town, the sampling units might be an individual person, a family, house-hold or a block in a locality.

Frame: In order to cover the population decided upon there should be some list, map or other acceptable material called the frame, which serves as a guide to the population to be covered.

### 4. Data to be collected

The data should be collected keeping in view the objectives of the survey. The tendency should not be to collect too many data.

### 5. The questionnaire or schedule

Having decided about the type of the data to be collected, the next important part of the sample survey is the construction of the questionnaire (to be filled in by the respondent) or schedule

Of enquiry (to be completed by the interviewer).  
The questions should be clear, brief, non-offending

## 6. Method of collecting information

The two methods commonly used for collecting data for human population are

1. Interview method
2. Mailed questionnaire method.

### 1. Interviewed Method -

In this method, the investigator goes from house to house and interviews the individuals personally.

### 2. Mailed Questionnaire method.

In this method, the questionnaire is mailed to the individuals who are required to fill it up and return it duly completed.

Although mail surveys are less costly there is scope for considerable non-response. On the other hand, interview method costs more and there are interviewer errors.

## 7. Non-respondents

Quite often (due to practical difficulties) the data cannot be collected for all the sampled units. For example the selected respondent may not be available at his place when the investigator goes there or he may fail or even refuse to give certain information when contacted. This incompleteness



is called non-response.

### 8. Selection of proper Sampling Design

A number of designs (plans) for the selection of a sample are available and a proper selection will guarantee good and reliable estimates.

### 9. Organisation of Field Work

The success of a survey to a great extent depends upon the reliable field work.

### 10. The Pretest

From practical point of view, a small pretest (i.e) trying out the questionnaire and field methods on a small scale has been found to be immensely useful.

### 11. Summary and Analysis of the Data

The analysis of the data may be broadly classified into the following heads.

#### a) Scrutiny and Editing of the data

The scrutiny or editing of the completed questionnaire will help in eliminating data that are clearly inconsistent.

#### b) Tabulation of data

The method of tabulation namely hand tabulation or machine tabulation will depend upon the quantity of the data.

#### c) Statistical Analysis

Different methods of estimation may be available for the same data. Appropriate formulae should be used to provide final estimates of the required information.

#### d) Reporting and conclusions

Finally, a report incorporating detailed statement of the different stages of the survey should be prepared.

#### 12. Information gained for future surveys

Any completed survey is helpful in providing a note of caution and taking lessons from it for designing future surveys.

#### Questionnaire Preparation

Questionnaire is also a form consist of questions relating to enquiry filled by respondents (who supply informations to the data collecting agency)

#### Principles involved in designing a questionnaire

1. The questionnaire must be carefully prepared, preferably by two or more people.
2. One might prepare the initial draft and pass it to two or more persons for review and suggestions. The revised questionnaire should be pretested under actual conditions.

#### Points to be kept in mind during its Preparation

##### 1. Covering letter

Every questionnaire should contain a covering letter. The covering letter should highlight the purpose of the study and assure the respondent that all information will be kept confidential.

##### 2. No. of questions should be kept in minimum

The fewer questions, the greater chance of



getting better response. Generally the no. of questions 15 to 40.

- 3. Questions should be simple and short
- 4. Questions can be of Yes/No type (or) multiple choice depending on the requirements.
- 5. Open-ended questions and double questions should be avoided.
- 6. Questions of Personal nature should be avoided. For ex - questions relating to sales, profit, Drinking habit should be avoided as far as possible.
- 7. The question should be formed in such a way that their answers do not require calculation.
- 8. Logical arguments - The questions should be logically arranged so that there is a continuity of response.
- 9. Cross checks and foot notes - The questionnaire should contain some questions which act as a cross check to the reliability of the information provided. For the purpose of the clarity certain questions which might create a doubt in the mind of respondents it is desirable to give foot notes.

Preparation of a Model Questionnaire :

Objective of the survey :

To enquire the expenditure habits of students residing in college hostels.

Name of the college \_\_\_\_\_

Name of the student \_\_\_\_\_

Course studying \_\_\_\_\_

Age —

Sex —

1. How much amount do you get from your father/ Guardian Per month ?
2. Do you get some scholarship? If so specify the amount
3. Is there any other source from which you get money regularly? If yes means mention the source
4. How much do you spent monthly on followings

Items	Amount (Rs)
Hostel fees Expenses	—
Room rent	—
other hostel Fees	—
clothing	—
Entertainment	—
Miscellaneous	—

5. Do you smoke? If so what is the daily expenditure on it?
6. Any other item on which you spend money?

**PILOT SURVEY:**

A pilot survey may be defined as a technique of collecting data on important points for the proper formulation of sample or census surveys. In the light of data collected from pilot survey some changes or adjustments may be introduced in the technique of data collection to be adopted in large scale surveys.



The following are main purposes

1. The pilot survey may be undertaken for testing the validity of the questionnaire
2. It may also provide data for estimation of various components of cost of different operations involved in the survey as well as the total cost. It is used to test the efficiency of the field investigators.
3. It also supply information to the data collecting agency about interview kind, travel kind etc.,
4. The pilot survey may also help in determining the most effective type and size of the sampling units.

### Schedules

Schedules are used in collecting primary data. It is a form which consist of a list of questions pertaining (relating) to enquiry and filled in by the trained investigators on the basis of information supplied by informants (or) respondents. Schedules are used in data collection where respondents are uneducated.

### Sampling distribution

If large number of samples of specified size are taken from some population and statistic such as mean, Variance etc is computed from its sample then we get a set of values of the statistic

(e) Let  $N$  = size of finite population

$n$  = sample size

Then the total no. of possible sample is  $N C_n = k$   
 (say) for each of these  $k$  samples, we can compute  
 some statistic  $t = t(x_1, x_2, \dots, x_n)$

Sample No	$t$	Statistic $\bar{x}$	$s^2$
1	$t_1$	$\bar{x}_1$	$s_1^2$
2	$t_2$	$\bar{x}_2$	$s_2^2$
$\vdots$			
$k$	$t_k$	$\bar{x}_k$	$s_k^2$

For each sample the set of values of the statistic constitutes the sampling distribution.

For example, the values  $t_1, t_2, \dots, t_k$  determine the

sampling distribution of the statistic  $t$

Mean of the sampling distribution of the statistic

$$\bar{T} = \frac{t_1 + t_2 + \dots + t_k}{k} = \bar{T} = \frac{\sum_{i=1}^k t_i}{k}$$

Variance of the sampling distribution of the statistic

$$V(t) = \frac{\sum_{i=1}^k (t_i - \bar{T})^2}{k}$$

## STANDARD ERROR

The standard deviation of the sampling distribution of the statistic is known as its standard error.

<u>Statistic</u>	<u>Standard Error</u>
1. Sample Mean $\bar{x}$	$\sigma/\sqrt{n}$
2. Observed sample proportion	$\sqrt{p_0/n}$
3. Sample std deviation	$\sqrt{\sigma^2/2n}$



4. Sample variance  $\frac{\sigma^2}{\sqrt{2/n}}$
5. Difference of two sample means  $(\bar{x}_1 - \bar{x}_2)$   $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
6. Difference of two sample std deviation  $(s_1 - s_2)$   $\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$
7. Difference of two sample of proportion  $\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$

### Utility of Standard Error

1. Std Error plays a very important role in the large sample theory and form the basis of the testing of hypothesis.
2. S.E helps us to determine the probable limits within which parameter may be expected to lie.
3. The reciprocal of the S.E is taken as the measure of reliability or precision of the sample.

### Sampling and Non-Sampling Errors

The error involved in the collection, processing and analysis of a data may be broadly classified under the following two heads.

- (i) Sampling Errors and
- (ii) Non-sampling Errors.

#### 1. Sampling Errors :

Sampling Errors have their origin in sampling and arise due to the fact that only a part of the population (ie sample) has been used to estimate population parameters and draw inferences about the

Population. Sampling errors are absent in a complete enumeration survey.

Sampling Errors (biases) are primarily due to the following reasons

1. Faulty selection of the sample
2. Substitution
3. Faulty demarcation of sampling units
4. Constant error due to improper choice of the statistics

Faulty Selection of the Sample

Some of the bias is introduced by the use of defective sampling technique for the selection of a sample. Ex. purposive or judgement sampling in which the investigator deliberately selects a representative sample to obtain certain results. This bias can be overcome by strictly adhering to a simple random sample.

Substitution

If difficulties arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population. This clearly leads to some bias since the characteristics possessed by the substituted will be different from the original.

Faulty demarcation of sampling units

Bias due to defective demarcation of sampling units is particularly significant in deal surveys such as



Agricultural Experiments in the field or crop cutting Surveys.

Constant Error due to improper choice of the statistics for estimation of Population Parameters.

2. Non-sampling Errors

The Non-sampling errors primarily arise at the stages of observation, ascertainment and processing of the data. These errors are present in both the complete enumeration and the Sample-surveys

Note: The data obtained in a complete census although free from sampling errors would still be subject to non-sampling errors. whereas data obtained, in a Sample Survey would be subject to both sampling and non-sampling errors.

More Important Non-Sampling Errors can be due to

1. Faulty planning (or) Definitions
2. Response Errors
3. Non-response Bias
4. Errors in coverage
5. Compiling Errors
6. Publication Errors.

1. Faulty Planning or Definitions

Here the errors due to

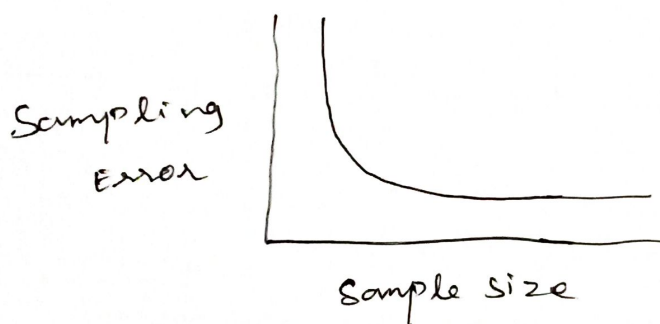
a) Data Specification being inadequate and inconsistent

For ex - if  $x_1, x_2, \dots, x_n$  is a sample of independent

Observations then the sample variance  $s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$  as an estimate of the population variance  $\sigma^2$  is biased whereas the statistic  $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  is an unbiased estimate of  $\sigma^2$ .

**Remark :**

- (i) Increase in the sample size (i.e) the number of units in the sample usually results in the decrease in sampling error.
- (ii) In fact, in many situations this decrease in sampling error is inversely proportional to the square root of the sample size as illustrated in the diagram



### Failure of respondent's Memory

Many of the questions in surveys refer to happenings or conditions in the past. There is a problem both of remembering the event and associating it with the correct time period.

### 3. Non-response Bias

Non-response biases occur if full information is not obtained on all the sampling units. In house-to-house surveys non-response usually results if the respondent is not found at home even after repeated calls or if he is unable to furnish the information on all the



Questions or if he refuses to answer certain questions

#### 4. Errors in coverage

If the objectives of the survey are not precisely stated in clear cut terms this may result in

1. The inclusion in the survey of certain units which are not to be included or
2. The exclusion of certain units which were to be included in the survey under the objectives.

For ex - in a census to determine the no. of individuals in the age group say, 20 years to 50 years more or less serious errors may occur. In deciding whom to ~~error~~ ~~due~~ to enumerate.

#### 2. Response Errors

These Errors may be due to any of the following reasons.

(i) Response errors may be accidental:

For ex - the respondent may misunderstand a particular question and accordingly furnish improper information un-intentionally.

(ii) Prestige bias:

An appeal to the pride or prestige of persons interviewed may introduce yet kind of bias called prestige bias.

Ex - upgrade his education, downgrade his age.

(iii) Self-interest:

Quite often in order to safeguard one's self-interest one may give incorrect information. Ex - a person may give an under estimate of his salary or production

and an over-statement of his expenses or requirements etc.

(iv) Bias due to interviewer :

Sometimes the interviewer may affect the accuracy of the response by the way he asks questions or records them.

(v) Failure of respondents memory :

Many of the questions in surveys refer to happenings or conditions in the past. There is a problem both of remembering the event and associating it with the correct time period.

5. Compiling Errors

Errors made in various operations of data processing such as editing and coding of the responses, Punching of cards, tabulation and summarising the original observations made in the survey are a essential source of error.

6. Publication Errors

Publication errors are the errors committed during presentation and printing of tabulated results. Basically due to the mechanics of Publication - the proofing error and like.