UNIT 12: SAMPLING AND POPULATION

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Concept of Population and Sample

- 12.3.1 Basic Principles of Sample Survey
- 12.3.2 Advantages of Sampling over Complete Enumeration
- 12.3.3 Different Steps in a Sample Survey
- 12.3.4 Errors in Surveys
- 12.3.5 Types of Sampling
- 12.3.6 Selection of Random Samples

12.4 Simple Random Sampling

- 12.4.1 Estimate of Population
- 12.4.2 Illustrated Examples

12.5 Stratified Random Sampling

- 12.5.1 Advantage and Disadvantages of Stratified Random Sampling
- 12.5.2 Allocation in Stratified Random Sampling

12.6 Systematic Random Sampling

- 12.6.1 Linear Systimatic Sampling
- 12.6.2 Circular Systematic Sampling
- 12.6.3 Advantages and Disadvantages of Systematic Sampling
- 12.6.4 Uses of Systematic Sampling

12.7 Cluster Sampling

- 12.7.1 Advantages and Disadvanges of Cluster Sampling
- 12.7.2 Optimum Cluster size

12.8 Let Us Sum Up

12.9 Further Readings

12.10 Answers To Check Your Progress

12.11 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- distinguish between a sample and population
- state the difference between parameter and a statistic
- discusss why sample surveys are adopted instead of complete enumeration in most cases
- acquire knowledge about the types of errors that one encounters which adopting sample surveys
- distinguish between random sampling and purposive sampling
- draw samples using simple random sampling technique
- discuss the advantages and disadvantages of using stratified random sampling from a heteregeneous population
- explain the process of linear systematic and circular systematic sampling and their relative merits and demerits
- evaluate the advantage of drawing cluster sampling and how to decide optimum cluster size.

12.2 INTRODUCTION

The use of sampling in making inference about a population is a practice that we are doing as a part of our everyday's life from unknown. When we have to make an inference about a large lot and we are not in a position to examine about a large lot and we are not in a position to examine each and every individual of the lot, we examine only a few numbers of the lot and make decission about the whole lot. For example a person wanting to purchase some rice will examine a handful of rice from a bagful of rice and on that basis make decisions about the whole bag of rice. Similarly in industry quality of a product is assured through sampling, the public opinion is ascertanied through opinion polls based on sampling. In all these cases inference is drawn about the whole from the knowledge of a part.

So what we see is that most of the enquiries that we make in practice are sampling enquires, sampling is inevitable because our resources are always limited, be it money or man hours. It is also preferred because it has

practical conveniencs and there are many situation where complete count is impossible.

In this unit we shall discuss how to use different sampling techniques to draw our samples according to the type of population and purpose of study. In section 9.3 we shall learn about basic concepts of sampling, its beneifits, errors etc. In the subsequent, chapters different sampling techniquies like simple random sampling, stratified random sampling, systematic sampling and cluster samplings are discussed along with their relatives advantages and disadvantages.

12.3 CONCEPT OF POPULATION AND SAMPLE

Population : The dictionary meaning of the word population is the number of people residing in a geographical area. But in statistics the word population is used in much wider sense. Thus population may be defined as a group of individuals with respect to certain characteristic under study. For example if we want to measure the average height of students of a college, then the students of that college will constitute a population. Population may be finite or infinite.

Sample : The part of the population which is selected from it in order to study its characteristis is called a sample. For example if we want to measure the standard of living of the people in a town of population 10,000, there if we estimate there characterestics from a properly selected 500 members of the population, then these 500 members of the town will constitute the sample.

Complete enumeration or census : It refers to the enumeration of each and every unit of the population. We all know about the census the government of India conducts in every ten years.

Sample survey : Where we enumerate units belonging to the sample only, it is sample survey.

Parameter : The basic objective of any sample survey is to know the characteristic of the population by examining only a part of it, by studying the characteristic of the population. So the term parameter is used to denote

the characteristic of the population. For example population mean, population variances are parameters. Parametrs are constant quantities.

Statistic : Values obtained from the study of the sample are known as statistic, sample mean, sample variance are the examples of statistic. statistics are constant quantities.

Sampling unit : The individuals that are to be sampled from the population to form the sample is called the sampling units. For example if we meant to measure the average income of the families in a city, then each household may be regarded as sampling unit, but if we want to know the average income of individuals in the same city, then the individuals will be the sampling units.

Sampling frame : Sampling frame is the complete list of sampling units, such sampling frames are difficult to construct and even difficult to keep it update. Simplest example of sampling france is the voters list.

Sampling error : It is the difference between the parameter and its estimate from the random sample, It is ineviable because only a part of the population is selected.

θ

Unbiased estimate : Let $T = t(x_1, x_2, ..., x_n)$ be a statistic for the sample values $x_1, x_2, ..., x_n$. This statistic is called an unbiased estimate of the parameter θ, if E (T) = .

Sampling variances : The variance of a statistic is called the sampling variance.

Standard error : The positive square root of sampling variance is the standard error.

12.3.1 Basic Principles of Sample Survey

There are two important laws on which the theory of sampling is based. They are principles of validity and optimization.

Principle of validity means that the sampling design should be such that valid estineates and teets abbout the parameter can be obtained.

Again optimization means either maximization or

minimization. According to this principle such sampling during should be preferred which gives maximum efficiency for a given cost of survey or the minimum cost for a specified level of efficiency.

12.3.2 Advantages of Sampling over Complete Enumeration

In most cases sample survey is preferred over complete enumeration because of the following conditions :

(i) In sample survey since only a part of the population is selected, there is reduction of cost either in terms of money or in terms of man hours. Although the cost per individual may be larger in sample survey, the total cost is yet expected to be smaller. In most of our cases resources are limited and the result of the survey is required within a specifeed times, it is always imperatives to adopt sample survey.

(ii) In certain types of enquirey highly trained personnel or specialised equipments must be used for collection of date. This is not possible in census method become a large group of investigators are to be employed and it will be highly expensive to train such a large group. So more accurate results sampling is the only way out.

(iii) In sample survey sampling errors cannot be avoided. On the otherhand though sampling error is absent in complete enumeration, large magnitude of non sampling errors cropain, whose magnitude can not be precisely determined and they greatly influence the accuracy of results. However sampling errors can be theoritically calculated and hence its effect on the accuracy of results can be minimized.

(iv) If the enumeration is by its nature distructive, these complete enumeration is not feasible. For example if we want to examine the quality of sweets in a sweet shop, we have to recourse to a sample. Also in cass of infinite or hypothetical population sampling is the only course available.

12.3.3 Different Steps in a Sample Survey

Conducting a large scale sample survey in values three main stages : (a) Planning stage, (b) Execution stage, (c) Analysis and reporting stage.

The planning stage consists of the following steps :

(i) **Defining the objectives :** The objectives of the survey should be clearly stated. The planner should also take into account the available resources, the time limit within which the study should be finished, the degree of accuracy required etc.

(ii) **Defining the population :** The population along with geographical and demographical boundaries must be clearly defined so that no ambiguity arise later on regarding the coverage of the study.

(iii) **Determination of the data to be collected :** This should be done in accordance with the objective of the survey. Once the decision is made about the nature of the data to be collected, one must prepare the questionnaire or a schedue for collection of data.

(iv) **Method of collection of data :** Decision is to be made whether the interview method or the mailed questionnaire method is to be adopted. Though the second method is less costty there is large scope for non-response and is praticable only among the educated masses.

(v) **Choies of sampling unit :** The sampling unit is the ultimate unit to be sampled for the purpose of the survey. For example depending on the requirement of the survey in a socio-economic survey the ultimate sampling unit may be a family or a member of a family.

(vi) Getting a sampling frame – Once the sampling unit is decided, the next step is to see whither the sampling frames of all the units of the population is available. An example of sampling frame can be the list of households from population census of

a region.

(vii) Designing the survey – This is the most important part of planning stage. A this stage decision is made about the type of sampling is to be used and about the need of pilot survey.

(viii) Training of personnel – The investigators and the supervisor should be properly trained before the survey is undertaken.

The execution stage involves identification of the sampled individual in the field and filling up questionnaire.

The first step in analysis and reporting stage is to edit completed questionnaires deleting data that are obviousely erroneous. The next step is to tabulate the data and then these tables are utilised for deriving necessary estimatis of population characteristics.

The next step is report writing. The report should incorporate a detacted statement regarding all the stages of the survey and all the statistical informations collected in a neat tabular form. It is necessary to report the amount of error to be expected in the estimates.

At the completion of the survey, the informations should be properly stored for possible use in designing future surveys.

12.3.4 Errors in Surveys

To appreciate the need for sample surveys, it is necessary to understand the role of sampling and non-sampling errors in sample survey and complete enumeration. The errors arising due to drawing inference about the population on the basis of sample is known as sampling errors, clearly sampling error is non-existent in complete enumeration, since the whole population is enumerated. Howerver some other kind of error crop-in, at the stage of as certainment and, precesing of data, which are termed as non-sampling errors and are common in both complete enumeration and sample surveys. We shall discuss about these two types of errors one by one.

Non sampling errors : As discussed above non-sampling errors are present in both complete enumeration as well as sample surveys.

Below we shall study different types of non-sampling errors.

(i) Respons errors : These errors have there origin in the respones or informations provided by the respondents. For example worng answers arising from pride, called error due to prestige, by virtue of which one may overstate ones educational qualification, or occupation or understate one's age. sometimes one may provide wrong answers for the protection of self interest. For example people generally tend to make an understatement of their incomes and overstatement of their expenses. Sometimes respondent may misunderstands a particular question and hence may provide wrong answers un willingly.

(ii) Errors arising from non-response : Often at the times of collection of data some respondents are not found at home even after repeated calls, or he or she simply refuses to furnish information. As a result of these non-respones the effective simple size decreases and certain characteristic of the population gets excluded, which gives rise to error.

(iii) Observational error : This type of error occurs when the variate value is abtained by observation, because of the influence of somes psychological factors. For example in case of eye condition of crop condition, the crop reporters invariably understimate the result.

(iv) Interviwer error : Sometimes the suggestions from the interviewer may influences the beliefs and prejudices of the respondents in interpreting some questions and thereby leading to some error.

Sampling Errors :

These errors have their origin in sampling and are absent in complete enumeration, some errors of this type are mentioned below.

(i) Error due to defective sampling technique : Samples case generally be drawn in two ways– (a) Random sampling and (b) Purposive sampling.

Random sampling is free from any type of errors because of the random nature of sampling. But in case of purposive sampling the sampler deliberately tries to select a representative sample to establish his own belief, which obviousey gives rise to errors. For example in a socio-economic survey of the city of Guwahati, the sampler may select only the people belonging to upper class if the investigator wants to establish that the people of the city of Guwahati are very rich.

(ii) **Error due to substitution :** Investigators often substitute one convenient member of the population when difficulties arise in enumerating another, which may lead to errors as the characteristics of the two persons may be quits different.

(iii) **Error due to wrong choics of statistic :** Such errors may arise when the investigator doesnot use appropriate statistic for the estimation of a particular parameter of the population. For example we know that sample mean square is an unbiased estimate of the population variance, but one may use sample variance in place of sample mean square leading to error.

12.3.5 Types of Sampling

Whenever we decide to take sample from a population it is necessary to choose some method in which sampling is to be done. There are various methods of choosing a sample from a population, depending on the nature of the data and the purpose of enquiry. However these various methods of sampling can be grouped under two broad heads : Probability sampling (alse known as random sampling) and non-probability (non random) sampling.

Probability sampling : Probability sampling methods are those in which every item in the universe has a known chance or probability of being chosen for the sample. This implies that the selection of the sample item is independent of the person making the study, i.e the items are chosen strictly at random.

Non Probability sampling : Non-probability sampling, also

called 'Purposive' or 'judgement' sampling is the process where the sample is selected on the basis of experience or according to the conveneince of the user. Hence unlike random sampling, here, the probability of selection of some of the units of the population are very high, and some are less. This method is not recommended for general use. However in case of experined investigator judgement sampling may give valuable results.

CHECK YOUR PROGRESS



Q 1: Fill in the blanks.

(i) If all the units of the population are enumerated these it is called ..... (sample/population)

(ii) A population consisting of unlimited number of units is called a ..... population. (finite/infinits)

Q 2: State whether the following statements are true or false.

(i) Sampling errors are present both in complete enumeration and sample survey. (T/F).

(ii) Judgement sampling is not recommended for general use. (T/F)

Q 3: In what situation sampling is inevitable.

Q 4: Mention in brief the objective of sampling.

Q 5: Distinguish between parameter and statistic.

12.3.6 Selection of Random Samples

Random samples are mainly selected either by (i) Lottery Method or (ii) Random number table method.

(i) Lottery Method : In this method all the units of the population are numbered and the numbers are written on some cards which are identical in shape, size and colour. These cards are then placed in a jar and shuffled properly. A blind-fold selection is made from these cards untill the desired sample size is acheived. However shuffling is necessary after each draw. This is a very reliable method of

relaction of random sample. But when the population is large this method is not practical.

(ii) Random number method : This is the most widely used method of selection of random samples. Here at first we number the population. Then we select at random any page of a random number table and pick up numbers in any row or column at random. Items corresponding to numbers thus selected constitute a random sample.

Random number tables are so constructed that each of the digits 0, 1, 2, ..., 9 occurs independently and approximately with same frequency. Some commonly used such tables are 'Teppets Random Number Table'. 'Fisher and Yates Table' etc.

12.4 SIMPLE RANDOM SAMPLING

The simplest and the most commonly used type of probability sampling is simple random sampling. In this sampling each unit of the population has the same probability of being included in the sample. Simple random sampling may be of two types – simple random sampling with replacement and simple random sampling without replacement, as discussed below :

Simple random sampling with replacement (SRSWR) :

In this type of sampling the members of the sample are drawn one by one from the population and after each draw the selected number is returned back to the population. As a result any member of the population may appear more than once in the sample. So if we want to draw a sample of size n from a population of size N, then each member of the population has the same probability of being selected and the total number of possible samples is N^n.

Simple random sampling without replacement (SRSWOR) :

In SRSWOR the sample units are drawn one by one and the members selected at any draw are not returned back to the population. Obviously in this case no member of the population can appear more than once in the sample, i.e. the samples units are distinct. If we want to draw a

sample of size n from a population of size N, the chance of selecting any unit at the first draw is $\frac{1}{N}$, at the 2nd draw it is $\frac{1}{N-1}$, at the third draw and so on. In SRSWOR the total number of possible samples are N_{g_n}.

12.4.1 Estimate of Population

Before going to find the estimate of population mean and population total we shall have to know some important notations as given below.

The N units of the population are denoted by $X_1, X_2, ..., X_N$

The n units of the sample are denoted by .

$\mu = \frac{1}{N}\sum_{i=1}^{n} X_i$ is the population mean.

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} (X_i - \mu)^2$ is the population variance.

is the population mean square.

$\frac{1}{N-1}\sum_{i=1}^{n} (X_i - \mu)^2$

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean.

$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$ is the sample variance.

$s'^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$ is the sample mean square.

Using the above notations we can state a very important theorem as stated below.

Theorem.

Statement : In simple random sampling sample mean is an unbiased estimator of the population.

In symbol, we can write,

$E(\bar{x}) = \mu$

Again we know that population total is $\sum_{i=1}^{n} X_i = N\mu$

$\therefore$ Estimate of population total is

i.e. $E(N\bar{x}) = N\mu$

Some useful Results :

(i) $\frac{n}{N}$ is the sampling fraction and is considered negligible if its valus is less than 0.05.

(ii) $\frac{N-n}{N-1}$ is called the finite population correction factor and may be ignored if N is vary large compared to n.

(iii) In SRSWR, the sampling variance of estimate of population mean is given by, $V(\bar{x}) = \frac{\sigma^2}{n}$.

(iv) In SRSWOR, the sampling variance of estimate of population mean is given by, $V(\bar{x}) = \frac{N-n}{N-1}\frac{\sigma^2}{n}$.

(v) $V(\bar{x})_{SRSWOR} < V(\bar{x})_{SRSWR}$

Note : The sampling design having less variance is always more efficient. Hence SRSWOR is more efficient than SRSWR sampling design.

12.4.2 Illustrated Examples

1. Suppose that 4 units of a population are $X_1 = 2$, $X_2 = 4$, $X_3 = 6$, and $X_4 = 8$. Draw all possible random samples of size 2 without replacement and calculate their means, show that $E(\bar{x}) = \mu$.

Sol^n : Here we have N = population size = 4

n = sample size = 2

Total no. of samples draw by SRSWOR is $4_{C_2} = 6$, as given below:

Serial No.	Sample	Sample mean (x_i)
1	(2. 4)	$\frac{2+4}{2}$
2	(2, 6)	4
3	(2, 8)	5
4	(4, 6)	5
5	(4, 8)	6
6	(6, 8)	7
Total		

$$E(\bar{x}) = \frac{1}{{}^4c_2}\sum_{i=1}^{n}\bar{x}_i = \frac{30}{6} = 5$$

$$\mu = \frac{X_1 + X_2 + X_3 + X_4}{4} = \frac{2+4+6+8}{2} = 5$$

$\therefore E(\bar{x}) = \mu$.



CHECK YOUR PROGRESS

Q 6: State true or false.

(i) When every unit of the population has the same probability of coming to the sample it is called sample random sampling. (T/F)

(ii) In SRSWOR the same item may appeax more than once in the sample.

Q 7: Fill in the blanks.

(i) $V(\bar{x})_{SRSWOR}$ = ...

(ii) In simple random sampling estimate of population total is .....

Q 8: Distinguish between SRSWR and SRSWOR.

12.5 STRATIFIED RANDOM SAMPLING

Till now we have discussed the technique of drawing simple random samples and also obtained estimates of the population parameters. But simple random sampling can give a representative sample only when the population under consideration is homogeneous with respect to the characteristic under study. But in practice we come across populations which are not homogeneous in nature. The sample technique used for such a population is stratified random sampling which by using available information concerning the population, attempts to design a more efficient sample than obtained by simple random sampling technique. While applying stratified random sampling technique, the procedure following is given below:

(i) The population to be sampled is subdivided into groups called strata which are non-overlapping and collectively exhaustive.

(ii) A sample random sample is then chosen independently from each

group, untill the required sample size is achieved.

However, for maximum efficiency the strata should be re formed that the units within a strata are as homogeneous as possible and a marked difference is possible belance the strata.

For illustration suppose our heterogeneous population of size N is divided into k strata, so that, there are

N_1 units in the 1st stratum

N_2 units in the 2nd stratum

. . . . . .

N_k units in the k^{th} stratum.

So that $N_1+N_2+.....+N_k = N$.

We now want to draw a sample of size n from this population so that we draw.

n_1 units from the 1st stratum

n_2 units from the 2nd stratum

. . . . . .

n_k units from the k^{th} stratum

such that $n_1+n_2+.....+n_k = n$. Only if the samples from each n_2 units from the 2nd stratum are drawn using SRS then the method is called stratified random sampling.

12.5.1 Advantages and Disadvantages of Stratified Random Sampling

Advantages :

(i) This method ensures batter representation of the characteristic of the population.

(ii) Decision of a big population into small sub-populations can being administrative convenience. For example the agency conducting the survey may employ different field officers for different parts of the population.

(iii) Sometimes different degree of accuracy is needed for different sections of the population, which may be possible only in stratified random sampling.

(iv) In case of skewed data like income data this method is very effective.

Disadvantages :

(i) It is important to deciside the basis of stratification well in advance, otherwise the results may be misleading.

(ii) Each strative most have, as far as possible homegeneous items, otherwise the results may not be reliable. If proper stratification is not done, the sample may have the effect of bias.

(iii) Because of the likelihood that the stratified sample will be more widely distributed geographically than a random sample, cost per observiation may be quite high.

12.5.2 Allocation in Stratified Random Sampling

In stratified random sampling, the allocation of sample to different strata is done by considering three factors, viz.

(i) the total number of units in the stratum,

(ii) the variability within the stratum,

(iii) the cost in taking observations per sampling unit in the stratum.

A good allocation is one where maximum precision is obtained with maximum resources. In otherwords, allocation of sample size should be such that to minimize the budget for a given variance or minimize the variance for a fixed budget, so that there is most effective use of available resources.

There are three methods of allocation of sample size to different strata in stratified sampling technique. They are

(i) Proportional allocation,

(ii) Neyman allocation,

(iii) Optimum allocation.

(i) Proportional allocation : This allocation was originally proposed by Bowley (1926). This procedure of allocation is very common in practice because of its simiplicity, where the size of the sample from each stratum varies with the size of the stratum. Larger the size of the stratum. Larger is the size

of the sample selected from that stratum. In proportion all allocation, one samples each stratum in proportion to its relative weight.

(ii) Neyman allocation : This allocation of total sample size to strata is called maximum variance allocation and is due to Neyman (1934). Here the allocation of samples among different strata is proportional to the product of the size and variability of the stratum. The mean difficulty in the allocation is that the value of S_i well be generally unknown.

(iii) Optimum allocation : In this method of allocation of the sample size in the respective strata with a view to minimize variance for a specified cost of the survey or to minimize the cost for a specified value of the variance.

Illustrated example :

1. The population of three towns are 40,000, 30,000 and 20,000 respectively. A stratified random sample of size 400 is to be drawn from this population using proportional allocation. Determine the sample size selected from each stratum.

Sol^n : Here, N_1= 40,000, N_2= 30,000, N_3= 20,000

$\therefore$ N = N_1+N_2+N_3 = 90,000

sample size = n = 400

Now, n_1 = sample size from 1st stratum

$$n_2 = 40{,}000 \times \frac{400}{90{,}000} = 133$$

$$n_3 = 40{,}000 \times \frac{400}{90{,}000} = 89$$

Obveovity $n_1 + n_2 + n_3 = 400 = n$.



CHECK YOUR PROGRESS

Q 9: State true or false.

(i) stratified sampling is preferred when the population is homogeneous. (T/F)

(ii) sample Random sampling method is more efficient than stratified random sampling method (T/F)

12.6 SYSTEMATIC RANDOM SAMPLING

Till now we have discussed sampling techniques where sampling unit were selected randomly. The sampling technique we are going to discuss now is also known as quasi-random sampling technique, because in this sampling technique only the first unit is selected at random. Once the interval starting point is determined the remainder of the items selected for the sample are predetermined by the sampling interval.

So a systematic sample is formed by selecting one unit at random and then selecting additional units at evenly spaced intervals untill the desired sample size is acheived. This method is popularly used in those cases where a complete list of the population from which sample is to be drawn is available. The list may be prepared in alphabitical, geogrephical, numerical or some other order. The items are then serially numbered. The first item is selected at random generally by using 'Lottery Method'. Subsequent items are selected by taking every k^{th} unit of the list, where k stands for sampling interval, i.e. $K = \frac{N}{n}$

where, n = sample size,

N = population size

Systematic sampling is simple and fool proof. We shall first discuss the sample selection procedures before going to discuss its advantages and desadvantages.

12.6.1 Linear Systimatic Sampling

As discussed above, a common procedure is the linear systematic sampling scheme. We suppose that the N units of the

population are linearly ordered from which a sample of size n is to be drawn. Here two cases may arise :

Case 1 : Let N be expressible in the form N = nk, where k is an integer, called the sampling interval.

In this procedure, a random number $i \leq k$ is selected. Here the sample will consists of every kth unit there after. Hence, under this procedure the sample comprises the units

.

The technique well generate k systematic samples with equal probability which may be shown in the following table.

Sample Number			
1	2		k
y_1	y_2		y_k
	y_{2+k}		y_{2k}
y_{1+2k}	y_{2+2k}		y_{3k}
.	.		.
.	.		.
.	.		.
$y_{1+(n-1)k}$	$y_{2+(n-1)k}$		y_{nk}

Case 2: N is not expressible in the form N = nk

Here the present sampling scheme will give rise to samples of unequal size. Here also we following the same procedure with the random selection of the 1st unit and selection of every kth unit thereafter since in this case $N \neq nk$, k is taken as an integer nearest to . Under this scheme the sample size is not necessary n and in some cases it may be (n-1).

To make this concept more clear let us consider a population of 12 units from which we are to draw a systamatic sample of size n.

i.e N = 12, n = 4

In this case N = nk where k = 3. Hence the possible samples will be,

$i\ (\leq k)$	Selected sample
1	(1, 4, 7, 10)
2	(2, 5, 8, 11)
3	(3, 6, 9, 12)

Now, let N = 11, n = 4.

So

So the samples will be.

$i\ (\leq k)$	Selected sample
1	(1, 4, 7, 10)
2	(2, 5, 8, 11)
3	(3, 6, 9)

i.e, when N is not a proper multiple of n all the samples are not of equal size. To overcome this difficulty of varying sample size we may use circular systematic sampling.

$N \neq nk$

$\therefore k = \frac{11}{4} = 2.7 \cong 3$

12.6.2 Circular Systimatic Sampling

So as discussed above the disadvantage of sample size being different from actual sample size required when , the procedure is slightly modified by which a sample of constant size is always abtained. The procedure consists in selecting a unit by a random start, from 1 to N and these thereafter selecting every k^{th} unit, k being an utiter nearest to , in a circular manner, untill a sample of n units is abtained.

For illustration let N = 11, n = 4, so that k = 3 (apprx.) so the possible samples are,

(1, 4, 7, 10), (2, 5, 8, 11), (3, 6, 9, 1), (4, 7, 10, 2), (5, 8, 11, 3), (6, 9, 1, 4), (7, 10, 2, 5), (8, 11, 3, 6), (9, 1, 4, 7), (10, 2, 5, 8), (11, 3, 6, 9).

12.6.3 Advantages and Disadvantages of Systematic Sampling

Advantages :

1) Systematic sampling is operationally more convencient than simple random sampling or stratified random sampling, because it requires less time and less work.
2) Systematic sampling may be more efficient than simple random sampling when the list from which the sampling units are drawn is arranged at random.

Disadvantages :

1) Systematic sampling becomes less representative if we are dealing with populations having 'hidden periodicitices'.
2) If the population is ordered in a systematic way with respect to the characteristic under study, then it is possible that only certain types of items will be included in the sample, or at least more of certain types of items than others. For instance, in a study of workers ages the list may be such that every tenth worker on the list gets above Rs. 2000 per month.

12.6.4 Uses of Systematic Sampling

Systematic sampling is operationally convenient and efficient in sampling some natural like forest areas for estimating volume of timber, areas under different types of couers etc.

CHECK YOUR PROGRESS



Q 10: State true or false.

(i) When the population consists of units arranged at random one would prefer systematic sampling. (T/F)

(ii) When the population size is not a proper multiple of sample size, one should prefer circular systematic sampling rather than ordinary systematic sampling. (T/F)

Q 11: In what situation systematic sampling is preferred over over sampling techniques.

Q 12: Complare systematic sampling with stratified sampling technique.

12.7 CLUSTER SAMPLING

In random sampling we have seen that the population is divided into a finite number of distict and identifiable elements known as sampling units. The smallest units into which the population is divided is called an element of the population. A group of such elements is called a cluster. So when the sampling unit is a cluster, rather than a single element of the population, the sampling procedure is known as cluster sampling.

Therefore, in cluster sampling the entire area containing the population is divided into some smaller area segments and each element of the population belongs to one and only one segment of the population. This procedure is also called ‘area sampling’.

Generally, in any sampling procedure the main problem lies in identification and location of an element. However once an element has been located, the time taken for surveying a few neighbouring elements is less. That is where the justification for formation of cluster lie. But as clusters are made-up of neigbouring elements, they may have similar characteristies. Hence it is always advisable to form small clusters and the number of clusters should also be more.

After dividing the population into clusters, the required number of clusters are selected by equal or unequal probabilities of selection. Again all elements in selected clusters are enumerated.

12.7.1 Advantages and Disadvantages of Cluster Sampling

Advantages :

1) Cluster sampling is less expensive and requires less times compared to other sampling techniques discussed so far. It is so because selection of neighbouring elements is easier, cheaper, faster and operationally more convencient.
2) This technique can be used when sampling frams of elements is not available and time consuming to prepare such a list.
3) It is less costly than SRS due to savings of time of journey, identification etc.

Disadvantages :

1) In general cluster sampling is less efficient than SRS. However, efficiency of cluster sampling increases with decrease in cluster size and increase in variation withen cluster.

Note : Various sampling procedures can be applied to cluster sampling by treating clusters.

12.7.2 Optimum Cluster Size

Efficiency of cluster sampling increases with the increase in the number of clusters. However, the cost of survey also increases with increase in the number of clusters. But in any survey our resources are always limited. That is we must try to minimize cost. Hence it is necessary to determine a balancing point by finding out the optimum cluster size and the number of clusters in the sample so that maximum efficiency can be obtained for a given cost or minimum cost is incurred for a given level of efficiency.



CHECK YOUR PROGRESS

Q 13: State whether the following statements are true or false.

(i) A cluster consists of a number of elimentary units of the population. (T/F)

(ii) Efficiency of cluster sampling increases as cluster size increases. (T/F)

Q 14: In what situation cluster sampling is preferred.



12.8 LET US SUM UP

- Population is a group of individuals with respect to certain characteristic under study.
- Sample is a part of the population selected from it in order to study its characteristic.
- Census or complete enumeration refers to enumeration of each and every unit of the population.
- Sample survey refers to enumeration of each and every unit of the sample.
- Two basic principles of sample survey are validity and optimization.
- Non-sampling errors are present in both census as well as sample survey, whereas sampling errors are present only in sample survey.
- Sampling in broadly classified into two categories– probability sampling and non-probability sampling.
- If each unit of the population has the same probability of being included in the sample, it is called sample random sampling. SRS may be of two types–

  Simple random sampling with replacment (SRSWR).

  Simple random sampling without replacment (SRSWOR).
- SRSWOR scheme is more efficient then SRSWR.
- Stratified random sampling is used when the population under consideration is not homogneous with respect to the characteristic

under study.

- stratified random sampling is more efficient than SRS.
- In systematic sampling techneque, only the first unit is selected at random and the rest gets automatically selected according to some pre-designed pathern.
- In systematic sampling when the population size N is a multiple of the sample size n, linear systematic sampling is used and when N is not a proper multiple of n, circular systematic sampling is to be used to overcome the difficulty of varying sample size.
- In cluster sampling, the population is divided into some non-overlapping clusters and required number of cluster are selected by equal or unequal probabilities of selection.
- Efficiency of cluster sampling increases with the increase in the number of clusters, but at the same time cost also increases. Hence, it is necessary to balance between them so that maximum efficiency can be abtained.



12.9 FURTHER READINGS

1) Cockran, W. G. Sampling Techniques (1963), John Wilay and Weley Eastern.

2) Gaon, A. N, Gupta, M. K, Dasgupta, B. Fundemantals of Statistics (Vol II). (1986). The World Press Private Limited.

3) Singh, Daroga, Chavdhury, F. S. Theory and Analysis of Sample Survey Designs. (1986). New Age International Publishers.

4) Sukhatme, P. V. and Sukhatme, B. V. Sampling Theory of Surveys with Applications (1970). FAO and Asia Publishing House.



12.10 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (i) Population census, (ii) infinite.

Ans to Q No 2: (i) True, (ii) True, (iii) False.

Ans to Q No 3: Sampling is inevitable under the following conditions.

(i) When the resources for survey, i.e. money and trained man power is limited,

(ii) When there is time constraint,

(iii) When the population is very large,

(iv) When the survey is destructive in nature,

(v) When survey is to be conducted in a very wide area.

Ans to Q No 4: The main objective of the survey is to obtain maximum information about the population at minimum cost, time or labour.

Ans to Q No 5: The difference between parameter and statistic is as follows:

(i) The parameter are population characteristics such as population population variance but statistics are sample characteristics such as sample mean, sample variance etc.

(ii) Parameters are constant quantitics but statistics are variables.

Ans to Q No 6: (i) True, (ii) False.

Ans to Q No 7: (i) $\frac{N-n}{N-1}\ \frac{\sigma^2}{n}$ (ii) $N\bar{x}$

Ans to Q No 8: If the units of the population are drawn one by one in such a way that the units drawn are returned back to the population after each draw, it is known as SRSWR. In this type of sampling the probability of drawing a unit at each draw remains $\frac{1}{N}$. In this case a unit may appear more than once in the sample.

On the otherhand if the units selected from the population are drawn one by one but are not returned back to the population, it is called SRSWR. Here the probability of drawing any unit at the r^{th} draw is $\frac{1}{N-r+1}$. In this type of sampling the same unit can never be included in more than one sample.

Ans to Q No 9: (i) False, (ii) True.

Ans to Q No 10: (i) True, (ii) False.

Ans to Q No 11: Systematic sampling is preferred when informations to be collected about trees in a forest, houses in a block, entries in a register, which are in serial order.

Ans to Q No 12: Systematic sampling resembles stratified sampling in the sense that the population is divided into strata of size k where one unit is drawn from each stration. But the fact that the strata should be homogeneous is not followed here.

Again no independent samples are drawn from each group. Also in systematic sampling we have only one sample from the whole population.

From above discussion it is clear that systematic sampling is quite different from stratified sampling.

Ans to Q No 13: (i) True, (ii) False.

Ans to Q No 14: The cluster sampling is preferred when

(i) the sampling frame is not available or too expensive and time consuming to prepare it,

(ii) the sampling units are distance apart,

(iii) the elimentary units may not be easily identifiable and localable.



12.11 MODEL QUESTIONS

Q 1: Give a brief account of advantages and disadvantages of sampling.

Q 2: What is simple random sampling? What are its advantages?

Q 3: What is stratified random sampling? Why stratified random sampling is more advantageous than SRS?

Q 4: Distinguish between random sampling and judgement sampling.

Q 5: Distinguish between linear and circular systematic sampling.

Q 6: Describe briefly the different sources of error in sample survey.

Q 7: What are the advantages of sample survey over complete enumeration.

Q 8: What is simple random sampling? Describe two methods of selection of selection of simple random sample.

Q 9: A simple random sample of 30 households was drawn from a city having 14848 households. The number of in the selected households are :

5, 6, 3, 3, 2, 3, 3, 4, 4, 3, 2, 7, 4, 3, 4, 4, 3, 3, 4, 4, 3, 3, 1, 2, 4, 3, 4, 3, 4, 3. Estimate the population total.

Q 10: Write short notes on: (i) Principles of sample survey, (ii) Non-sampling error. (iii) Cluster sampling, (iv) Systematic sampling.

*** ***** ***