UNIT 6

STRATIFIED RANDOM SAMPLING

Structure

6.1 INTRODUCTION

In the Units 1 to 5 of Block 1, we presented the concepts and related theories of two sampling schemes, namely, Simple Random Sampling (SRS) scheme and Varying Probability Sampling (VPS) scheme, which are considered to be the fundamental classes of sampling schemes.

The other sampling schemes belonging to these two classes are either extensions or advancements of these two schemes. We have already mentioned at several places in Units 4 and 5 that Varying Probability Sampling (VPS) schemes and particularly Probability Proportional to Size (PPS) sampling scheme that Probability Proportional to Size with Replacement and Probability Proportional to Size without Replacement are comparatively quite complicated than Simple Random Sampling (SRS) schemes due to their changing probability structure after each draw, which needs a new definition of the variable values at each draw. Keeping in view this practical difficulty of Varying Probability Sampling (VPS) schemes, in the literature of survey sampling most of the sampling schemes are developed assuming that they follow the law of Equal Probability Selection Method (EPSEM) similar to Simple Random Sampling (SRS) scheme.

As mentioned above, in this unit, our aim will be to discuss some basics of the Stratified random Sampling (STRS), such as, concept and meaning of stratification of a population; kinds/types of stratification; reasons of stratification of an unstratified population, examples of stratified populations which are frequently observed in the real world, etc. We shall then describe the procedure(s) of selecting random sample(s) from a stratified population for the purpose of estimation of some population parameters. Particularly, we shall show how a suitable estimator can be defined for estimating the population mean. We shall discuss some of the properties of this estimator and shall obtain the expression of the variance of the estimator. The estimator of this variance would also be obtained so as to obtain a numerical value of it based on the sample values.

Section 6.2 of this unit mainly explains the meaning of stratification of a finite population into a number of sub-populations and some other related concepts, like stratum and strata, stratified population. In this section we shall also present some examples of such populations which are either available in the stratified form or stratified by the sampler himself/herself using certain characteristic(s) of the population as per the objectives of the survey. Section 6.3 deals with the problem of selecting a random sample from the population which is given in the stratified form. The process of selecting the random sample of required size in such cases is described with its steps. Section 6.4 is devoted to the usual problem of estimating parameters ‘Population Total’ and ‘Population Mean’ on the basis of the selected stratified sample. Suitable estimators for both the parameters have been defined and some of their properties have been derived and highlighted. In Section 6.5, we have attempted to obtain the expression of the sampling variance of the population mean. On the basis of the expression of the sampling variance, we have tried to explain why sometimes stratification process is advantageous for getting more efficient estimators as compared to the SRSWOR scheme. Section 6.6 has been devoted to getting an unbiased estimator of the sampling variance so obtained.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ discuss the meaning of the terms “Stratification”, “Stratum and Strata” “Stratified Population” in relation to Stratified Random Sampling (STRS) scheme;
- ❖ explain reasons and purposes behind stratifying a finite population either geographical/spatial region-wise or by the sampler himself/herself as per the objectives of the survey in order to derive more precise results as compared to an unstratified population;
- ❖ present some real-world populations which are stratified following the abovementioned methods of stratification;
- ❖ describe the procedure of selecting a random sample of fixed size from a stratified population, that is, a population divided into a number of sub-populations;
- ❖ explain how the selected sample can be utilized for the purpose of estimating some population parameters, such as, population mean and population total and what are the estimators for these parameters;

- describe and derive some properties of these parameters;
- obtain the expression for the sampling variance of the estimator of population mean and explain the criterion of appropriate stratification in order to get more precise results based on the sample selected; and
- obtain the estimate of the sampling variance of the estimator of population mean.

6.2 STRATIFICATION AND STRATIFIED POPULATIONS

In order to proceed for selecting a random sample from a stratified population and dealing with such a sample for estimation purposes, it is necessary to understand the meanings of a naturally stratified population; stratification of a given population, if it is not stratified and benefits derived from stratification process of an unstratified population. All these concepts shall be explained with the help of some real-world finite populations, either naturally stratified or stratified as per need of the sampler in order to get some benefits over unstratified populations.

6.2.1 Meaning of Stratification

In Sub-section 1.2.2(a) of Unit 1 of Block 1, we have already given the concepts and definitions of the terms **"Population"** and "**Units"** or more specifically, **"Elementary Units"** of the population in sampling theory. According to discussions made there, by a population in sampling theory we mean a 'collection', 'gathering' or 'group' of some "specific objects", belonging to the population under consideration, which must be known to the researcher before he/she plans to conduct a sample survey for collecting necessary or desired information on these specific objects. If there is only a finite and countable number of distinct specific objects in the population, it is termed as a **"Finite Population"**. The "specific objects" of the defined population on which relevant information are gathered as per objectives of the survey, are generally called '**Units'**, '**Elements'** or more specifically '**Elementary Units'**.

Further, you are familiar that in "Simple Random Sampling" (SRS) scheme, a random sample of pre-determined size is selected from the population using a sampling frame of all the units belonging to the population. In other words, in such a method, the entire population is considered to be one population, ignoring all types of divisions of its units into different classes, if any. Diagrammatically, the process of selection of units of the population under Simple Random Sampling (SRS) scheme can be depicted as follows:

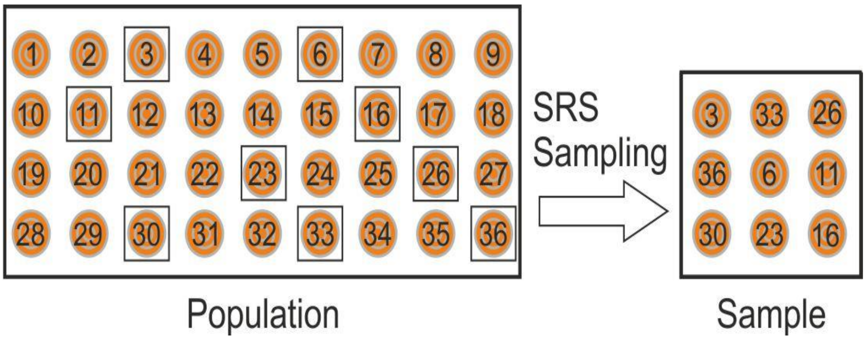

Fig: 6.1: Sampling Process Under Simple Random Sampling Scheme

The population consists of N objects, shown by the circles ◎. A random sample of size n is selected from the population under SRS scheme. Let us assume that the circles, shown by "◎ covered with square □" in the population, are randomly selected in n draws to constitute the sample. Clearly, this diagram shows that the units in the sample are selected from all parts of the population and that the population is not divided into different sub-groups or classes. Thus, in order to utilize the Simple Random Sampling (SRS) scheme, we must have following three things:

(i) There should be a pre-defined population consisting of some "object of interest". The Population must be well-defined in respect of its boundaries and the specific objects of interest which have to be observed or contacted for gathering the required information on characteristic under study.

(ii) An exhaustive list of all the objects belonging to the given population, that is, a complete and exhaustive sampling frame for the population objects without omissions and/or duplications of objects. Let the sampling frame consists of N objects, implying that there are exactly N objects of interest in the population under consideration.

(iii) The number of objects which are to be included in the sample of a given size. Let it be n. Then, using the sampling frame of objects, n objects of the population have to be selected one by one in n successive draws using some chance mechanism. Thus, one gets a SRS sample of size n.

Contrary to the situation, where we have only one population in hand for sampling purpose and we select only one sample from this population; sometimes, we face a different situation as regards the nature of the population. In many of the practical situations, we face the cases where the defined population is either naturally or administratively divided into a number of distinct classes (sub-populations) or needs to be divided artificially into such classes for some reasons; such that there is no empty sub-population and must have some objects of interest.

The basic question in such cases which arises is that how to a sample of pre-determined size, n, can be selected from such divided population if we need to fetch answers of some quarries related to the entire population or need to estimate any parameter of the population concerned.

In fact, "Stratified Random Sampling" (STRS) scheme helps us in answering this important question. In order to explain the basic theories and concepts which applies to a population, divided into distinct sub-populations, we shall first define and explain few of the terms which are as follows:

(a) Stratification:

Given a finite population of size N, say, if one wishes, due to his/her convenience or due to some scientific reasons, to classify the elementary units of the population into some distinct and non-overlapping classes or sub-populations following some rules of division, then this process of division of the given population is known as **"Stratification"**. Thus, stratification is merely a process of division of a population into a number of non-overlapping and distinct sub-populations such that each and every elementary unit of the population belongs to one and only one sub-population.

(b) Stratum and Strata:

Each sub-population, obtained after stratification of a finite population is called a **"stratum"**. **"Strata"** is the plural word used for stratum. The **'stratum size'** refers to the number of units belonging to a stratum. If strata are non-overlapping and distinct with each other, obviously, the population size must be equal to the sum of sizes of all the strata. Thus, if the population size is N and the population is stratified into five non-overlapping sub-populations, then the condition

$$N = N_1 + N_2 + N_3 + N_4 + N_5 = \sum_{i=1}^{5} N_i$$

must be satisfied. Here, N_i stands for the size of the i^{th} stratum.

(c) Stratified Populations:

Clearly, a stratified population is that population which consists of k non-overlapping and distinct sub-populations satisfying the condition $N = \sum_{i=1}^{k} N_i$.

It can be seen that in real situations, most of the populations are available in stratified form. In fact, whenever a survey sampler wishes to select a simple random sample from a population, he/she comes across with a population which is already stratified into a number of sub-populations; the stratification is generally done by the administrators/ local governments for their convenience in order to keep separate administrative records for the sub-populations so that these can be handled with all ease.

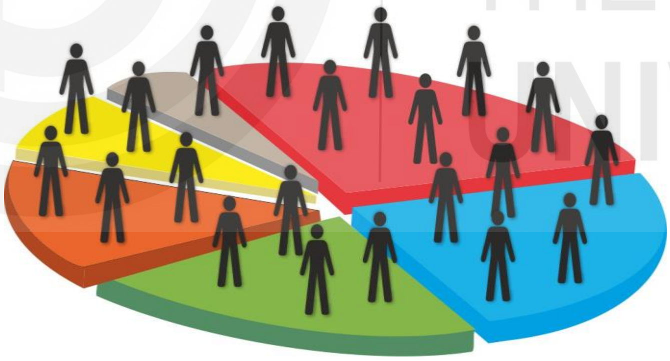

Fig. 6.2: Stratified Population

On the other hand, sometimes the population under consideration might be unstratified, but the survey sampler expects more gain in accuracy in results, derived from the sample values, in comparison to Simple Random Sampling scheme; if he/she stratifies the population into a number of strata. Accordingly, he/she tries to divide the undivided population into a number of strata using some rule(s) of division. One of such rules which is most commonly used is to divide the elementary units in such a way that a particular stratum consists of similar kind of units so that the stratum is as homogeneous as possible in respect of the characteristic under study. By following this concept of stratification, in fact, the sampler divides the population into strata such that each stratum is internally homogeneous but are heterogeneous amongst them. We shall see afterwards why this rule of stratification is most popular and what benefits the sampler have in respect to the accuracy of the sample results.

6.2.2 Examples of Stratified Populations

(a) Geographical/Spatial Stratification:

(i) There is a lot of examples of sub-division (stratification) of a population either over several geographical regions or over different spaces within these regions done by the administrators. This type of stratification is done mainly due to the reasons of administrative conveniences in maintaining the available important records on some characteristics of the concerned population, quick recording of these information at regular intervals and their constant updating for smaller segments of the population. As for example, consider the population of human beings all over the world. However, the overall world population is so huge that practically it might be considered to be an infinite population, though it is countable; and therefore, it is impossible to prepare a sampling frame for this huge population for sampling purposes, as we do in SRS process. Moreover, we know that the total world population reside into a number of countries; countries being the sub-populations (strata) of the total world population. No doubt, this is an example of geographical stratification which is done only to divide the total population into small segments (sub-populations) for administrative convenience of governments/administrators of countries and there is no other purpose behind such stratification. Let us now think of a particular stratum of human population, that is, a particular country, its size might be practically too large. The countries also stratify their respective populations internally into different states/counties for the convenience in maintaining important vital, economic, social and many other records related to their inhabitants. Generally, the states/counties have their own independent systems of collecting, maintaining and presenting these data. Further, states are generally stratified into districts which are then stratified into cities, towns, villages. The ultimate stratification of cities/towns/villages into more small sub-populations provides us wards, sectors and mohalla's. You can realize that all these division of human population are examples of geographical stratification and/or spatial stratification which are done by the concerned governments/administrators according to their own convenience in handling the necessary records.

(ii) The population of indoor patients in hospitals is generally admitted into different wards according to their kind of disease. These wards are, in fact, strata into which the indoor patients are accommodated by the hospital administrations according to their sizes and disease. Generally, these wards provide independent indoor facilities to the patients, like, appropriate paramedical staff, number of beds for patients, storage of necessary medicines, sanitary facilities, etc. This is an example of spatial stratification within the hospital.

(iii) Similar to above example, spatial stratification is followed by schools, where strata are different sections (standards) of students admitted; universities, where different departments/faculties/ institutes are strata consisting of students and teaching staff of a particular discipline; institutions/ organizations having a number of sections (divisions) each division consisting of experts/workers providing independently different kinds of services and so on.

It can be visualized that the stratification process in all the above examples is mainly because of spatial reasons. However, you can see that in most of the spatial stratification cases, care is taken by the administration to include similar kind of services/facilities provided in a particular stratum and two different strata must differ significantly in this respect.

(b) Stratification Made with Some Specific Purposes

You have seen that one way of stratification is to divide the population geographically or over space by governments/administrators according to their administrative conveniences. You have also seen that such a stratification method is mainly adopted due to converting the entire population into small sub-populations so as to make these small size sub-populations handy in tackling them easily in all administrative purposes.

However, even if the data/information are available in the form of geographically / spatially divided strata; often the concerned researcher / investigator sometimes feels that he/she would get sample results with more accuracy, if the population would have been divided into strata on the basis of some different characteristic(s) of the units other than their geographical/spatial characteristic. In such situations, generally he/she tries to re-arrange the units into newly formed strata following some scientifically justified criterion depending upon the objectives of the study.

Below we give some examples where stratification of a population may be done with some specific purposes as per demand of the survey / study:

(i) Let a manufacturing company desired to recruit some applicants for its computer division, consisting of various types of vacancies in the division with different slabs of salaries. However, as per the essential qualification for these posts, each applicant is supposed to have at least graduate level knowledge and workable experience of computers and software packages. The designations of the posts are as follows:

1) Executive Director;

2) Software Engineer / Supervisor;

3) Programme Developer;

4) Computer Technician/Operator; and

5) Office Assistant.

The company, after recruiting the desired applicants, appointed the desired number of candidates on these posts keeping in view their theoretical as well as practical knowledge of software and computer. This is an example in which the population of selected candidates, found suitable for one or more of these posts, was categorized into any one of the five posts, after having done a thorough scrutiny of their theoretical knowledge and workable experience of handling software packages and computer operation along with their suitability for the post. So, in this example, criterion of stratification of candidates into five different sub-populations was the skill, experience and specific knowledge the candidate had. The stratification process, therefore, is made purposefully fulfilling some criterion.

(ii) Let us consider the population of all the teaching staff in a university. Let the teaching positions in the university be Professor, Associate Professor, Assistant Professor, Guest Faculty and Research Scientist and all these posts belong to different institutes / faculties / departments and schools of the University. Thus, the total population of teachers is available in the stratified form, where strata are institutes / faculties / departments and schools. You know that usually, in almost all the universities, the criterion of stratification of the concerned population of teachers is to appoint them in such a department which deals with that discipline to which teachers have excelled themselves with sufficient teaching and research experiences.

Let us now assume that an investigator decides to conduct a sample survey to estimate the *'average monthly expenditure'* of the teaching community of the University. Obviously then, he/she decides to select a random sample of teachers of appropriate size out of the entire teaching community. But, due to stratification of teachers over departments, perhaps it would not be feasible and simple to get a single sampling frame of the entire teaching community of the University for selecting a simple random sample of appropriate size. Moreover, he/she may expect to observe a large variability in the characteristic *'monthly salary'* of these teaching categories in each stratum because departments generally include all categories of teachers within them, and hence, a large variability in expenditure pattern too, highly correlated to monthly salary, is expected over the categories of teachers. In order to make the sample a true representative of the variability in monthly salaries, therefore, the sampler might think to put teachers in five distinct strata such that each stratum consists of only one category of teachers. If stratification is done in this way, obviously each stratum becomes internally almost homogeneous in terms of variable 'Monthly Salary' and heterogeneous amongst themselves. Following this stratification process, if separate samples are selected from each stratum and then mixed up into one sample, the extent of variability existing in the population would also be reflected in the sample. No doubt, the sampler is then required to prepare separate sampling frames for all the five categories of teachers.

Remark 6.1: As mentioned earlier, mostly human populations are readily available in stratified form; stratification done either over geographical regions or over space. Examples considered above make it clear that geographically or spatially stratification are acceptable as such to the researchers, dealing with some projects, without making any changes in the criterion of stratification. But sometimes they might feel that if stratification would have been done on the basis of other characteristics; it would certainly improve the accuracy of the results derived from the sample values. In such cases he/she has to stratify the total population into strata; stratification characteristic being his/her own choice as per the objective of the research. Thus, a geographically / spatially stratified human population might be re-structured into newly constructed strata made on the basis of religions, castes, ages, family sizes, highest education levels, association with political parties, types of jobs, etc.

On the basis of what has been discussed and described in this section, answer the following Self-Assessment Question:

SAQ 1

What do you mean by stratification of a finite population? What is a stratum?

6.3 SELECTING A STRATIFIED RANDOM SAMPLE OF FIXED SIZE

After knowing about the different stratification methods of a population along with their examples, the next thing which is important to know is that how a random sample of fixed size can be selected from a stratified population.

Let us assume that the finite population under consideration is of size N. Let it be divided into k sub-populations following any criterion of stratification, that is, the given population consists of k strata in all. Suppose that the size of the i^{th} stratum is N_i (i = 1, 2, 3, …, k), such that

$$N_1 + N_2 + N_3 + \ldots + N_k = \sum_{i=1}^{k} N_i = N \qquad \ldots (6.1)$$

Clearly, when the population is stratified in a number of strata of different sizes, instead of only one population of size N, we have to deal with k sub-populations in order to get a sample of size n from the whole population having size N. The question arises of how it can be done. The answer is provided in the next sub-section.

6.3.1 Stratified Random Sampling Scheme

The sampling procedure followed to select a random sample of pre-fixed size from a stratified population is termed as ***"Stratified Random Sampling (STRS)"*** scheme. The steps of Stratified Random Sampling Scheme are as follows:

Step I: Given the finite population of size N, stratified into k non-overlapping strata with respective sizes $N_1, N_2, N_3, \ldots, N_k$ such that

$$\sum_{i=1}^{k} N_i = N \text{ and } N_i > 0 \text{ for all i;}$$

Treat these k sub-populations as k different populations with varying sizes from which we wish to select k independent simple random samples, one sample from each stratum, using Simple Random Sampling simple random sampling without replacement (SRSWOR) scheme of units. Let the size of the sample to be selected from the given population be n and the size of the sample to be selected from the i^{th} stratum be n_i for i = 1, 2, 3 …, k, pre-fixed in advance. Here, all n_i must be greater than zero and the condition $\sum_{i=1}^{k} n_i = n$ must be satisfied. Obviously, then we observe that $0 < n_i < N_i$ for all i.

Step II: With the help of available sampling frame of the first stratum, select n_1 units from the stratum using Simple Random Sampling without Replacement scheme from the sub-population of size N_1.

Step III: Similarly, from the available sampling frame of the second stratum, select n_2 units using Simple Random Sampling without Replacement scheme from the sub-population of size N_2.

Step IV: Repeat this process for third stratum, fourth stratum, …, k^{th} stratum in order to select respective random samples of sizes $n_3, n_4, \ldots, n_k$ from sub-populations of respective sizes $N_3, N_4, \ldots, N_k$.

The following diagram explains the process of sampling in Stratified Random Sampling scheme:

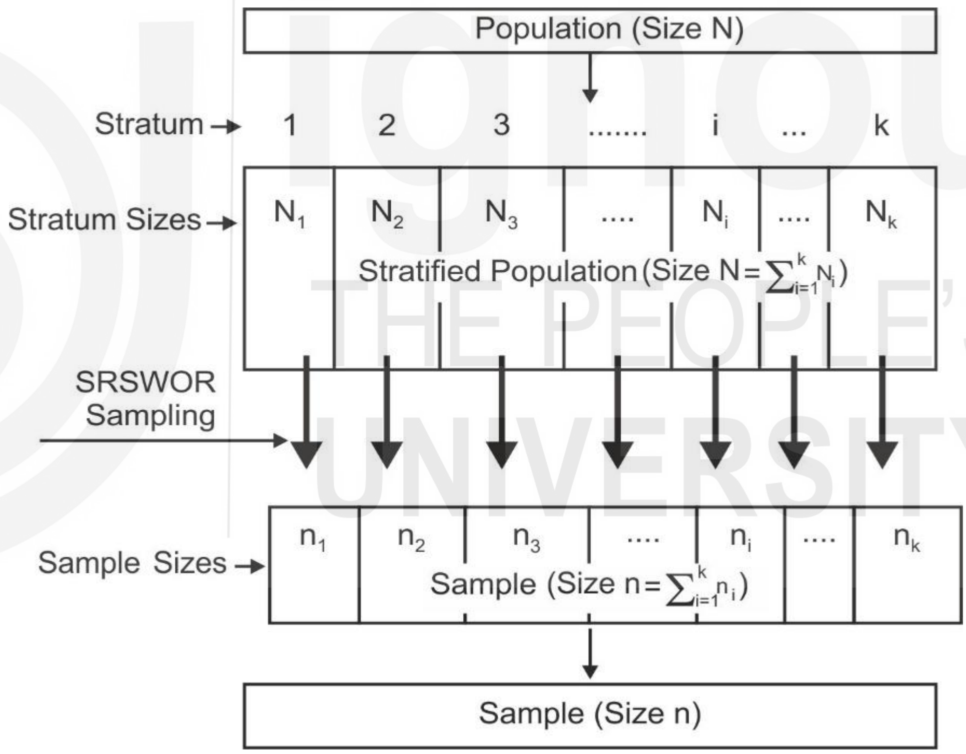

Fig. 6.3: Process of Selecting a Stratified Random Sample

Since, k random samples of desired sizes are selected independently from all the strata, finally we get a random sample of size $\sum_{i=1}^{k} n_i$ which is equal to the pre-fixed size of the sample, which is n.

Thus, since in case of stratified population, we have k sub-populations each having a separate sampling frame instead of only a single sampling frame for the total population, we are bound to select k distinct samples of different sizes, one sample selected from each stratum. Our combined sample of size $\sum_{i=1}^{k} n_i = n$ assures us to be a good mixture of all types of units even if the population have a large variability in respect of characteristic under study. In this sense, as compared to Simple Random Sampling without Replacement scheme, Stratified Random Sampling scheme provides a sample with greater representativeness of the variability existing in the population.

Remark 6.2:

Even though in Stratified Random Sampling, we select independent samples from each stratum; our aim is not to use these separate samples for estimating some parameter of a specific stratum, but we are mostly eager to estimate or infer about a parameter of the given population of size N. For instance, using the sample of size n_3, selected from the third stratum, one can estimate the population mean, population ratio, population variance, etc., of the third sub-population using Simple Random Sampling without Replacement scheme; but generally, it is not the motto in any survey when the population is stratified. Instead in most of the cases, selecting independent samples from each stratum in a stratified population, the aim of the survey / project is to estimate /predict /infer something about the parameter(s) of the concerned population and not of the sub-populations in which the population is divided.

Remark 6.3: As mentioned above that the ultimate sample size, n, is generally pre-fixed by the sampler/researcher prior to the beginning of the survey on the basis of (i) the total cost of the survey, (ii) time allotted for completing the survey and (iii) the number of skilled personnel to be deployed for completing the enumeration work, etc. Since, in Stratified Random Sampling scheme samples are selected from each and every stratum, so that the total of these separately selected samples equals to ‘n’, the question arises how many units have to be selected from each stratum so that the condition,

$$\sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k = n$$

is satisfied.

For example, n being 200, the sizes of the samples selected from four strata, that is, (n_1, n_2, n_3, n_4) might be (50, 50, 50, 50); (80, 20, 40, 60); (70, 90, 15, 25) and so on. Thus, it seems that given the pre-fixes sample size n, there might be arbitrariness in choosing the sample sizes, which might lead to different results. But, in fact, this is not so.

In order to avoid the arbitrariness in deciding the sample sizes to be selected from each stratum, and, hence, making the selection process of samples to be objective-specific and scientifically justified; some rules for deciding the sizes of the samples for each stratum are available in the literature of the survey sampling. However, we shall discuss this matter in the next unit in this block.

Now, you may try to answer the following Self-Assessment Question:

SAQ 2

“The stratified random sampling scheme, in fact, is an extension of simple random sampling scheme”. Do you agree with this statement and if yes, in what sense?

6.4 ESTIMATING POPULATION MEAN

As usual, if a random sample of fixed size is selected under any sampling scheme, the researcher utilizes the information gathered in the sample for the purpose of estimating/predicting some of the population parameters. The same thing is common with the sample selected under stratified random sampling scheme. In this section, our aim is to discuss and describe theories and methods of estimation of some population parameters on the basis of a random sample gathered under the Stratified Random Sampling scheme.

First of all, we shall define some notations and their meanings which are required to derive the results needed for proposing estimators for population parameter in Stratified Random Sampling scheme.

6.4.1 Terms and Notations

(i) Population Related Terms and Notations

The given population consists of N units. These N units are stratified into k distinct and non-overlapping sub-populations (strata) such that the i^{th} sub-population consists of N_i units (i = 1, 2, 3, …, k). We use the following notations with meaning of each of them:

Y_{ij} : Value of the study variable Y on the j^{th} unit belonging to the i^{th} stratum

(j = 1, 2 , …, N_i; i = 1, 2, …, k).

$\bar{Y}_i$: Mean of the i^{th} Stratum in the Population. Thus, $\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}$.

S_i^2 : Population Mean Square in the i^{th} Stratum.

Thus, $S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2$.

$\bar{Y}$: Overall Population Mean.

Thus,

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{N_i} Y_{ij} = \frac{1}{N}\sum_{i=1}^{k} N_i\bar{Y}_i = \sum_{i=1}^{k}\frac{N_i}{N}\bar{Y}_i = \sum_{i=1}^{k} W_i\bar{Y}_i;$$

since $\sum_{j=1}^{N_i} Y_{ij} = N_i\bar{Y}_i$ and $W_i = \frac{N_i}{N}$; is the proportion of units in the i^{th} stratum.

Obviously,

$$\sum_{i=1}^{k} W_i = \frac{N_1 + N_2 + N_3 + \ldots + N_k}{N} = 1, \text{ since } \sum_{i=1}^{k} N_i = N.$$

Y_{Total}: Overall Population Total.

Thus, $Y_{Total} = N\bar{Y} = \sum_{i=1}^{k} N_i\bar{Y}_i$.

S^2 : Overall Population Mean Square.

Thus, $S^2 = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y})^2$.

(ii) Sample Related Terms and Notations

y_{ij} : The value of the study variable Y on the j^{th} unit selected at the r^{th} draw in the sample selected from the i^{th} stratum. Obviously, here j = 1, 2, …, n_i.

$\bar{y}_i$: Mean of the sample selected from the i^{th} stratum.

Thus, $\bar{y}_i = \frac{1}{n_i}\sum_j^{n_i} y_{ij}$.

s_i^2 : Sample Mean square in the i^{th} stratum in the sample.

Thus, $s_i^2 = \frac{1}{n_i - 1}\sum_j^{n_i}(y_{ij} - \bar{y}_i)^2$.

6.4.2 Estimation of Population Mean

We know that the population mean is given by

$$\bar{Y} = \sum_{i=1}^{k} W_i \bar{Y}_i,$$

which is, in fact, a weighted mean of the strata means, $\bar{Y}_i$, with weights W_i where i = 1, 2, 3, …, k. Since, the random sample selected from the i^{th} stratum is drawn from the stratum using SRSWOR scheme, no doubt, the sample mean $\bar{y}_i$ will be an unbiased estimator of the stratum mean, $\bar{Y}_i$, for all i, as per the result derived in **Theorem 5** under Sub-section 2.4.1, Unit 2 of Block 1. Thus, we can say that for all i, we have $E(\bar{y}_i) = \bar{Y}_i$.

We also know that N_i is the size of the i^{th} stratum, it is treated to be a constant. Therefore, $W_i = \frac{N_i}{N}$ will also be a constant quantity.

We observe that the population mean $\bar{Y}$ is a weighted mean of the kind $\sum_{i=1}^{k} W_i \bar{Y}_i$;

where W_i is a constant and $\bar{Y}_i$ is the mean of the i^{th} stratum; the ultimate estimator of $\bar{Y}$ may be considered to be a function of similar kind based on sample means. Let us, therefore, consider the function

$$\bar{y}_{STRS} = \sum_{i=1}^{k} W_i \bar{y}_i.$$

Since, this function of sample means is a similar type of function as $\sum_{i=1}^{k} W_i \bar{Y}_i$; we may consider this function as an estimator of the population mean $\bar{Y}$. We shall, therefore, observe whether it is an unbiased estimator for $\bar{Y}$ or not. Let us discuss and prove the following theorem for this purpose:

Theorem 1: The estimator $\bar{y}_{STRS}$ is an unbiased estimator for population mean $\bar{Y}$.

Proof: Let us consider the expected value of the estimator $\bar{y}_{STRS}$. We have

$$E(\bar{y}_{STRS}) = E\left(\sum_{i=1}^{k} W_i \bar{y}_i\right) = E(W_1\bar{y}_1 + W_2\bar{y}_2 + W_3\bar{y}_3 + ... + W_k\bar{y}_k)$$

$$= W_1 E(\bar{y}_1) + W_2 E(\bar{y}_2) + W_3 E(\bar{y}_3) + \ldots + W_k E(\bar{y}_k), \quad \ldots (6.2)$$

Remembering that W_i's are constants and applying the additive property of expectation. Now, we know that

$$E(\bar{y}_i) = \bar{Y}_i \text{ for all } i = 1, 2, \ldots, k.$$

Using this result, we can write the equation (6.2) as

$$E(\bar{y}_{STRS}) = W_1\bar{Y}_1 + W_2\bar{Y}_2 + W_3\bar{Y}_3 + \ldots + W_k\bar{Y}_k = \sum_{i=1}^{k} W_i\bar{Y}_i = \bar{Y} \qquad \ldots (6.3)$$

This completes the proof of the theorem.

6.4.3 Estimation of Population Total

Sometimes, it is required to estimate the population total, given by

$$Y_{Total} = N\bar{Y} = \sum_{i=1}^{k} N_i\bar{Y}_i.$$

An immediate estimator of the parameter, Y_{Total}, therefore, can be considered to be

$$\hat{T}_{Total} = \sum_{i=1}^{k} N_i\bar{y}_i$$

which has similar structure as that of population total except that stratum mean is replaced by the sample mean of the sample selected from the i^{th} stratum. Now, let us discuss and prove the following theorem that the estimator $\hat{T}_{Total}$ is an unbiased estimator for population total:

Theorem 2: An unbiased estimator of population total, denoted by $\hat{T}_{Total}$, is given by $\sum_{i=1}^{k} N_i\bar{y}_i$ or $N\bar{y}_{STRS}$.

Proof: Due to the definition of population total, Y_{Total} and the result that

$$E(\bar{y}_i) = \bar{Y}_i,$$

we can see that

$$E(\hat{T}_{Total}) = E\left(\sum_{i=1}^{k} N_i\bar{y}_i\right) = \sum_{i=1}^{k} N_i E(\bar{y}_i) = \sum_{i=1}^{k} N_i\bar{Y}_i = Y_{Total}.$$

Hence the theorem.

Example 1: Given the following sample data:

Stratum	N_i	n_i	$\sum y_i$
I	110	20	4000
II	60	15	10500
III	30	10	60000

(i) Estimate $\bar{Y}_1$, $\bar{Y}_2$, $\bar{Y}_3$

(ii) Estimate $\bar{Y}$

Solution: It is known that sample mean is an estimator of corresponding population mean.

(i) We know that $\bar{y}_i = \frac{1}{n_i}\sum_{j}^{n_i} y_{ij}$. Therefore, estimator of $\bar{Y}_1$ will be

$$\bar{y}_1 = \frac{1}{20} \times 4000 = 200$$

Similarly, we have estimates of $\bar{Y}_2$ and $\bar{Y}_3$, respectively,

$$\bar{y}_2 = \frac{1}{15} \times 10500 = 700 \quad \text{and} \quad \bar{y}_3 = \frac{1}{10} \times 60000 = 6000$$

Thus, the estimate of $\bar{Y}_i$ (i = 1, 2, 3) on the basis of given data is

$$\therefore \bar{Y}_1 = 200, \qquad \bar{Y}_2 = 700, \qquad \bar{Y}_3 = 6000.$$

(ii) We know that $\bar{y}_{STRS}$ is unbiased for $\bar{Y}$. Thus

$$\bar{y}_{STRS} = \frac{\sum N_i \bar{y}_i}{N} = \frac{110 \times 200 + 60 \times 700 + 30 \times 6000}{200}$$

$$= \frac{22000 + 42000 + 180000}{200} = \frac{244000}{200} = 1220$$

Therefore, estimated value of $\bar{Y}$ on the basis of the given data is 1220.

Try to answer the following Self-Assessment Question, with the help of contents of this section:

SAQ 3

Define the estimator for population mean of a stratified population and show that it is an unbiased estimator.

6.5 SAMPLING VARIANCE OF THE SAMPLE MEAN ESTIMATOR

Obviously, being an estimator, $\bar{y}_{STRS}$ varies over sample to sample and, hence, it would have some variability which can be measured with the help of its variance. It is, therefore, our next step is to deduce an expression of the variance of the estimator of population mean, that is, to find the expression of $V(\bar{y}_{STRS})$.

6.5.1 Variance of the Estimator

Let us consider the following theorem which provides the expression of the sampling variance of the estimator $\bar{y}_{STRS}$:

Theorem 3: The sampling variance of the estimator $\bar{y}_{STRS}$ under Stratified Random Sampling Scheme is given by:

$$V(\bar{y}_{STRS}) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i} \right) W_i^2 S_i^2 \qquad \ldots (6.4)$$

Proof: We know that the i[th] stratum (i = 1, 2, 3,…, k), in fact, is treated to be a population of size N_i from which a Simple Random Sampling without Replacement sample of size n_i is selected. Therefore, sample mean estimators, $\bar{y}_i$s are random variables. Moreover, since strata are non-overlapping populations from which samples of specific sizes are selected independently, $\bar{y}_i$s are independent random variables.

We have already shown in the Theorem 8 under Sub-section 2.5 of Unit 2 that the sampling variance of the sample mean estimator $\bar{y}$ under Simple Random Sampling without Replacement scheme is given by

$$V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 \qquad \text{… (6.5)}$$

where N and n are respectively population and sample sizes and S^2 is the population mean square.

Since, in Stratified Random Sampling scheme, instead of one population, we have k sub-populations which are independent to each other, therefore, using the result (6.5), for the i^{th} stratum the sampling variance of the estimator $\bar{y}_i$ can be written as

$$V(\bar{y}_i) = \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2 \qquad \text{… (6.6)}$$

since, for the i^{th} stratum, the population and sample sizes are respectively N_i and n_i and the population mean square is S_i^2.

We know that

$$\bar{y}_{STRS} = \sum_{i=1}^{k} W_i \bar{y}_i,$$

where W_i's are constants and $\bar{y}_i$s are independent random variables.

Therefore, we have

$$V(\bar{y}_{STRS}) = V\left(\sum_{i=1}^{k} W_i \bar{y}_i\right) = \sum_{i=1}^{k} W_i^2 V(\bar{y}_i) \qquad \text{… (6.7)}$$

Using the additive theorem of variance of independent random variables; given by

$$V\left(\sum_{L=1}^{r} a_L X_L\right) = a_1^2 V(X_1) + a_2^2 V(X_2) + \ldots + a_r^2 V(X_r),$$

for constants $a_1, a_2, \ldots, a_r$ and independent random variables $X_1, X_2, \ldots, X_r$.

Now, applying the result (6.6) in the expression (6.7), we have

$$V(\bar{y}_{STRS}) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) W_i^2 S_i^2$$

Hence the theorem.

Example 2: A population is divided into two strata and the units in each stratum are given below:

Strata	
Higher Income Group	**Lower Income Group**
3, 4, 6, 11	1, 3, 6, 2

(i) Find the mean of each stratum.

(ii) Find the mean square of each stratum.

Solution: Let us calculate the mean and mean square of each stratum as:

(i) Mean of the I[st] and II[nd] strata:

$$\bar{Y}_1 = \frac{3+4+6+11}{4} = \frac{24}{4} = 6$$

$$\bar{Y}_2 = \frac{1+3+6+2}{4} = \frac{12}{4} = 3$$

(ii) Mean square of the I[st] and II[nd] strata:

$$S_1 = \frac{1}{(N_1-1)}\sum_{j=1}^{N_1}\left(Y_{1j} - \bar{Y}_1\right)^2$$

$$= \frac{1}{3}\left[(3-6)^2 + (4-6)^2 + (6-6)^2 + (11-6)^2\right]$$

$$= \frac{1}{3}\left[9+4+0+25\right] = \frac{38}{3} = 12.67$$

$$S_2 = \frac{1}{(N_2-1)}\sum_{j=1}^{N_2}\left(Y_{2j} - \bar{Y}_2\right)^2$$

$$= \frac{1}{3}\left[(1-3)^2 + (3-3)^2 + (6-3)^2 + (2-3)^2\right]$$

$$= \frac{1}{3}\left[4+0+9+1\right] = \frac{14}{3} = 4.67$$

Example 3: From the following data find the stratified sample mean $\bar{y}_{STRS}$ and its variance $Var(\bar{y}_{STRS})$.

Strata	N_i	n_i	S_i	$\bar{y}_i$
1	80	29	12	80
2	160	39	8	30
3	260	32	4	10
	500	100		

Solution: For calculating the mean, we prepare the following table:

Strata	N_i	n_i	S_i	N_iS_i	$\bar{y}_i$	$N_i\bar{y}_i$
1	80	29	12	960	80	6400
2	160	39	8	1280	30	4800
3	260	32	4	1040	10	2600
Total	500	100		3280		13800

$$\bar{y}_{STRS} = \frac{\sum_{i=1}^{3} N_i\bar{y}_i}{N} = \frac{13800}{500} = 27.6$$

For calculating the variance, we prepare the following table:

Strata	N_iS_i	$(N_iS_i)^2$	$(N_iS_i)^2/n_i$	$(N_i-n_i)/N_i$	$\frac{(N_iS_i)^2}{n_i} \times \frac{(N_i-n_i)}{N_i}$
1	960	$(960)^2$	$(960)^2/29$	51/80	20259.31
2	1280	$(1280)^2$	$(1280)^2/39$	121/160	31770.26
3	1040	$(1040)^2$	$(1040)^2/32$	228/260	29640.00
Total	500	100			81669.57

$$\text{Var}(\bar{y}_{STRS}) = \frac{1}{N^2}\sum_{i=1}^{k}\frac{(N_i - n_i}{N_i} \times \frac{(N_i S_i)^2}{n_i} = \left(\frac{1}{500}\right)^2 (81669.57) = 0.32$$

Remark 6.4: The expression (6.4) can be written as follows:

$$V(\bar{y}_{STRS}) = \left(\frac{1}{n_1} - \frac{1}{N_1}\right)W_1^2 S_1^2 + \left(\frac{1}{n_2} - \frac{1}{N_2}\right)W_2^2 S_2^2 + \ldots + \left(\frac{1}{n_k} - \frac{1}{N_k}\right)W_k^2 S_k^2$$

In this expression, sample sizes, n_is population sizes, N_is and weights, W_i^2s; being constants for each stratum are known, but population mean squares, S_i^2s are not known. Further, we observe that the sampling variance of the estimator, $V(\bar{y}_{STRS})$, is directly proportional to all the mean squares, S_i^2s, that is, $V(\bar{y}_{STRS}) \propto S_i^2$ for all i.

This is an indication that the variability of the estimator is possible to reduce to certain extent by making each stratum internally homogeneous as far as possible, subject to the given constraints of the survey, such as cost of the survey, time allotted for the survey, employment of trained and skilled personnel, etc. In other words, if the units of the population, possessing almost similar nature in respect of the study characteristic, are put in the same stratum; it could be possible to make the estimator $\bar{y}_{STRS}$ more and more efficient. This is a criterion which is generally followed by the sampler while he/she forms different strata himself/herself on the basis of certain characteristics. We have already explained this criterion with the help of a real example (Example (ii) in the Sub-section 6.2.2(b)).

However, if the strata are made internally homogeneous as far as possible, these would be more and more heterogeneous among themselves; since, the total variability of the given population cannot be changed and, hence, reducing the variability within each stratum, obviously we make strata which are more and more heterogeneous among themselves with respect to the study characteristic.

In view of these discussions, we can say that

"Strata should be constructed in such a way that each stratum is, as far as possible, internally homogeneous in respect to the study variable which automatically will increase the heterogeneity between strata."

Remark 6.5: We have seen that the estimator for the population total is given by $\hat{T}_{Total} = \sum_{i=1}^{k} N_i \bar{y}_i$.

Therefore, we have

$$V\left(\hat{T}_{Total}\right) = V\left(\sum_{i=1}^{k} N_i \bar{y}_i\right)$$

$$= \sum_{i=1}^{k} N_i^2 V(\bar{y}_i)$$

$$= \sum_{i=1}^{k} N_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2$$

$$= N^2 V(\bar{y}_{STRS}).$$

Remark 6.6: The expression of $V(\bar{y}_{STRS})$ is seen to be a function of unknown mean squares of all the strata. Therefore, similar to SRSWOR scheme, here also we need to derive an expression of the estimate of $V(\bar{y}_{STRS})$ on the basis of sample mean squares for samples drawn from each stratum. The next section deals with this problem.

6.6 ESTIMATING VARIANCE OF SAMPLE MEAN

We can obtain an expression for the estimator of the sampling variance on the basis of the sample in hand. Let us consider the following theorem for this purpose:

Theorem 4: An unbiased estimator of the sampling variance of $\bar{y}_{STRS}$, denoted as $\hat{V}(\bar{y}_{STRS})$, is given by

$$\hat{V}(\bar{y}_{STRS}) = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 s_i^2, \qquad \ldots (6.8)$$

where s_i^2 stands for the sample mean square of the i^{th} stratum (i = 1, 2,3,…, k).

Proof: Since, we know that sample mean square, s^2 is an immediate estimator of population mean square, S^2 in Simple Random Sampling without Replacement scheme; it suggests that the expression

$$\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 s_i^2$$

would be the immediate estimator of $V(\bar{y}_{STRS})$, keeping in view the structure of $V(\bar{y}_{STRS})$. Let us now show that this estimator of $V(\bar{y}_{STRS})$ is unbiased estimator.

We have $E\left[\hat{V}(\bar{y}_{STRS})\right] = E\left[\sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 s_i^2\right]$

$$= \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 E\left(s_i^2\right)$$

$$= \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 S_i^2 = V(\bar{y}_{STRS}).$$

This completes the proof of the theorem.

Now you may try to answer the following Self-Assessment Question:

SAQ 4

Obtain the sampling variance of the weighted estimator of population mean. Show that an unbiased estimator of the sampling variance is given by

$$\hat{V}(\bar{y}_{STRS}) = \sum_{i=1}^{k}\left(\frac{1}{n_i} - \frac{1}{N_i}\right)W_i^2 s_i^2;$$

where symbols have their usual meanings.

6.7 SUMMARY

In this Unit, we have discussed:

- The meaning of some important terms, like, 'stratification', 'stratum', 'strata' and a 'stratified population' in the context of Stratified Random Sampling Scheme.
- Some examples of such populations which are either available in the stratified form or re-stratified in a different way by the sampler on the basis of some other characteristic in view of the need for it in order to get better results.
- The method of selection of a stratified a random sample selected from a stratified population using Simple Random Sampling without Replacement scheme.
- The problem of estimation of some population parameters utilizing the Stratified Random sample.
- Suitable estimators for population mean and population total along with their salient properties.
- The sampling variance of the estimator of population mean and an important criterion of stratification process.
- An unbiased estimator of the variance in order to estimate the sampling variance of the sample mean estimator from the selected sample.

6.8 TERMINAL QUESTIONS

1. In a school, admissions of some students were taken for the standards to which they have to be admitted. At the time of admission, each candidate was allotted a token number as well as an alphabet signifying a particular characteristic of the student. Stratify them according to their standards irrespective of the characteristic they possess and find the number of strata as well as the size of each stratum:

Token No.	Standard	Token No.	Standard	Token No.	Standard
1	I (a)	12	I (e)	25	IV (b)
2	IV (c)	13	III (c)	26	I (a)
3	III (b)	14	III (b)	28	III (a)
4	I (c)	15	IV (a)	30	I (c)
5	II (a)	16	III (d)	29	II (d)
6	I (e)	17	III (a)	32	I (b)
7	I (b)	18	II (c)	33	IV (a)
8	III (c)	19	IV (a)	34	III (e)
9	III (a)	20	III (e)	35	I (e)
10	II (c)	21	II (a)	36	II (a)
11	II (a)	22	I (d)	37	IV (d)

2. Mention the reasons for stratification of a finite population. Explain with an example what do you mean by geographical region-wise stratification. For what reasons, sometimes stratification is done purposefully on the basis of some other characteristic instead of geographical regions. Give an example.

3. For the data given in **1** given above, stratify the population of students according to the particular characteristic they possess. Hence, find the number of strata and the size of each stratum.

4. Consider the following stratified population:

Stratum (i)	I	II	III	IV	V	VI
Stratum Size (N_i)	16	8	40	12	20	24

Let a sample of size 42 is to be selected from the population. In the absence of any specific rule for deciding the sample size to be selected from each stratum, the researcher decides to select samples of equal sizes. What would be the sizes of samples to be selected from each stratum?

5. Consider the following stratified population:

Stratum (i)	I	II	III	IV	V	VI
Stratum Size (N_i)	12	18	20	25	30	40
Stratum Mean ($\bar{Y}_i$)	35.5	16.5	24.0	28.2	14.0	12.7

Let a sample of size 42 is selected from the population such that sample means for the sample selected from strata I, II, III, IV, V and VI are found to be 22.2, 15.4, 30.0, 20.8, 15.5 and 12.4 respectively. What is the value of overall population mean? Find an unbiased estimator of it.

6. Consider the following stratified population:

Stratum (i)	I	II	III	IV	V	VI
Stratum Size (N_i)	12	18	20	25	30	40
Stratum Mean ($\bar{Y}_i$)	35.5	16.5	24.0	28.2	14.0	12.7
Stratum Mean Square (S_i^2)	5.3	2.0	10.5	18.8	22.0	19.6

Find the sampling variance of the estimator of population mean for a sample of size 42, when the sample size for the stratum I to VI are all equal.

6.9 ANSWERS / SOLUTIONS

Self-Assessment Questions (SAQs)

1. **Hint:** For answering the question, you may go through the content of Sub-section 6.2.1 of Section 6.2. Also see the (a) and (b) parts of the Sub-section 6.2.1.

2. **Hint:** Yes, agree with the statement. The reason for accepting the statement is that the process of selecting separate and independent samples of pre-fixed sizes from each stratum one uses the Simple Random Sampling without Replacement scheme. However, in Simple Random Sampling without Replacement scheme, we select only one sample from the given population, but in Stratified Random Sampling scheme, we have a number of sub-populations after stratifying the given population into a number of strata (sub-populations), we are forced to select a number of independent samples, drawn one from each stratum, in order to constitute the required sample. Thus, in fact, Stratified Random Sampling scheme is nothing, but Simple Random Sampling without Replacement scheme repeated a number of times independently for all the strata. Thus, Stratified Random Sampling scheme is an extension of the Simple Random Sampling without Replacement scheme applied to a number of populations.

3. **Hint:** For the answer to this question, you are referred to Sub-section 6.4.2.

4. **Hint:** For the answer to the question, you are referred to the Theorem 3 under Sub-section 6.5.1 and Theorem 6 under Section 6.6.

Terminal Questions (TQs)

1. In order to find the answer of the numerical part, we have to divide the students into a number of standards; standards being the different strata. We shall represent each student by the token number allotted to him/her. Since, there are four standards I, II, III and IV, we have in total 4 strata which are distinct and non-overlapping. On the basis of token number of each student and the standard for which they are admitted, we have the total population of students stratified in four strata as follows:

Strata (Standard)	Token Numbers	Number of Students
I	10, 13, 15, 16, 20, 3, 26, 30, 32, 35	10
II	14, 19, 1, 2, 9, 29, 36	7
III	12, 17, 18, 4, 5, 7, 8, 22, 28, 34	10
IV	11, 6, 21, 25, 33, 37	6

Thus, there are 4 strata and sizes of the strata are: I – 10; II – 7, III – 10 and IV – 6. The population size is equal to 10+7+10+6 = 33.

2. **Hint:** For the answer of the first part of the question, you are referred to sub-section 6.2.1 and particularly to 6.2.1(c). For answering the second part, you are suggested to go through the sub-section 6.2.2(a). You are referred to sub-section 6.2.2(b) for the answer of the last part.

3. The division of students into different strata according to possession of particular characteristic is as follows:

Strata (Presence Of The Characteristic)	Token Numbers	Number of Students
(a)	10, 14, 18, 19, 2, 6, 8, 21, 26, 28, 33, 36	12
(b)	12, 16, 5, 25, 32	5
(c)	11, 13, 17, 1, 4, 9, 30	7
(d)	20, 7, 29, 37	4
(e)	15, 3, 22, 34, 35	5

Thus, we have 5 strata each for five levels of the characteristic used for stratification. Moreover, the sizes of the stratum: (a) are 12; (b) are 5; (c) are 7; (d) are (4) and (e) are 5.

4. The data given in the exercise are given as follows:

There are 6 strata whose sizes (N_i) are also given. We have to select a sample of size 42 from this population with selecting equal size samples from each stratum.

Stratum (i)	I	II	III	IV	V	VI
Stratum Size (N_i)	16	8	40	12	20	24

Since the sample size is 42 and there are 6 strata, therefore, the sample size to be selected from each stratum will be 42/6 = 7. Thus, we have the following table showing the sample size to be selected from each stratum:

Stratum (i)	I	II	III	IV	V	VI
Sample Size (n_i)	7	7	7	7	7	7

5. Given the data in the exercise, we want to compute the overall population mean, which is given by, $\bar{Y} = \sum_{i=1}^{k} W_i \bar{Y}_i$.

So, using the table, we get the overall population mean

$$\bar{Y} = \left(\frac{12}{145}\right) \times 35.5 + \left(\frac{18}{145}\right) \times 16.5 + \left(\frac{20}{145}\right) \times 24 + \left(\frac{25}{145}\right) \times 28.2$$

$$+ \left(\frac{30}{145}\right) \times 14 + \left(\frac{40}{145}\right) \times 12.7 = 19.558.$$

We know that an unbiased estimator of $\bar{Y}$ is given by

$$\bar{y}_{STRS} = \sum_{i=1}^{6} W_i \bar{y}_i ,$$

where $\bar{y}_i$s are sample means of sample selected from respective strata. Using values of W_i and given values of sample means, $\bar{y}_i$, we have

$$\bar{y}_{STRS} = \left(\frac{12}{145}\right) \times 22.2 + \left(\frac{18}{145}\right) \times 15.4 + \left(\frac{20}{145}\right) \times 30.0$$

$$+ \left(\frac{25}{145}\right) \times 20.8 + \left(\frac{30}{145}\right) \times 15.5 + \left(\frac{40}{145}\right) \times 12.4 = 18.101.$$

We observe that the estimated value is quite close to the population mean. In this sense, the estimate is highly efficient.

6. Given the data related to 6 strata, we have N_i's, $\bar{Y}_i$s and S_i^2s. We are required to compute the value of the sampling variance of the estimator $\bar{y}_{STRS}$. We know that it is given by

$$V\left(\bar{y}_{STRS}\right) = \sum_{i=1}^{k} \left(\frac{1}{n_i} - \frac{1}{N_i}\right) W_i^2 S_i^2.$$

In order to compute it, therefore, we need to compute the values of W_i^2 first. We present the needed values in the following table:

Stratum (i)	I	II	III	IV	V	VI
Weight ($W_i = N_i / N$)	12/145 =0.083	18/145 =0.124	20/145 =0.138	25/145 =0.172	30/145 =0.207	40/145 =0.276
Stratum Mean ($\bar{Y}_i$)	35.5	16.5	24.0	28.2	14.0	12.7
Stratum Mean Square (s_i^2)	5.3	2.0	10.5	18.8	22.0	19.6
Sample Sizes (n_i) in Case (i)	7	7	7	7	7	7

Similarly for this case, we have $n_i = 7$ for i = 1, 2, 3, 4, 5, 6.

Then, we have

$$V(\bar{y}_{STRS}) = \left(\frac{1}{7}-\frac{1}{12}\right)(0.083)^2 \times 5.3 + \left(\frac{1}{7}-\frac{1}{18}\right)(0.124)^2 \times 2.0 +$$

$$\left(\frac{1}{7}-\frac{1}{20}\right)(0.138)^2 \times 10.5 + \left(\frac{1}{7}-\frac{1}{25}\right)(0.172)^2 \times 18.8 +$$

$$\left(\frac{1}{7}-\frac{1}{30}\right)(0.207)^2 \times 22.0 + \left(\frac{1}{7}-\frac{1}{40}\right)(0.276)^2 \times 19.6$$

$$= 0.0022 + 0.0027 + 0.0186 + 0.0572 + 0.1032 + 0.1760$$

$$= 0.3599.$$