These days, the Computers are programmed to generate random numbers. Such random numbers are called *pseudorandom numbers*.

14.2 PROBABILITY OR RANDOM SAMPLES

A sample is called a *random sample* if the probability of selection for each unit in the population is known prior to sample selection. The important kinds of random samples which differ in the manner in which the sampling units are selected, are discussed in the subsections that follow:

14.2.1 Simple Random Sample. A sample is defined to be a *simple random sample* (SRS) if it is selected in such a manner that (i) each unit in the population has an equal probability of being included in the sample and (ii) each possible sample of the same size has an equal probability of being the sample selected.

Suppose a finite population contains N units and a sample of n units is to be selected. If we sample with replacement, the number of all *possible* samples of size n that could be selected is N^n, as the first unit of the sample can be selected in N different ways, the second unit can also be selected in N ways and so on. When we sample without replacement, then the number of all possible samples when the *order* of the units is considered, is the number of *permutations* of n units from N, *i.e.* ${}^N P_n = N(N-1) \ldots (N-n+1)$. But in practical problems, we ignore the *order* in which the n units are drawn. Then the number of different samples of n units that can be selected when the *order* is disregarded, is the number of *combinations* of n units from a finite population of N units, *i.e.*, $\binom{N}{n} = \dfrac{N!}{n!\,(N-n)!}$. Thus there are $\binom{N}{n}$ samples that could be selected and these samples occur with equal probabilities.

As an illustration, suppose we wish to select random samples of size 2 from a population of say, 5 students, identified as A, B, C, D and E. If we sample *with replacement*, then there are $(5)^2 = 25$ possible samples, which are listed below:

AA	BA	CA	DA	EA
AB	BB	CB	DB	EB
AC	BC	CC	DC	EC
AD	BD	CD	DD	ED
AE	BE	CE	DE	EE

14.2.2. Stratified Random Sample. A sample of size n is de... to be a *stratified random sample* if it is selected from a population which has been divided into a number of non-overlapping groups or sub-populations, called *strata*, such that part of the sample is drawn at random from *each stratum*. Stated differently, let a heterogeneous population containing N units be divided into k subpopulations or strata of sizes $N_1, N_2, ..., N_k$ in such a manner that all units in each stratum are believed to be very similar with respect to the measurements of interest. Then a *stratified random sample* of size n will be composed of the simple random samples of predetermined sizes $n_1, n_2, ..., n_k$ $(\sum n_i = n)$ drawn independently from the strata 1, 2, ..., k respectively. It is to be emphasized that good stratification requires that each of these strata should be internally homogeneous but externally the strata should differ from one another.

A population may be stratified according to the size such as large, medium and small; to the administrative grouping such as provinces and districts; to the geographic area such as urban and rural; to the natural characteristics such as age-group, sex, family size, occupation, education, tastes of the consumers, etc. The advantages of stratified random sampling are low cost, greater accuracy and a better coverage. Stratified random sampling is used when (i) the variations among strata are greater than the variations within strata, (ii) information about some parts of the population is desired.

The purpose of stratified random sampling is three-fold. *Firstly*, the strata obtained by subdividing the heterogeneous population into homogeneous groups, adequately represent the population so the information concerning individual stratum is gathered. *Secondly*, it provides improved estimates of the population characteristics. *Thirdly*, it reduces the variance of the estimator.

Allocation of Sample Sizes. By *allocation* of a sample we mean the way the total sample size n is distributed among the various strata into which the population has been divided. Four methods of allocating the sample numbers are available. They are:

(a) *Equal Allocation.* The allocation is called *equal* when from each stratum, equal number of sampling units is selected. That is the total sample size n is distributed equally among all the k strata. Thus the stratum sample size n_i for equal allocation is

$$n_i = \frac{n}{k}, \quad \text{for } i = 1, 2, ..., k$$

This is the simplest method of allocation.

(b) *Proportional Allocation.* The allocation is said to be *proportional* when the total sample size n is distributed among the different strata in proportion to the sizes of strata. In other words, the allocation is proportional if

$$n_i = n \cdot \frac{N_i}{N}, \text{ for } i = 1, 2, \ldots, k$$

where N_i is the population size of the *ith* stratum, n_i is the *ith* stratum sample size and N is the total size of the population. A sample of size n_i from stratum i is drawn by random numbers and investigated. This way of allocation is the next simplest method and it is the most frequently used method. The advantage of proportional allocation is that it does not require information either on the stratum variance or on the costs of sampling units in different strata.

Example 14.5. Suppose a population of $N=9$ is stratified into 3 strata with the following measurements:

Stratum I	$X_{11}=1, X_{12}=2, X_{13}=4$
Stratum II	$X_{21}=6, X_{22}=8$
Stratum III	$X_{31}=11, X_{32}=15, X_{33}=16, X_{34}=19$

If two measurements are drawn from each stratum for the sample, state how many samples of size 6 could be chosen from this population? List these samples and compute the mean for each sample.

Here the population consists of 3 strata and from each stratum 2 units are to be selected to make up a sample of $n=6$. Assuming the sampling without replacement, we can choose $\binom{3}{2}=3$ possible subsamples from the first stratum, $\binom{2}{2}=1$ possible subsample from the second stratum and $\binom{4}{2}=6$ possible subsamples from the third stratum. Each of the possible subsamples from stratum I is to be associated with the subsample from stratum II and then these combinations are further associated with each of the possible subsamples from stratum III. Hence, in all there are 3 × 1 × 6, *i.e.* 18 possible samples of size 6 with 2 measurements from each stratum. The first sample consists of (1, 2, 6, 8, 11, 15) of which (1, 2) is from stratum I, (6, 8) is from stratum II and (11, 15) is from stratum III. The 18 possible samples are listed below and the sample means appear in the last column.

Sample No.	Sample data from stratum			Total	Sample Means
	I	II	III		
1	1, 2	6, 8	11, 15	43	7.17
2	1, 2	6, 8	11, 16	44	7.33
3	1, 2	6, 8	11, 19	47	7.83
4	1, 2	6, 8	15, 16	48	8.00
5	1, 2	6, 8	15, 19	51	8.50
6	1, 2	6, 8	16, 19	52	8.67
7	1, 4	6, 8	11, 15	45	7.50
8			11, 16	46	7.67
9			11, 19	49	8.17
10			15, 16	50	8.33
11			15, 19	53	8.83
12			16, 19	54	9.00
13	2, 4	6, 8	11, 15	46	7.67
14			11, 16	47	7.83
15			11, 19	50	8.33
16			15, 16	51	8.50
17			15, 19	54	9.00
18			16, 19	55	9.17

Example 14.6. Select a stratified random sample of size $n = 5$ by proportional allocation from the following population. Find the sample mean and the estimate of the population mean.

Stratum I	$X_{11} = 12, X_{12} = 14, X_{13} = 13, X_{14} = 8.$
Stratum II	$X_{21} = 25, X_{22} = 30, X_{23} = 40, X_{24} = 35, X_{25} = 24, X_{26} = 28$

To select a stratified random sample of $n = 5$ by using proportional allocation, i.e. $n_i = n \cdot \frac{N_i}{N}$, the sample size is allocated as

$$n_1 = n \cdot \frac{N_1}{N} = 5 \times \frac{4}{10} = 2, \text{ and}$$

$$n_2 = n \cdot \frac{N_2}{N} = 5 \times \frac{6}{10} = 3.$$

Using a table of random numbers, we select the following subsamples from

Stratum I: $X_{11} = 12,\ X_{12} = 14,$

Stratum II: $X_{22} = 30,\ X_{23} = 40,\ X_{25} = 24.$

Now the sample mean of sample n is

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i \bar{x}_i = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}$$

$$= \frac{1}{5}[12 + 14 + 30 + 40 + 24] = \frac{120}{5} = 24$$

An *estimator* (a term to be defined later) of the population mean by the stratified sampling, denoted by $\bar{X}_{st}$, is given by

$$\bar{X}_{st} = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{X}_i = \sum_{i=1}^{k} W_i \bar{x}_i,$$ where $\bar{x}_i$ is the *ith* sub-sample mean and $W_i (= N_i/N)$ is termed the *weight* of the *ith* stratum.

Hence the *estimate* of the population mean is

$$\bar{x}_{st} = \frac{N_1}{N}\bar{x}_1 + \frac{N_2}{N} \cdot \bar{x}_2 \text{ where } \bar{x}_1 = \frac{12 + 14}{2} = 13, \text{ etc.}$$

$$= \frac{4}{10}(13) + \frac{6}{10}\left(\frac{94}{3}\right) = \frac{240}{10} = 24.$$