14.2.3 Systematic Random Sample. A sample of size n is defined to be a *systematic random sample* if it is obtained by choosing one unit at random from the first k units and thereafter selecting every *kth* unit after the N units in the population have been serially numbered from 1 to N or arranged in a systematic fashion. The letter k, called the *sampling interval*, stands for some integer nearest to $\frac{N}{n} = \frac{\text{Population size}}{\text{Sample size}}$ and the sample is generally expressed by saying "a 1 in k sample". For example, when the number of the unit selected at random from the first k units is i ($i = 1, 2, ..., k$), the systematic sample of size n will contain the units with numbers i, $i + k$, $i + 2k$, ..., $i + (n - 1)\,k$. Suppose that $i=7$ and k = 20, then a systematic sample consisting of every 20th unit, will be composed of the units numbered 7, 27, 47, 67 and so on. Thus in systematic sampling, sampling units are selected at uniform interval after a random start.

When the sampling interval corresponds to some periodic or cyclic characteristic in the population, the systematic sampling will result in a non-representative sample. For example, suppose every 20th shop in a big bazaar is a corner shop and the sampling interval is also 20. If the random start coincides with a corner shop, then the sample will include all the corner shops, and the sample will be highly non-representative as different characteristics (say, dealing in beverages) are associated with the corner shops. One way of avoiding of this sort of periodicity is to take a *fractional* sampling interval and then to round off the numbers obtained, *e.g.* if sampling interval is 20.7 and we start at 7, then the subsequent numbers are 27.7, 48.4, 69.1, 89.8, etc. and the rounding off results in 28, 48, 69, 90, etc. The sample will then consist of the units with serial numbers 7, 28, 48, 69, 90, etc.

If we think of the population as being divided into n strata, each consisting of k units, then a systematic sample resembles a stratified random sample with one unit from each stratum. The advantages of systematic sampling are that it saves much time and effort, it is economical, it is easily selected and conveniently worked out. Furthermore, a systematic sample is more representative of the population sampled as the sample is more evenly distributed across the population.

14.2.4. Cluster Sample. A random sample is said to be a *cluster sample* if it consists of *first* selecting at random groups of individual units, called *clusters* (treated as sampling units) into which a population can be divided and *then* including in the sample either all the units from each of the chosen clusters, or selecting a random sample of the units which the cluster comprises. In other words, suppose that a population is first divided into M smaller groups (equal or unequal in size), called *clusters*, such as the blocks of the cities, households, classes, etc. A random sample consisting of a number of clusters (say, m) is then selected from all these clusters, where each chosen cluster is either subsampled or all the subunits in the sampled clusters are included in the sample. Such a sample is called a *cluster sample*. Cluster sampling requires that the clusters should be as internally dissimilar as possible and different clusters should be very similar.

The procedure is called *one-stage cluster sampling* when all the units which each of the sampled clusters comprises, are included. If each of the sampled clusters is subsampled, then the sampling plan is called *two-stage cluster sampling* or *subsampling*. The plan is called *multistage cluster sampling* when more than two stages are involved in taking the sample. When the clusters relate to geographical areas, the sampling is known as *area sampling*.

As each cluster is treated as a single sampling unit in the selection process, the clusters are therefore called the *primary sampling units* (*psu*), while the subunits composing a cluster, are called the *secondary sampling units* (*ssu*).

The advantage of cluster sampling is savings in cost and time, *i.e.* the cost of sample selection and travel expenses of interviewers are considerably low. Cluster sampling is used when (i) the sampling frames of adequate coverage are not available, (ii) the variations among clusters are smaller than the variations within clusters. It is to be noted that cluster sampling is mostly used in statistical quality control.

14.2.5. Multistage Sample. A sample is called a *multistage* sample when it is selected in stages, the sampling units at each stage being subsampled from the larger units selected at the previous stage. Here a population is divided into a number of units, called *first-stage units*, which are subsampled. Each of the selected second stage units, is further divided into *third-stage units*, from which a subsample is again selected and so on. For example, in a sample survey, we select a random sample of n_1 districts (first-stage units), then we take a subsample of n_2 villages (second-stage units) from each of the selected districts, again, we select a sub-sample of n_3 households (third stage units) from each of the selected villages and so on. In a multistage sample, the sample size is the

number of units included in the sample at the final stage in the sampling. The advantage of multistage sampling is that only the parts of the population chosen at any stage need to be listed for sampling at the next stage. This technique is commonly used in large-scale surveys.

14.2.6. Multi-phase Sample. A sample is said to be a *two-phase sample* when certain items of information are collected from all the units in a sample and other items of usually more detailed information are collected from a subsample of the units composing the original sample. Two -phase sampling, also called *Double Sampling*, was proposed in 1938 for the first time by Neyman. When more phases are added, it becomes *multi-phase sampling*. It is important to note that in multi-phase sampling, the *same* units are used at each phase, whereas in multi-stage sampling, the units are different at different stages of sampling. The advantages of multi-phase sampling are that it can prove cheaper and it reduces the burden on respondents.

14.2.7. Sequential Sampling. This is another method of sampling where the sample size is not fixed in advance but sampling units are drawn on by one or in lots, and the decision is based on a definite rule relating to the sampling units themselves. That is, we draw one unit at a time and after each drawing, we make a decision whether to accept the lot or group, whether to reject it or whether to continue sampling. A graphic or tabular procedure is generally used to find when sampling should terminate. This technique was developed during World War II by Abraham Wald (1902–50) and is good for reaching decisions rapidly.