Thus, eight single hills per plot should be measured to satisfy the requirement that the estimate of the treatment mean would be within 5% of the true value 95% of the time.

15.1.3 Sampling Design

A *sampling design* specifies the manner in which the *n* sampling units are to be selected from the whole plot. There are five commonly used sampling designs in replicated field trials: simple random sampling, multistage random sampling, stratified random sampling, stratified multistage random sampling, and subsampling with an auxiliary variable.

15.1.3.1 Simple Random Sampling. In a simple random sampling design, there is only one type of sampling unit and, hence, the sample size (*n*) refers to the total number of sampling units to be selected from each plot consisting of *N* units. The selection of the *n* sampling units is done in such a way that each of the *N* units in the plot is given the same chance of being selected. In plot sampling, two of the most commonly used random procedures for selecting *n* sampling units per plot are the *random-number technique* and the *random-pair technique.*

15.1.3.1.1 The Random-Number Technique. The random-number technique is most useful when the plot can be divided into *N* distinct sampling units, such as *N* single-plant sampling units or *N* single-hill sampling units. We illustrate the steps in applying the random-number technique with a maize variety trial where plant height in each plot consisting of 200 distinct hills is to be measured from a simple random sample of six single-hill sampling units.

□ STEP 1. Divide the plot into *N* distinctly differentiable sampling units (e.g., *N* hills/plot if the sampling unit is a single hill or *N* 1 × 1 cm sub-areas per plot if the sampling unit is a 1 × 1 cm area) and assign a number from 1 to *N* to each sampling unit in the plot.

For our example, because the sampling unit is a single hill, the plot is divided into $N = 200$ hills, each of which is assigned a unique number from 1 to 200.

□ STEP 2. Randomly select *n* distinctly different numbers, each within the range of 1 to *N*, following a randomization scheme described in Chapter 2, Section 2.1.1.

For our example, $n = 6$ random numbers (each within the range of 1 to 200) are selected from the table of random numbers, following the procedure described in Chapter 2, Section 2.1.1. The six random numbers selected may be:

Sequence	Random Number
1	78
2	17
3	3
4	173
5	133
6	98

☐ STEP 3. Use, as the sample, all the sampling units whose assigned numbers (step 1) correspond to the random numbers selected in step 2. For our example, the six hills in the plot whose assigned numbers are 78, 17, 3, 173, 133, and 98 are used as the sample.

15.1.3.1.2 The Random-Pair Technique. The random-pair technique is applicable whether or not the plot can be divided uniquely into N sampling units. Hence, the technique is more widely used than the random-number technique. We illustrate the procedure with two cases—one where the plot can be divided into N distinct sampling units and another where clear division cannot be done.

Case I is one with clear division of N sampling units per plot. For illustration, we use the example in Section 15.1.3.1.1. Assuming that the plot consists of 10 rows and 20 hills per row ($N = 200$ hills), the steps involved in applying the random-pair technique to select a random sample of $n = 6$ single-hill sampling units are:

☐ STEP 1. Determine the width (W) and the length (L) of the plot in terms of the sampling unit specified, such that $W \times L = N$. For our example, the sampling unit is a single hill; and $W = 10$ rows, $L = 20$ hills, and $N = (10)(20) = 200$.

☐ STEP 2. Select n random pairs of numbers, with the first number of each pair ranging from 1 to W and the second number ranging from 1 to L; where W and L are as defined in step 1.

For our example, $n = 6$ random pairs of numbers are selected by using the table-of-random-number procedure described in Chapter 2, Section 2.1.1, with the restrictions that the first number of the pair must not exceed W (i.e., it must be within the range from 1 to 10) and the second number of the pair must not exceed L (i.e., it must be in the range from 1 to 20). The six random pairs of numbers may be as follows:

7, 6
6, 20
2, 3
3, 9
9, 15
1, 10

□ STEP 3. Use the point of intersection of each random pair of numbers, derived in step 2, to represent each selected sampling unit. For our example, the first selected sampling unit is the sixth hill in the seventh row, the second selected sampling unit is the twentieth hill in the sixth row, and so on. The location of the six selected single-hill sampling units in the plot is shown in Figure 15.1.

Case II is one without clear division of N sampling units per plot. For illustration, consider a case where a sample of six 20 × 20-cm sampling units is to be selected at random from a broadcast-rice experimental plot measuring 4 × 5 m (after exclusion of border plants). The steps involved in applying the random-pair technique to select a random sample of $n = 6$ sampling units are:

□ STEP 1. Specify the width (W) and length (L) of the plot using the same measurement unit as that of the sampling unit. For our example, the centimeter is used as the measurement unit because the sampling unit is defined in that scale. Thus, the plot width (W) and length (L) are specified as 400 cm and 500 cm. Note that with this specification, the division of the plot into N distinct sampling units cannot be made.

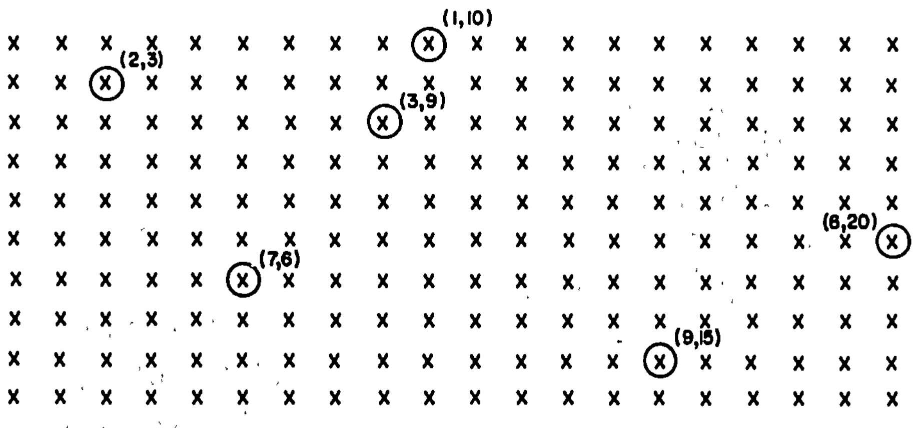

Figure 15.1 The location of six randomly selected sample hills, using the random-pair technique, for a plot consisting of 10 rows and 20 hills per row.

□ STEP 2. Select n random pairs of numbers, following the table-of-random-number procedure described in Chapter 2, Section 2.1.1, with the first number of the pair lying between 1 and W and the second number lying between 1 and L.

For our example, the six random pairs of numbers may be:

253, 74
92, 187
178, 167
397, 394
186, 371
313, 228

□ STEP 3. Use the point of intersection of each of the random pairs of numbers (derived in step 2) to represent the starting point of each selected sampling unit. For our example, we consider the starting point to be the uppermost left corner of each sampling unit. Thus, with the first random pair of (253, 74) the first selected sampling unit is the 20 × 20-cm area whose uppermost left corner is at the intersection of the 253 cm along the width of the plot and the 74 cm along the length of the plot (see Figure 15.2). The rest of the selected sampling units can be identified in the similar manner. The locations of the six selected 20 × 20-cm sampling units in the plot is shown in Figure 15.2.

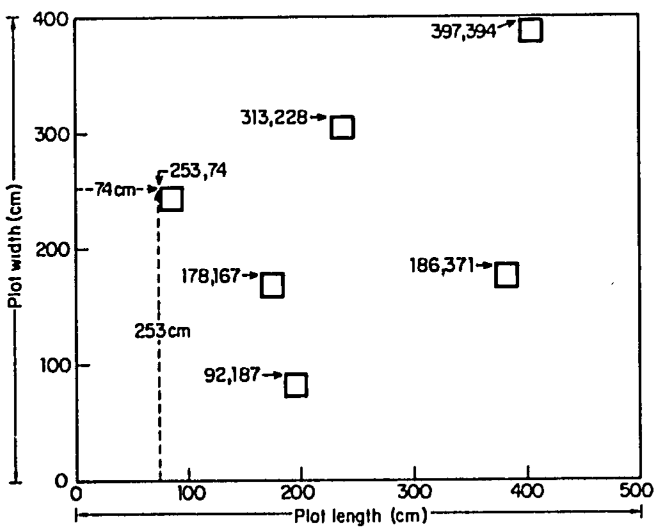

Figure 15.2 The location of six randomly selected 20 × 20-cm sampling units, using the random-pair technique for a plot measuring 4 × 5 m.

15.1.3.2 Multistage Random Sampling. In contrast to the simple random sampling design, where only one type of sampling unit is involved, the multistage random sampling design is characterized by a series of sampling stages. Each stage has its own unique sampling unit. This design is suited for cases where the best sampling unit is not the same as the measurement unit. For example, in a rice field experiment, the unit of measurement for panicle length is a panicle and that for leaf area is a leaf. The use of either the panicle or the leaf as the sampling unit, however, would require the counting and listing of all panicles or all leaves in the plot—a time-consuming task that would definitely not be practical.

In such cases, a single hill may still be used as the basic sampling unit. However, to avoid difficulty of measuring all leaves or all panicles in each sample hill, a multistage random sampling design could be used to identify a few sample leaves or sample panicles that need to be measured in each sample hill. Thus, such a design provides more than one type of sampling unit and, subsequently, more than one stage of sampling.

In the measurement of panicle length, for example, a two-stage sampling design with individual hills as the primary sampling unit and individual panicles as the secondary sampling unit can be employed. This would involve the application of a simple random sampling design twice—once to the primary sampling unit (the hill) and another to the secondary sampling unit (the panicle). To get this, a simple random sample of n_1 hills would first be taken from the plot (first-stage sampling) and a simple random sample of n_2 panicles would then be taken from each of the selected n_1 hills (second-stage sampling). This would result in $n = (n_1)(n_2)$ sample panicles per plot.

The extension of the multistage random sampling design to three, four, or more stages is straightforward. For example, in the case where leaf area is to be measured, a three-stage sampling design, with individual hills as the primary sampling unit, individual tillers as the secondary sampling unit, and individual leaves as the tertiary sampling unit, would be appropriate.

The selection of the sample is done separately and independently at each stage of sampling, starting with the first-stage sampling, then the second-stage sampling, and so on, in the proper sequence. At each sampling stage, the random selection procedure follows that of the simple random sampling design described in Section 15.1.3.1. For example, in the case of the two-stage sampling design for the measurement of the panicle length, the selection process starts with the random selection of n_1 single-hill sampling units from the plot. Then, for each of the n_1 sample hills, the total number of panicles is determined and the random-number technique is applied to select a random sample of n_2 panicles from the particular hill. This random selection process is repeated n_1 times, separately and independently for each of the n_1 sample hills, resulting in the total of $n = (n_1)(n_2)$ panicles, on which panicle length is to be measured.

15.1.3.3 Stratified Random Sampling. In a stratified random sampling design, sampling units within a plot are first grouped into k strata before a set of m sampling units is selected randomly from each stratum. Thus, the total number of sampling units per plot (n) is equal to $(m)(k)$.

The stratified random sampling design is useful where there is large variation between sampling units and where important sources of variability follow a consistent pattern. In such cases, the precision of the sample estimate can be improved by first grouping the sampling units into different strata in such a way that variability between sampling units within a stratum is smaller than that between sampling units from different strata. Some examples of stratification criterion used in agricultural experiments are:

- *Soil Fertility Pattern.* In an insecticide trial where blocking was based primarily on the direction of insect migration, known patterns of soil fertility cause substantial variability among plants in the same plot. In such a case, a stratified random sampling design may be used so that each plot is first divided into several strata based on the known fertility patterns and sample plants are then randomly selected from each stratum.
- *Stress Level.* In a varietal screening trial for tolerance for soil salinity, areas within the same plot may be stratified according to the salinity level before sample plants are randomly selected from each stratum.
- *Within-Plant Variation.* In a rice hill, panicles from the taller tillers are generally larger than those from the shorter ones. Hence, in measuring such yield components as panicle length or number of grains per panicle, panicles within a hill are stratified according to the relative height of the tillers before sample panicles are randomly selected from each position (or stratum).

It should be noted at this point that the stratification technique is similar to the blocking technique, described in Chapter 2, Section 2.2.1. It is effective only if it can ensure that the sampling units from the same stratum are more similar than those from different strata. Thus, the efficiency of the stratified random sampling design, relative to the simple random sampling design, will be high only if an appropriate stratification technique is used.

15.1.3.4 Stratified Multistage Random Sampling. When the stratification technique of Section 15.1.3.3 is combined with the multistage sampling technique of Section 15.1.3.2, the resulting design is known as stratified multistage random sampling. In it, multistage sampling is first applied and then stratification is used on one or more of the identified sampling stages.

For example, consider the case where a rice researcher wishes to measure the average number of grains per panicle through the use of a two-stage sampling design with individual hills in the plot as the primary sampling unit and individual panicles in a hill as the secondary sampling unit. He realizes that the number of grains per panicle varies greatly between the different panicles of the same hill. Hence, if the m panicles from each selected hill were selected entirely at random (i.e., a multistage random sampling design), the high variability between panicles within hill would cause the precision of the sample estimate to be low. A logical alternative is to apply the stratification technique by dividing the panicles in each selected hill (i.e., primary sampling unit) into k strata, based on their relative position in the hill, before a simple random sample of m panicles from each stratum is taken separately and independently for the k strata.

For example, if the panicles in each selected hill are divided into two strata based on the height of the respective tillers—the taller and shorter strata—and a random sample of three panicles taken from each stratum, the total number of sample panicles per plot would be $(2)(3)(a) = 6a$, where a is the total number of randomly selected hills per plot. In this case, the sampling technique is based on a two-stage sampling design with stratification applied on the secondary unit. Of course, instead of the secondary unit (panicles) the researcher could have stratified the primary unit (i.e., single-hill) based on any source of variation pertinent to his experiment (see also Section 15.1.3.3). In that case, the sampling technique would have been a two-stage sampling design with stratification of the primary unit. Or, the researcher could have applied both stratification criteria—one on the hills and another on the panicles—and the resulting sampling design would have been a two-stage sampling with stratification of both the primary and secondary units.

15.1.3.5 Subsampling With an Auxiliary Variable. The main features of a design for subsampling with an auxiliary variable are:

- In addition to the character of interest, say X, another character, say Z, which is closely associated with and is easier to measure than X, is chosen.
- Character Z is measured both on the main sampling unit and on the subunit, whereas variable X is measured only on the subunit. The subunit is smaller than the main sampling unit and is embedded in the main sampling unit.

This design is usually used when the character of interest, say X, is so variable that the large size of sampling unit or the large sample size required to achieve a reasonable degree of precision, or both, would be impractical. To improve the precision in the measurement of X, without unduly increasing either the sample size or the size of sampling unit, the subsampling with an auxiliary variable design can be used.

Improvement is achieved by measuring Z from a unit that is larger than the unit where X is measured. By using the known relationship between Z and X, it is as if X were measured from the large unit. With the proper choice of the auxiliary variable Z, a large increase in the degree of precision can be achieved with only a small increase in the cost of measuring Z from a larger unit. This means that Z must be chosen to best satisfy two conditions:

- Z must be closely associated with X and its relationship known.
- Measurement of Z must be with minimum cost.

For example, weed count is usually one of the characters of primary interest in evaluating the effect of weed infestation. Weed count is, however, highly variable and requires a relatively large sampling unit to attain a reasonable degree of accuracy. Furthermore, the task of counting weeds is tedious and time consuming, and its cost increases proportionally with the size of the sampling unit. On the other hand, weed weight, which is closely related to weed count, is simpler to measure and its measurement cost is only slightly affected by the size of the sampling unit. Thus, weed weight offers an ideal choice as the auxiliary variable for weed count.

To count weeds in a replicated field trial, the following sampling plan, based on the subsampling design with weed weight as the auxiliary variable, may be used:

- A sample of n 60 × 60-cm sampling units is randomly selected from each plot.
- From each of the n units, weed weight (Z) and weed count (X) is measured on a subsample (say 20 × 20-cm subunit) while the rest of the weeds in the main sampling unit is used only for measuring weed weight (Z).

15.1.4 Supplementary Techniques

So far, we have discussed sampling techniques for individual plots, each of which is treated independently and without reference to other plots in the same experiment. However, in a replicated field trial where the sampling technique is to be applied to each and all plots in the trial, a question usually raised is whether the same set of random sample can be repeated in all plots or whether different random processes are needed for different plots. And, when data of a plant character are measured more than once over time, the question is whether the measurements should be made on the same samples at all stages of observation or should rerandomization be applied.

The two techniques aimed at answering these questions are *block sampling* and *sampling for repeated measurements.*

15.1.4.1 Block Sampling. Block sampling is a technique in which all plots of the same block (i.e., replication) are subjected to the same randomization scheme (i.e., using the same sample locations in the plot) and different sampling schemes are applied separately and independently for different blocks. For example, for a RCB experiment, the total number of times that the randomization process is applied is the number of replications (r). Consider a case where the researcher wishes to measure panicle number in a RCB trial with eight treatments and four replications. He decides to use a simple random sampling design with the single-hill sampling unit and a sample size of six. With block sampling, he needs only to apply the randomization scheme four