CHAPTER 15

Sampling in Experimental Plots

Plot size for field experiments is usually selected to achieve a prescribed degree of precision for measurement of the character of primary interest. Because the character of primary interest—usually economic yield such as grain yield for grain crops, stover yield for forage crops, and cane yield for sugar cane—is usually the most difficult to measure, the plot size required is often larger than that needed to measure other characters. Thus, expense and time can be saved if the measurements of additional characters of interest are made by sampling a fraction of the whole plot. For example, make measurements for plant height from only 10 of the 200 plants in the plot; for tiller number, count only 1 m^2 of the 15 m^2 plot; and for leaf area, measure from only 20 of the approximately 2,000 leaves in the plot.

There are times, however, when the choice of plot size may be greatly influenced by the management practices used or the treatments tested. In an insecticide trial, for example, relatively large plots may be required to minimize the effect of spray drift or to reduce the insect movement caused by insecticide treatments in adjacent plots. In such cases, plot size would be larger than that otherwise required by the character of primary interest. Consequently, even for the primary character such as grain yield, it may still be desirable to sample from a fraction of the whole plot.

An appropriate sample is one that provides an estimate, or a sample value, that is as close as possible to the value that would have been obtained had all plants in the plot been measured—the plot value. The difference between the sample value and the plot value constitutes the *sampling error.*

Thus, a good sampling technique is one that gives a small sampling error. In this chapter we deal with the basic features of sampling technique as applied to replicated field trials (plot sampling) and the development of an appropriate sampling technique for a given field experiment.

15.1 COMPONENTS OF A PLOT SAMPLING TECHNIQUE

For plot sampling, each experimental plot is a *population. Population value*, which is the same as the plot value, is estimated from a few plants selected from each plot. The procedure for selecting the plants to be measured and used for estimating the plot value is called the *plot sampling technique*. To develop a plot sampling technique for the measurement of a character in a given trial, the researcher must clearly specify the sampling unit, the sample size, and the sampling design.

15.1.1 Sampling Unit

The *sampling unit* is the unit on which actual measurement is made. Where each plot is a population, the sampling unit must necessarily be smaller than a plot. Some commonly used sampling units in replicated field trials are a leaf, a plant, a group of plants, or a unit area. The appropriate sampling unit will differ among crops, among characters to be measured, and among cultural practices. Thus, in the development of a sampling technique, the choice of an appropriate sampling unit should be made to fit the requirements and specific conditions of the individual experiments.

The important features of an appropriate sampling unit are:

- *Ease of Identifications.* A sampling unit is easy to identify if its boundary with the surrounding units can be easily recognized. For example, a single hill is easy to identify in transplanted rice because each hill is equally spaced and is clearly separated from any of the surrounding hills. In contrast, plant spacing in broadcasted rice is not uniform and a single hill is, therefore, not always easy to identify. Consequently, a single hill may be suitable as a sampling unit for transplanted rice but not for broadcasted rice.
- *Ease of Measurement.* The measurement of the character of interest should be made easy by the choice of sampling unit. For example, in transplanted rice, counting tillers from a 2 × 2-hill sampling unit can be done quite easily and can be recorded by a single number. However, the measurement of plant height for the same sampling unit requires independent measurements of the four hills, the recording of four numbers and, finally, the computation of an average of those four numbers.
- *High Precision and Low Cost.* Precision is usually measured by the reciprocal of the variance of the sample estimate; while cost is primarily based on the time spent in making measurements of the sample. The smaller the variance, the more precise the estimate is; the faster the measurement process, the lower the cost is. To maintain a high degree of precision at a reasonable cost, the variability among sampling units within a plot should be kept small. For example, in transplanted rice, variation between single-hill sam-

pling units for tiller count is much larger than that for plant height. Hence, although a single-hill sampling unit may be appropriate for plant height, it may not be so for tiller count. Thus, for transplanted rice, the use of a 2×2-hill sampling unit for tiller count and a single-hill sampling unit for the measurement of plant height is common.

15.1.2 Sample Size

The number of sampling units taken from the population is *sample size*. In a replicated field trial where each plot is a population, sample size could be the number of plants per plot used for measuring plant height, the number of leaves per plot used for measuring leaf area, or the number of hills per plot used for counting tillers. The required sample size for a particular experiment is governed by:

- The size of the variability among sampling units within the same plot (sampling variance)
- The degree of precision desired for the character of interest

In practice, the size of the sampling variance for most plant characters is generally not known to the researcher. We describe procedures for obtaining estimates of such variances in Section 15.2. The desired level of precision can, however, be prescribed by the researcher based on experimental objective and previous experience. The usual practice is for the researcher to prescribe the desired level of precision in terms of the margin of error, either of the plot mean or of the treatment mean. For example, the researcher may prescribe that the sample estimate should not deviate from the true value by more than 5 or 10%.

With an estimate of the sampling variance, the required sample size can be determined based on the prescribed margin of error, of the plot mean, or of the treatment mean.

15.1.2.1 Margin of Error of the Plot Mean. The sample size for a simple random sampling design (Section 15.1.3.1) that can satisfy a prescribed margin of error of the plot mean is computed as:

$$n = \frac{(Z_\alpha^2)(v_s)}{(d^2)(\bar{X}^2)}$$

where n is the required sample size, Z_α is the value of the standardized normal variate corresponding to the level of significance α (the value Z_α can be obtained from Appendix B), v_s is the sampling variance, $\bar{X}$ is the mean value, and d is the margin of error expressed as a fraction of the plot mean.

For example, a researcher may wish to measure the number of panicles per hill in transplanted rice plots with a single hill as the sampling unit. Using data from previous experiments (see Section 15.2.1) he estimates the variance in panicle number between individual hills within the same plot (v_s) to be 5.0429 —or a *cv* of 28.4% based on the average number of panicles per hill of 17.8. He prescribes that the estimate of the plot mean should be within 8% of the true value. The sample size that can satisfy the foregoing requirement, at the 5% level of significance, can be computed as:

$$n = \frac{(1.96)^2(5.0429)}{(0.08)^2(17.8)^2}$$

$$= 9.6 \approx 10 \text{ hills/plot}$$

Thus, panicle number should be counted from 10 single-hill sampling units per plot to ensure that the sample estimate of the plot mean is within 8% of the true value 95% of the time.

15.1.2.2 Margin of Error of the Treatment Mean. The information of primary interest to the researcher is usually the *treatment mean* (the average over all plots receiving the same treatment) rather than the *plot mean* (the value from a single plot). Thus, the desired degree of precision is usually specified in terms of the margin of error of the treatment mean rather than of the plot mean. In such a case, sample size is computed as:

$$n = \frac{(Z_\alpha^2)(v_s)}{r(D^2)(\bar{X}^2) - (Z_\alpha^2)(v_p)}$$

where n is the required sample size, Z_α and v_s are as defined in the equation in Section 15.1.2.1, v_p is the variance between plots of the same treatment (i.e., experimental error), and D is the prescribed margin of error expressed as a fraction of the treatment mean. Take note that, in this case, additional information on the size of the experimental error (v_p) is needed to compute sample size.

To illustrate, consider the same example we used in Section 15.1.2.1. For an experiment with four replications, the researcher wishes to determine the sample size that can achieve an estimate of the treatment mean within 5% of the true value. Using an estimate of $v_p = 0.1964$ (see Section 15.2.1), sample size that can satisfy this requirement at the 5% level of significance can be computed as:

$$n = \frac{(1.96)^2(5.0429)}{4(0.05)^2(17.8)^2 - (1.96)^2(0.1964)}$$

$$= 8.03 \approx 8 \text{ hills/plot}$$

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.

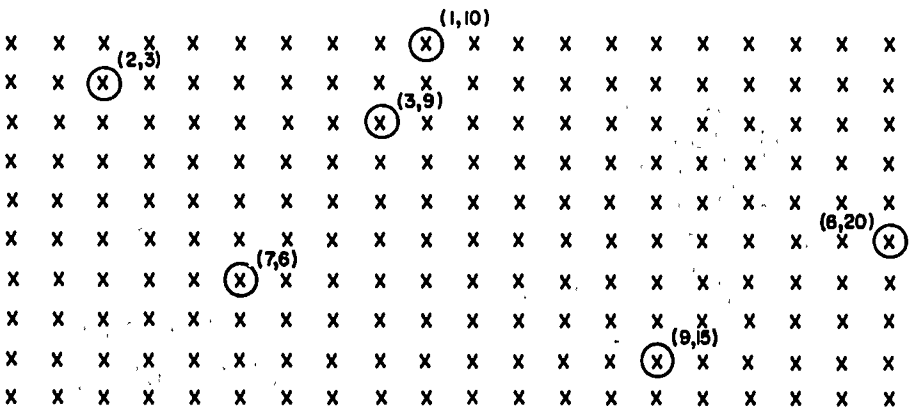

Figure 15.1 The location of six randomly selected sample hills, using the random-pair technique, for a plot consisting of 10 rows and 20 hills per row.

□ 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.

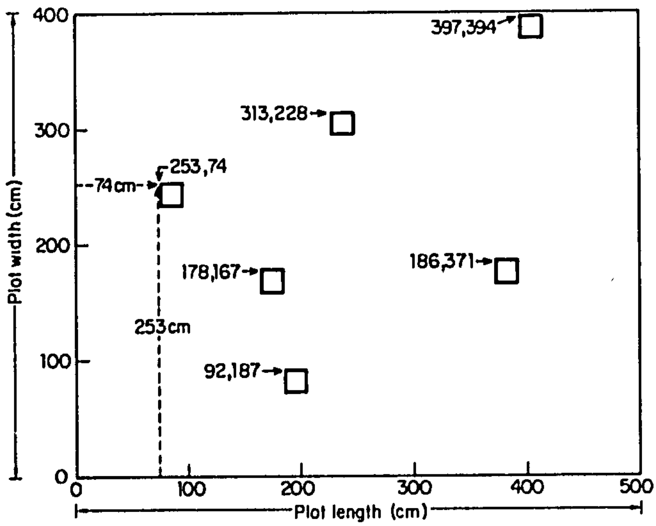

Figure 15.2 The location of six randomly selected 20 × 20-cm sampling units, using the random-pair technique for a plot measuring 4 × 5 m.