Chapter 4
Sampling and Experimentation

4.1 Introduction

This chapter serves as a bridge between descriptive statistics and inferential statistics. In the preceding chapters we focused on descriptive methods (descriptive statistics) which are used to explore the distribution of the values of a variable among the units in the sample. In most applications the sample is selected from a well–defined population and the ultimate goal is to use the data to make inferences about the distribution of the values of the variable (or variables) among all of the units in the population. Recall that the **population** is the collection of all of the units that are of interest and the **sample** is a subset of the population for which we have or will obtain data. Thus the purpose of inferential statistics is to use the data, which characterize the sample, to make inferences about the population.

We will concentrate on making inferences about particular aspects or characteristics of the population which can be quantified in terms of population parameters. A **parameter** is a numerical characteristic of the population. Recall that a **statistic** is a numerical characteristic of the sample. Thus, parameters and statistics are analogous quantities which quantify certain aspects of the population (parameter) or the sample (statistic). Since the goal of inference is to characterize certain aspects of the population, the first thing we need to do is to choose a variable that is suitable for inference in the sense that the values of the variable contain information about relevant characteristics of the population. Recall that a **variable** is a characteristic of a unit. Given a suitable variable, we can measure or observe the values of the variable for the units in the sample. These values can then be used to determine the value of a statistic and this statistic can be used to make inferences about the corresponding population parameter. For example, we might use the statistic as an estimate of the corresponding parameter or we might use the statistic to assess the evidence for a particular conjecture about the value of the parameter. Notice that once the sample is obtained and the data are collected we can determine the value of the statistic. On the other hand, we will never know the value of the parameter unless we take a census, *i.e.*, unless the sample is the whole population.

4.2 Sampling

Sampling is the process of obtaining a sample from a population. Our ultimate goal is to use the sample (which we will examine) to make inferences about the population (which we will not examine in its entirety). If the sample is selected from the population in an *appropriate fashion*, then we can use the information in the sample to make reliable and quantifiable inferences about the population. When the sample is obtained we will use the distribution of the variable among the units in the sample to make inferences about the distribution of the variable among the units in the population. If the distribution of the variable in the sample was exactly the same as the distribution of the variable in the population, then it would be easy to make inferences about the population; but, this is clearly too much to ask. Therefore we need to determine how to select a sample so that the sample is representative of the population.

The first step in deciding whether a method of choosing a sample will yield a representative sample requires a distinction between two populations. Before we obtain a sample we need to decide exactly which population we are interested in. The **target population** is the collection of all of the units that we want to make inferences about. We then need to determine which population our sample actually comes from. The **sampled population** is the collection of all of the units that *could* be in the sample. Notice that the sampled population is determined by the method used to select the sample.

Ideally the sampling method is chosen so that the sampled population is exactly the same as the target population and we can refer to this collection as the population. In practice, there may be some differences between the target population and the sampled population. When the sampled population is not identical to the target population we cannot be confident that the sample (which comes from the sampled population) will be representative of the target population. Furthermore, we cannot be confident that the statistic (which is based on a sample from the sampled population) will be suitable for inference about the parameter (which corresponds to the target population).

If there is a difference between the sampled population and the target population, in the sense that the distribution of the variable in the sampled population is different from the distribution of the variable in the target population, then a sample (obtained from the sampled population) is said to be **biased** for making inferences about the target population. If we use a biased sample to make inferences about the target population, the resulting inferences will not be appropriate. For example, a statistic based on a biased sample, may provide a suitable estimate of the corresponding parameter in the sampled population; but, it may not provide a suitable estimate of the corresponding parameter in the target population. Therefore, if the sampled population is different from the target population, then we must modify our goals by redefining the target population or we must change the sampled population by modifying our sampling method, since we want these two populations to be the same so that our inferences will be valid for our target population. It may be possible to change the method of obtaining our sample so that all of the units in the target population could be in our sample and these two populations are the same. If it is not possible to change the sampling method, then we must change our goals by restricting our inferences to the sampled population. In any case, once a sampling method has been chosen, the sampled population is determined and we should restrict our inferences to this sampled population. In conclusion, when making inferences from a sample we must carefully consider the restrictions imposed by the sampling method, since statistical theory can only justify inferences about the sampled population.

Example. Medical malpractice insurance. An insurance company that provides medical malpractice insurance is interested in determining how common it is for a medical doctor to be involved in a malpractice suit. The company plans to obtain a random sample of 500 doctors from the listing in a professional association directory. We will not assume that doctors are required to belong to this association.

In this example a medical doctor is a unit. The implied target population is all medical doctors in the region that is of interest to this insurance company, *e.g.*, if the company is considering offering insurance to all medical doctors with a medical practice in the US, then the target population is all medical doctors with a medical practice in the US. The sampled population is all medical doctors who are listed in the professional association directory. We know that doctors are not required to belong to this association. Furthermore, it may be possible for doctors who do belong to this medical association to not be listed in the current directory. There may also be doctors listed in the directory who are no longer in medical practice or who do not practice in the region of interest. Therefore information based on a random sample from doctors listed in the directory may not be appropriate if the goal is to describe the population of all medical doctors practicing in the region of interest. For example, if medical doctors who have never been sued for malpractice are more likely to be listed than those who have been sued for malpractice, then information based on a random sample from doctors listed in the directory may not be appropriate if the goal is to describe the population of all medical doctors in the region of interest. Therefore, inferences based on a random sample of doctors selected from those listed in this directory should be restricted to only those doctors who are listed in the directory (the sampled population).

Assuming that we have defined a method of selecting a sample so that the sampled population is the same as the target population, we next need to consider exactly how we should select the units that constitute the sample. Since we are assuming that the sampled and target populations are the same, we do not need to worry about the type of bias described above. However, we might introduce bias if we do not select the units for the sample in an appropriate fashion. The approach to sampling that we will adopt is called random sampling. The idea behind random sampling is to eliminate potential bias (intentional or unintentional) from the selection process by using *impersonal random chance* to select the sample. In addition to eliminating bias random sampling also provides the basis for theoretical justification and quantification of inferences based on the sample.

All of the sampling situations we consider can be viewed as being abstractly the same as the simple situation of selecting a sample of balls from a box of balls.

Example. Balls in a box. A box contains a collection of balls. In this situation a unit is a ball and the collection of balls in the box is the population. Our sample will consist of n balls selected from the box.

When a ball is selected we need to "measure" it, *i.e.*, we need to determine the value of the variable for this ball. Before we can do so we need to define a suitable variable. The definition of a variable consists of a description of the variable and an indication of its possible values. For example, suppose that the balls in the box are of various colors; say red, blue, and green. If we are interested in the characteristic color, then we might define the qualitative variable "color of the ball" with possible values of red, blue, and green. On the other hand, suppose that the balls in the box are of various weights. If we are interested in the characteristic weight, then we might define the quantitative variable "weight of the ball" with possible values equal to potential weights in grams.

Given a variable, suitable parameters and statistics are defined to correspond to the variable. With respect to the color of a ball, the proportion of red balls in the box (in the population) is a parameter and the proportion of red balls among the n balls selected from the box (in the sample) is a statistic. With respect to the weight of a ball, the mean weight of the balls in the box (in the population) is a parameter and the mean weight of the n balls selected from the box (in the sample) is a statistic.

The simplest type of random sample is called a simple random sample. A **simple random sample of size n** is a sample of n units selected from the population in such a way that every possible sample of n units has the same chance of being selected. This definition of a simple random sample can be refined to distinguish between two versions of simple random samples. If we require that the possible samples of n units are such that a particular unit can occur at most once in a sample, then we refer to the sample as being a **simple random sample of size n, selected without replacement.** On the other hand, if we allow a particular unit to occur more than once in the sample, then we refer to the sample as a **simple random sample of size n, selected with replacement.**

Example. Balls in a box (revisited). To obtain a **simple random sample of size n, selected without replacement** from the balls in our box, we first mix the balls in the box and select one ball at random (so that each ball in the box has the same chance of being selected). We then remove the selected ball from the box giving us one ball in our random sample. Then we mix the remaining balls in the box and select one ball at random (again so that each ball remaining in the box has the same chance of being selected) and remove the selected ball from the box giving us two balls in our random sample. This process of choosing a ball at random and removing it from the box is continued until n

balls have been selected. These n balls form the simple random sample of size n, selected without replacement.

To obtain a **simple random sample of size n, selected with replacement** from the balls in our box, we first mix the balls in the box and select one ball at random (so that each ball in the box has the same chance of being selected). We then measure or observe the value of the variable for the selected ball giving us the value of the variable for one of the balls in our random sample. Then we return the selected ball to the box, mix the balls in the box, and select one ball at random (again so that each ball in the box has the same chance of being selected) and measure or observe the value of the variable for the selected ball giving us two values of the variable corresponding to the two balls selected for our random sample. This process of choosing a ball, measuring the value of the variable for the ball, and returning the ball to the box is continued until n balls have been selected and measured. These n measurements (values of the variable) correspond to the balls that form the simple random sample of size n, selected with replacement.

If the population from which we wish to select a random sample is not too large, then it is possible to envision actually labeling each unit in the population, placing these labels on a collection of balls, placing these labeled balls in a box, and selecting a simple random sample of these balls as described above. In fact, state lotteries, where a simple random sample of numbers is selected from a collection of allowable numbers (the units), are conducted in this way. If you have ever observed the complicated mechanisms used to select winning lottery numbers, you know that it is difficult to convince people that a method of "drawing balls from a box" yields a proper simple random sample.

We will now discuss a simple alternative to sampling from an actual population of balls based on a sequence of random digits. A **sequence of random digits** is a list of the ten digits $0, 1, 2, \dots, 9$ with the following two properties.

1. For any given position in the list, each of the ten digits, $0, 1, 2, \dots, 9$, has the same chance of being in that position.
2. The entries in the list are independent of each other. That is, knowing the values in any part of the list would provide us with no information about any other values in the list beyond the information implied by property 1.

These two properties also hold if we think of the sequence of random digits as a list of two digit numbers $(00, 01, 02, \dots, 99)$ or as a list of three digit numbers $(000, 001, \dots, 999)$ or as a list of such numbers with any fixed number of digits.

To use a sequence of random digits to select a simple random sample we first need to assign suitable numerical labels to all of the units in the population. A list of all of the units in the population along with their labels is called a **sampling frame**. To insure that we get a random sample we need to use labels that have the same number of digits. That is, if N is the number of units in the population, then

1. If $N \leq 10$, we should use N of the one digit labels $0, 1, 2, \ldots, 9$.
2. If $11 \leq N \leq 100$, we should use N of the two digit labels $00, 01, 02, \ldots, 99$.
3. If $101 \leq N \leq 1000$, we should use N of the three digit labels $000, 001, \ldots, 999$, and so on.

One place to find a sequence of random digits is in a random number table. A **random number table** is simply a table containing a sequence (list) of random digits. These tables are usually formatted into rows and columns with periodic spaces. This arrangement has no significance, it only serves to make the table easier to use. An alternative to finding and using a random number table is to use a computer program or a calculator to generate random digits or random numbers. We will first discuss the use of a random number table to select a simple random sample.

After suitable labels have been assigned to the units in the population, as discussed above, we then proceed to the random number table to determine the labels that correspond to the units to be included in the sample. We choose a starting point in the table and go through the table (the list) in a systematic fashion reading the appropriate number of digits as we go. For example, we can select some row as our starting point. Then, reading across the row we make a note of the first digit, or digits if the labels have more than one digit, then the second digit or digits, and so on. If we reach the end of the row before obtaining n labels, we simply go on to the next row. If we come upon a digit or digits that was not used as a label, we simply skip it. If we want to sample without replacement, we also skip any label that we have already selected. This process is continued until n labels have been selected from the random number table. The units in the sampling frame that correspond to the n labels we selected from the random number table form the simple random sample.

Table 1. Random digits.

05797	43984	21575	09908	70221	19791	51578	36432	33494	79888
10395	14289	52185	09721	25789	38562	54794	04897	59012	89251
35177	56986	25549	59730	64718	52630	31100	62384	49483	11409
25633	89619	75882	98256	02126	72099	57183	55887	09320	73463
16464	48280	94254	45777	45150	68865	11382	11782	22695	41988

Table 1 contains a small portion of the random number table in the book *A Million Random Digits with 100,000 Normal Deviates*, Rand Corporation, (1955). To illustrate the use of a random number table to select a simple random sample we will select a sample of $n = 10$ students from the students in the Stat 214 example who were registered in section 1. The $N = 36$ students listed on the first page of Table 1 of Chapter 1 form the population. If we use the line numbers in this table as labels, then this page of Table 1 is our sampling frame. Since there are 36 units in this population, we will use two digit labels. Starting at the beginning of the first row of random digits in Table 1 and reading two digits at a time across the first row and the beginning of the second row yields the labels: 05, 79, 74, 39, 84, 21, 57, 50, 99, 08, 70, 22, 11, 97, 91, 51, 57, 83, 64, 32, 33, 49, 47, 98, 88 10, 39, 51, 42, 89, 52, 18, 50, 97, 21, 25. Therefore, if we want a simple random sample of size 10, selected with replacement, then the students listed in lines 5, 21, 8, 22, 11, 32, 33, 10, 18, and 21 form our sample. If we want a simple random sample of size 10, selected without replacement, then we skip the second 21 and the students listed in lines 5, 21, 8, 22, 11, 32, 33, 10, 18, and 25 form our sample.

In this particular example when we read numbers from the random number table we had to skip a lot of pairs of digits since these numbers were not used as labels. In this example $N = 36$ and since $2 \times 36 = 72 < 100$ we can assign two labels to each unit. That is, assign 01 and 37 ($37 = 36 + 1$) as labels for unit 1 (the student listed in line 1), assign 02 and 38 ($38 = 36 + 2$) as labels for unit 2, and so on, assigning 36 and 72 as labels for unit 36. If we use these labels we only have to read through part of the first row to get the valid labels 05, 39, 21, 57, 50, 08, 70, 22, 11, 51, 57, and 64. Translating these labels to line numbers (by subtracting 36 if the number is greater than 36) shows that, using this method, the students listed in lines 5, 3, 21, 21, 14, 8, 34, 22, 11, and 15 form our sample of size 10, selected with replacement and the students listed in lines 5, 3, 21, 14, 8, 34, 22, 11, 15, and 28 form our sample of size 10, selected without replacement.

An alternative to finding and using a random number table is to use a computer program or a calculator to generate suitable random numbers. Computer programs will usually provide a list of random numbers with the desired number of digits automatically. Many calculators provide random numbers that are between zero and one. To obtain a random number with the appropriate number of digits from a random number between zero and one, simply read off the appropriate number of digits from the beginning of the number. For example, if a calculator provided the number .12345678 and we needed a three digit number, the number would be 123. You should be aware that the algorithms or methods that computer programs and calculators use to generate a sequence of random numbers vary in their quality. Some of these algorithms are not very successful at generating a valid sequence of random numbers.

When we take a simple random sample, all of the possible samples have the same chance of being selected. There are situations where it is not appropriate for all of the possible samples to have the same chance of being selected. Suppose that there are two or more identifiable subsets of the population (subpopulations). If we obtain a simple random sample from the whole population, then it is possible for the resulting sample to come entirely from one of the subpopulations, or for the sample not to contain any units from one or more of the subpopulations. If we know or suspect that the distribution of the variable of interest varies among the subpopulations, then a sample which does not contain any units from some of the subpopulations will not be representative of the whole population. Therefore, in a situation like this we should not use a simple random sample to make inferences about the whole population. Instead we should use a more complex kind of random sample. One possibility is to use a sampling method known as **stratified random sampling** which is described below in the context of a simple example.

Suppose we wish to estimate the proportion of all registered voters in the United States who favor a particular candidate in an upcoming presidential election. We might expect there to be differences in the proportion of registered voters who favor this candidate among the various states. For example, we might expect support for this candidate to be particularly strong in his or her home state. Because we are interested in the proportion of all registered voters in the United States who favor this candidate, we want to be sure that all of the states are represented fairly in our sample.

We can use the states to define **strata** (subpopulations), take a random sample from each state (stratum), and then combine these samples to get a sample that is representative of the entire country. This is an example of a stratified random sample. The simplest type of **stratified random sample** is obtained as described in the following three steps.

1. Divide the population into appropriate strata (subpopulations).
2. Obtain a simple random sample within each stratum.
3. Combine these simple random samples to get the stratified random sample from the whole population.

To obtain a representative sample in the opinion poll example, we would need to determine the number of registered voters in each state and select simple random samples of sizes that are proportional to the numbers of registered voters in the states.

4.3 Experimentation

The sampling approach to data collection discussed in the preceding section is often used to perform an observational study. The steps involved in conducting an observational study based on a random sample are summarized below.

1. Obtain a random sample of units from the population of interest.
2. Obtain the data. That is, determine the values of the variable for the units in the sample.
3. Use the data to make inferences about the population. More specifically, use the distribution of the variable in the sample to make inferences about the distribution of the variable in the population from which the sample was taken.

In an **observational study** we obtain a sample of units, observe the values of a variable, and make inferences about the population. The purpose of such an observational study is to observe the units in the sample and, based on these observations, to make inferences about what we would observe if we examined the entire population. On the other hand, in an **experimental study** we manipulate the units and observe their response to this manipulation. In the experimental context, a particular combination of experimental conditions is known as a **treatment**. The purpose of an experiment is to obtain information about how the units in the population would respond to a treatment; or, to compare the responses of the units to two or more treatments. The response of a unit to a particular treatment is determined by measuring the value of a suitable **response variable**.

The steps involved in conducting a simple experimental study based on a random sample are summarized below.

1. Obtain a random sample of units from the population of interest.
2. Subject the units in the sample to the experimental treatment of interest.
3. Obtain the data. That is, determine the values of the response variable for the units in the sample.
4. Use the data to make inferences about the how the units in the population would respond to the treatment. More specifically, use the distribution of the response variable in the sample to make inferences about the distribution of the response variable in the population from which the sample was taken. In this context it may be easiest to think of the population as the hypothetical population of values of the response variable which would result if all of the units in the population were subjected to the treatment.

We will now discuss the basic ideas of experimentation in more detail in the context of a simple hypothetical experiment. Suppose that a new drug has been developed to reduce the blood pressure of hypertensive patients. The treatment of interest is the administration of the new drug to a hypertensive patient. The change in a patient's blood pressure will be used as the response variable. For this example the plan of the simple experiment described above is summarized in the steps below.

1. Obtain a random sample of n hypertensive patients.
2. Measure the blood pressure of each patient before the new drug is administered.
3. Administer the new drug to each of these patients.
4. After a suitable period of time, measure the blood pressure of each patient.
5. For each patient determine the change in his or her blood pressure by computing the difference between the patient's blood pressure before the drug was administered and the patient's blood pressure after the new drug was administered. This change or difference will serve as the response variable for assessing the effects of the new drug. In this example the relevant population is the hypothetical population of changes in blood pressure that we would observe if all of the hypertensive patients in the population from which the sample was selected had been subjected to this experiment.

Suppose that we actually conducted this experiment. Furthermore, suppose that the data indicate that the hypertensive patients' blood pressures tend to decrease after they are given the new drug, *i.e.*, suppose that the data indicate that most of the patients experienced a meaningful reduction in blood pressure. We can conclude that there is an association between the new drug and a reduction in blood pressure. This association is clear, since the patients (as a group) tended to experience a decrease in blood pressure after they received the new drug. Can we conclude that the new drug caused this decrease in blood pressure? The support for the contention that the new drug caused the decrease in blood pressure is not so clear. In addition to the new drug there may be other factors associated with the observed decrease in blood pressure. For example, the decrease in blood pressure might be explained, in whole or in part, as the physical manifestation of the psychological effect of receiving medication. In other words, we might observe a similar decrease in blood pressure if we administered a placebo to the patients instead of the new drug. It is also possible that some other aspects of the experimental protocol are affecting the patients' blood pressures. The way that this experiment is being conducted does not allow us to separate out the effects of the many possible causes of the decrease in blood pressure. If we hope to establish a cause and effect relationship between taking the new drug and observing a decrease in blood pressure, then we need to use a comparative experiment.

In a **randomized comparative experiment** baseline data is obtained at the same time as the data concerning the treatment of interest. This is done by randomly dividing the available units (patients) into two or more groups and comparing the responses for these groups. In the drug example there is one treatment of interest, administer the new drug. Therefore, in this situation we only need two groups, a control group and a treatment group. The units (patients) in the **control group** do not receive the treatment (do not receive the new drug). The units (patients) in the **treatment group** do receive the treatment (do receive the drug). During the course of the experiment we need to keep all aspects of the experiment, other than the treatment itself, as similar as possible for all of the units in the study. The idea is that, if the only difference between the units in the control group and the units in the treatment group is that the units in the treatment group received the treatment, then any observed differences between the responses of the two groups must be caused by the treatment. In the drug example it would be a good idea to administer a placebo to the patients in the control group, so that they do not know that they did not receive the new drug. It would also be a good idea to "blind" the patients and the people administering the drug or placebo by not telling them which patients are receiving the placebo and which patients are receiving the new drug. The purpose of such blinding is to eliminate intentional or unintentional effects due to patient or administrative actions which might affect a patient's response. The steps for conducting such a **randomized comparative experiment** are given below.

1. Randomly divide the group of available patients into two groups: a group of n_1 patients to serve as the control group and a group of n_2 patients to serve as the treatment group. These two groups are random samples of sizes n_1 and n_2 from the group of available patients. The samples sizes n_1 and n_2 may be different.
2. Administer the placebo to the patients in the control group and administer the new drug to the patients in the treatment group.
3. Obtain the data. That is, measure the response variable for each of the $n_1 + n_2$ patients in the experiment. For example, we could determine the change (difference) in each patient's blood pressure as measured before and after administration of the placebo or new drug.
4. Compare the responses of the patients in the treatment group to the responses of the patients in the control group and make inferences about the effects of the new drug versus the placebo.

In this example there are two hypothetical populations of changes in blood pressure. The hypothetical population of changes in blood pressure that we would observe if all of the available hypertensive patients were subjected to this experiment and given the placebo and the hypothetical population of changes in blood pressure that we would observe if all of the available hypertensive patients were subjected to this experiment and given the new drug. Notice that, strictly speaking, our inferences in this example only apply to the hypertensive patients who were available for assignment to the groups used in the experiment. If we want to make inferences about a larger population of hypertensive patients, then the group of available patients used in the study should form a random sample from this larger population.

The experiment described above is designed to compare the effects of the new drug to the effects of a placebo. Suppose that we wanted to compare the effects of the new drug to the effects of a standard drug. To make this comparison we could design the experiment with three groups: a control group, a treatment group for the new drug, and a treatment group for the standard drug. If our only goal is to compare the two drugs (treatments), then we could eliminate the placebo control group and run the experiment with the two treatment groups alone.

Example. Cloud seeding. The data referred to in this example are given in Simpson, Olsen, and Eden (1975), *Technometrics*, **17**, 161–166. These data were collected in southern Florida between 1968 and 1972 to determine whether injection of silver iodide into cumulus clouds tends to increase rainfall. Fifty–two days that were deemed suitable for cloud seeding were randomly divided into two groups of 26 days. An airplane, equipped to inject silver iodide into a target cloud, was flown through the target cloud. For one group of 26 days the device used to inject the silver iodide was loaded and the target cloud was seeded. For the other group of 26 days the device used to inject the silver iodide was not loaded and the target cloud was not seeded. On all 52 days the airplane flew through the target cloud. Furthermore, the pilots and technicians on the plane were not aware of whether the device used to inject the silver iodide was loaded or not. For each day the amount of rainfall (total volume of rain falling from the cloud base), measured in acre–feet, was determined.

In this example there are 52 days that were deemed suitable for cloud seeding. Each of these days is a unit and this group of 52 days is the group of "available units" which were used in the experiment. The response variable is the amount of rainfall measured after the airplane was flown through the cloud. The two relevant hypothetical populations for which inferences could be made in this example are: the collection of 52 rainfall amounts which would have been obtained if the plane had been flown through the cloud but the cloud had not been seeded with silver iodide, and the collection of 52 rainfall amounts which would have been obtained if the plane had been flown through the cloud and the cloud had been seeded with silver iodide.

We can define two population mean rainfall amounts (parameters) corresponding to these populations, *i.e.*, the mean of the 52 rainfall amounts which would have been obtained if the cloud was not seeded on each of the 52 days and the mean of the 52 rainfall amounts which would have been obtained if the cloud was seeded on each of the 52 days. The two sample mean rainfall amounts (statistics) based on the rainfall amounts recorded for each of the two groups of 26 days, *i.e.*, the mean of the 26 rainfall amounts recorded on the 26 days when the cloud was not seeded and the mean of the 26 rainfall amounts recorded on the 26 days when the cloud was seeded, could be used to make inferences about the corresponding population mean rainfall amounts.

Since the 52 days on which this experiment was conducted did not form a random sample from some larger population of days suitable for cloud seeding, we cannot justify extending our inferences beyond these 52 days. We might reasonably argue that our inferences apply to similar days in the area where the study was conducted; but, we cannot use statistical theory to justify extrapolations to days other than these 52.

4.4 Summary

Reliable and quantifiable inferences about a population (about the population distribution of a variable) require careful consideration of the definition of the relevant population and of the method used to obtain the data on which the inferences are based.

A sampling study is conducted by selecting a random sample of units from a population, observing the values of a variable for the units in the sample, and then making inferences or generalizations about the population. More specifically, the distribution of the values of the variable among the units in a random sample is used to make inferences about the distribution of the variable among the units in the population. The first consideration in planning or interpreting the results of a sampling study is the determination of exactly which units could be in the sample. The collection of all units which could be in the random sample is known as the sampled population and this sampled population is the relevant population for inferences based on the sample. The second consideration concerns the proper selection of the units which constitute the sample. We cannot properly quantify inferences unless the sample is a properly selected random sample from the population.

An experimental study differs from a sampling study in that the units used in the experimental study are manipulated and the responses of the units to this experimental manipulation are recorded. For an experimental study the relevant population or populations are hypothetical populations of values of the variable defined by the experimental treatment(s) and corresponding to all of the units available for use in the experiment. That is, the relevant population(s) is the population of values of the variable which would be observed if all of the available units were subjected to the experimental treatment(s). In the context of a comparative experiment we cannot properly quantify inferences unless the units are assigned to the treatments being compared using an appropriate method of random assignment. This random assignment of units to treatments is analogous to the random sampling of a sampling study.