APPENDIX C

How to Choose a Random Sample

As we have seen in this book, it is extremely important to be able to choose a random sample. Suppose that we want to choose a random sample of size n from a population of size N. How can we accomplish this?

The first step is to number the population from 1 to N in any arbitrary manner. Then we will choose a random sample by designating n elements of the population that are to be in the sample. To do this, we start by letting the first element of the sample be equally likely to be any of the N elements. The next element is then chosen so that it is equally likely to be any of the remaining $N - 1$ elements, the next so that it is equally likely to be any of the remaining $N - 2$ elements, and so on, until we have amassed a total of n elements, which constitute the random sample.

To implement this scheme, it seems that we would always have to keep track of which elements had already been selected. However, by a neat trick, it turns out that this is not necessary. Indeed, we can arrange the N elements in an ordered list and then randomly choose not the elements themselves but rather the positions of the elements that are to be put in the random sample. Let us see how it works when $N = 7$ and $n = 3$. We start by numbering each of the 7 elements in the population and then arranging them in a list. Say the initial order is

$$1, 2, 3, 4, 5, 6, 7$$

We now choose a number that is equally likely to be 1, 2, 3, 4, 5, 6, or 7; say 4 is chosen. This means that the element in position 4 (element number 4 in this case) is put in the random sample. To indicate that this element is in the random sample and to make certain that this element will not be chosen again, we interchange in the list the element in position 4 with the one in position 7. This results in the new list ordering

$$1, 2, 3, 7, 5, 6, \underline{4}$$

where we have underlined the element that is in the random sample. The next element to be put in the random sample should be equally likely to be any of the elements in the first 6 positions. Thus we select a value that is equally likely to be 1, 2, 3, 4, 5, or 6; the element in that position will become part of the random sample. And to indicate this and to leave the first 5 positions for the elements that have not yet been chosen, we interchange the element in the position chosen with the element in position 6. For instance, if the value chosen was 4, then the element in position 4 (that is, element number 7) becomes part of the random sample, and the new list ordering is

$$1, 2, 3, 6, 5, \underline{7}, \underline{4}$$

The final element of the random sample is equally likely to be any of the elements in positions 1 through 5, so we select a value that is equally likely to be 1, 2, 3, 4, or 5 and interchange the element in that position with the one in position 5. For instance, if the value is 2, then the new ordering is

$$1, 5, 3, 6, \underline{2}, \underline{7}, \underline{4}$$

Since there are now three elements in the random sample, namely, 2, 7, and 4, the process is complete.

To implement the foregoing *algorithm* for generating a random sample, we need to know how to generate the value of a random quantity that is equally likely to be any of the numbers $1, 2, 3, \ldots, k$. The key to doing this is to make use of *random numbers* that are the values of random variables that are uniformly distributed over the interval $(0, 1)$. Most computers have a built-in random number generator that allows one to call for the value of such a quantity. If U designates a random number—that is, U is uniformly distributed over the interval $(0, 1)$—then it can be shown that

$$I = \text{Int}\,(kU) + 1$$

will be equally likely to be any of the values $1, 2, \ldots, k$, where Int (x) stands for the integer part of x. For instance,

$$\text{Int}\,(4.3) = 4$$
$$\text{Int}\,(12.9) = 12$$

and so on.

Program A-1 uses these to generate a random sample of size n from the set of numbers $1, 2, \ldots, N$. When running this program, you will be asked first to enter the values of n and N and then to enter any four-digit number. For this last request, just type and enter any number that comes to mind. The output from this program is the subset of size n that constitutes the random sample.

Example C.1

Suppose we want to choose a random sample of size 12 from a population of 200 members. To do so, we start by arbitrarily numbering the 200 members of the population so that they now have numbers 1 to 200. We run Program A-1 to obtain the 12 members of the population that are to constitute the random sample.

```
THIS PROGRAM GENERATES A RANDOM SAMPLE OF K
OF THE INTEGERS 1 THRU N
ENTER THE VALUE OF N
? 200
ENTER THE VALUE OF K
? 12
Random Number Seed (–232,768 to 32,767)? 355
THE RANDOM SAMPLE CONSISTS OF THE FOLLOWING 12 ELEMENTS
90 89 82 162 21 81 182 45 38 195 64 1
```

■