

## Random Numbers

In Section 1.1.1 we discussed the lottery method for randomly selecting a random number between 1 and  $N$ . In our description, we sampled *without replacement*. When a labeled ball was drawn out of the barrel it was not replaced before the next ball was drawn. We sometimes also sample *with replacement*. Each time a ball is drawn, the number on that ball is noted and the ball replaced before remixing and drawing the next ball. In this way the same random number can occur two or even more times. This idea will be important in our discussion of random numbers. Of course, in surveys, we always sample without replacement because there is no point in interviewing the same person twice.

### 1 Random number tables

The lottery method is a clumsy physical process for choosing random samples. Often it is convenient to use a ready-made table of random numbers. A random number table is a table of digits. The digit given in each position in the table was originally chosen randomly from the digits 1, 2, 3, 4, 5, 6, 7, 8, 9, 0 by a random process in which each digit is equally likely to be chosen. A possible mechanism for doing this is the lottery method with replacement using ten labeled balls. Table 1 gives part of such a table: another part is given in the Appendix A1 of the text. The digits are often grouped in fives as shown here. The frequencies with which each digit occurs in the fifty digits of Table 1 are given in Table 2. Thus, the digit “1” appears in Table 1 five times, the digit “2” appears seven times, “3” appears four times, etc.

**Table 1 :** Random Digits

12429	63527	74608	01549	00793
28354	61218	95782	63940	58128

**Table 2 :** Frequency of Occurrence of Each Digit in Table 1

<i>Digit :</i>	1	2	3	4	5	6	7	8	9	0
<i>Frequency :</i>	5	7	4	5	5	4	4	6	5	5

We will now learn to use the tables to accomplish a few small tasks.

**Example 1.** First we will randomly sample 4 people (or labels) from 8 so we need 4 random digits without replacement from 1, 2, 3, . . . , 8. We simply read off random digits ignoring those that are out of range or recur (we are

sampling without replacement) until we get four of them. Going from left to right across the top row of Table 1 we get

1 2 4 [2] [9] 6 3 5 ... .

(Here, we have placed square brackets around numbers that are repeats of previously appearing numbers or are out of range.) Taking the first four usable numbers we get 1, 2, 4, 6 and the random sample consists of the individuals with those labels.

**Example 2.** We now wish to sample 6 individuals from 60 (labeled 1 to 60). We continue taking pairs, ignoring repetitions and those out of range, until we have 6 of them. Beginning from the left of the top row of Table 1 and taking the digits in pairs, we get

12 42 [96] 35 27 [74] 60 [80] 15 49 ...

and we take the individuals labeled 12, 42, 35, 27, 60, and 15.

**Example 3.** Consider an agricultural experiment to compare four treatments (e.g. fertilizers) laid out in 3 blocks of land each containing 4 plots. For each block of land we want to randomly allocate one of each of the four treatments to a plot. This is called a randomized block design. To do this we want to find a random ordering (permutation) of the numbers 1, 2, 3, 4, representing treatments to apply to the plots in a block. Again we will work across rows and ignore numbers outside the range 1 to 4 and repetitions within a block to obtain:

1 2 4 [2] [9] [6] 3 || [5] 2 [7] [7] 4 [6] [0] [8] [0]  
1 [5] [4] [9] [0] [0] [7] [9] 3 || 2 [8] 3 [5] 4 [6] 1.

Our three permutations are (1, 2, 4, 3), (2, 4, 1, 3) and (2, 3, 4, 1). Using “*treat.*” to abbreviate treatment, the plot layout might look like:

Block 1	Block 2	Block 3
plot: 1 2 3 4	plot: 1 2 3 4	plot: 1 2 3 4
<i>treat.:</i> 1 2 4 3	<i>treat.:</i> 2 4 1 3	<i>treat.:</i> 2 3 4 1

Any systematic fertility trend from left to right is counteracted somewhat by the random ordering of treatments.

**Example 4.** We can use the tables to simulate the tossing of a “fair” coin. If we denote the digits 1, 2, 3, 4, 5 as being heads (*H*) and the digits 6, 7, 8, 9, 0

as being tails ( $T$ ), at each “toss” we are equally likely to get a head or a tail. Let us “toss” 21 times and use the second row of our table this time. Then  $3 = H$ ,  $2 = H$ ,  $8 = T$ ,  $3 = H$ , etc. The whole sequence of twenty one tosses is:

$H T H H H T H H H T T H T T H T H T H T H.$

The following is a real sequence of 21 tosses:

$T T H T T H H H H H H T H T H T T T H H H.$

Sometimes with actual random sequences, as here, you get what looks like too long a run of heads or tails. When this type of behavior occurs with an artificially constructed sequence one might suspect the randomness of the random number table. For truly random numbers, the choice of a digit should not be affected by what is chosen previously. There are statistical tests for checking whether the lengths of run are consistent with random behavior.

**Warning:** One shouldn’t keep using the same part of a random number table over and over again as we have been doing. We should use a large table and find the starting point randomly, even if only by the “look away and stick a pin in it” method.

## 2 Pseudo-random numbers

It should be clear from the examples in Section 1 that it is not the lottery method of producing the numbers that matters but the properties of the numbers produced. They should have the properties we would expect “random” numbers to have. For example, we would like the numbers to satisfy a wide range of criteria such as approximately equal frequencies of the digits taken one at a time, two at a time etc. Also, the frequency of the lengths of runs like those in a heads/tails sequence should be “reasonable”. “Random” numbers produced on a computer or calculator are not random at all. They are generated using a formula, so that the sequence produced is actually predetermined. For this reason they are more properly called pseudo-random numbers. However, the methods for producing them can be tuned so that they appear random in any way you care to view them. In particular, over a long sequence, each number (or pair or triple of numbers) appears approximately equally often. Many random number generators produce a decimal number between zero and one. How do we turn this into an integer between 1 and  $N$  for sample selection as in Section 1? If  $u$  is the random number produced by your calculator, compute

$$i = 1 + u \times N$$

and use the portion of the answer to the left of the decimal point.<sup>1</sup> You can now produce a stream of pseudo-random numbers. For sampling without replacement ignore repeated values.

### Exercises

1. Simulate the experiment of tossing a drawing pin (thumb tack) in which the two outcomes “point up” and “point down” occur with probability 0.6 and 0.4 respectively. Perform 15 “tosses”.
2. Consider an experiment in which patients attend a clinic for ankylosing spondylitis (a form of arthritis). You want to compare four treatments, namely aspirin at a given dose, indocid at a given dose and naprosyn at two different levels of dosage. Show how to use Table 1.5.1 to randomly assign each incoming patient to one of the four treatments.
3. Suppose there are 100 sets of accounts and 20 of these contain fraudulent transactions. Simulate the sampling of 10 sets of accounts one at a time without replacement, noting whether or not each set contains a fraudulent transaction. How many fraudulent sets of accounts did you find?
4. Given an  $8 \times 8$  grid of 64 square plots, use the random numbers to select 6 of the plots at random without replacement.

---

<sup>1</sup>For example if  $N = 500$  and  $u = 0.2134$ , then  $i = 1 + 0.2143 \times 500 = 108.15$  so that our random number between 1 and 500 is 108. [Note:  $u$  should never be exactly zero or one but if it is to the precision of your calculator then  $u = 0$  should become a random number of 1 and  $u = 1$  should become  $N$ . Also  $u$  needs to be recorded to many more significant figures than  $N$ .]