

3.1 Introduction

Chapter 3 introduces the topic of probability and random sampling to the reader. Neither **Excel** nor **XLSTAT** offers any analyses directly related to probability but both have features that we will briefly mention to help users take random samples from populations. **XLSTAT** offers the **Data Sampling** option that allows the user to randomly sample from within a data set. The **RANDBETWEEN** function and the **Random Number Generation** tool within **Excel** allow the user two methods of drawing random selections from a population of N items. All three procedures can be used to generate random samples and are discussed below.

The following examples from *Statistics* are solved with **Excel** and **XLSTAT** in this chapter:

Excel Companion		Statistics Text
Exercise	Page	
3.1	56	Example 3.22
3.2	58	Example 3.21
3.3	60	Example 3.21

3.2 Random Sampling

3.2.1 The Data Sampling Option

The random sampling section in the text defines to the reader what a random sample is and gives a method of generating a random sample that utilizes the random number table found in the Appendix of the text. **XLSTAT** offers the **Data Sampling** option that accomplishes the same task. We illustrate with the following example.

Exercise 3.1: As an example we turn to Example 3.22 from the *Statistics* text:

A designed experiment in the medical field involving human subjects is referred to as a *clinical trial*. One recent clinical trial was designed to determine the potential of using aspirin in preventing heart attacks. Volunteer physicians were randomly divided into two groups: the *treatment* group and the *control* group. Each physician in the treatment group took one aspirin tablet a day for one year, while the physicians in the control group took an aspirin-free placebo made to look identical to an aspirin tablet. Since the physicians did not know to which group – treatment or control, they were assigned, the clinical trial is called a *blind study*. Assume that 20 physicians volunteered for the study. Randomly assign half the physicians to the treatment group and half to the control group.

Solution:

We begin by creating an **Excel** data set that is made of a column of physician ID numbers. For this exercise, we will use the ID numbers 1 – 20. When created, the data set should look similar to the one shown in Figure 3.1.

Figure 3.1

	A
1	Physician ID
2	1
3	2
4	3
5	4
6	5
7	6
8	7
9	8
10	9
11	10
12	11
13	12
14	13
15	14
16	15
17	16
18	17
19	18
20	19
21	20

To take a random sample, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 3.2. To randomly divide the physicians into two groups, we click on the **Preparing data** menu and select the **Data sampling** option shown in Figure 3.3.

This opens the **Data sampling** menu shown in Figures 3.4. We first need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 1 – 21, with row 1 being the variable label. We specify the location in the **Data** box and check the **Variable labels** box to indicate the first row of data represents the variable name, **Physician ID**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described. In the **Sampling:** pull-down menu, we select **Random without replacement** as our sampling type. Because we want to split the data set into two groups of ten physicians, we need to randomly sample a single group of ten physicians from the set of 20 physicians. We accomplish this by specifying **1** as the **Number of samples:** and **10** as the **Sample size** in the **Data sampling** menu. We click **OK** to view our results.

Figure 3.2

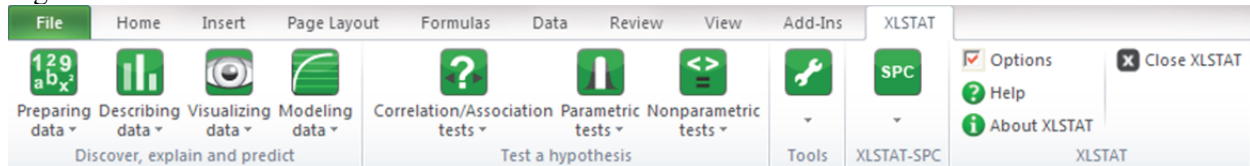


Figure 3.3

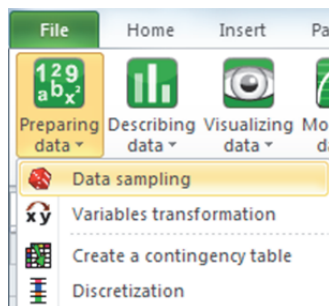
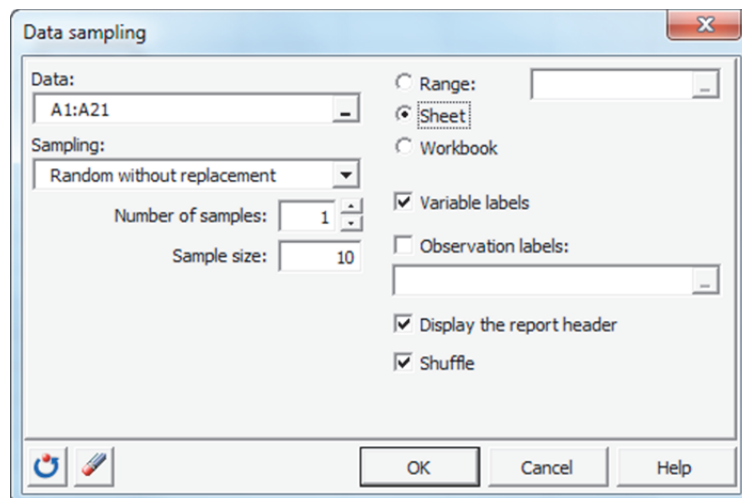


Figure 3.4



The results of our sampling are shown in Figure 3.5. Please note that the sample you select will most likely be different due to the randomness of the sampling procedure that is being used.

Figure 3.5
Sampled data:

Physician ID
11
18
10
16
1
12
7
9
4
15

In our sample, the treatment group would be comprised of physicians 1, 4, 7, 9, 10, 11, 12, 15, 16, and 18. The remaining physicians – 2, 3, 5, 6, 8, 13, 14, 17, 19, and 20 – would comprise the control group.

The **Data sampling** option is useful when the entire data set can be displayed in an **Excel** workbook. Sometimes the data we are interested in sampling from comes from an extremely large population which makes the **Data sampling** option difficult to work with. In these cases, there are two **Excel** functions that might be better choices.

3.2.2 The RANDBETWEEN Function

The **RANDBETWEEN** function offered in **Excel** provides a simpler method of drawing a random sample from a population of known population size. We illustrate this function using Example 3.21 found in the *Statistics* text.

Exercise 3.2: Suppose you wish to randomly sample five households from a population of 100,000 households to participate in a study.

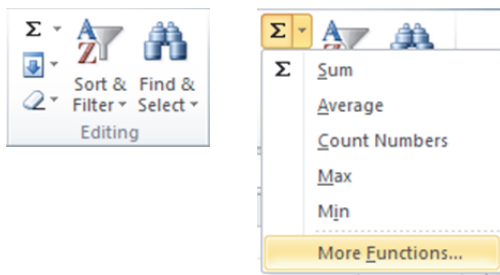
- How many different samples can be selected?
- Use a random number generator to select a random sample.

Solution:

We are shown in the text that there are 8.33×10^{22} possible random samples of five homes that can be selected from a population of 100,000 homes. In part b we are asked to take one of these samples.

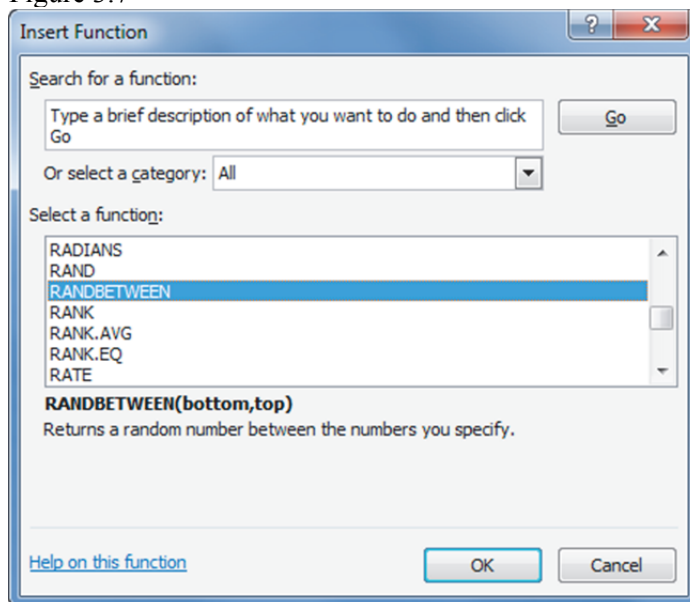
We begin by opening up **Excel** and placing the cursor on any cell in the blank worksheet. We click on the **Home** tab and then select the arrow next to the Σ Icon in the **Editing** group of options shown in Figure 3.6.

Figure 3.6



We then select the **More Functions** option to get the **Insert Function** menu shown in Figure 3.7 Click on the arrow to select that **All** categories are being used and scroll down until you reach the **RANDBETWEEN** function. Click OK to open the **RANDBETWEEN** function.

Figure 3.7



The user is asked to furnish the two values, called **Bottom** and **Top**, that represent the smallest and largest values that we want to sample from. In our example, we wish to sample five homes from within a group of 100,000 homes. We specify a Bottom value of 1 and a Top value of 100,000 in Figure 3.8. Click OK. **Excel** returns a random integer between the values specified. In our example, **Excel** returned the value of 48,051 (see Figure 3.9). We would sample the home that was labeled as 48,051 as our first home in the sample.

Figure 3.8

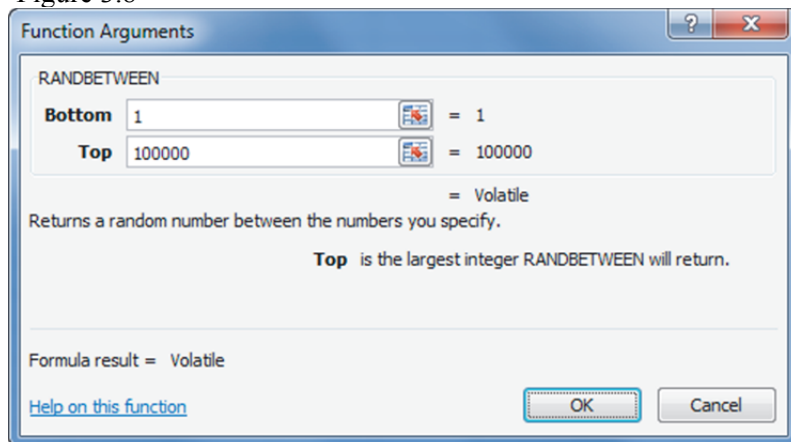
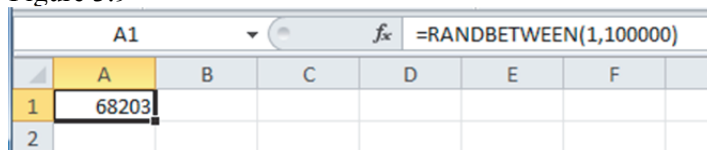


Figure 3.9



Note that both the Bottom and Top numbers are possible for selection when using the **RANDBETWEEN** function. In our example, we would need to repeat this process four additional times to generate a random sample of five homes. The **RANDBETWEEN** function is a useful tool when small samples sizes are desired. When larger samples are necessary, we look to the **Random Number Generation** tool that **Excel** offers. It is explained in the next section.

3.2.3 The Random Number Generation Tool

We will again utilize Example 3.21 found in the *Statistics* text to demonstrate how to use the **Random Number Generation** tool within **Excel**.

Exercise 3.3: Suppose you wish to randomly sample five households from a population of 100,000 households to participate in a study.

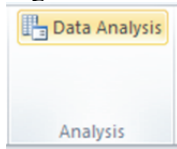
- How many different samples can be selected?
- Use a random number generator to select a random sample.

Solution:

We are shown in the text that there are 8.33×10^{22} possible random samples of five homes that can be selected from a population of 100,000 homes. In part b we are asked to take one of these samples.

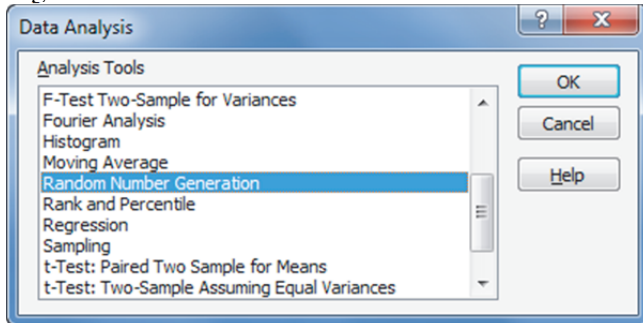
We begin by opening up **Excel** and placing the cursor on any cell in the blank worksheet. We click on the **Data** tab and then click on the **Data Analysis** Icon found in the **Analysis** group as shown in Figure 3.10.

Figure 3.10



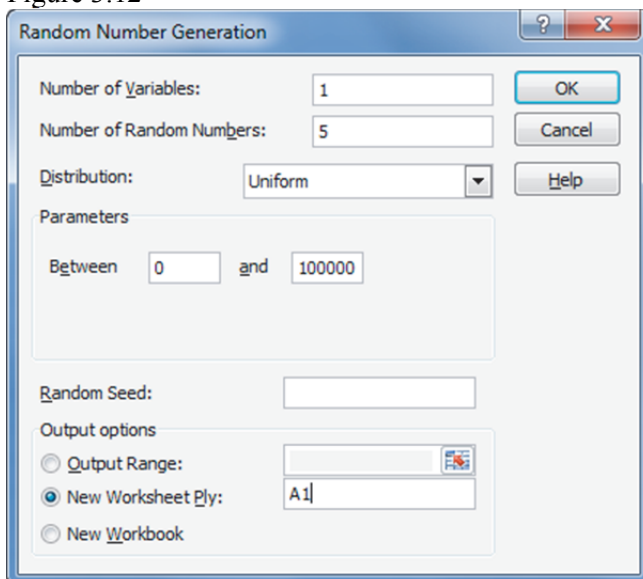
This will open up the **Data Analysis** menu shown in Figure 3.6. Scroll down the list and select the **Random Number Generation** tool. Click OK.

Figure 3.11



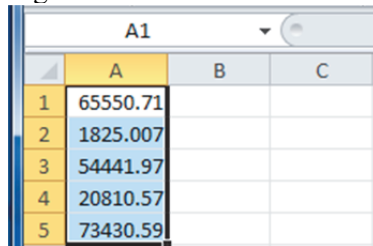
The **Random Number Generation** menu is opened as shown in Figure 3.12. In this example, we are asked to take a single sample of five homes from the 100,000 homes available. We enter the value **1** as the **Number of Variables** to use, indicating we want just one sample. We enter the value **5** as the **Number of Random Numbers**, indicating our sample size. To choose a random sample as described in the text, we select the **Uniform Distribution** type and indicate the values **1** and **100,000** as the values to select **Between**. We have the option of choosing a specific random number sequence by entering a value in the **Random Seed** box, but we choose to let **Excel** decide and will leave it blank. The **Output Options** allow us to select where the random numbers will be located within the **Excel** workbook. In Figure 3.12, we have specified that **Excel** list the random numbers beginning in cell **A1**. We click on OK to select the random numbers

Figure 3.12



The random numbers selected by **Excel** are shown in Figure 3.13. Notice that the random digits are not expressed as integers but contain decimal values as part of the random numbers.

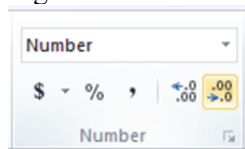
Figure 3.13



	A	B	C
1	65550.71		
2	1825.007		
3	54441.97		
4	20810.57		
5	73430.59		

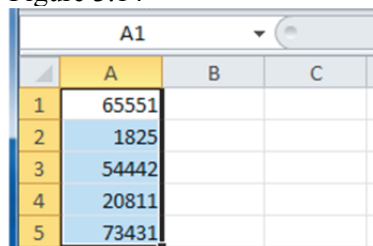
To remove these decimals, we can format the numbers selected and decrease the decimals of the values until we are left with the rounded integers. Click on the **Home** tab and Select the **Decrease Integers** button found in the **Numbers** group as shown if Figure 3.14.

Figure 3.14



Continue clicking on the **Decrease Integers** button until all decimals are removed from the random numbers and you are left with values similar to those shown in Figure 3.14. These would be the homes that should be sampled to generate a random sample of five homes from the 100,000 contained in the population.

Figure 3.14



	A	B	C
1	65551		
2	1825		
3	54442		
4	20811		
5	73431		

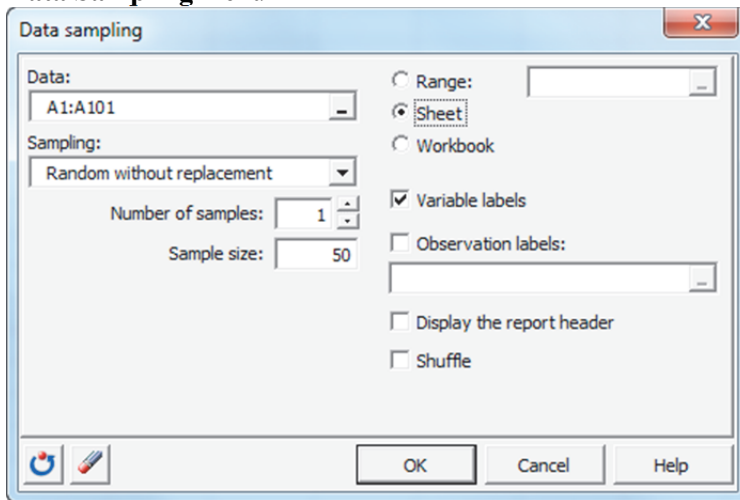
It needs to be noted that both the **RANDBETWEEN** and **Random Number Generation** tool can result in repeated selections of random numbers. It may be desirable to select more random numbers than necessary to replace repeated random numbers when and if they occur.

3.3 Technology Lab

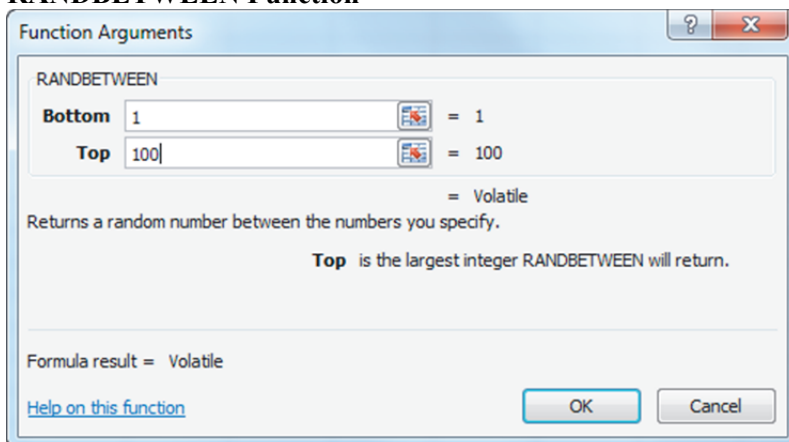
The Technology Lab consists of problems for the student to practice the techniques presented in each lesson. Each problem is taken from the homework exercises within the *Statistics* text and includes an **Excel** data set (when applicable) that should be used to create the desired output. The completed output has been included with each problem so that the student can verify that he/she is generating the correct output.

Kiwifruit as an iron supplement. Massey University (New Zealand) researchers have designed a 16-week study to determine the effectiveness of gold kiwifruit as an iron supplement in women with an iron deficiency (*BMC Public Health, Vol. 10, 2010*). Half the women in the study will receive an iron fortified breakfast cereal with a banana, and half will receive an iron fortified breakfast cereal with gold kiwifruit. The researchers are planning for 100 women to participate in the study. Number the women from 1 to 100. Then use the random sampling techniques discussed in this chapter to select 50 women to receive the kiwifruit breakfast. The remaining 50 women will receive the banana breakfast.

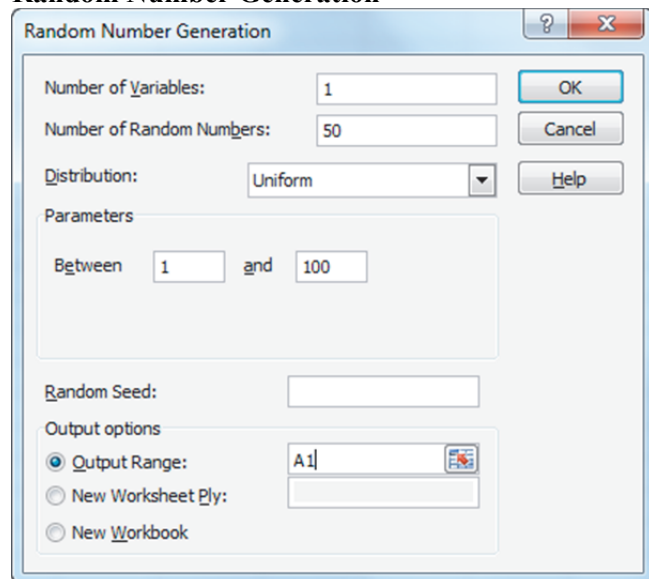
Data Sampling menu



RANDBETWEEN Function



Repeat until 50 unique women have been selected.

Random Number Generation

Random Number Generation

Number of Variables: 1 OK

Number of Random Numbers: 50 Cancel

Distribution: Uniform Help

Parameters

Between 1 and 100

Random Seed:

Output options

☒ Output Range: A1

☐ New Worksheet Ply:

☐ New Workbook

Check for duplicate results and sample additional women as needed.

Note: Due to the nature of random sampling, each result should vary from previous samples taken. For this reason, we have supplied only the menus needed to generate the results, rather than the results themselves.