Using *Microsoft Excel* to Illustrate the Central Limit Theorem

by Randall Gallaher, Lewis & Clark Community College

Course

MATH 145 (Elementary Statistics)

Description

This module will help students comprehend the Central Limit Theorem. *Microsoft Excel* will be used to generate random samples of a given size of digits 0 through 9. The mean of each sample is then calculated. A tally of the sample means that fall within given intervals is then found. This tally is then used to construct a graph, which should be approximately normal depending on the sample size.

Transferability

This module will transfer to any introductory statistics class or any class that utilizes the central limit theorem.

Faculty Technology Skill

Faculty will need to have some familiarity with spreadsheet programs. The specific spreadsheet program used in the module is *Microsoft Excel.*

Faculty Equipment

The faculty member will need a computer containing the program *Microsoft Excel*. If the faculty member wishes to use this module as a demonstration, he or she will also likely need a computer projection system.

Student Equipment

This module is intended to be used as a demonstration to students. If this is done, the students will need no equipment. If this module is used as an activity, then students will need a computer containing the program *Microsoft Excel.*

Cost

Assuming that the school already has *Microsoft Excel*, no additional expenses are required.

Improvement on Teaching and Learning

By viewing this demonstration, students will hopefully gain a better understanding of the central limit theorem. They should see that as the sample size gets bigger and bigger, the sampling distribution will be closer and closer to normal. The use of technology allows students to see many more examples of sampling distributions than would be possible without it.

Nontechnology Comparison

The same quality of learning obtained from this module is not really possible without it. The time required for the tedious computations impedes the ability to get a full

appreciation of the central limit theorem. By the time the computations are done, the students tend to lose focus of what they are ultimately looking for. Use of the spreadsheet enables students to focus more on interpretation and understanding the central limit theorem.

How to Use in the Classroom

This module should be used after discussion of normal probability distributions. After initial discussion of the central limit theorem, this module could also probably be used to demonstrate the full meaning of it.

Using *Microsoft Excel* to Illustrate the Central Limit Theorem

Activity: Elementary Statistics Module

Step 1. Generate random samples of specified size *n* from digits 0 through 9.

Let's begin with samples of size $n = 5$.

Excel has a random number function, RAND(), that will generate random numbers from 0 to 1, including 0 but not including 1. We can generate random numbers from 0 to 10, including 0 but not including 10 by multiplying RAND() by 10. *Excel* also has an integer function, INT(), which will find the greatest integer value smaller than the argument value between the parentheses. For example INT(4.98) will return a value of 4. Thus, the command INT(10*RAND()) will return the integer values from 0 to 10, including 0 but not including 10. That is, INT(10*RAND()) will return a random integer from 0 to 9, including both 0 and 9.

In cell A1, type in =INT(10*Rand()), and press **Enter**. Highlight cells A1 through E1. Click **Edit,** and choose the **Fill** and **Right** options. When this is done, cells A1 through E1 will be filled with integer values from 0 to 9, inclusive. This represents our sample of size 5.

Note: As further entries are made in this spreadsheet, the random values previously determined will be regenerated. That is, the values currently in cells A1 to E1 will change as further entries are added.

Step 2. Compute the sample mean $\bar{x}$ of this sample.

We will use *Excel's* AVERAGE() function to compute $\bar{x}$. In cell G1, type in =AVERAGE(A1:E1) and press **Enter**. The average of the values in cells A1 to E1 now appears in cell G1.

Step 3. Generate an $\bar{x}$ distribution.

Highlight cells A1 through G100. Click on **Edit** and choose the **Fill** and **Down** options. The 100 entries down the G column constitute a sample $\bar{x}$ distribution.

Step 4. Find the mean and standard deviation of this $\bar{x}$ distribution. That is, find $\bar{x}_x$ and S_x.

Click on cell I1; type in =AVERAGE(G1:G100), and press **Enter**. Click on cell I3; type in =STDEV(G1:G100), and press **Enter**. The mean of the $\bar{x}$ distribution now appears in cell I1, and the standard deviation of $\bar{x}$ distribution now appears in cell I3.

Step 5. Tally the number of $\bar{x}$ values that fall from 0 to 1, 1 to 2, 2 to 3, . . . , 9 to 10.

This requires some slick maneuvering. *Excel* does not have a function that will allow us to directly make such tallies. Thus, we will find these counts with a two-step process. First, we will use the COUNTIF() function.

Click on cell I5, type in =COUNTIF(G1:G100,"<1"), and press **Enter.**
Click on cell I6, type in =COUNTIF(G1:G100,"<2"), and press **Enter.**
Click on cell I7, type in =COUNTIF(G1:G100,"<3"), and press **Enter.**
Click on cell I8, type in =COUNTIF(G1:G100,"<4"), and press **Enter.**
Click on cell I9, type in =COUNTIF(G1:G100,"<5"), and press **Enter.**
Click on cell I10, type in =COUNTIF(G1:G100,"<6"), and press **Enter.**
Click on cell I11, type in =COUNTIF(G1:G100,"<7"), and press **Enter.**
Click on cell I12, type in =COUNTIF(G1:G100,"<8"), and press **Enter.**
Click on cell I13, type in =COUNTIF(G1:G100,"<9"), and press **Enter.**
Click on cell I14, type in =COUNTIF(G1:G100,"<10"), and pre**ss Enter.**
Click on cell J5, type in =I5, and press **Enter.**
Click on cell J6, type in =I6-I5, and press **Enter.**
Highlight cells J6 through J14; click on **Edit,** and choose options **Fill** and **Down.**

The values in cells J6 through J14 now contain the desired tallies.

Step 6. Construct a chart of the tallies.

Highlight cells J5 through J14. Click the **Insert** menu and the **Chart** option. From the Chart type: menu, choose **XY (Scatter),** and from the Chart sub-type, click on the icon in the second row second column. (When you click on this icon, the text below the images will say, "Scatter with points connected by smoothed Line without markers.") Click **Finish,** and the chart will be created.

The graph should appear relatively mound shaped and symmetrical (that is, approximately normal).

Recall that when further entries are made in unused cells, the random numbers are regenerated. Click on an unused cell (such as K1). Press **Delete**. The random numbers are regenerated, and the chart is adjusted accordingly. Repeat this process as often as you like to see that the $\bar{x}$ distributions are approximately normal.

Step 7. Repeat this process with randomly generated samples of size n = 10, 20, 30, 40, 50, and so on.

We do not need to start from scratch in order to repeat this process with randomly generated samples of size n = 10. We simply need to insert five additional columns of randomly generated numbers within the five we already have. Highlight column B by clicking on the "B" at the top of the column. Click the **Insert** menu and the **Columns** option. A blank column is inserted. Insert four more blank columns using the same procedure. Columns B through F should now be blank. Highlight columns A through F by clicking on the "A" at the top of the A column and moving to the right until columns

A through F are highlighted. Click on the **Edit** menu, and choose options **Fill** and **Right**. The blank columns should be filled in with random numbers from 0 to 9.

Scroll over to column L. The entries in column L are now the sample means of the samples of size 10. Likewise, the entries in columns N and O are updated using the information from the samples of size 10. Also, the chart of the tallies is updated.

Use the exact same procedure for randomly generated samples of size 20, 30, 40, etc.