# DEMONSTRATING THE CENTRAL LIMIT THEOREM IN THE CLASSROOM: AN EXCEL EXERCISE[1]

Hilde Patron[2], William J. Smith[3], and David Boldt[4]

**ABSTRACT**

The Central Limit Theorem (CLT) states that regardless of the underlying distribution, the distribution of the sample means approaches normality as the sample size increases. Understanding the CLT is central to other concepts presented in a business statistics class; however, the CLT remains one of most misunderstood concepts in introductory statistics. This paper describes the steps for developing a Microsoft Excel spreadsheet that allows students to visualize the CLT. In addition to introducing students to creative uses of spreadsheets, we show that student learning is enhanced by incorporating this pedagogical exercise in the classroom.

Key Words: Central Limit Theorem, Spreadsheet Applications, Student Learning

JEL Classification: A22

## Introduction

The Central Limit Theorem (CLT) is one of the most important concepts introduced in a business or economics statistics class and is the foundation for much of the decision-making tools used in the business world. The power of the CLT is that it greatly reduces the information and effort required to make decisions. Basically, the theorem says that even if the population distribution is not normally shaped, the sampling distribution, or the distribution of the sample means, will be approximately normal (Brightman, 1986). The larger the sample size, the more normal the sampling distribution will become. For sample sizes larger than 30, the sampling distribution from almost any population will be normally shaped. Furthermore, since convergence of the sampling distribution to a normal distribution does not depend on the size of the population from which the sample is drawn, larger samples are not required for larger populations to achieve normality. As pointed out by Hardy (2002), other important results of the Central Limit Theorem are:

1. The sample mean is itself a random variable with its own distribution.
2. The mean of the probability distribution of the sample mean is the same as the mean of the probability distribution from which the sample is drawn.

[1] We would like to thank Lee Hoke, University of Tampa, for his helpful suggestions on improving earlier versions of this paper.
[2] Assistant Professor of Economics, Department of Economics, University of West Georgia, 1601 Maple Street, Carrollton, GA 30118.
[3] Assistant Professor of Economics, Department of Economics, University of West Georgia, 1601 Maple Street, Carrollton, GA 30118.
[4] Professor of Economics, Department of Economics, University of West Georgia, 1601 Maple Street, Carrollton, GA 30118.

3. The sample mean is less variable than individual observations from the population from which the sample is drawn.
4. The standard deviation of the sample mean is equal to $\frac{s}{\sqrt{n}}$.

Without an understanding of the Central Limit Theorem, it would be difficult for a researcher to persuade himself or herself that population statistics could be effectively derived from sample statistics, especially when the population is known to be non-normally distributed. A major obstacle to understanding the concepts surrounding the CLT is that students typically are faced with a single sample and do not understand that the sample mean is itself a random variable.

In this paper, we outline a set of steps to produce a very intuitive example of the CLT using Microsoft Excel. There are examples of pedagogical tools that use spreadsheets or programming languages such as Java or JavaScript to present the CLT in a visual manner. Our Excel-based application improves on the currently available tools by simplifying the construction of the spreadsheet and by providing instructions that makes purpose-specific modification of the spreadsheet easy. A spreadsheet-based approach is an excellent way of demonstrating the CLT as it reinforces as well as builds on the spreadsheet knowledge students have gained in previous courses. In addition, the construction of the spreadsheet could be used as part of the students' assignment, whereas other web-based tools require substantial programming knowledge to duplicate, and thus are not practical for students to produce themselves for a classroom exercise. In the section that follows, we discuss a variety of web-based tools that provide useful demonstrations of the CLT. Subsequently, we outline a step-by-step procedure for building a CLT demonstration model using an Excel spreadsheet. After describing the model development, we then present results of an empirical examination of the effectiveness of this Excel-based simulation of the CLT on student learning. The final section of the paper summarizes the paper and provides insights into other possibilities for other Excel-based tools that might be integrated into an introductory business or economics statistics course.

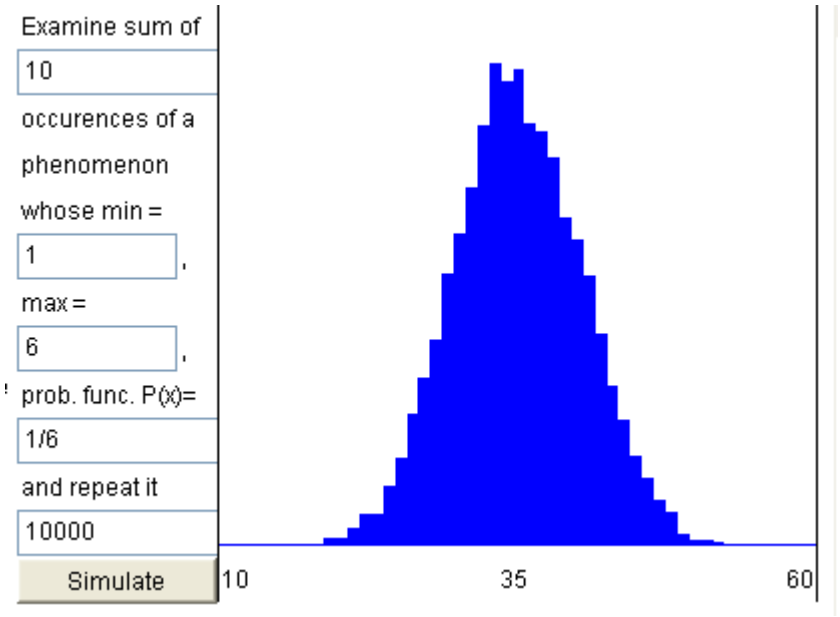
**Using Online Sources to Demonstrate the Central Limit Theorem**

There are a variety of on-line sources that allow teachers and students to simulate an assortment of distributions and to visualize the effect of these different distributional assumptions in the context of the CLT. However, almost without exception, on-line sources either build their CLT demonstration applications using something other than a simple spreadsheet or may not provide the source code. Furthermore, because these educational web applications are written in languages such as Java™, JavaScript™, PHP or some other web-based programming language, their usefulness as a pedagogical tool is somewhat diminished in a business statistics course. It is typically far beyond the capacity of introductory statistics students to produce code that would replicate what can be found at these sites. Furthermore, web-based applications require an internet connection to be effectively used in a classroom environment. A demonstration built in Excel can be used with or without an internet connection. As long as the instructor or student has access to a computer with Excel, they have access to all they require to build, modify and use their own CLT demonstration tool.

Although web-based demonstration tools have their limitations, many provide a very sophisticated assortment of controls that allow the user to change distributions or parameters of a distribution with a mouse click. As an example, Janro Elonen (2002) has developed a Java™

applet that produces a simulation of dice rolls (see Figure 1). A number of other similar teaching tools are also accessible via the internet.[5]

**FIGURE 1**




**Using Excel to Demonstrate the Central Limit Theorem**

Excel, as well as other spreadsheet programs, has the ability to produce random numbers under a variety of distributional assumptions. Because the power of the CLT is its ability to handle non-normal distributions our examples focus on these types of distributions.

To simulate the CLT there are four basic steps: First, we must simulate collecting a finite sample of size n from an infinite population (N). We do this by generating n random numbers. Second, to generate a graph approximating the sampling distribution, we repeat the first step over multiple columns in the Excel spreadsheet. Third, once the multiple samples are generated, we calculate the sample mean of each sample. Fourth, we produce a graph or histogram of the means (Middleton, 2004).

To demonstrate how this could be used in class, we will first consider a manufacturing example. In this example, suppose a wire manufacturing plant is producing 50 foot electrical extension cords. The actual lengths of extension cords are uniformly distributed and range between 49.5 feet and 50.5 feet. Although the distribution of the entire population of extension cords manufactured by this plant is uniformly distributed, our example will demonstrate that the

[5] Users of Elonen's CLT applet can modify various parameters of the distribution, such as the number of observations in the sample of rolls, the number of sides on the dice (i.e., the pdf), and the number of times the sampling is repeated (Elonen, 2002). A similar on-line tool has been developed by Charles Stanton (2008), a Professor of Mathematics at California State University as San Bernardino. His demonstration tool also uses dice as an example. The Applet allows the user to choose among a pre-specified number of dice and number of rolls. As the user plays the applet over and over, the sampling distribution is built. There are several other examples of on-line tools (Annis, 2008, Rogers et al., 2001, Scott et al., 2008, Wachsmuth, 2003, and Tsai and Wardell, 2006).

means of the samples generated will be approximately normally distributed. Furthermore, by using Excel, we can dynamically demonstrate to a class how repeating the sampling produces similarly shaped sampling distributions each time we draw a new set of random samples. Appendix A provides the step-by-step instructions for developing the MS Excel application.

**Evaluation of the CLT Exercise Tool**

In this section, we evaluate our teaching tool using the students' performance on classroom exercises developed in earlier sections. We seek to answer three questions:

1. Do students feel that experience with the simulation improves their understanding of the Central Limit Theorem?
2. Does interaction with the simulation really improve student understanding?
3. Is active interaction with the simulation more or less effective than passive interaction?

To answer these questions we first randomly assign classes to different treatment groups. Because the treatment involves classroom instruction, it is infeasible to randomly assign individual students to groups, because individual students cannot be randomly excluded or included in lectures. The groups' sizes were matched as closely as possible, however since the control and treatment groups were based on classes, we had to allow for slight differences in the number of students participating in each group.

Group 1: consists of 30 students who observed the instructor demonstrate the CLT simulation. They were also present during the regular lecture about the CLT.

Group 2: consists of 33 students who developed the simulation of the CLT in Excel by themselves based on the instructions in this article (no teacher demonstration). Students in this group did not observe the instructor demonstration of the CLT simulation and may not have attended the regular theory lecture on the CLT.

Group 3: consists of 43 students who were not present during the lecture on the CLT or the in-class demonstration of the excel simulation of the CLT. Also, these students did not develop the simulation of the CLT in Excel by themselves.

Even though student perception of the benefits from this exercise may or may not reflect changes in actual performance, determining a baseline for student perception is of potential importance for evaluating future improvements in this Excel-based exercise, therefore we survey the students about their perceived benefits from this exercise. Students who watched the simulation were asked a simple question: "Did witnessing the simulation in Microsoft Excel by your professor improve your understanding of the CLT?" Of the respondents, 29 out of the 30 (96.67%) answered "yes." Students who developed the Excel exercise themselves were also asked a similar question: "Did completing the Excel simulation help you better understand the implications of the Central Limit Theorem?" Of the respondents, 25 of the 33 students in this group (75.76%) answered "yes". Students in this group were also asked if the instructions to

complete the file were easy to follow, and if they thought that the exercise was a good complement to the regular CLT lecture. Of the responses, 75.76% and 78.79%, respectively, were positive. It would appear that students feel more confident of their understanding of the CLT after viewing or completing the simulation.

Although confidence is important when implementing new statistical concepts, confidence alone can be a dangerous thing. To answer the second question, i.e., whether the simulation actually improved student understanding, all three groups were required to take a quiz. The quiz was administered online through the campus edition of WebCT Vista. Out the 106 students registered in the class, 90 completed the quiz. The quiz consisted of several questions in addition to those focused on the CLT. However, one of the questions focused distinctively on concepts developed in the CLT exercises, and more specifically on the shape of the sampling distribution of the sample means as the sample size increases.

We used the score on this response to measure both the student understanding of the CLT and the benefits associated with the Excel teaching tool we developed. If the student answered the question correctly, he/she scored a 1. Otherwise that score was labeled 0. This variable is called a BINOMIAL RESPONSE. Our goal was to determine whether the students in group 1 performed significantly better than the other two groups. Students in this group were identified with a dummy variable GROUP 1, which equals 1 if the student was in group 1 and 0 otherwise. We also included in our models for control purposes the student's actual grade point average (GPA). Since this is a classroom-based environment, we cannot randomize treatments within a classroom, thus we estimate the model both with and without GPA to address the potential for unobserved ability or motivation-related factors. Table 1 presents summary statistics of these variables for the common sample used in the estimations.

**Table 1:**
**Descriptive Statistics**

| | Mean | Median | Max. | Min. | St. Dev. | Sum | Obs. |
|---|---|---|---|---|---|---|---|
| **BINOMIAL RESPONSE (full sample)** | 0.72 | 1.00 | 1.00 | 0.00 | 0.45 | 65.00 | 90 |
| **For Group 1 only** | 0.90 | 1.00 | 1.00 | 0.00 | 0.31 | 26.00 | 29 |
| **For Group 2 only** | 0.65 | 1.00 | 1.00 | 0.00 | 0.49 | 20.00 | 31 |
| **For Group 3 only** | 0.63 | 1.00 | 1.00 | 0.00 | 0.49 | 19.00 | 30 |
| **GPA (full sample)** | 2.89 | 2.85 | 4.00 | 1.97 | 0.49 | 259.89 | 90 |
| **For Group 1 only** | 3.14 | 3.11 | 4.00 | 1.97 | 0.50 | 91.04 | 29 |
| **For Group 2 only** | 2.84 | 2.81 | 3.60 | 2.12 | 0.45 | 88.00 | 31 |
| **For Group 3 only** | 2.70 | 2.61 | 3.85 | 2.00 | 0.43 | 80.85 | 30 |
| **GROUP 1** | 0.32 | 0.00 | 1.00 | 0.00 | 0.47 | 29.00 | 90 |
| **GROUP 2** | 0.34 | 0.00 | 1.00 | 0.00 | 0.48 | 31.00 | 90 |
| **GROUP 3** | 0.33 | 0.00 | 1.00 | 0.00 | 0.47 | 30.00 | 90 |

An analysis of Table 1 shows that the proportions of students in groups 1, 2 and 3 who answered the CLT question correctly are approximately 90%, 65% and 63%, respectively. This is a first indication that students in group 1 performed better than students in the other two groups. In what follows we explore this possibility more formally. We first do hypotheses tests for differences in proportions. Letting $\pi_1$ ($\pi_2$) denote the proportion of students in group 1 (2) who answered the quiz question correctly we formulate the following null and alternative hypotheses:

$$H_0: \pi_1 = \pi_2$$

$$H_\alpha: \pi_1 \neq \pi_2$$

To test the hypothesis we calculate a z test,

$$z = \frac{\pi_1 - \pi_2}{\sqrt{\frac{\pi_p(1-\pi_p)}{n_1} + \frac{\pi_p(1-\pi_p)}{n_2}}}$$

where $\pi_p$ represents the proportion of students in both groups who answered the question correctly and $n_1$ ($n_2$) denotes the number of students in group 1 (2). We do similar pair wise tests for difference of proportions between the three groups of students. Results are summarized in the table below:

**Table 2:**
**z-Test for Difference of Proportions in Binomial Responses**
(p-values in Parenthesis)

| | **Z test (p-value)** |
|---|---|
| **GROUP 1 and GROUP 2** | -2.30 (0.02) |
| **GROUP 1 and GROUP 3** | -2.13 (0.03) |
| **GROUP 2 and GROUP 3** | -0.10 (0.92) |

As Table 2 shows, the proportion of students in group 1 who answered the CLT question correctly is significantly different (higher) than the proportion of students from groups 2 and 3. That is, we clearly reject the null hypotheses that $\pi_1 = \pi_2$ or $\pi_1 = \pi_3$. However, we cannot reject the null hypothesis that the proportions of students in groups 2 and 3 who answered the CLT question correctly are the same. This suggests that students in group 1 (students who were present when the instructor demonstrated the CLT simulation) performed better than students who did the simulation themselves or who did not have any experience with the simulation, but that firsthand experience with the simulation without instructor guidance does not improve student performance.

To determine the benefits of the simulation in terms of student learning, we first calculated simple correlation coefficients between the performance variable (BINOMIAL RESPONSE) and the explanatory variables. As Table 3 illustrates, there is a positive correlation between students in group 1 and the BINOMIAL RESPONSE. Furthermore, this relationship becomes stronger when we exclude from the sample the second group of students. This suggests to us that interaction with the Excel exercise has a positive impact on student learning, especially

so when the student witnesses the instructor doing the exercise, more so than when the student performs the simulation by himself/herself. This is similar to the findings of other studies of computer-based statistical demonstrations of the CLT (Hagtvedt et. al. 2007).

**Table 3:**
**Correlation Coefficients**
(p-values in Parenthesis)

| BINOMIAL RESPONSE | Full Sample | Restricted Sample (Groups 1 & 3) | Restricted Sample (Groups 1 & 2) | Restricted Sample (Groups 2 & 3) |
|---|---|---|---|---|
| **GROUP 1** | 0. 27<br>(0.01) | 0.31<br>(0.02) | 0.30<br>(0.02) | NA |
| **GROUP 2** | -0.12<br>(0.24) | NA | -0.30<br>(0.02) | 0.01<br>(0.93) |
| **GROUP 3** | -0.14<br>(0.19) | -0.31<br>(0.02) | NA | -0.01<br>(0.93) |
| **GPA** | 0.36<br>(0.00) | 0.32<br>(0.01) | 0.32<br>(0.01) | 0.42<br>(0.00) |

Even though the simple correlation suggest a positive relationship between watching the instructor do the Excel simulation and learning, to measure this relationship appropriately we ran logit models. Table 4 shows the results from the logit model. Columns 1 and 2 show the estimated coefficients and marginal effects of the model estimated using all observations. Columns 3 and 4 show the coefficients and marginal effects of a logit estimation using a restricted sample of 59 students. This restricted sample excludes students from group 2. Columns 5 and 6 show the coefficients for the restricted sample for the 60 students in groups 1 and 2.

The first column in Table 4 shows that there is a positive and significance relationship between BINOMIAL RESPONSE and students in GROUP 1. The second column further shows that students who observe the instructor demonstrate the Excel simulation have a 26 percent higher chance of answering the question correctly. For the restricted samples (columns 3 through 6), the coefficients of GROUP 1 are still positive and significant. This seems to suggest that watching the instructor do the demonstration or doing the demonstration oneself has a similar positive impacts on student learning.

**Table 4:**
**Logit Models**
(p-values in parenthesis)

| | **Dependent Variable: BINOMIAL RESPONSE** | | | | | |
|---|---|---|---|---|---|---|
| | Full Sample | | Restricted Sample (Includes groups 1 & 3) | | Restricted Sample (Includes groups 1 & 2) | |
| | Coef. | Marginal effect | Coef. | Marginal effect | Coef. | Marginal effect |
| **CONST.** | 0.57 | 0.11 | 0.55 | 0.09 | 0.60 | 0.10 |
| | (0.03) | (0.01) | (0.15) | (0.10) | (0.11) | (0.07) |
| **GROUP 1** | 1.59 | 0.26 | 1.61 | 0.26 | 1.56 | 0.25 |
| | (0.02) | (0.00) | (0.02) | (0.01) | (0.03) | (0.01) |
| **Log Likelihood** | -49.53 | | -29.36 | | -29.81 | |
| **LR Statistic** | 7.30 | | 5.94 | | 5.58 | |
| **Actual 1s and 0s correctly predicted** | 72.22% | | 76.27% | | 76.67% | |
| **Number of Observations** | 90 | | 59 | | 60 | |

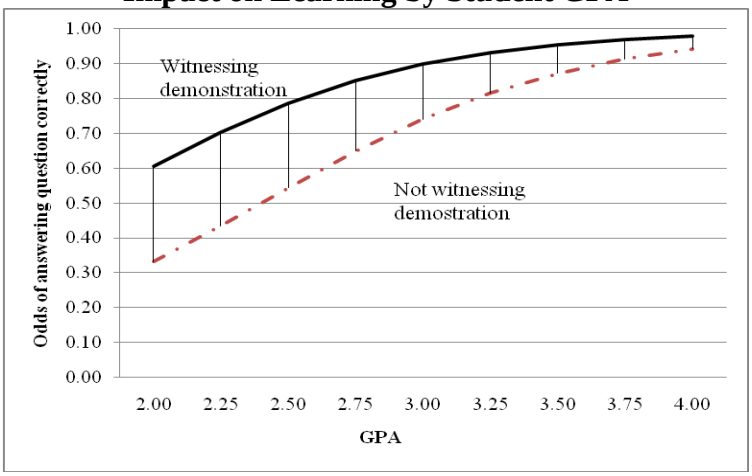
Even though the results provide evidence that viewing the simulation increases learning, it is possible that those students who were present when the instructor demonstrated the CLT were the same students who came to class regularly and who were highly motivated. To account for this motivation (as well as the student's natural ability) we included the student's incoming grade point average in the estimations. Results are summarized in Table 5. Columns 1 and 2 show the estimates using the full sample of students. GPA has indeed a positive and significant effect on the probability that a student answers the question correctly. Furthermore, by including GPA in the models, GROUP 1 loses strength and significance. The marginal effect of GROUP 1 on BINOMIAL RESPONSE is now only 18 percent. Since the variables GROUP 1 and GPA are positively correlated (correlation coefficient =0.31, p-value=0.02), it is likely that including GPA underestimates the effect of GROUP 1 on BINOMIAL RESPONSE. Results of the restricted model (columns 3 through 6) are similar to the full sample model.

**Table 5:**
**Expanded Logit Models**
(p-values in parenthesis)

| | Dependent Variable: BINOMIAL RESPONSE | | | | | |
|---|---|---|---|---|---|---|
| | Full Sample | | Restricted Sample (Includes groups 1 & 3) | | Restricted Sample (Includes groups 1 & 2) | |
| | Coef. | Marginal effect | Coef. | Marginal effect | Coef. | Marginal effect |
| **CONSTANT** | -4.20 | -0.73 | -2.96 | -0.46 | -3.42 | 0.52 |
| | (0.02) | (0.01) | (0.17) | (0.16) | (0.11) | (0.10) |
| **GROUP 1** | 1.13 | 0.18 | 1.14 | 0.18 | 1.26 | 0.19 |
| | (0.11) | (0.06) | (0.14) | (0.12) | (0.09) | (0.07) |
| **GPA** | 1.75 | 0.31 | 1.32 | 0.20 | 1.43 | 0.22 |
| | (0.01) | (0.00) | (0.10) | (0.08) | (0.06) | (0.05) |
| **Log Likelihood** | -45.15 | | -27.89 | | -27.87 | |
| **LR Statistic** | 16.06 | | 8.88 | | 9.46 | |
| **Actual 1s and 0s correctly predicted** | 75.56% | | 76.27% | | 76.67% | |
| **Obs.** | 90 | | 59 | | 60 | |

We conclude from the various models that in fact student performance increases with experience with the CLT Excel exercise. There is also some evidence that observing the instructor demonstration of the exercise increases the chances of a correct response. We also find that poorer students (those with a lower GPA) have more to gain from witnessing the presentation. Figure 2 depicts the difference in the expected probability of correct responses by GPA. Students with lower GPAs experience a greater increase in the probability of a correct response if they observe the classroom demonstration than students with higher GPAs.

**Figure 2:**
**Impact on Learning by Student GPA**



**Conclusions**

The Central Limit Theorem is a critical concept that business students should master because of its application in areas such as hypothesis testing, regression analysis, and in the business world in general (e.g., Six Sigma). This paper provides a set of instructions for developing a Microsoft Excel-based demonstration tool that can be used either in the classroom or in a student assignment. In either case, it can be used to provide insight into the inner-workings of this fundamental concept. Furthermore, by focusing this exercise on Excel, this pedagogical tool accomplishes two goals in many business statistics classes:

1. it demonstrates a statistical concept, and
2. it does so in a commonly used tool that most business school students are required to master in their degree programs.

By using Excel as the medium of instruction, students are able to learn this important concept, while practicing basic spreadsheet skills, and developing a familiarity with more advanced spreadsheet features like formulas and sampling. The benefit to the instructor is that this exercise accomplishes several pedagogical goals in a reasonably simple package, while remaining flexible enough to accommodate a variety of different real-world applications. Our intent in the future is to set up Excel modules to demonstrate other statistical concepts such as interval estimation as well as hypothesis testing and make these modules publicly available. In addition, we hope to further test the effect of this teaching tool on student performance.

## Appendix A: Step-By-Step Procedures for Using Excel to Demonstrate the CLT

### *Step 1: Creating the Random Samples*

Since the extension cord lengths are uniformly distributed, we need to first create a uniform random number between 49.5 and 50.5 feet. This is accomplished by using the RAND function in Excel. Let us assume for now that we intend to calculate the means for 30 samples of size 10. In Figure A1 below, the function =49.5+RAND() is entered in cell B2 to generate a uniform random number between 49.5 and 50.5. (Note that generating a uniform random number with a wider range can be accomplished by multiplying the RAND() portion of the function by the desired range width. Adding a number other than 49.5 will change the lower bound. The general function for any uniform distribution is =X+Y*RAND() where X is the choice of the lower bound of the distribution and Y is the choice of the range of the distribution.)

**Figure A1**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | SAMPLE # | Ob1 | Ob2 | Ob3 | Ob4 |
| 2 | 1 | =49.5+RAND() | | | |
| 3 | 2 | =49.5+RAND() | | | |
| 4 | 3 | =49.5+RAND() | | | |
| 5 | 4 | =49.5+RAND() | | | |
| 6 | 5 | =49.5+RAND() | | | |
| 7 | 6 | =49.5+RAND() | | | |
| 8 | 7 | =49.5+RAND() | | | |
| 9 | 8 | =49.5+RAND() | | | |
| 10 | 9 | =49.5+RAND() | | | |
| 11 | 10 | =49.5+RAND() | | | |
| 12 | 11 | =49.5+RAND() | | | |
| 13 | 12 | =49.5+RAND() | | | |
| 14 | 13 | =49.5+RAND() | | | |
| 15 | 14 | =49.5+RAND() | | | |
| 16 | 15 | =49.5+RAND() | | | |
| 17 | 16 | =49.5+RAND() | | | |
| 18 | 17 | =49.5+RAND() | | | |

### *Step 2: Copying*

The cell B2 is then copied down to cell B31, and the contents of these thirty cells are copied to the adjacent cells in column C through column K. Each row from B2:K2 to B31:K31 represents thirty uniform random samples.

### *Step 3: Calculate the Sample Means*

To calculate the sample means for each of the thirty samples, we simply enter the function =AVERAGE (B2:K2) in cell L2, and then copy the function in L2 into all the cells down to L31. The result should look like Figure A2 (sample 17-30 not shown). It should be noted that entering anything into an empty cell and pressing Enter or pressing the F9 key will cause each of the samples to be re-calculated, creating thirty new samples each time. The previous section provides a very clear example of how the distribution of the sample means approximates a normal distribution, even though the underlying data generating process is uniformly distributed, not normally distributed. The following section provides some useful extensions to the exercise above to expand the discussion beyond the uniform distribution.

**Figure A2**

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SAMPL | Ob1 | Ob2 | Ob3 | Ob4 | Ob5 | Ob6 | Ob7 | Ob8 | Ob9 | Ob10 | Sample Means | |
| 2 | 1 | 49.89 | 49.61 | 49.60 | 50.44 | 50.10 | 49.55 | 49.82 | 50.01 | 49.77 | 50.14 | 49.89 | |
| 3 | 2 | 50.33 | 50.42 | 49.82 | 49.77 | 49.52 | 49.62 | 49.65 | 50.18 | 50.14 | 50.44 | 49.99 | |
| 4 | 3 | 50.19 | 49.92 | 50.41 | 49.77 | 50.13 | 49.98 | 49.92 | 50.23 | 50.46 | 50.00 | 50.10 | |
| 5 | 4 | 50.25 | 50.13 | 50.13 | 49.87 | 49.93 | 49.56 | 50.09 | 50.49 | 50.43 | 49.93 | 50.08 | |
| 6 | 5 | 50.16 | 50.06 | 49.76 | 49.63 | 49.77 | 49.60 | 50.39 | 50.02 | 50.13 | 50.49 | 50.00 | |
| 7 | 6 | 49.61 | 50.35 | 50.06 | 50.34 | 50.37 | 49.77 | 49.90 | 50.01 | 49.89 | 49.77 | 50.01 | |
| 8 | 7 | 50.35 | 50.14 | 49.65 | 50.21 | 49.71 | 50.22 | 50.20 | 50.40 | 50.48 | 49.89 | 50.13 | |
| 9 | 8 | 50.11 | 49.62 | 50.05 | 49.61 | 49.79 | 50.25 | 49.79 | 49.75 | 49.73 | 50.10 | 49.88 | |
| 10 | 9 | 50.13 | 50.02 | 50.00 | 49.94 | 50.05 | 50.31 | 50.09 | 50.25 | 49.89 | 50.19 | 50.09 | |
| 11 | 10 | 50.37 | 49.98 | 50.15 | 49.86 | 50.19 | 49.62 | 50.40 | 49.62 | 49.62 | 49.60 | 49.94 | |
| 12 | 11 | 49.69 | 49.60 | 49.67 | 49.96 | 49.67 | 50.10 | 50.06 | 50.43 | 50.35 | 50.23 | 49.98 | |
| 13 | 12 | 50.22 | 50.22 | 49.53 | 49.98 | 50.14 | 49.91 | 49.93 | 49.55 | 50.17 | 49.94 | 49.96 | |
| 14 | 13 | 50.11 | 49.80 | 50.23 | 49.68 | 50.11 | 49.97 | 50.30 | 50.41 | 50.38 | 49.66 | 50.07 | |
| 15 | 14 | 50.16 | 50.27 | 49.64 | 50.29 | 50.32 | 50.23 | 50.12 | 49.51 | 50.45 | 50.23 | 50.12 | |
| 16 | 15 | 50.13 | 50.07 | 50.33 | 49.75 | 50.00 | 49.66 | 49.93 | 49.62 | 50.42 | 50.03 | 49.99 | |
| 17 | 16 | 50.38 | 50.34 | 49.66 | 49.58 | 50.36 | 50.29 | 49.68 | 50.05 | 49.59 | 49.57 | 49.95 | |

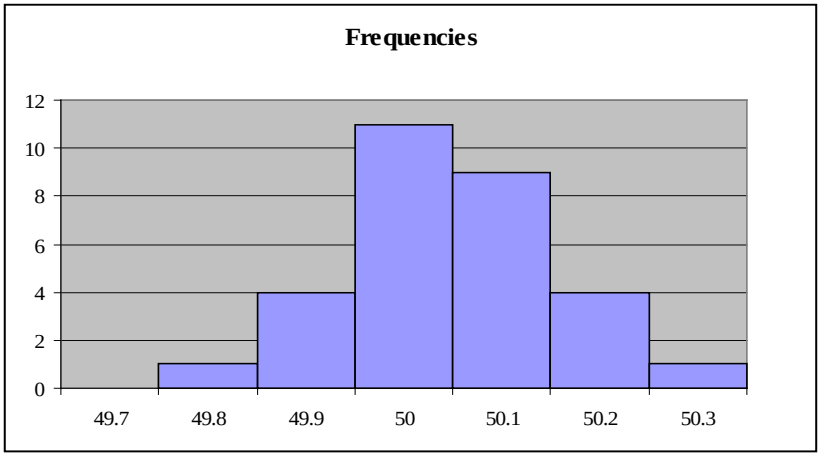
***Step 4: Producing the Histogram (Graph) of the Sample Means***

Although a histogram can be produced using a variety of methods, only one method is used here for simplicity. In Excel, place your curser in cell M2, press Shift and press the down arrow until cells M2 through M8 are selected (seven cells should be highlighted). Then type the following:

=FREQUENCY(L2:L31,{49.7, 49.8, 49.9, 50, 50.1, 50.2, 50.3})

After you have typed this function, hold down Ctrl+Shift and press Enter. This calculates the frequency of the sample means in each of the bins defined in the function above, starting with 49.7 and going through 50.3. There are seven bins in this example (49.7 to 50.3) so, seven cells were required. Type the bins used in your frequency function starting with the value 49.7 in cell N2 and going down next to each frequency calculation for use as labels. Once the frequencies are calculated, and the bin labels are typed in, use the Chart Wizard to graph the histogram. The bin values are used as the labels for the X-axis. The results should look something like Figure A3 below. To remove the space between the bars in the graph, right click on a bar in the graph and select Format Data Series, then select the Options tab, and set the Gap width to "0." To resample, press F9.

**Figure A3**




The previous section provides a very clear example of how the distribution of the sample means approximates a normal distribution, even though the underlying data generating process is uniformly distributed, not normally distributed. The following section provides some useful extensions to the exercise above to expand the discussion beyond the uniform distribution.

***Step 5 (OPTIONAL): Extensions of the CLT Lesson with Other Distributions***

Excel has several options for different distributional assumptions. This could be very helpful in trying to simulate a specific process with a known population distribution, such as flipping a coin (Bernoulli). The values of a Bernoulli distributed random variable can take on two values, 0 and 1. The Bernoulli distribution is useful in predicting discrete outcomes like success versus failure. To simulate a random draw from a Bernoulli distribution with a 0.5 probability of success or failure uses the following formula in cell:

=ROUND((RAND()),0)

Recall that, in Excel, the RAND function returns a uniform random number between 0 and 1. By using the ROUND function and eliminating the decimal, a uniform random variable is converted to one that is Bernoulli distributed, with a mean of 0.5 (a coin toss).

A discrete distribution of any range could be simulated by adding the minimum value to and multiplying the RAND() portion of the equation above by R-1, where R is the range. For example, the roll of a die, a discrete distribution that takes on integer values between 1 and 6 can be simulated by the following:

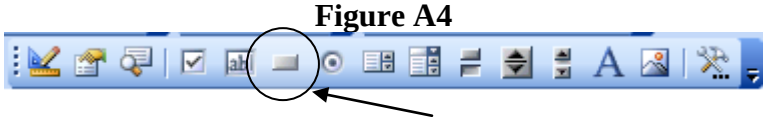
=ROUND(1+5*(RAND()),0)

***Step 6 (OPTIONAL): Using Visual Basic and Buttons to Simplify the Process***

In addition to building a CLT demonstration spreadsheet, there are simple methods for increasing the interactivity of the spreadsheet, thus making it easier to use and modify. Buttons and scrollbar can be used to change parameters of a distribution or even to populate (re-populate) items in a spreadsheet. Included in Excel's repertoire of tools is the Control Toolbox (Figure A4). The Control Toolbox allows the user to quickly add control features such as buttons, and assign these buttons a specific task. To access the control toolbox, select View, Toolbars, and

select Control Toolbox. To add a button, right click on the Button form on the Control Toolbar, then click and drag on the spreadsheet the size button desired.

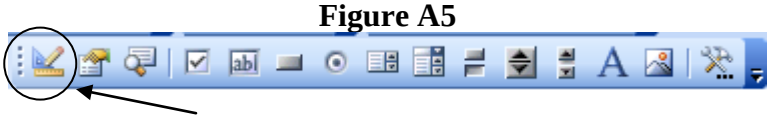
**Figure A4**



Once the button is created, specific commands can be assigned to it by double clicking on the button. Below is code that can be used to fill in the region of cells from B2 to L31 with a Bernoulli random variable with the click of the button. Only the bolded portion is required, since the un-bolded code appears by default.

```
Private Sub CommandButton1_Click()
Range("$B$2:$L$31") = "=ROUND((Rand()),0)"
        Range("$B$2:$L$31").Select
End Sub
```

**Figure A5**



Once the code has been added, close the code window and click on the Design Mode button (Figure A5) to toggle out of design mode and to activate the button for use. These steps can be repeated for a variety of different distributions to allow the user to choose different distributions without having to go through the creation process outlined earlier. By using the same range of cells and different distributions, quick comparisons can be made between the properties of one distribution and another.

**References**

Annis, C. "Central Limit Theorem," 2008. Accessed at http://www.statisticalengineering.com/central_limit_theorem.htm, Accessed January 2008.

Brightman, H. J. 1986 *Statistics in Plain English*. Cincinnati: South-Western Publishing Co.

Elonen, Jarno. 2002. "Central Limit Theorem Explained," Finland, 2002, Accessed at http://elonen.iki.fi/articles/centrallimit/index.en.html#demo, Accessed August, 2007.

Hagtvedt, R., Jones, G. T., and Jones K. 2007. "Pedagogical Simulation of Sampling Distributions and the Central Limit Theorem," *Teaching Statistics*. 29.3, (August): pp 94-97.

Hardy, M. E. 2002. "Repeated Simulated Sampling in Excel as a Tool for Teaching Statistics." *Journal of Computing Sciences in Colleges*. *17*.5, (April): pp. 167-74.

Middleton, M. R. 2004. *Data Analysis Using Microsoft Excel*. Belmont, CA: Duxbury Press.

Rogers, T., Rogers, M., and Rogers, S. R. 2001 "The Central Limit Theorem – How to Tame Wild Populations," Retrieved from http://intuitor.com/statistics/CentralLim.html, Retrieved September 2007.

Scott, D., Benway, J., Lu, J., Tang, Z., Shea, A., Quinones, M., Baggerly, K., Austin, J., Swartz, M., and Swartz, R., 1998. " Rice Virtual Lab in Statistics," Accessed at http://onlinestatbook.com/rvls.html, Accessed September 2007.

Stanton, C. "The Central Limit Theorem: Java Probability Applets."Accessed at http://www.math.csusb.edu/faculty/stanton/probstat/clt.html, Accessed January 2008.

Tsai, W. and Wardell, D. 2006 . *An Interactive Excel VBA Example for Teaching Statistical Concepts*, *INFORMS Transactions on Education* 7:1. Accessed at http://ite.pubs.informs.org/Vol7No1/TsaiWardell/ Accessed January, 2008.

Wachsmuth, Bert G. " 2003. "Seton Hall University Math & Computer Science Department Thinklets." Accessed at http://www.math.shu.edu/thinklets/Math/Statistics/CLT/clt.html, Accessed September 2007.