USING EXCEL TO CREATE INTERACTIVE STATISTICS WORKSHEETS

Laura McSweeney
Fairfield University
Math and Computer Science Department
Fairfield, CT 06824
lmcsweeney@mail.fairfield.edu

Abstract

In statistics education, the trend has moved to using technology to help students develop a deeper conceptual understanding of statistics [1]. In this paper, I'll share how Microsoft Excel can be used as more than a computational tool by creating interactive statistical worksheets that students can use for exploratory learning.

Introduction

Technology can clearly be used to help students automate the many calculations that are inherent in statistics. However, technology can be used as more than a computational tool. It can be used to improve the conceptual learning of statistics as well. Java applets are common applications available on the web that allow students to learn interactively using statistical tools. However, lack of Java programming experience can prevent instructors from designing their own interactive tools. In addition, computers in on-campus computer laboratories may not have the appropriate plug-ins to run existing applets downloaded from the web. Usually, one can achieve similar results by using the built-in features of Microsoft Excel to construct interactive worksheets. These worksheets can be designed to allow students to explore how parameters affect the scale and shape of a distribution or how they influence a model or formula. The worksheets can be designed to allow students to make connections between different topics in the course. Students can also use these dynamic worksheets to make and test conjectures, which again brings learning to a higher level.

While the specific worksheet developed in this paper is geared towards a statistical application, the methods can be used to create interactive worksheets for other courses from general to upper level mathematics courses.

Goals of the Interactive Statistical Worksheets

In introductory statistics classes, the goals of the course often include topics such as: introducing important distributions like the binomial, normal and t-distribution; examining the effects of changing the parameter(s) in the different distributions; demonstrating how a single point can influence the measures of central tendency and linear regression; demonstrating the relationship between the shape of a distribution and the measures of central tendency; exploring the main concepts of hypothesis testing and motivating ideas and formulae like the Central Limit Theorem, properties of expected values and variance, and confidence intervals. In the past, I tried to motivate these ideas in my classes by using a few well-chosen data sets. Despite these in-class examples, many students did not seem to grasp or retain the key ideas.

I wanted to find a way that students could actively explore lots of data sets in a short amount of time. After a presentation by Jeff Libby and Bart Stewart of the U.S. Military Academy who demonstrated how they used Excel in their dynamical systems courses [2], I was inspired to create interactive statistical worksheets in Excel. I was able to create statistical worksheets that students could use to explore the topics listed above. In addition, using Excel would not require students to purchase additional statistical software. It also gave me the flexibility to create worksheets without macros, since the anti-virus programs in our on-campus computer laboratories do not always allow files with macros to be opened.

Typically, the worksheets are used to introduce a particular topic. The students investigate a specific interactive worksheet for homework or as an in-class activity. Then I engage the students in a discussion of their findings and answers to the questions in the worksheets. During this interaction with the class I can quickly determine if they have absorbed the main points of the worksheets. These explorations also give the students a good conceptual foundation on which the computational part can be built.

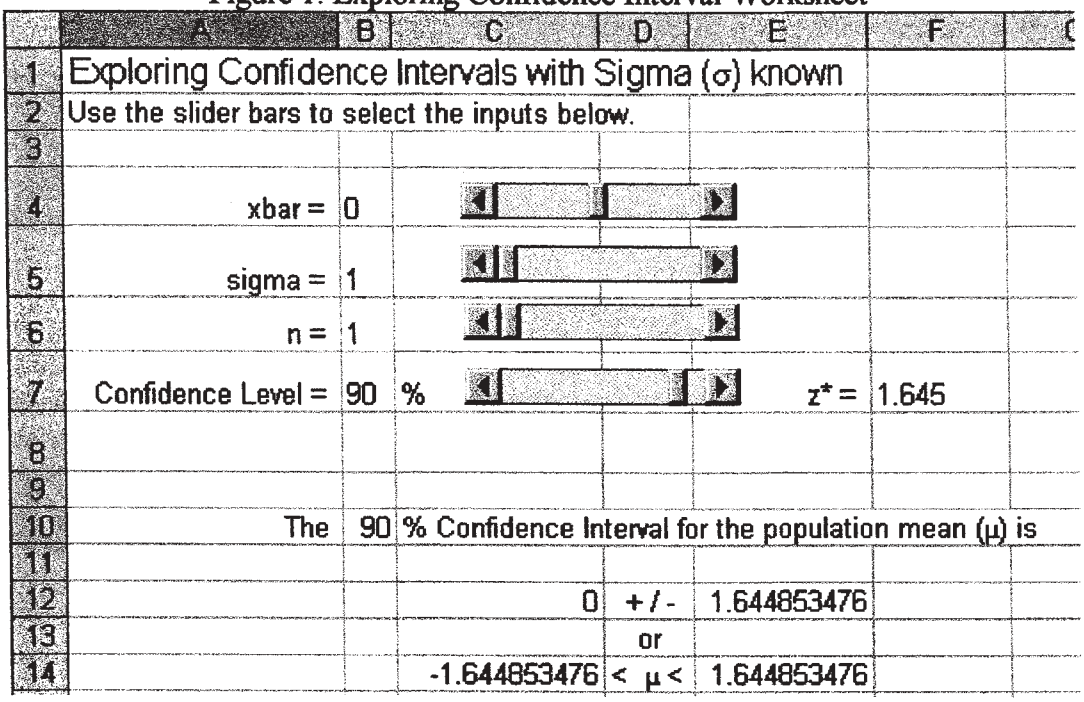
Example of the Confidence Interval Worksheet

The "Exploring confidence interval" worksheet, shown below in Figure 1, allows students to investigate the effects that the sample mean (xbar = $\bar{x}$), the population standard deviation (sigma = σ), the sample size (N) and the confidence level have on the width of confidence intervals. Students can explore what happens to the confidence interval when they change the different parameters using the scroll bars in the worksheet. Students can then examine why their observations make sense mathematically when they look at the confidence interval formula. Using the accompanying set of questions on the worksheet, students are able to make connections between what they observe and the computations.

In order to create this worksheet in Excel (on a PC), one needs to start Excel and open a blank worksheet. The Forms Toolbar is needed to add scroll bars to the worksheet. To add the Forms toolbar to the worksheet (Figure 2), simply go to View/Toolbar and select Forms from the list. The Forms Tool Bar should appear on your worksheet.

We will need 4 scroll bars for this particular worksheet, one for each of the 4 variables. To add a scroll bar to worksheet, highlight the scroll bar icon  on the Forms Toolbar by clicking on the icon. Back in the worksheet, click and drag the cursor to create a scroll bar. Repeat this process 3 more times to add the remaining scroll bars on the worksheet.

Figure 1: Exploring Confidence Interval Worksheet



	A	B	C	D	E	F
1	Exploring Confidence Intervals with Sigma (σ) known					
2	Use the slider bars to select the inputs below.					
3						
4	xbar =	0				
5	sigma =	1				
6	n =	1				
7	Confidence Level =	90	%		z* =	1.645
8						
9						
10	The	90	% Confidence Interval for the population mean (μ) is			
11						
12			0	+/-	1.644853476	
13				or		
14			-1.644853476	< μ <	1.644853476	

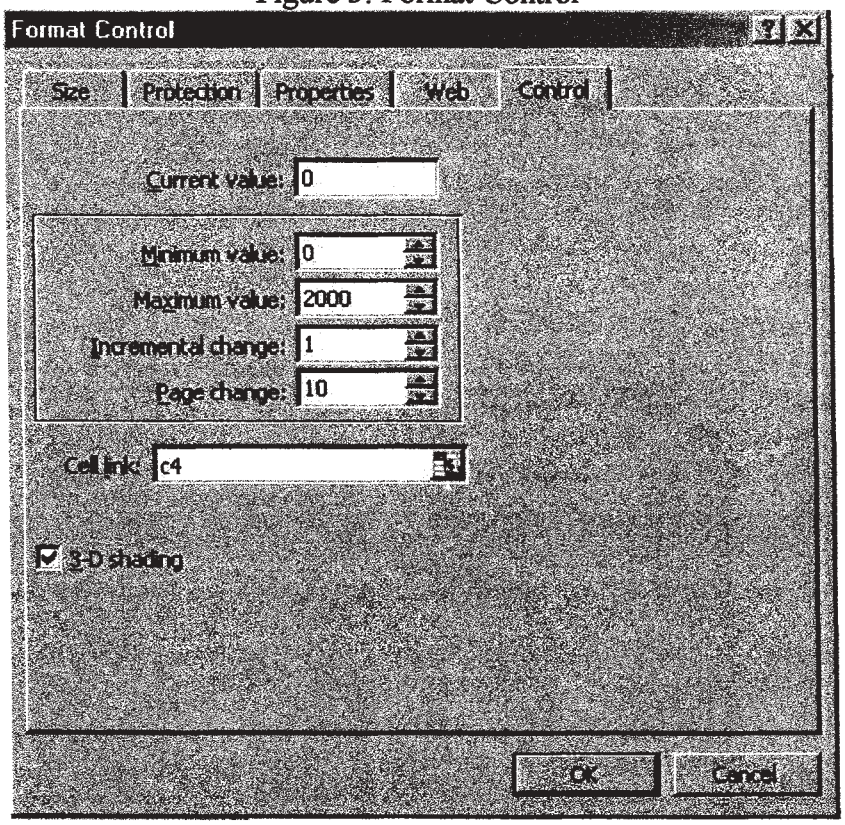
Figure 2: Forms Toolbar

Next we need to set the control parameters for each scroll bar, shown in Figure 3. This will attach the scroll bar to the value of a cell. We'll first do this for the scroll bar for the sample mean, xbar. Place cursor over the scroll bar for the sample mean and right click. Select Format Control and fill in the values shown in Figure 3 and then click OK.

This particular scroll bar will control the value in cell C4, designated by Cell Link, and the values will change from 0 to 2000 by increments of 1. Each click on the arrows on the ends of the scroll bar will change the value in C1 by 1 (Incremental change) while each click within the scroll bar will change the value using a step size of 10 (Page change). The minimum and maximum values must be integers between 0 and 30000 inclusive.

Now, since we want the sample mean to be able to take on negative values, we click in cell B4 and type = C4 – 1000. This will make the value of B4 equal to the value in cell C4 (which is determined by the scroll bar) minus 1000. Therefore, the sample mean will take on integer values between -1000 and 1000. Notice that we could have allowed the sample mean to take on fractional values by adjusting the formula in cell B4.

Figure 3: Format Control



We repeat this procedure for the other scroll bars using the following values:

Scroll bar for parameter	Minimum Value	Maximum Value	Incremental Change	Page Change	Cell Link
sigma	0	1000	1	100	B5
N	1	1000	1	100	B6
Confidence Level	1	99	1	10	B7

Notice that since the parameters N and the confidence level are assumed to be positive and integer-valued, the parameter's value is simply determined by the scroll bar's value. For simplicity, we assume sigma is also integer-valued, although, we could allow other values by using the procedure used for setting the control parameter for the sample mean.

We need to create some intermediate calculations. Notice that most of these results are hidden from view in the worksheet (Figure 1). This is done by making the font color the same as the background color.

	Formula	Description
CI Level (hidden)	F5 = B7/100	Confidence level as a decimal
alpha/2 (hidden)	F6 =(1-F5)/2	The area in one tail (upper or lower)
z* (not hidden)	F7=NORMSINV(1-F6)	Calculates the z-score > 0 which has area below it equal to 1 – (alpha/2).
width (hidden)	F9 = F7*B5/SQRT(B6)	Width of the CI

Lastly, to complete this worksheet, we enter the following formulae in addition to the text shown in Figure 1. To include Greek symbols in the text, you can go to Insert/Symbol and select the appropriate symbol from the list.

B10 = B7 (The confidence level as a percent)
C12 = B4 (The sample mean)
E12 = F9 (The width of the confidence interval)
C14 = B4 – F9 (The lower bound of the confidence interval)
C16 = B4 + F9 (The upper bound of the confidence interval)

As mentioned previously, some of the values are hidden in the final worksheet. In this particular worksheet, the cells which will be hidden from view are C4, E5:F5, E6:F6, E9:F9. To hide these values, we simply change the font color in these cells to match the background color by highlighting them, right-clicking and selecting Format Cells from the menu. Next, choose the Font tab and for the Color option, select the color of your background from the drop-down menu. Click OK when you are done.

I also use the Textbox tool  in the Drawing Toolbox to add text to each worksheet. To add a textbox, simply click on the Textbox tool icon and then in the worksheet you can click and drag the cursor to form a textbox. To add text, simply click in the textbox and type. Typically in these text boxes, I outline the main points I want students to focus on and give them questions that they should investigate.

Other worksheets

The methods used to construct the confidence interval worksheet can be used to create additional worksheets. One may also need to use other statistical functions that are built in to Excel. Examples of other worksheets created include:

- The effect of skewness on the measures of central tendency,
- The effect of outliers on the measures of central tendency,
- Exploring the binomial distribution,
- Exploring the normal distribution,
- Exploring the t-distribution,
- Exploring linear regression and outliers,
- Exploring the Central Limit Theorem and
- Exploring properties of expected value, variances and standard deviation.

All these worksheets are available from the website

www.faculty.fairfield.edu/lmcsweeney.

References

[1] American Statistical Association's Curriculum Guidelines for Undergraduate Programs in Statistical Science, (2003).
(http://www.amstat.org/education/Curriculum_Guidelines.html)

[2] Libby, Jeff and Bart Stewart. "Enhancing 'Point and click' Technology with the Rigors of 'Chalk and Blackboard' Drills", Technology, Pedagogy and Course Redesign II Conference at Fairfield University, 2002.