Illustrating sampling distribution of a statistic:
MINITAB REVISITED

H. Dean Johnson & Marc A. Evans
Washington State University, Pullman, USA
<dean_johnson@wsu.edu>

Understanding the concept of the sampling distribution of a statistic is essential for the understanding of inferential procedures. Unfortunately, this topic proves to be a stumbling block for students in introductory statistics classes. In efforts to aid students in their understanding of this concept, alternatives to a lecture-based mode of instruction have been introduced in the literature with some of these approaches using in-class activities, simulations using statistical software, and web-based applets. In this paper, the use of statistical software, for the purpose of illustrating sampling distributions, is revisited through the use of Minitab macros, an approach that has not been observed in the literature. The result is a user-friendly way for students in introductory statistics classes to explore the concept of the sampling distribution of a statistic. Although the focus of this article will be on the sampling distribution, the methods described in this paper are applicable to instruction of other statistical concepts. These include confidence intervals and power.

Introduction

In the statistics education literature related to introductory statistics instruction, many researchers advocate the use of computer simulation as a means of teaching difficult and abstract concepts, e.g. the sampling distribution of a statistic (see for example Shibli 1990; Prybutok, Bajgier, and Atkinson 1991). It can be argued that the benefits derived from using this approach can be explained using constructivist theory (Mills, 2002). Constructivist theory suggests that one does not acquire new knowledge through passive means, e.g. reading a textbook, listening to a lecture, and memorising facts (Von Glasersfeld, 1987). For one to acquire new knowledge, one must interact with the new knowledge. Taking the constructivist viewpoint, it can be argued that statistics students benefit from the simulations by being able to construct their own knowledge of a statistical concept through the simulation experience.

It is important, however, that computer simulations are implemented in such a way as to not interfere with the learning that might otherwise take place. Today, a popular means of implementing simulations is through the World Wide Web in which students go through an interactive program to generate the simulations. Such programs are indeed useful in that they are user-friendly and provide a clear visual presentation of concepts. One drawback of many of these interactive programs is that, in using the program,

the student do not see many aspects of the simulation on screen, e.g. the actual population values, the observations in each random sample and the mean for each sample. With applets illustrating the sampling distribution for a statistic, for example, students generally only see the parent population distribution and the sampling distribution of the statistic. Being distanced from the details of the simulation in this manner takes away from the learning of the concepts.

In this paper, an attractive alternative to interactive web programs is through the use of Minitab (2004) statistical software to perform and display results from simulations. The idea of using Minitab, for the purpose of teaching abstract concepts through simulations, is not a new one. The purpose of this paper is to illustrate a very user-friendly approach to implementing simulations in Minitab, an approach that has not been observed in the literature. Using the approach described in this paper, students are able to see the details of the simulations on screen, in a spreadsheet. They can see the actual values in the population, the values in each random sample, and the resulting sample statistics obtained from the different samples. Using this approach enables students to move closer to the details of the simulation thus providing a better understanding of the concepts.

Running simulation macro in Minitab

In an introductory statistics course, students are exposed to numerous statistical inference procedures. To have any understanding of these procedures, it is essential that students have a good understanding of the concept of the sampling distribution of a statistic. To aid in the understanding of this concept, simulations are often used. Following is an illustration of how simulations can be easily implemented in Minitab.

The purpose of the simulation below is to illustrate the sampling distribution of the sample mean. Random samples of size $n = 5$ are selected (with replacement) from a population consisting of values the authors have associated with the cards in a standard deck of 52 cards. For each randomly selected sample, the sample mean is calculated and results are displayed in a Minitab worksheet.

Step 1

In the data window in Minitab, enter population data as illustrated in Figure 1.

At this point, the student can go through the menu options in Minitab to calculate the mean and standard deviation, as well as construct a histogram to examine the shape of the population distribution. These steps could be incorporated into the macro but it is preferable to have the students go through these steps on their own. For the example population, the distribution is discrete uniform.

Figure 1. Population data entered into Minitab.

Step 2

Go to the session window in Minitab and enable the command line editor. This can be done by clicking editor and then enable commands in the menu options. When this is done, the command prompt, "MTB>", will appear in the session window .

Step 3

Run the simulation macros. These macros can be found in the Appendix. ASCII format and a .mtb extension in the file name were used for the macro files. We use these macros only for the purpose of illustrating statistical concepts, not for teaching students how to write a program in Minitab. As such, students are not required to write their own macro, only run the simulation program provided to them. However, it would be appropriate to discuss with the class what actions are being implemented by the macro.

The macro is implemented by going through the following sequence of menu options in Minitab.

1.   File
2.   other files
3.   run an exec
4.   open the simulation program and execute the program 1 time.

The user is asked to specify (in the session window) the sample size and how many samples to select from the population as illustrated below. As shown in Figure 2, 500 samples, each of size 5, are selected and the mean is calculated for each sample.

The interactive process above serves two important functions. One, having students specify the sample size and the number of samples reinforces what is being done in the simulation. Two, the interactive process allows students to rerun easily the simulation for different sample sizes, preventing difficulties in the implementation of the simulation from interfering with the learning of the concepts.

Step 4

Observe the simulation results. The results of the simulation can be seen by going to the data window as shown in Figure 3.

The mean and standard deviation of the sample means can now be calculated and a histogram can be constructed to illustrate the properties of the
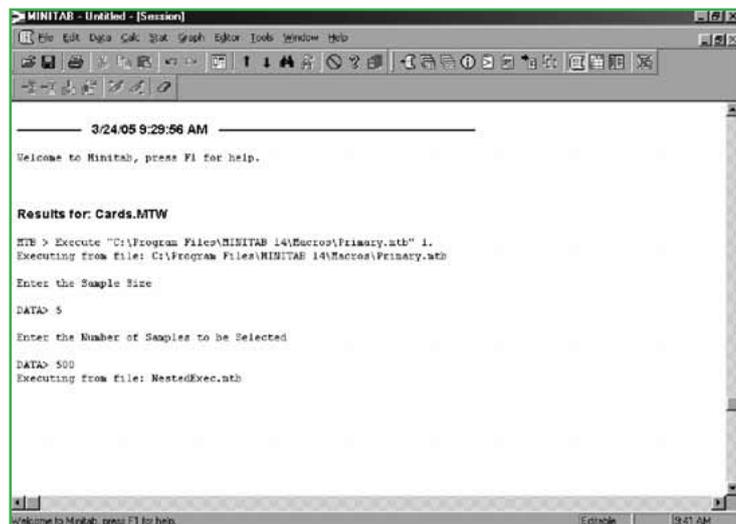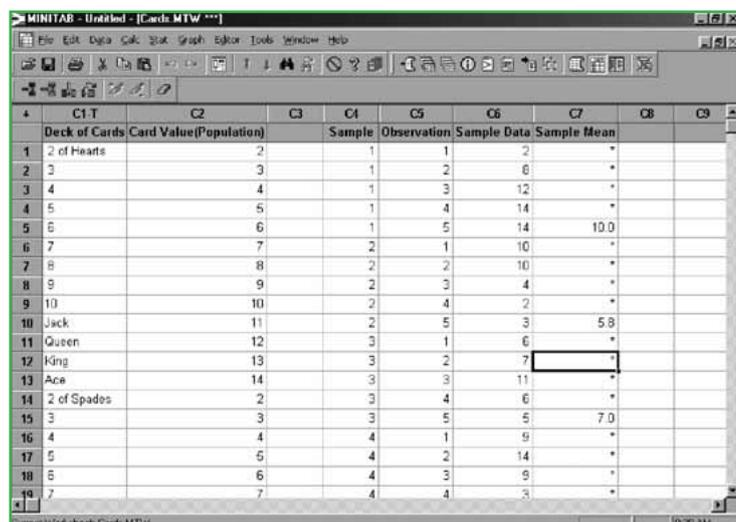

Figure 2. Simulation initiated in Minitab session window.

Figure 3. Simulation results displayed in data window.

sampling distribution of the sample mean. This is easily done through using the menu options in Minitab, with the results of one simulation shown in Figure 4.

Sampling distribution activity

For students to reap the benefits of the simulation activity presented in this paper, it is important that the simulation is accompanied with questions related to the simulation, and it is important that these questions are graded. As stated in Garfield's ten principles of learning statistics (Garfield, 1995), Students learn to value what they know will be assessed. In an introductory statistics class at Washington State University, the simulation is made part of a graded assignment. Each week in this class, students are required to attend
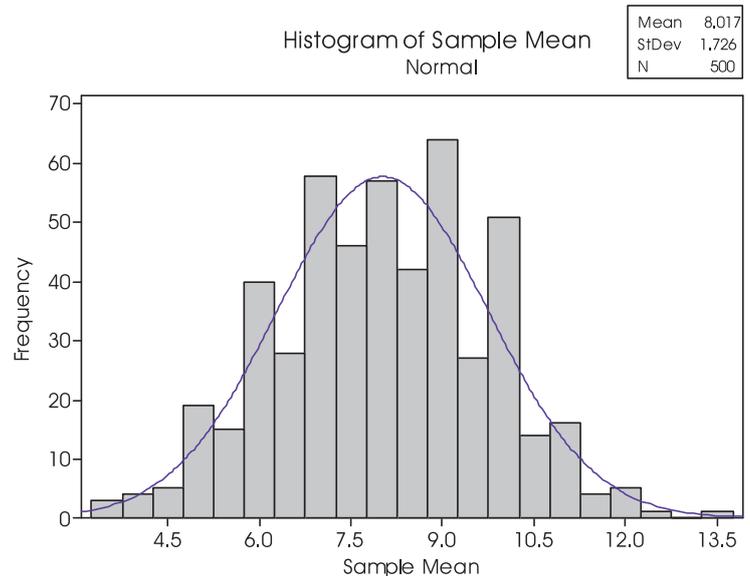

Figure 4. Approximation of the sampling distribution of the sample mean based on 500 samples (n=5) from the population of card values.

Lab activity

Purpose of the Lab

To reinforce the concept of sampling distribution of the sample mean through the use of a Minitab simulation.

Steps

1. Enter the values for the population (the card values for a standard deck of 52 cards) into Minitab.

2. Look at the distribution for the population of card values from which you will be selecting random samples. Describe the shape of the distribution. Calculate the mean and standard deviation for the population.

3. Using the Minitab simulation macro, draw 500 random samples, each of size 5, from the population of cards values and calculate the mean for each sample. Construct and examine the distribution of the sample means. This is an approximation of the sampling distribution of $\bar{x}$ when $n=5$, i.e. the distribution of values taken by $\bar{x}$ in all possible samples of the same size ($n=5$) from the same population (the population of card values).

4. Looking at the individual card values in the parent population and looking at the sample means obtained from the different samples, which seem to vary more?

5. From what we learned about the sampling distribution of $\bar{x}$ in class, the mean of the sampling distribution of $\bar{x}$ you obtained should be close to $\mu = 8$, the mean for the population of card values.

Also, the standard deviation for the sample means you obtained should be close to

$$\frac{\sigma}{\sqrt{n}} = \frac{3.742}{\sqrt{5}} = 1.673$$

the standard deviation of the original population divided by the square root of the sample size. Calculate the mean and standard deviation for the sample means and see if this is the case.

6. In class, we learned that if the original population distribution is discrete uniform, the sampling distribution of $\bar{x}$, for $n=5$, will look approximately normal. Is this what you observed in your simulation?

three one-hour lectures as well as a two-hour lab session. The assignment is completed in one of the lab sessions. The following is an assignment one could consider using in a statistics lab session. It is assumed in the following lab that the steps involved in running the Minitab simulation (presented earlier) have already been introduced to the students.

Conclusion

In teaching introductory statistics courses, the concept of the sampling distribution of a statistic proves to be one of the most difficult for students. Yet, this concept is fundamental to understanding statistical inference. The approach presented in this paper is an effective way of approaching the teaching of this difficult and abstract concept to introductory students. The following are several reasons we provide for the success observed from using this approach.

1. Constructivist theory states that students learn best, not through passive means as in a lecture, but through active means when they are able to interact with the concept and construct their own knowledge. Going through the simulation in Minitab enables students to construct their own knowledge of the sampling distribution of a statistic as opposed to having the concept passed on to them through lecture.

2. Students are sampling from a population for which they are familiar. Almost everyone is familiar with a deck of cards. This is very different from telling students that they are sampling from, for example, a normal distribution with $\mu=5$ and $\sigma=1$, as is often seen in applets and in textbooks.

3. Students are not distanced from the simulation process. With the approach presented in this paper, they see the actual population values of the parent distribution, the values in each sample selected, and the sample means calculated from each sample. Seeing all of these details is important for students in introductory statistics courses.

4. It is very easy to implement the approach. This is important because if students have to struggle to implement the approach, this will interfere with the learning that takes place.

In this paper, the focus has been on demonstrating the use of a Minitab macro for the purpose of illustrating the concept of the sampling distribution of the sample mean. There are also benefits from using Minitab macros for illustrating other statistical concepts, e.g. confidence intervals and power.

One issue is the length of time it takes for the macro to run. This time can vary considerably depending on the speed of the machine being used. Given this, it is important to perform a test run of the macro before using it in a class. For example, it took 25 seconds for the simulation to run on the computer used by the authors. This is longer than it takes for web applets to run, which return results instantaneously. However, given the benefits to students from utilising the Minitab macros, this increase in the amount of time for the simulation to run is compensated by an increased understanding of the concept of the sampling distribution of a statistic.

Many approaches have been suggested in the literature for teaching the concept of sampling distribution of statistic. The approach presented in this paper is another to consider.

References

Mills, J. D. (2002), Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education [Online], 10*(1). (www.amstat.org/publications/jse/v10n1/Mills.html)

Minitab (2004). (Version 14) [Computer software]. State College, PA: Minitab Inc.

Prybutok, V. R., Bajgier, S. M. & Atkinson, M. (1991). Statistics instruction: The next generation. *Journal of Computers in Mathematics and Science Teaching, 10*, 15–27.

Shibli, M. A. (1990), A two-stage exercise on the binomial distribution using Minitab. *Collegiate Microcomputer, 8*(1), 55-60.

Von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of Representation in the Teaching and Learning of Mathematics* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix: Minitab macros

PRIMARY.MTB:

```
FILE "TERMINAL";
  NOBS 1.
Note
Note Enter the Number of Samples to be Selected
Note
Set c99;
  FILE "TERMINAL";
  NOBS 1.
let k1 = c100(1)
let k2 = c99(1)
Name c4 "Sample"
Name c5 "Observation"
Name c6 "Sample Data"
Name c7 "Sample Mean"
set c104
 k1(1)
end
let c102 = k2*c104
Execute "NestedExec.mtb" k2.
Erase c99-c104
Erase c3
```

NestedExec.mtb:

```
Sample k1 C2 C3;
  Replace.
set c103
  (1:k1)
end
stack c102 c4 c4
stack c103 c5 c5
stack c3 c6 c6
Let K3 = Mean(C3)
Let C101(k1) = k3
stack c101 c7 c7
let k2 = k2 - 1
let c102 = k2*c104
```