May 2017

A Study of the Effect of Using Simulations on Students' Learning of Inferential Statistics in the Elementary Statistics Classes in the Mathematics Department of the University of Wisconsin Milwaukee

Alexa Schut
University of Wisconsin-Milwaukee

A STUDY OF THE EFFECT OF USING SIMULATIONS ON STUDENTS' LEARNING OF INFERENTIAL STATISTICS IN THE ELEMENTARY STATISTICS CLASSES IN THE MATHEMATICS DEPARTMENT OF THE UNIVERSITY OF WISCONSIN MILWAUKEE

by

Alexa Schut

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Mathematics

at

The University of Wisconsin-Milwaukee

May 2017

ABSTRACT

A STUDY OF THE EFFECT OF USING SIMULATIONS ON STUDENTS' LEARNING OF INFERENTIAL STATISTICS IN THE ELEMENTARY STATISTICS CLASSES IN THE MATHEMATICS DEPARTMENT OF THE UNIVERSITY OF WISCONSIN MILWAUKEE

by

Alexa Schut

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Kevin McLeod

This thesis reports the results of a studying into the use of simulation-based teaching in Introductory Statistics Class to analyze the effectiveness of this teaching strategy. We give a brief overview of the more recent research into the impact of using computer simulations in an introductory statistics course in order to deepen student understanding of inferential statistics along with the a look at a similar study recently conducted at another university. We then give a review of our study conducted in Math Stat 215 classes at UW-Milwaukee to evaluate whether or not the use of simulations in this introductory statistics class will help cause students to reach a deeper understanding of the underlying concepts of inferential statistics. We conclude with the possible impact that this research can have on future studies at the university.

TABLE OF CONTENTS

LIST OF TABLES

1. Introduction

1.1 Background

Over the past few years, there has been a strong call to incorporate more technology in the classroom. No matter the age of the students, instructors have been asked to start using progressively more technology to help students in their learning and so that they can be technologically fluent for their future employment. With this demand for students to have more technological access in the classroom, the way in which we teach college statistics has been changing. Many college level, introductory statistics courses have embraced some level of technology such as graphing calculators, computer software, or online applets, but the question still remains as to how useful this technology is in helping students understand inferential statistics. More recently, the use of simulation oriented applets for teaching inferential statistics has become popular, but again there has been limited research up to this point on the true impact of simulations on student learning. The present study therefore looked at the effect of introducing simulations into one introductory statistics class at the University of Wisconsin Milwaukee.

Before we begin to look at the recent research and literature on the use of simulations that has been done, the state of Wisconsin has standards for mathematics for kindergarten to twelfth grade. Though these standards do not directly apply to college level statistics, we can use the high school standards as a reference for standards of statistical thinking. One of the state standards for statistics in high school focuses on "Making Inferences and Justifying Conclusions" (Wisconsin Department of Public Instruction, 2011). This standard is described as being able to "understand and evaluate random processes underlying statistical experiments" (Wisconsin Department of Public Instruction, 2011). This means that students need to view statistics as a way to draw inference about a particular population based on a random samples, and that students are able to use simulations to look at patterns in data and test conclusion (Wisconsin Department of Public Instruction, 2011). We can see that these standards are pushing students to have a deeper understanding of what statistics mean in the real world, where data is often messy and the message can be obscured. The use of statistical tools like simulations can help describe variability, and help students to make informed decisions (Wisconsin Department of Public Instruction, 2011). Another part of this standard is for students to "make inferences and justify conclusions from sample surveys, experiments, and observational studies" (Wisconsin Department of Public Instruction, 2011). This standard is asking for students to "recognize the purposes of and differences among sample surveys, experiments, and observational studies: explain how randomization relates to each…Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling…Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant…Evaluate reports based on data" (Wisconsin Department of Public Instruction, 2011).

1.2 Literature Research

Because of the increased call for technology, there has been some literature written on simulations and a few recent studies on the impact of using computer simulations in the classroom. These highlighted what benefits and disadvantages arise from using computer based simulations in the classroom. The GAISE report (Guidelines for Assessment and Instruction in Statistics Education Report) is an educational report produced by the American Statistical Association. According to the GAISE report, “any introductory [statistics] course should take as its main goal helping students to learn the basic elements of statistical thinking” (Martha Aliaga, 2012). Because of this goal, many have been turning to using technology and specifically simulations in the classroom to push students to think more about the meaning of statistics and less about rote calculation of values. It has been the focus in the past for some statistics courses to have students spend large portions of their time on calculations while the interpretations of what they are finding was minimalized. In a short study about using simulations for analyzing the brain power of dolphins to teach students about inferential statistics, Strayer and Matuszewski noted that often statistics classes become all about memorizing some steps and formulas for hypothesis testing that gave a conclusion about the null hypothesis, but the students barely understand the concept of inferential statistics (Jeremy Strayer, 2016). While students focus on following a given sequence of steps, using a calculator to get the numerical answers, they miss the bigger picture of what is going on behind all the work they were told to do by their teacher and textbook. Too often students are focused on simply getting the “right” answer for deciding if they should reject or not reject an idea and see no need to understand the meaning of the values they found as long as they have achieved the correct solution. While it is possible to show students the deeper meaning of inferential statistics without simulations, some statistics educators over the past few years have started to see the benefit of using simulations to teach this topic. One of the leading proponents of simulations based statistical education in colleges, Nathan Tintle noted six reasons why the simulations-based method can be very effective in creating a deeper understanding of inferential statistics. He listed that simulations

- clearly present "the overarching logic of inference,"
- strengthen the "ties between statistics and probability/mathematical concepts,"
- encourage students to "focus on the entire research process,"
- facilitate students to think about "advanced statistical concepts,"
- allow for more time to "explore, do and talk about real research and messy data," and
- allow students to act on a "firmer foundation on which to build statistical intuition" (Nathan Tintle, 2015).

Overall, Tintle seems to have a desire for students to focus less on just getting through the calculations, and to have them focus on what it means to truly understand and use statistics. Allan Rossman and Beth Chance, who worked closely with Nathan Tintle on a course that relies on simulations to teach statics, noted in their research that students need to realize that statistics is about more than getting a 'yes' or 'no' answer, but rather about finding a conclusion that fits with the context of the situation and explains how strong of a conclusion has been reached (Allan Rossman, 2000). They want their students to have a deeper understanding of statistics than just answering the questions on a test or reciting facts with no real understanding what they mean in the real world. "Our goal is not only for students to be able to interpret conclusions reported in scholarly and popular literature, but also to be able to explain them clearly to people who are not familiar with statistics" (Allan Rossman, 2000). Students need to be prepared to take what they learning their classes and apply it to the real world. Therefore a statistics course should have the goal of students leaving with the ability to hold a conversation with another person about statistical ideas that show they're understanding the results of inferential statistics as real life conclusions.

One of the main goals of using simulations in a classroom is to create a deeper conceptual understanding of inferential statistics. Students need to be able to do more than just calculate a p-value and then recall the rules for whether or not they should reject their null hypothesis. There must be a step from this work to the real use of this information. This is where a conceptual understanding of inferential statistics is needed, and this step is often left out in favor of focusing on the mathematical calculations in statistics. But given the advancements in technology we can bring the conceptual understanding of statistics to the forefront of the class. "Technology has also expanded the range of graphical and visualization techniques to provide powerful new ways to assist students in exploring and analyzing data and thinking about statistical ideas, allowing them to focus on interpretation of results and understanding concepts rather than on computational mechanics" (Beth Chance, 2007). Through the use of simulations students will be able to use larger, messier datasets, choose the appropriate tools and strategies, focus on the input and output rather than a formula, manipulate situations to see relationships, and provide in-depth interpretations of what the data is showing (Beth Chance, 2007). This will allow students to approach statistics as they would in real life, by doing complete studies. The use of computers has allowed statisticians to analyze real data sets and focus on interpretation rather than number crunching. Similarly, “computers have freed us to simplify our curriculum, so that we can put more emphasis on core ideas like randomized data production and the link between randomization and inference, less emphasis on cogs in the mechanism” (Mia Stephens, 2014). Students can collect, analyze, and make conclusions about data on their own questions/topics, allowing students to feel empowered in statistics and experience the full practice of statistics. We can focus on testing their ability to handle data sets over a full analysis rather than just small questions of specific moments, and we can hope that the students will see that the data can reveal possible unexpected but complete stories of real world situations (Beth Chance, 2007). With simulations, focus in the classroom can be turned to the students and their experiences with data rather than focused on rote calculations. The class should focus on statistical thinking, a focus on data and concepts rather than theory and recipes, and on active learning (Allan Rossman, 2000). We want the course to emphasize the actual data and what it shows rather than just the theorems and procedures, and in order to achieve that goal when possible the calculations and graphs should be done using technology (Martha Aliaga, 2012). Through the use of simulation-based learning, we hope to convince not only students but instructors that a true understanding of statistics is more important than “the ability to force data through a black-box model” (Schaffner, 2015).

For example, consider what might happen when students calculate a p-value. If they are in a standard class, the students would need to set up the complete test, then calculate the standardizing statistics, and then find a table of values that they would search through to get the p-value. In this process, they are focused on plugging into a formula and then reading a table. In a recent research report, Stephens, Carver, McCormack noted that when we use statistical tables like this, “[It] puts a heavy burden on our students, who must struggle through difficult theory before they're taught how to make decisions and draw inferences from data. Even then, do they understand what a p-value is or what a confidence interval represents?” (Mia Stephens, 2014). Students in a traditional class are expected to understand the value in the context of the problem, but using simulations there can be a stronger emphasis on this part of the situation rather than on being able to correctly use the statistical tables in a book. We can focus on conceptually understanding the p-value, so that students can understand how the value is connected to the null hypothesis and the probability of “observed or more extreme” values (Cetinkaya-Rudel, 2015). This allows us to have students understand that the p-value goes beyond just being a value to compare to the significance level to make a yes or no decision, and it becomes a measure of the strength of evidence obtained from the sample (Allan Rossman, 2000). This idea of focusing on the concepts and data analysis is recommended by both the GAISE Report and SET (The Statistical Education of Teachers written by the American Statistical Association). They both advocate for the use of technology in the classroom to help shift the class time to be about “developing concepts and analyzing data” (Christine Franklin, 2015).

This example illustrates that students should be focused on a deep understanding of what their data is outputting rather than simply stopping when they get a final value as “the answer.” Students should be spending their time on explanations and justifications for the conclusions they draw about their final outputs (Beth Chance, 2007). What is the point of teaching them the calculations if they do not understand why they are doing them, and what the conclusions are telling them in the context of their study? If students can understand the reasons for what they are doing and can see the important concepts then the particular procedures that need to be performed should be easy to learn (Martha Aliaga, 2012). This can allow for a more effective use of instructional time in class, since the focus can be on students understanding of key ideas with a few techniques rather than trying to cover several techniques with the concepts tacked on at the end (Martha Aliaga, 2012).

As instructors we should always be looking for ways to teach statistics in a way that reaches students and to focus on the important concepts behind statistics. Using simulations can allow an instructor to introduce inference before having to cover the details of probability distributions (Cetinkaya-Rudel, 2015), (Karsten Maurer, 2016). Students can use simulations to focus on how the process works as a whole and what their finds mean in context, rather than first having to learn theory and procedure. With the use of technology, students can simplify work such as finding p-values, by find the values through an applet that provides a more accurate and timely answer rather than using tables (Beth Chance, 2007). Some recent educators have even gone so far as to say “that previously standard topics in an introductory course (e.g., short-cut methods for calculating standard deviation) are no longer necessary to discuss in class” (Beth Chance, 2007). Technology affords students away to process calculations with more efficiency and accuracy which allows them to focus on what it the values mean in the context of the study. After students have seen and worked with calculations a few times, the use of simulations can ease the process and shift focus on to what the values mean. Students should use calculations that are vital to understand and avoid computations that are divorced from understanding (Martha Aliaga, 2012). Technology also allows students to use large data sets that match real life situations rather than artificial smaller data sets (Martha Aliaga, 2012). Real life data sets allow students to see the applications of statistics in action and will prepare students for their future work. As the GAISE report notes, statisticians' approach to their work has been changed by the use of technology which means we should also change the way that we teach statistics (Martha Aliaga, 2012). Simulations allow instructors to show students the process in a dynamic way that allows students to ask questions and test their own ideas (Andrew Zieffler, 2007), which will enable them to see the overarching ideas behind statistics before they look at the each topic more deeply.

Another of the benefits of using simulations is being able to reach students who are visual learners. Often students find that mathematical topics over all seem much easier once they have been given a visual example of the topic. With the use of simulations students can interact with data and see visual representations. Not only are students able to see the data in action, but instructors have an easier way of presenting the data to students. Students are able to take data sets and investigate different 'what if' questions by direct observation through the computer simulations that would normally be time consuming (Beth Chance, 2007). Using computer applets students can understand different concepts and relationships through repetition, changing parameters, and seeing the behavior of the data rather than relying on theory (Beth Chance, 2007). The visuals are especially useful for the more abstract topics that can be covered such as what a confidence level really means for a confidence interval, how the mean and median of datasets interact, or how probability is long run random behavior rather than a guarantee. Students can use the simulation-based applets to see large scale simulations and how they can be used to approximate probabilities (Christine Franklin, 2015). This introduction to probabilities through simulations has recently been pointed to as a more effective way of teaching versus formal probability as has been previously taught since it allows students to visualize and manipulate data that gives a concrete example (Allan Rossman, 2000). Applets can also provide a way to continue using visuals to teach statistics throughout the semester rather than only for the initial introduction of what statistics are (Allan Rossman, 2000). This can provide a concrete connect that carries through the whole of the class and provides the students a holistic approach to the study of statistics that starts each study with graphical representations of the data before jumping into the analysis.

As has already been mentioned the use of simulation-oriented teaching can be a helpful teaching strategy. With simulations, instructors have the ability to use visuals that are easy, simple, and relevant to the course to help students understand the material, as mentioned above. Simulations also allow for a classroom that is more focused on the student rather than on teacher lectures. The GAISE Report talks about how statistics instructors should be focusing on alternate methods of instruction such as projects, lab work, group work, and discussions, so that students can be actively involved in their learning (Martha Aliaga, 2012). Along with this shifting focus to student centered learning, there is the challenge of engaging students in the materials so that they can truly understand the underling concepts of statistics while getting hands on experience with performing statistical studies. Technology when used effectively can enhance the classroom, help to shift the focus to student driven learning such as lab work, and provide hands-on working with data sets that the students find engaging and relevant to their lives. Technology may also be used to introduce students to bootstrap resampling, which involves repeated sampling from a data set with replacement to approximate the target population. This results in large numbers of bootstrap samples that can then be used to estimate such things as standard error and confidence intervals (Mia Stephens, 2014). This strategy can be quite useful since confidence intervals are a very helpful, though underutilized, statistical practice for drawing conclusions about data sets that can be more broadly used than a test of significance (Allan Rossman, 2000). Though this transition to technology may require more work from instructors and close monitoring that the technology remains helpful rather than a burden, there have been positive results from using technology in the classroom in recent studies (Martha Aliaga, 2012). Pairing computer simulations with tactile simulations (cards, dice, coins, spinners, etc.) will allow students to experience statistics on a personal level rather than through a standard lecture style. The use of technology in the classroom can also teach students how to use technology effectively as they explore data through projects and open-ended problems (Martha Aliaga, 2012). It also provides students with an introduction to a “modern data-analytic approach to statistical thinking” that they will be able to then use as they transition into the working world (Christine Franklin, 2015).

A large part of why students take an introductory statistics class is to prepare for their future work once they leave school. This means that our statistics courses should focus in on providing students with real world interactions using statistics. As students enter the industrial world, the ability to use technology for analysis in data is a growing demand that students need to be trained to meet (Schaffner, 2015). As a tool used by statisticians, simulation can be useful to students as both a visual and as a way to build a deep understanding of the material in a way similar to what would be expected of them outside of the classroom (Andrew Zieffler, 2007). With the introduction of simulations in the classroom, students are able to work with real data sets, automate calculations, appropriately modify graphics and statistical models, and explore questions that may arise in a given situation (Beth Chance, 2007). The use of real data in the classroom is very important as it allows learning to be authentic and relevant, which will engage the students and provide them with the opportunity to analyze a situation that they may encounter outside of the classroom (Martha Aliaga, 2012). As mentioned above the use of simulations can also allow for group work, projects, and discussions that allow students to more easily collaborate with one another which can improve oral and written communication skills so as to effectively convey the conclusions that they are drawing from the data (Beth Chance, 2007). This a very important part of training students as they go out into the working world so that they can effectively communicate with others who may not know how to think statistically. The use of technology in the classroom can also cause students to become more independent learners by using technology outside of the classroom, which can allow the instructor to focus more on data analysis and group work rather than lectures (Beth Chance, 2007). As mentioned before, the use of simulations can produce a deeper conceptual understanding of statistics that results in students who are able to think and write statistically. Students who are able to work through large real world data sets using statistically focused technology are afforded the opportunity to work collaboratively, learn from their peers, discover their own studies that they can analyze, and see the importance of the conclusions they draw (Martha Aliaga, 2012). Being able to work with the a dataset that they find interesting and that they can use for a complete statistical analysis provides students with an engaging project that they are able monitor the whole way through the process, which results in a deeper understanding of what they have done, why they did it, and what their conclusions mean as an impact on the world.

1.3 Concerns When Using Simulations

Though there has been a push for the use of simulations in the classroom as of late, some instructors have expressed concerns over the use of the simulations in the classroom. Some of these concerns were about using the technology correctly, rather than the use of technology in itself. Schaffner noted that there is often a “disconnect between software that is useful for teaching versus doing statistics” (Schaffner, 2015). He went on to describe how the first attempts at using technology for activities and demonstrations in the classroom caused it to be harder to get through the required material so that it became a burden to fit everything in to the allotted class time, which is why he felt that it was better to switch to a mostly flipped classroom model so that the technology and classroom style were compatible for learning (Schaffner, 2015). When using technology such as simulations in the classroom, it may be necessary to restructure the class so that the use of technology is more integrated and a stronger student centered classroom style may need to be implemented to effectively use technology while still covering the full material of a typical introductory statistics course. Another concern that has been voiced in some of the more recent research into simulations is about the need to plan the use of simulations out carefully and then monitor student interaction close as the simulations are used (Beth Chance, 2007). As teachers prepare to use technology in the classroom, the initial focus should be on why they are using the technology and how it will be implemented in the room so that careful planning can be used in implementing the technology (Beth Chance, 2007). As instructors plan their use of technology, they must make sure that the material is accessible, interesting, and useful for their students so that students may learn from the data they provide rather than simply having the students use technology for the sake of using the technology (Martha Aliaga, 2012). The technology should also be used as more than just a way to perform calculations, but rather to explore different concepts and enhance the student learning through concrete examples (Martha Aliaga, 2012). The technology should also be easily accessible to the students so that they can easily interact with the simulations and are not caught up in trying to learn a complicated process for the simulations. The simulations are meant to aid in the process so that the students can focus on understanding their conclusions and the purpose of the study, so a complicated program can be counterproductive. In a recent study Watkins, Bargagliotti, and Franklin looked at two misconceptions that in particular can show up because of the use of simulations: that students believed that you needed to draw multiple samples in order to make a valid statistical inferences when using computer simulations for SDM (the Sample Distribution of the Mean); and students have trouble distinguishing randomness due to random selection of data from the randomness from using small numbers of replications (Ann Watkins, 2014). As they looked into the issue they found that when looking at the sampling distribution of the mean, their students thought that the sample size would changes how the population mean and the mean of the sample means relate, and how the standard deviation of the population relates to the standard deviation of the sample means. Their students thought that they needed to change the sample size to be larger in order for the mean of the sample means to be equal to the population mean (though this relationship holds for any sample size) and for the standard deviation of the sample means to be equal to the population standard deviation divided by the sample size (this relationship also holds for any sample size) (Ann Watkins, 2014). They were lead to this conclusion because of the errors that can occur when you use simulations of repeated samples rather than looking at the theory directly (Ann Watkins, 2014). These issues can be caused by using simulations if instructors are not careful in the presentation of the material, and should be directly addressed with students to avoid some confusion. So as long as the instructor is mindful of this issue and takes the time to address it, the students will not be misinformed about sample sizes in this situation (Ann Watkins, 2014). The final concern that came up in the research for this study was that students would not be able to connect the values and visuals given in the simulations with the process used to achieve these values (Allan Rossman, 2000). Rossman suggested the possibility of using physical manipulatives to avoid this sort of confusion (Allan Rossman, 2000). It could also be avoided by having students go through the full process by hand a few times along with using the applets, so that they can compare the theoretical process with the results from the applets.

1.4 Summary of Another Recent Study on Simulations in the Classroom

In 2016, Instructors at Iowa State University in Ames, IA conducted an experiment where they looked at two classes and introduced simulations with the StatKey software package into the classroom to teach inferential statistics. In their study they found a significant improvement in students understanding of confidence intervals using simulation-based teaching (Karsten Maurer, 2016). They looked at two sections of their introductory statistic course for the spring of 2014. There were 101 participating students in the study. For both sections, students were randomly assigned to either a traditional approach to inferential statistics (Cohorts C and D) or a class that used simulations (Cohorts A and B) (Karsten Maurer, 2016). They used co-teaching and alternated the instructors so that all the cohorts received even teaching instruction. For the first half of the class the two classes were taught in the standard way and then split into four classes when they began to study inferential statistics. The topics covered in the course were univariate and bivariate descriptive statistics, linear regression, experimental design, basic probability rules, the binomial distribution, the normal distribution, sampling distributions, and inference on means and proportions (Karsten Maurer, 2016). Both classes were taught in a lecture format with weekly group lab assignments, weekly homework, a midterm exam, and a final exam, but no textbook was used.

After the first nine weeks, in which all of the students were receive the exact same lectures, Cohorts A and B switched to using computer simulation for teaching inferential statistics while Cohorts C and D continued with the traditional approach (Karsten Maurer, 2016). The students in Cohorts A and B used StatKey software package for the simulations-oriented inference statistics. They did have a few simulations with Cohorts C and D, but fewer than in in Cohorts A and B; in fact, they only used the software to show that sampling distributions can approximate the normal distribution in certain situations (Karsten Maurer, 2016). The assignments for both classes, though different, were kept as similar as possible when the classes covered similar topics. It was noted that all of the cohorts learned how to use normal theory for inferential statistics, however the simulation-based cohorts also learned the concepts of inference using simulations before learning the theory-based approach (Karsten Maurer, 2016). To evaluate learning among the students, the researchers used the ARTIST (Assessment Resources Tools for Improving Statistical Thinking site created by educators at the University of Minnesota) scaled question sets (each 10 multiple choice questions focused on critical thinking and statistical literacy) and administered as part of the final exam (Karsten Maurer, 2016). They also looked at two questions from the final exam that focused on conducting statistical inference in an applied setting using the theory-based approach (Karsten Maurer, 2016). For the ARTIST questions and the final exam question that focused on confidence intervals, their data showed that the simulation-based inference group had higher average scores than the traditional group. But the simulation group scored lower on average on the hypothesis testing question of the final exam. As for variability, the simulation group had larger variability than the traditional group for all questions except for those focused on confidence intervals (Karsten Maurer, 2016). In the end they noted no significant change for hypothesis testing between the two groups, but they did find that there was a 7.146% improvement for questions about confidence intervals (Karsten Maurer, 2016). They did note that there were some issues with their study such as the differing levels of experience of the students that were assigned to each cohort, which results in a higher Type 1 error (Karsten Maurer, 2016).

2. Method

2.1 Research Design

For our study, we started with a smaller scale initial test to see if there should be further research into using simulations in the classroom at the University of Wisconsin Milwaukee. We focused on the Elementary Statistics course taught in the Mathematical Science Department that fulfills a general Mathematics requirement for many students. Two classes were taken for this study with simulations introduced into one class. There were two main ways that the computer simulations were introduced. Firstly the applets were shown in class with a brief demonstration of the working components of the applet and secondly an example that related to the current material of the course.

2.2 Simulations Group and Control Group Intervention

We focused on two different Mathematical Statistics 215 classes that were taught by the same instructor. Both class meet on Tuesdays and Thursdays for 75 minutes over a 14 week fall semester. The control group meet at 8:00 am and the simulation group meet at 9:30 am. The control class had 15 students in the class that consented to be a part of the study and the simulation class had 28 students that consented to be a part of the study. Both classes concentrated on an introduction to inferential statistics and covered topics such as descriptive statistics, measures of center and standard deviation, correlation and regression, basic probability, normal distribution, confidence interval, and hypothesis testing. The content and examples for the class were similar and often matched exactly. Both classes were taught in a lecture format with weekly quizzes, three semester exams, a final exam, and effort based homework from the textbook.

In the intervention (simulations) class, we introduced a few simulation-based applets in the classroom using them for examples. The students then completed two to three different problems that involved using the computer applets for each homework assignment. The homework focused on walking them through the applets again and then having them draw conclusions from the values, graphics, and trends that they saw in the applets. The simulation-based problems were still graded based on effort. The topics that were covered in the simulation homework were:

- how the sample means and median relate for different shaped data sets;
- a conceptual understanding of what the mean, median and standard deviation of a data set are and what the values would look like for a given data set;
- relating a line of best fit to the data to see how outliers can change the line;

- understanding how the correlation coefficient relates to the spread of the data; how probability relates to long run behavior;
- using sample distribution to approximate the area under a normal curve in certain condition;
- how the area under a normal curve relates to probability;
- relating the mean of sample means to the population mean;
- relating the standard deviation of the sample means to the population standard deviation;
- how confidence levels relate to the width and precision of a confidence interval;
- how p-value relates to the observed value or more extreme of a normal curve.

For these problems, and the examples we used in class, we selected applets from two different sites. The first site was the Rossman and Chance Applets from Hope College (http://math.hope.edu/isi/) that were created to accompany a newer textbook that teaches introductory statistics through simulations. This site was mostly used for the applets that focused on regression and correlation, hypothesis testing, and confidence intervals. We used the Descriptive Statistics Applet, Regression/Correlation Applet, Correlation Guessing Game, One Proportion Applet, One Mean Applet, Theory-Based Inference Applet, Matched Pairs Applet and the Multiple Means Applet. The second site was the RVLS applets (http://onlinestatbook.com/stat_sim/index.html) which was used to look at how the mean and median of data sets compare based on the shape and skew of the data and what happens when you change sampling sizes. For this site we only used the Sampling Distribution Applet.

2.3 Measurement of Understanding

The means of measuring understanding that we used in our study included a pre/posttest, survey, and select exam questions (See Appendix A). In order to evaluate the impact of simulations on the students' level of understanding we constructed measures of students' understanding, and feelings towards the applets. For both the simulations class and the control class, the students were given a short test near the start of the semester and then again at the end of the semester. The test consisted of four questions. The students were asked several questions

- what statistics is and why we study statistics,
- how they would set up an experiment with 20 people and two variables,
- what the role of randomization is and why it is important to use random samples, and
- when they would expect the mean to be greater than or less than the median.

We compared the students' responses between the two times to see if they had improved in understanding or not. We looked at each class separately to see if the simulations helped with understanding in this particular areas.

Along with the post test, the simulations class was also asked to fill out a survey at the end of the semester about what they thought about the simulations used in class. The students were asked if they liked the applets, if they thought they were helpful in learning the material, if they would recommend the applets to be used for future students, what they liked about the applets, and what they disliked. We first looked at how many students responded overall positively or overall negatively to the applets, then at more specific examples of what the students said was good and what was bad about the applets.

The final way we evaluated understanding in the study, was by looking at select questions from the semester exams and the final exam. Both classes were given the same exams throughout the semester, and we looked at questions that related to the simulation-based homework problems that we gave to the simulations class to compare understanding. We created a rubric that looked at each selected question and evaluated the depth of the responses students gave. For each question, we assigned a value of 0, 1, or 2:

- 0 meant that the response was completely wrong and/or showed a strong misunderstanding of the topic;
- 1 meant that their response showed some understanding of the topic but did not show a strong/complete understanding; and
- 2 meant that they showed that they understood the topic well and could express that.

After the data was entered and tallied, each question was looked at individually and class averages were compared. For each question, we took the class average of the simulation class and subtracted the class average of the control class. Then using the Rossman and Chance applets that we had the students using, we found the two sided p-values and the one sided p-value for this observed statistic. We used the Non-pooled t-statistic to find these two p-values in this case. We then evaluated if the p-values were statistically significant at a 5% significance level to see if any of the differences and if the simulation class had a higher average in those case than the control class.

Below is a summary of the questions that were evaluated:

- Exam 1 #4c) How the relationship between the mean, and median can tell you about the shape of a distribution.
- Exam 2 #2d) Relating the correlation coefficient and coefficient of determination to the line of best fit.
- Exam 2 #3b) Relating probability to the long run behavior of a random process.
- Exam 3 #1d) The relationship between confidence level, interval width, and precision of the estimate of a single population mean.
- Exam 3 #2) Performing Hypothesis Tests for one sample mean.
- Exam 3 #3) Performing Hypothesis Tests for one sample mean.
- Exam 3 #4) Performing Hypothesis Tests for one sample mean.
- Exam 3 #5bc) Find the probability of a situation using the area under a normal curve using statistical tables.
- Final #1b) Calculating and Interpreting the Standard Deviation of a Sample Dataset.
- Final #1c) Performing Hypothesis Tests for one sample mean.
- Final #1d) Finding and Interpreting a 99% Confidence Interval
- Final #2d) How the relationship between the mean, and median can tell you about the shape of a distribution.
- Final #3c) Interpreting and using the coefficient of determination for linear regression.
- Final #3d) Using the correlation coefficient to describe the relationship between two variables.

- Final #6d) The relationship between confidence level, interval width, and precision of the estimate of a single population mean.
- Final #7a) Performing Hypothesis Tests for one sample mean.
- Final #8ab) Find the probability of a situation using the area under a normal curve using statistical tables.

3. Results

3.1 Results for the Rubric Scores

To start we will look at the data that we collected through out the semester from their in class exams and final exam. The exams were the same for both classes and students were given the same amount of time. After the semester ended, selected questions were used to evaluate students understanding. Each of these questions was compared to a rubric that as mentioned before assigned values of 0, 1, or 2 to the answers students gave. There was only one exception to these assignments: for the questions that focused on being able to use areas under a curve to find probability using a normal curve and statistical tables, we assigned a value of 1 if they were able to correctly convert to the standard normal curve, but did not understand how this related to the statistical tables or how to continue after they had done the initial work. A value of 2 was assigned if the student could successfully find the standardizing score and use the statistical tables correctly. This question was evaluated different to judge whether or not students who used the simulation-based approach would have a deeper understanding of what the statistical table values actually represented and how they connected area to probability, or if using the simulations would result in students being unable to use statistical tables compared to the standard approach to teaching statistics.

To protect the students' privacy, each participating student was assigned an ID number that will be used for the rest of the analysis. The students in the control class were given ID numbers in the 700's and the students in the simulations class were given ID numbers in the 800's. On the following two pages, there are tables summarizing the scores that each student in each class received. There are also averages and standard deviations provided for each question. Students' overall average score and standard deviations are provided also. Along with an overall average for each class.

Table 1. Simulation Class's Rubric Scores for In class and Final Exams

ID Number	E. 1 #4c	E. 2 #2d	E. 2 #3b	E. 3 #1d	E. 3 #2	E. 3 #3	E. 3 #4	E. 3 #5bc	F. #1b	F. #1c	F. #1d	F. #2d	F. #3c	F. #3d	F. #6d	F. #7a	F. #8ab	Average	Median	Std. Dev.
828	1	0	0	0	0	0	0	0	2	0	0	0	2	0	1	0	0	0.353	0	0.681
822	1	1	1	2	1	2	2	2	0	1	2	1	1	2	2	2	2	1.471	2	0.606
811	0	1	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0.294	0	0.456
825	2	2	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	0.588	0	0.691
809	1	2	1	2	1	1	2	2	1	1	2	1	1	0	1	1	1	1.235	1	0.546
810	2	1	0	2	0	0	0	2	1	2	0	2	1	0	2	1	1	1.000	1	0.840
815	2	1	2	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0.471	0	0.776
819	1	2	1	0	0	0	0	1	1	0	0	2	1	0	0	0	2	0.647	0	0.762
806	1	2	1	2	0	0	0	1	0	1	2	0	1	2	2	1	2	1.059	1	0.802
807	1	1	1	0	0	0	0	2	2	0	2	2	1	0	2	0	2	0.941	1	0.872
820	2	1	1	2	0	0	0	2	2	1	2	1	1	1	1	0	0	1.000	1	0.767
830	0	1	1	1	0	0	0	0	0	1	2	1	1	0	2	0	2	0.706	1	0.749
813	2	2	0	2	0	2	2	0	0	1	2	2	0	1	2	1	2	1.235	2	0.876
803	2	2	1	2	0	2	2	2	2	2	2	2	2	1	2	2	2	1.765	2	0.546
831	2	2	1	1	0	0	0	0	0	0	0	2	0	0	1	0	2	0.647	0	0.836
804	0	2	1	2	0	0	0	2	0	1	0	2	1	1	2	1	1	0.941	1	0.802
818	2	2	1	2	1	1	1	0	1	1	2	2	2	1	2	1	2	1.412	1	0.600
801	1	2	1	2	0	0	0	2	1	2	1	0	1	1	2	2	2	1.176	1	0.785
827	2	2	1	2	0	0	0	2	2	1	2	1	2	2	2	1	2	1.412	2	0.771
826	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0.294	0	0.456
829	0	2	0	1	0	0	0	1	2	1	1	0	1	1	2	0	0	0.706	1	0.749
814	2	2	1	2	2	2	0	2	1	2	2	2	2	2	2	2	2	1.765	2	0.546
805	2	1	1	2	0	0	0	0	1	2	2	2	1	0	2	1	2	1.118	1	0.832
824	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.235	0	0.424
802	0	2	1	1	0	0	0	0	2	2	1	2	0	1	2	2	0	0.941	1	0.872
817	0	0	1	2	0	0	0	1	2	2	1	2	1	0	2	1	2	1.000	1	0.840
808	1	1	1	1	0	2	2	2	2	1	2	0	1	0	2	0	2	1.176	1	0.785
821	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0.235	0	0.424
Average	1.000	1.467	0.933	1.467	0.200	0.467	0.333	0.933	1.067	1.133	1.067	1.200	1.000	0.733	1.600	0.867	1.333	0.988		
Median	1	2	1	2	0	0	0	1	1	1	1	2	1	1	2	1	2			
Std. Dev.	0.894	0.718	0.249	0.618	0.542	0.806	0.699	0.929	0.854	0.806	0.854	0.909	0.730	0.680	0.712	0.806	0.869			

Table 2. Control Class's Rubric Scores for In class and Final Exams

ID Number	E. 1 #4c	E. 2 #2d	E. 2 #3b	E. 3 #1d	E. 3 #2	E. 3 #3	E. 3 #4	E. 3 #5bc	F. #1b	F. #1c	F. #1d	F. #2d	F. #3c	F. #3d	F. #6d	F. #7a	F. #8ab	Average	Median	Std Dev
704	2	2	1	0	2	2	2	2	1	0	1	1	1	1	0	2	2	1.294	1	0.749
702	1	0	1	2	0	2	0	0	1	0	0	1	1	2	1	0	1	0.765	1	0.730
711	1	2	1	1	0	0	0	0	1	2	2	2	2	1	2	2	0	1.118	1	0.832
707	0	0	2	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0.353	0	0.588
703	2	2	1	2	2	2	2	2	1	1	0	2	2	2	2	2	2	1.706	2	0.570
715	0	1	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1	0.353	0	0.478
701	2	2	1	1	2	2	2	2	0	1	2	2	2	1	2	1	2	1.588	2	0.600
705	2	2	0	0	0	0	1	0	1	0	2	1	1	2	0	0	0	0.706	0	0.824
712	1	1	1	0	0	2	0	0	1	0	2	0	0	0	0	2	0	0.588	0	0.771
710	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.176	0	0.513
709	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0.176	0	0.381
713	2	1	2	2	0	0	0	0	2	1	2	2	0	0	2	1	0	1.000	1	0.907
714	2	0	0	0	0	0	0	2	0	0	2	1	1	0	0	0	2	0.588	0	0.844
706	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.118	0	0.322
708	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.059	0	0.235
Average	1.267	0.933	0.867	0.667	0.400	0.667	0.467	0.600	0.600	0.400	0.867	0.867	0.733	0.600	0.667	0.733	0.667	0.706		
Median	1	1	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0			
Std Dev	0.772	0.854	0.618	0.789	0.800	0.943	0.806	0.879	0.611	0.611	0.957	0.806	0.772	0.800	0.869	0.854	0.869			

3.2 Evaluation of the Significance of the Data

We found that there was a significantly higher class average for the simulations class for Exam 2 question #2d (Relating the correlation coefficient and coefficient of determination to the line of best fit), Exam 3 question #1d (The relationship between confidence level, interval width, and precision of the estimate of a single population mean), Final Exam question #1b (Calculating and Interpreting the Standard Deviation of a Sample Dataset), Final Exam question #1c (Performing Hypothesis Tests for one sample mean), Final Exam question #6d (The relationship between confidence level, interval width, and precision of the estimate of a single population mean), and Final Exam question #8ab (Find the probability of a situation using the area under a normal curve using statistical tables). All of these had a two sided p-value and one-sided p-value of 5% or lower and had a higher class average in the simulations class. None of the data showed any significant results where the control class had a higher average than the simulations class did, though there were some questions where the control class's class average was higher. One the next page there is a table summarizing the results for each class and the corresponding one-sided and two-sided p-value for the observed statistic of the simulation class average subtract the control class average.

Table 3. Summary Statistics of Rubric Scores

Simulations Mean	1.000	1.467	0.933	1.467	0.200	0.467	0.333	0.933	1.067	1.133	1.067	1.200	1.000	0.733	1.600	0.867	1.333
Simulations Median	1.000	2.000	1.000	2.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	2.000	1.000	1.000	2.000	1.000	2.000
Simulations Std Dev	0.894	0.718	0.249	0.618	0.542	0.806	0.699	0.929	0.854	0.806	0.854	0.909	0.730	0.680	0.712	0.806	0.869
Sample Size	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000	28.000
Control Mean	1.267	0.933	0.867	0.667	0.400	0.667	0.467	0.600	0.600	0.400	0.867	0.867	0.733	0.600	0.667	0.733	0.667
Contral Median	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000
Contral Std Dev	0.772	0.854	0.618	0.789	0.800	0.943	0.806	0.879	0.611	0.611	0.957	0.806	0.772	0.800	0.869	0.854	0.869
Sample Size	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000	15.000
Two Sided P-value of Mean	31.44%	4.99%	40.00%	0.24%	39.54%	49.25%	59.30%	25.40%	4.50%	0.19%	50.30%	22.50%	28.00%	58.80%	0.10%	62.20%	2.30%
One Sided P-value of Mean	84.28%	2.49%	34.83%	0.12%	80.23%	75.37%	70.43%	12.75%	2.27%	0.10%	25.20%	11.31%	14.02%	29.47%	0.08%	31.05%	1.20%

3.3 Pre and Post Test Changes

Students in both the simulations class and the control class were given a test at the beginning of the semester and at the end of the semester. The test consisted of four questions. In the control group, we found that most of the students understanding of the material did not change greatly. If they had a strong handle of the topic at the beginning of the semester it was still that way. Similarly, if they did not know the answer at the beginning of the class, they were still unclear by the end of the semester. Below are some of the notable answers that students gave from each of the classes and what this means for their understanding.

The first question that students answered was about the reason for studying statistics and what statistics is. Student 701 started by noting that,

> "[statistics is the study of] numbers and graphs, helps us better understand data and make observations"

and by the end of the semester said,

> "[statistics was about] probability and we study it to help solve word problems and configure data."

We noticed that this seemed to be a transition from viewing statistics as a tool to draw conclusions about data to being focused on solving homework problems. Student 707 started by saying that,

> "[statistics] deals with numbers (stats) – companies use it and are also used in studies for data"

though they later to stated that,

> "[statistics is] probability and studies to figure out certain thing/answers."

We noticed that they were getting closer to what statistics are though they still only focus on trying to get answers rather than having a deeper understanding of the real world. Student 801 started out by saying that,

> "[statistics is the] study of mathematical quantities that reflect on a study of a population. It is important to study statistics as it helps us better understand the world around us,"

but later noted stated that,

> "[statistics is the] study of probability and the means of scientific testing."

Student 817 originally noted that they did not know what statistics is but by the end of the semester they started to see that statistics is about using probabilities to make conclusions which is a step closer what we hope students see statistics are about.

The second question was about setting up their own experiment using twenty subjects with two variables. Student 701 started by simple saying they would do a blind study though by the end of the semester they listed out an experiment that involved a control group, experimental group, independent and dependent variable, and randomization. The students in simulation class, over all did not seem to have much change in how they approached this situation. Student 807 talked about using random samples both times. Student 806 and several other students talked about splitting into two groups based on variables and then comparing the two groups.

The third question was about the importance of randomization and why you need to use random samples. Student 701 started by that noting that random means there is no certain process which prevent bias by the end of the semester they still talked about avoiding bias and mentioned how each subject has to have equal likelihood of being chosen. Student 827 initially noted that randomization is used to represent a population and that if you do not use random samples the subjects may be similar to each other so they would not provide useful data. At the end of the semester they stated that using random samples “allows for the probability of being chosen equal among all people” and this will give accurate and unbiased data.

The final question asked about the relationship between the mean and median, what that would be mean about the data. Student 701 started by saying they were not sure what the relationship is, but ended by noting the relationship is effected by outliers and the skew of the data. Student 809 started by saying that you would expect it when the sample size is larger but later said that,

> “[they would expect the relationship when] there are large outliers or small outlier present in the data.”

We noticed that this student started without a real understanding of the relationship but realized that outlier can affect the values in question and how they relate. Student 801 originally thought the relationship was based on the split of the data but later noted that it was connected to the skew of the data.

3.4 Simulations Group Student Evaluations

At the end of the semester students in the simulation-based class were given a post semester evaluation of the use of simulation in the classroom and how they felt about them after a semester. Students were asked to describe if they thought the simulations were helpful to understand course content, describe if they thought they were helpful to understand the homework, if they would recommend the use of them to other students, and what they liked and disliked about the simulations. 25 students completed the survey and of those students, 11 had positive feedback for the simulations and 14 had negative feedback for the simulations. Here is some of what the students said that shows the dominate types of feedback we got back, both positive and negative. Student 801 noted that most of the applets were fairly easy to use and that it gave a good way to check their understanding of the solutions they found using the theoretical approach. Student 802 felt that the applets were not helpful because they felt all they did was "click buttons" though they did find the applets were easy to use. Student 803 said,

> "[The applets helped] visualize what happens to some variables when we manipulate others."

Student 806 felt that the simulations did not add anything to the class that they could not learn from the standard approach to the course. Student 807 found the applets confusing and hard to follow. Student 809 thought that they were not necessary unless you personally thought they were helpful,

> "[but otherwise they were a] waste of time because we were never tested on them…"

Student 813 thought that the applets were helpful because, "[you could see] step by step how to solve problems with accurate graphs of the data."

Student 817 thought that the applets worked well but should only be a resource rather than required for the course. Student 818 noted that the applets highlighted what they did not know, but they would only want to use it as a resource. Student 819 thought that the applets were useful because they provide visuals of the processes. Student 826 thought that the applets cleared up topics that they found confusing in class. Students 831 enjoyed the applets that gave them a chance to try the samples out on my own. Student 829 noted that they did not like the applets because

"unless I see the math and equations used I don't understand the data."

Overall the negative comments that we received focused on how the students felt that it was extra work in the class. We believe that this feeling could be because we started off by having the simulations added on to the standard class rather than a full integration. The students seemed to view the applets as just work to get through rather than as tool to show them details they may have missed or misunderstood. The positive comments we got, seemed to match the goals we had for providing visual learning opportunities and a way to create a clearer understanding of the material that may not be obvious in a standard lecture on statistics. And even some of the negative comments showed that the students were too focused on the calculations rather than the conclusions that they came to.

4. Conclusions

4.1 Interpretation of the Results

As we saw above, there were a few areas of inferential statistics that students who received the simulation-based approach in the introductory to statistics course improved in. The main areas that the rubric scores showed improvement in where dealing with confidence intervals and the meaning of standard deviations. The majority of questions that we reviewed about these topics showed a significant improvement in understanding if the student was in the simulation-based class. The difference in understanding for confidence intervals is consistent with the results from the study conducted at Iowa State University (Karsten Maurer, 2016). We believe that the simulations helped students to focus on the meaning of the interval rather than on how to calculate the values for the interval. The applets also allowed for them to change the confidence level to see the relationships between the level, width, and precision. We noted that the visual representation would also be helpful to students to see what the standard deviation of a data set really is and how it relates the data to the average of the data. Students were able to use the applets to see a data set, find the mean of the data set, and then see what the range of one standard deviation below and above the mean is. This shows that the simulations were helpful for the students in these two particular areas.

There were other questions in which we saw significant difference between the responses from the two classes. These questions were about the relationship between the regression line and the correlation, hypothesis testing, and finding the area under a curve to find probability. In this case we saw a significant difference between the two classes, however other questions that focused on the same material did not have a significant difference between the two classes. So we achieved mixed results for these topics as to whether the simulations helped to improve student understanding.

We can see that based on our results from the pretests and posttests that though the students may not have shown much change in their understanding of randomization and setting up experience, there seemed to be some improvement in how the simulations based class understood the purpose of statistics, and the relationship between a mean and median of a data set. We can see that the applets may have helped students by allowing them to see visuals of the relationship between the mean and median based on the skewness of data and outliers. Students also seemed to have gained a deeper understanding of what statistics is and why we want to study statistics. This goes back to our original goal for the use of simulations to have students gain a conceptual understanding of the statistical processes. Students were starting to see the bigger picture of why you need to know these process and what the goal of the work they were doing is. We hope that this is because with the use of simulations, the students focus more on the overarching goals and concepts of the course rather than on the calculations and formulas of the course. With these increases in understanding, even if they are quite modest, there is some evidence that the use of simulations is helping student understanding of statistics. We conjecture that if more simulations were incorporated into a class, or if simulations were incorporated more tightly into the course structure, we would see greater increases in student understanding.

In the simulations class, we gave the students post surveys about the simulations. As we saw above, the overall feedback was slightly more negative than positive. But the students who gave positive feedback had detailed reasons about how they liked the visuals it provided, the chance to work with data hands on, the ease of the applets themselves, and the ability to explore topics that they were unsure about. This covers several of the reasons listed at the start of the paper about why you would want to use simulations to help student understanding. While the students who gave negative feedback focused on not liking the extra work that was required and felt that it was not worth the effort. So though there was more negative feedback, it seemed that it was not about the applets hindering their learning but rather that they preferred the traditional approach and thought that the applets were just extra work on top of that.

4.2 Limitations for this Study

There are some limitations in this study that we should be aware of as we look at the results. One of the main limitations of this study is that students were not randomly assigned to the classes, but registered for the class of their choice through the university's usual registration process. As a result, we were not able to control the class sizes (the simulation-based class turned out to be about twice the size of the control class), or the prior statistical knowledge of the students. There is also the potential that students in one class were more willing to reach out for help from the class instructor, tutors, and supplemental instructors. If one class was more open to asking for help than another that could result in the class showing an increase of understanding because of the outside help. We also had no way of monitoring how often students studied, or the level of dedication the students had towards the class. As we looked through the pretests and posttest, some students left questions blank or wrote down one to two word responses. This might have influenced the data collection, if students were not fully participating in the testing. Therefore they could have not improved according to the data even if their understanding had because they did not care put in the effort on the test. We were also missing several students from the analysis because they either missed the pretest or posttest or both. A particular issue that arose in the data collection was missing students that caused there to be holes in the data.

Another limitation that we had with this study was the extent to which we were able to introduce simulations. Because this was the first study of this type to be done at UWM in Math Stat 215, we were limited on how much we could modify within the course. We were only able to use the simulations as an aid in the classroom rather than teaching through simulations. But in order to cover all the required material through the traditional approach, we could not spend as much time on applet based activities in class. We also were able to only assign a few problems that were focused on the applets, but in order to conform to the other sections of the class there was still a focus on textbook problems in the homework.

There are a few modifications that should be made for future studies to avoid some of the issues of this study. As noted, we were limited in how we could use simulations in the classroom. Know that we have seen some promising starts for the use of simulations, it would make sense that the next attempt at using simulations in the classroom would take the use of simulations a step farther. It would focus on using more simulation in the classroom, getting students using the applets in class, more focus on simulations in the homework, and possible simulation-based projects or simulation-based group work. This would allow students to become more comfortable with the applets and allow the instructor to show more concepts through the simulations. It would also help if the pretest and posttest were given with some incentive to the students so that they would give the tests their full effort. While having the test being graded may not be the best since students are not expected to know the answers at the start of the semester, instructors could offer effort based scores to ensure students fully engage with the tests.

5. Significance and Implications for Future Studies

After doing this study we needed to evaluate whether this would be a path that should be followed or if simulations should not be used again. Based on the results of this study and the research that we referenced at the beginning of the paper indicates that simulations would be a useful tool in statistics that would warrant more research into. We achieved significant improvement in our data when we look at confidence interval and standard deviations. Which shows that the applets were effective in some areas of statistics so they can be utilized to help students with this understanding. We also noticed some increased understanding of the concept of why we study statistics and measures of central tendency that indicate that simulations were helping students with understanding the overarching concepts of the class. This indicates that the use of simulations in a classroom can be helpful, and we would recommend that they be used in these types of class in the future. It would also be helpful to continue studying the impact of the simulation-based approach by conducting further studies, particularly in Math Stat 215 at UW-Milwaukee.

References

Allan Rossman, B. C. (2000). *Teaching the Reasoning of Statistical Inference: A "Top Ten" List*. Retrieved March 20, 2017, from http://www.rossmanchance.com/papers/topten.html

Andrew Zieffler, J. G. (2007). *Studying the Role of Simulation in Developing Students' Statistical Reasoning*. Retrieved March 15, 2017, from http://iase-web.org/documents/papers/isi56/IPM40_Zieffler.pdf

Ann Watkins, A. B. (2014). *Simulation of the Sampling Distribution of the Mean Can Mislead*. Retrieved March 18, 2017, from http://ww2.amstat.org/publications/jse/v22n3/watkins.pdf

Beth Chance, D. B.-Z. (2007). *The Role of Technology in Improving Student Learning of Statistics*. Retrieved March 20, 2017, from http://escholarship.org/uc/item/8sd2t4rr

Cetinkaya-Rudel, M. (2015, December 9). *Teaching Computation as an Argument for Simulation Based Inference*. Retrieved March 20, 2017, from https://www.causeweb.org/sbi/?p=971#more-971

Christine Franklin, A. B. (2015, May). *Statistical Education of Teachers*. Retrieved March 21, 2017, from http://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

Jeremy Strayer, A. M. (2016, April). *Statistical Literacy: Simulations with Dolphins*. Retrieved March 10, 2017, from http://www.nctm.org/Publications/Mathematics-Teacher/2016/Vol109/Issue8/Statistical-Literacy_-Simulations-with-Dolphins/

Karsten Maurer, D. L. (2016). *Comparison of Learning Outcomes for Simulation-based and Traditional Inference Curricula in a Designed Educarional Experiment*.

Retrieved March 22, 2017, from http://escholarship.org/uc/item/0wm523b0#page-1

Martha Aliaga, G. C. (2012). *Guidelines for Assessment and Instruction in Statistics Education: College Report*. Retrieved March 21, 2017, from http://www.amstat.org/asa/files/pdfs/GAISE/2005GaiseCollege_Full.pdf

Mia Stephens, R. C. (2014). *From Data to Decision-making: Using Simulation and Resampling Methods to Teach Inferential Concepts*. Retrieved March 17, 2017, from http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_8B3_STEPHENS.pdf

Nathan Tintle, B. C. (2015, December 29). *Combating Anti-Statistical Thinking Using Simulation-Based Methods Throughout the Undergraduate Curriculum*. Retrieved March 19, 2017, from http://www.tandfonline.com/doi/abs/10.1080/00031305.2015.1081619?journalCode=utas20

Schaffner, A. (2015, May 9). *Some Thoughts and Experiences Using Simulation/Randomization Based Methods in Introductory Statistics Courses and in the Undergraduate Statistics Curriculum*. Retrieved March 20, 2017, from https://www.causeweb.org/sbi/?p=828#more-828

Wisconsin Department of Public Instruction. (2011, September). *Wisconsin Standards for Mathematics*. Retrieved March 22, 2017, from https://dpi.wi.gov/sites/default/files/imce/standards/pdf/common-core-math-standards.pdf

Appendix A. Assessment Tools

Pre and Post Test Question:

1) What is statistics and why do we study statistics?

2) How would you set-up an experiment with twenty subjects who have to choose between two variables?

3) What is the role of randomization in statistical studies? Why is random sampling so important? What is a random sample?

4) When would you expect the mean to be greater than or less than the median?

Survey Questions:

1) Did you feel that the simulations helped in your understanding of the content of the course? Explain.

2) Do you think that this was a useful tool as you worked on your homework? Explain.

3) Would you recommend this technique for future students? Explain.

4) What did you like about the simulations?

5) What did you dislike about the simulations?

Exams Given:

MATH STAT 215 MIDTERM 1

1) For the sample shown below, find the mean and standard deviation.

12 13 18 19 20 22 36

(a) Compute the standard deviation using the definitional formula (using the table below). What does this tell you about the values?

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$

(b) Calculate the z-score for the highest observation. What this value means?

2) A sample of 25 UWM student in a Math Stat 215 class GPAs was taken are recorded below.

2.40 2.57 2.59 2.61 2.66

2.68 2.72 2.73 2.74 2.74

2.74 2.77 2.78 2.84 2.91

2.92 2.96 2.98 3.01 3.05

3.09 3.18 3.22 3.34 3.46

(a) Is this an observational study or an experiment?

(b) What are the subjects of the study?

(c) What is/are the variable/s of the study? Classify the variables as quantitative or qualitative.

(d) Create a dot plot of the information. Is the data skewed, if so which direction (left or right)?

3) For the sample below, compute the standard deviation using the computing formula.

3 5 8 9 10 12 14

4) For the dataset below:

5 6 7 9 10 11 12 12 12 19

(a) Find the 5-number summary.

(b) Determine the mean and the mode of this data set.

(c) Using the mean, median, and mode, talk about whether the distribution has any skewness, and if so, in what direction.

5) For the dataset below, do a complete box-and-line plot analysis. This analysis includes finding the quartiles, minimum, maximum, interquartile range, upper and lower limits, adjacent values, and possible outliers. (List all of the values you find out and label them)

5 16 17 20 22 23 25 25 26 28

6) An experiment was done to examine the effects of a study guide and study time on test scores. Assume all the students in this experiment were identical in all other ways. The experiment was done in such a fashion:

-- Either a student received a study guide, or the student did not;

-- different students spent either 6 hours, 8 hours, or 10 hours studying for the test.

(a) What is the response variable in this experiment?

(b) What are the factors in this experiment?

(c) List the levels of each factor.

(d) How many possible treatments can be created in this experiment?

(e) List each possible treatment.

(f) In the general sense, and not specific to this experiment, what is the difference between a blind experiment and a double-blind experiment?

7) For each of the situations below, identify the sampling design used. (Convenience Sampling, Stratified Random Sampling, Cluster Sampling, Systematic Random

Sampling, Judgement Sampling, Simple Random Sampling, Quota Sampling)

EXPLAIN.

(a) To study student opinions about campus security, I randomly sample 50 students from each dormitory.

(b) Out of the first 50 employees on the employee roster, the 13th was randomly sampled. From there, every 50th person after employee # 13 was also sampled.

(c) A number table tells me to sample the 59th, 43rd, 17th, 82nd, and 29th members of a local church to survey them on plans for expanding the parish's grade school.

(d) To find out whether or not customers of a local store would prefer to see a new brand of chocolate chip cookie sold at the store, I sample the first 10 customers willing to try a free sample.

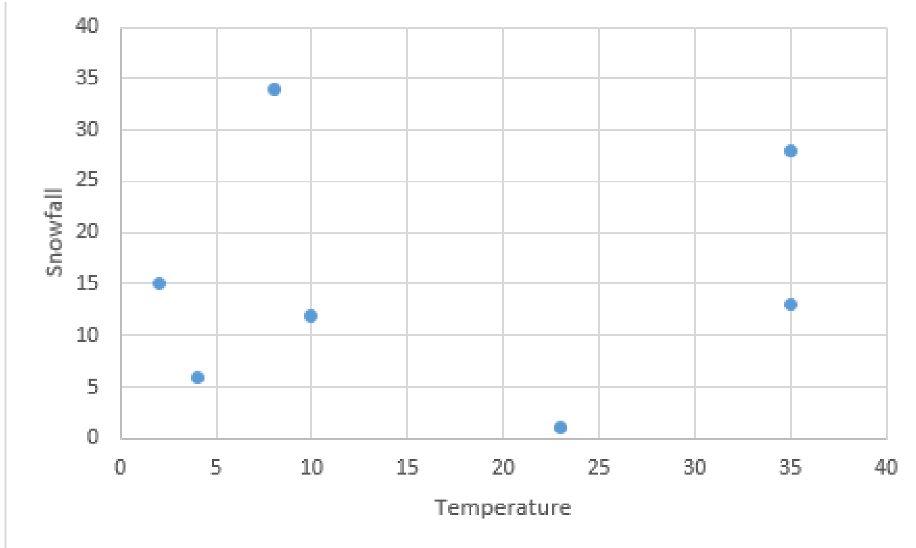
(e) Thirty people were sampled from each of 5 different income demographics to get their opinion of the quality of the service their broker provides.

8) Tim is a male student who is 5'10" tall, weighs 165 pounds, and is majoring in economics. Identify the quantitative variables I sampled on, and the qualitative variables I sampled on (there are two of each).

MATH STAT 215 MIDTERM 2

1) Suppose that you wanted to predict snowfall for this coming season. So for one week you record the temperature each day and the snowfall on that day. Below is the data you have collected for those days.

Temperature	Snowfall
35	28
23	1
10	12
35	13
8	34
4	6
2	15



a) Does finding a regression line for the data seem reasonable? (Explain)

b) Determine the regression equation for the data, and draw its graph on the scatterplot above. You may use the table below if you want to organize your information.

x	y	X^2	Y^2	XY

c) Use the regression equation to predict the snowfall for a 28 degree day.

d) Use the regression equation to figure out what temperature it is if you get 10 inches of snow.

2) Now that you have found the regression equation for using temperature to predict snowfall, you show it to your friend. Your friend questions the validity of your equation. You decide to check to see if your equation is valid or not.

a) Find the Correlation Coefficient. Show all work.

b) Find the Coefficient Determination. Show all work.

c) Compare your Correlation Coefficient to the Slope of the Regression Line you found in question 1 (b)? What is the same? What is different?

d) Based on your values for (a) and (b), do you think that this regression line is very helpful for telling you about the relationship? (Explain)

3) Suppose that you are playing the Game of Life with some friends. In order to move forward in the game, you must spin a spinner that has the numbers 1 to 10 on it (so 10 sections). Each section is the same size.

a) Find the following probabilities:

1) P(landing on 10)

2) P(landing on 4)

3) P(landing on 1)

4) P(landing on 7)

b) What does the P(landing on 10) mean (do not use the words percentage, or probability in your answer)?

Consider these possible events,

A = the number is even

B = the number is odd

C = the number is more than 8

D = the number is less than 5

d) Describe each of the following events in words and determine the probability of it happening.

1) P(Not A)

2) P(A or C)

3) P(A and C)

e) Determine which of the following are Mutually Exclusive and sketch each of the corresponding Venn Diagram

1) A and B

2) A and (not C)

3) B and D

4) A, B, and D

4) According to the UWM Degree Data, a frequency distribution for the number of different Degrees for Undergraduate, Masters, and PHD levels, is as shown.

The Level of the Degrees are labeled 1 = Undergraduate, 2 = Masters, and 3 = PHD

Level of the Degrees	1	2	3
# of Degrees	94	64	33

Let X denote the Level of a randomly selected Degrees.

a) What are the possible outcomes for the random variable X?

b) Use random-variable notation to represent the event that the Degree you pick is a Master's Degree.

c) Determine the probability distribution of the random variable X.

d) What is P(X = 3), and interpret your answer in terms of percentages.

e) Find the Mean of the Discrete Random Variable X.

f) Find the Standard Deviation of the Discrete Random Variable X.

5) Suppose you are attempting to win a Free Throw contest at a Panther Game. The likelihood of you making the shot is 35%. You need to make a certain amount of shots out of 10 to win a T-shirt.

(a) Use the binomial distribution computing formula to determine the probability that in a sample of 10,

(i) You make exactly 2shots.

(ii) You make exactly 8 shots.

(iii) You make all 10 shots.

(c) Determine the mean and standard deviation of the binomial variable.

(d) If you were to make a Histogram of the Probability distribution of the Binomial Variable, which way would the graph be Skewed? (Explain)

(e) Use the Binomial Distribution Table below to verify your answers for question (a). You will circle the solutions in the chart for (i), (ii), and (iii).

MATH STAT 215 MIDTERM 3

1) The number of research journals in a university library is assumed to be normally distributed with a standard deviation $\sigma = 10.2$. If, for a sample of 200 university libraries, the sample mean number of research journals is 180.4, determine the following confidence intervals to approximate μ, the number of research journals in a typical university library. SHOW ALL WORK AND INTERPRET EACH.

(a) 99 % confidence interval

(b) 95 % confidence interval

(c) 90 % confidence interval

(d) What is the relationship between confidence level, interval width, and precision of the estimate of a single population mean?

2) The number of pine trees on a mountain slope in western Wyoming varies normally. If the population of numbers of pine trees on mountain slopes in western Wyoming has a standard deviation $\sigma = 22.6$, and a sample of 40 mountain slopes has a mean of 242.4 pine trees per slope. You will perform a hypothesis test at the 5 % level of significance whether the mean number of pine trees per mountain slope is different from 250. Make sure you show all 6 steps.

(a) Using the Critical Value Approach

(b) Using the P-value Approach

3) The mean price of a gallon of E-85 fuel at Milwaukee County gas stations varies normally with a standard deviation $\sigma = \$0.08$. A local reporter is trying to estimate the actual retail price of E-85 in the Milwaukee area. He samples 60 gas stations and finds a mean retail price of $1.82 per gallon for E-85 fuel.

(a) Test at the $\alpha = 0.05$ level the assumption that the mean price of a gallon of E-85 fuel is above $1.80. Make sure you show all 6 steps. (Use the approach of your choice – Indicate which you choose at the start of the steps)

(b) What is the Type I error in this situation? What is the Type II error this situation?

4) The local bus service planner wants to determine whether he can reduce non-rush hour service on a given route. He samples 20 days worth of ridership between 10 AM and 3 PM for the route, and finds a sample mean number of riders of 182.4 and a sample standard deviation of 8.8 riders.

Test at the 5% level of significance whether or not the mean number of riders on this route between 10 AM and 3 PM is less than 185. Assume the number of riders per day between 10 AM-3 PM varies normally. Make sure you show all 6 steps. (Use the approach of your choice – Indicate which you choose at the start of the steps).

5) The number of problems in a typical homework section in a math textbook is normally distributed with a mean μ of 54.2 and a standard deviation σ of 12.6.

(a) Determine the mean and standard deviation of the sampling distribution of the sample means for samples of size 60.

(b) What is the likelihood that the mean number of homework problems in a homework section will be less than 53?

(c) What is the likelihood that the mean number of homework problems in a homework section will be between 53 and 56?

6) A 95% confidence interval was set up for the mean number of students in a typical UWM World History lecture. The confidence interval is shown below.

$\mu = (90.2, 93.6)$

(a) Find the sample mean value that was used to set up this interval.

(b) What is the margin of error of this interval?

(c) If $\sigma = 5.5$ was used in creating the above confidence interval, then determine the sample size, n, necessary to create a new confidence interval with a margin of error of 1.2.

MATH STAT 215 FINAL EXAM

1) Below are the number of loads of laundry done per day at a local laundromat washer for the past 12 days.

13, 15, 17, 9, 10, 22, 14, 16, 15, 13, 8, 15

a) Answer the following questions about this situation.

- Is this an observational study or an experiment?
- What is the subject of the study?
- What are the variables of the study? Classify the variables as Quantitative or Qualitative.

b) Calculate the sample mean and sample standard deviation for the number of loads above. What does the standard deviation represent?

c) Suppose that the owner assumes that on average 15 loads are done a day. Based on the sample above and the information you calculated in part (b). Perform a hypothesis test at the 5% level whether the mean number loads washed per day is more than 15. Show all 6 steps and all work.

d) Create a 99% confidence interval based on the values from part (b) to estimate the value of the population mean. Interpret the interval.

2) Use the data set below to answer the following questions.

30 32 35 39 42 46 48 50 52 55 75

(a) Find the 5-number summary

(b) For the dataset, do a complete box-and-line plot analysis. This analysis includes finding the range, interquartile range, and the upper and lower limits.

20 50 80

(c) Determine the mean and the mode of this data set.

(d) Using the mean, median, and mode, talk about whether the distribution has any skewness, and if so, in what direction.

3) A linear relationship was fit to a data set that sampled on two variables: number of hours spent behind the wheel commuting to work per week, and number of hours spent per week on hobbies.

(a) Which variable is the response variable (y), and which variable is the explanatory or stimulus variable (x)?

(b) The linear relationship fit to the data had the best-fit equation $y = -1.25x + 4.65$. Identify the slope and the y-intercept of the equation. Then explain what each of the numbers in this equation represents in the context of this problem.

(c) The regression process yielded a value of $r^2 = 0.975$ for this trend line. Tell me what r^2 is called, what it represents, and what it tells you about the linear relationship imposed on this data set.

(d) What is the correlation coefficient for this data set? What does it tell you about the correlation between the two variables?

4) Below is a distribution of a random variable:

x	Frequency	P(X = x)
1	12	
2	15	
3	22	
4	17	
Total		

a) Fill in the last column and bottom row of the table, by finding the probability of randomly picking that value for X.

b) Find the Mean and Standard Deviation of this Discrete Random Variable.

c) In general, what does it mean for two events to be mutually exclusive?

5) If the probability of a car in a parking being an import is 0.40, then determine, for a sample of size 10,

(a) the probability of exactly 2 cars being imports (USE THE BINOMIAL PROBABILITY COMPUTING FORMULA);

(b) the mean and standard deviation of this binomial variable.

6) The number of movies on TV in a given week is assumed to be normally distributed with a standard deviation $\sigma = 8.74$. If, for a sample of 60 weeks, the sample mean number of movies on TV is 68.2, determine the following two-sided confidence intervals to approximate μ, the mean number of movies on TV in a week. **Interpret the Intervals.**

(a) 90 % confidence interval (show all work)

(b) 95 % confidence interval (show all work)

(c) 99 % confidence interval (show all work)

(d) What is the relationship between confidence level, interval width, and precision of the estimate of a single population mean?

7) The number of "fruit scummies" sold daily at the local campus cafe is assumed to be equal to 40 with a standard deviation of 7.8.

7*) If a sample of 12 days of sales shows a sample mean number of fruit scummies sold equal to 43.1, test at the 5% level whether the mean number of fruit scummies sold per day is fewer than 40. Make sure to show all 6 steps. Show all work.

(b) What is the type 1 error in this case? What is the type 2 error in this case?

(c) What sample size would you need to use have a 90% confidence interval with a margin of error of 5?

8) If the mean number of people riding the route 62 bus between 8-9 AM is $\mu = 216.3$, with a standard deviation $\sigma = 10.4$, determine the probability of each of the following and write your answers using the correct probability notation:

(a) between 200 and 220 riders being on the bus between 8-9 AM on a given day;

(b) more than 240 riders on the bus between 8-9 AM on a given day.

9) For each of the following situations:

(i) Determine if you should use a one mean or two mean approach

(ii) If is it one mean, choose the z-score or the t-score OR if it is two mean, choose pooled, non-pooled, or paired.

(iii) State the Null and alternative Hypothesis for each situation.

(a) The mean price of a gallon of E-85 fuel at Milwaukee County gas stations is said to be \$1.90 and varies normally with a standard deviation σ = \$0.08. A local reporter is trying to estimate the actual retail price of E-85 in the Milwaukee area. He samples 60 gas stations and finds a mean retail price of \$1.82 per gallon for E-85 fuel.

(b) The local bus service planner wants to determine whether he can reduce non-rush hour service on a given route and is told that the average number of riders is 190. He samples 20 days worth of ridership between 10 AM and 3 PM for the route, and finds a sample mean number of riders of 182.4 and a sample standard deviation of 8.8 riders.

(c) A researcher is examining the nutritional intake of two sub-Saharan populations to see if there is a difference in malnutrition levels between adult women in the two populations. He samples on daily iron intake of adult women, assuming the standard deviations of the two populations will be equal. He finds that, from a sample of 111 adult women drawn from the first population, the sample mean daily iron intake is 15.4 mg, while from a sample of 91 adult women drawn from the second population, he observes a sample mean daily intake of 17.2 mg.

(d) A study is conducted to compare the level of pain tolerance between men and women. The researcher takes two groups each with 50 people, and matches each man & woman together based on a mild stress test results. They then performed shock therapy on each participant while recording the stress level of each on a scale of 1 to 10 and found the average difference between men and women to be 1.5 with a standard deviation of 0.4.