

TRADITIONAL LECTURE VERSUS AN ACTIVITY APPROACH FOR  
TEACHING STATISTICS: A COMPARISON OF OUTCOMES

by

Jennifer L. Loveland

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Mathematical Sciences

Approved:

---

Dr. Kady Schneiter  
Major Professor

---

Dr. Brynja Kohler  
Committee Member

---

Dr. Yan Sun  
Committee Member

---

Dr. John Stevens  
Committee Member

---

Dr. Patricia Moyer-Packenham  
Committee Member

---

Dr. Mark McLellan  
Vice President for Research and  
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2014

Copyright © Jennifer L. Loveland 2014  
All Rights Reserved

**ABSTRACT**

Traditional Lecture Versus an Activity Approach for Teaching Statistics:  
A Comparison of Outcomes

by

Jennifer L. Loveland, Doctor of Philosophy  
Utah State University, 2014

Major Professor: Dr. Kady Schneiter  
Department: Mathematics and Statistics

Many educational researchers have proposed teaching statistics with less lecture and more active learning methods. However, there are only a few comparative studies that have taught one section of statistics with lectures and one section with activity-based methods; of those studies, the results are contradictory. To address the need for more research on the actual effectiveness of active learning methods in introductory statistics, this research study was undertaken.

An introductory, university level course was divided into two sections. One section was taught entirely with traditional lecture. The other section was taught using active learning methods and a minimal amount of lecture. Both sections were taught by the same instructor during the same semester. The experiment was repeated the next semester.

Students' exam scores were analyzed to determine if the activity-based teaching approach led to higher student comprehension and understanding of statistical concepts, and the ability to apply statistical procedures. Surveys were also administered to students to ascertain if the lecture or activity-based approach led to higher, more positive student attitudes toward statistics.

Analysis of the data did not show that the activity-based teaching method led to higher student comprehension or procedural ability. Neither teaching method led to significantly higher student attitudes. Student comments indicated a positive response to the activity-based methods, but the responses also indicated a student desire for more teacher-centered time in the activity course.

(266 pages)

## **PUBLIC ABSTRACT**

### **Traditional Lecture Versus an Activity Approach for Teaching Statistics: A Comparison of Outcomes**

by

Jennifer L. Loveland, Doctor of Philosophy  
Utah State University, 2014

Many educational researchers have proposed teaching statistics with less lecture and more active learning methods. However, not enough research has been done to show that teaching a statistics course with active learning methods is actually effective. This research study was undertaken to provide evidence that the active learning methods are effective in teaching introductory statistics.

An introductory, university level course was divided into two sections. One section was taught entirely with traditional lecture. The other section was taught using active learning methods and a minimal amount of lecture. Active learning methods included group work, hands-on activities to gather data, and periods during which the instructor acted as a guide to lead the students to discovering statistical concepts. Both sections were taught by the same instructor during the same semester. The experiment was repeated the next semester.

Students' exam scores were analyzed to determine if the activity-based teaching approach led to higher student comprehension and understanding of statistical concepts, and the ability to apply statistical procedures. Surveys were also administered to students to ascertain if the lecture or activity-based approach led to higher, more positive student attitudes toward statistics.

Analysis of the data did not show that the activity-based teaching method led to higher student comprehension or procedural ability. Neither teaching method led to significantly higher student attitudes. Student comments indicated a positive response to the new activity-based methods. However, student responses indicated a desire for more teacher-centered time in the activity course. In particular, students wanted the teacher to introduce the topic at the beginning of each class before they started working in groups.

## ACKNOWLEDGMENTS

I would like to express gratitude to my advisor, Dr. Kady Schneider, and to the rest of my committee members, Drs. John Stevens, Brynja Kohler, Yan Sun, and Patricia Moyer-Packenham, for all their help and ideas. I also appreciate all the other professors and graduate students who have contributed to my doctoral experience.

I would like to thank the Mathematics and Statistics Department at Utah State University for allowing me to conduct my research with their classes. I am also grateful to the department for giving me the opportunity to teach various courses throughout my graduate career; it has been a great learning experience.

Many thanks to my family on all sides who are always willing to help and support me, especially my parents who never stop believing in me. Of course, I never could have completed such an monumental undertaking without the support of my loving husband who rarely complained about boredom while I was working.

Jennifer L. Loveland

## CONTENTS

	Page
<b>ABSTRACT . . . . .</b>	<b>iii</b>
<b>PUBLIC ABSTRACT . . . . .</b>	<b>iv</b>
<b>ACKNOWLEDGMENTS . . . . .</b>	<b>v</b>
<b>LIST OF TABLES . . . . .</b>	<b>ix</b>
<b>LIST OF FIGURES . . . . .</b>	<b>x</b>
<b>INTRODUCTION . . . . .</b>	<b>1</b>
Background of the Problem . . . . .	1
Purpose of the Study . . . . .	3
Research Questions . . . . .	5
Defining Terms . . . . .	5
<b>LITERATURE REVIEW . . . . .</b>	<b>7</b>
Why Are Reform Efforts Needed? . . . . .	7
Proposed Methods for Teaching Statistics . . . . .	13
Obstacles to Using the Reform Oriented Methods . . . . .	25
Previous Research Studies . . . . .	27
Possible Assessment Instruments . . . . .	34
<b>METHODS . . . . .</b>	<b>39</b>
Research Design . . . . .	39
Participants and Setting . . . . .	40
Procedures . . . . .	42
Data Sources and Instruments . . . . .	49
Data Analysis Methods . . . . .	54
<b>RESULTS . . . . .</b>	<b>61</b>
Implementation of Teaching Methods and Development of Instructional Material . . . . .	61
Compare Exam Scores by Teaching Method . . . . .	74
Compare Student Attitudes by Teaching Method . . . . .	84
Results from Teacher Observations . . . . .	89
Results from Students' Comments and Surveys . . . . .	101
<b>DISCUSSION . . . . .</b>	<b>117</b>
Limitations and Future Research Possibilities . . . . .	117
Final Conclusions . . . . .	120
<b>REFERENCES . . . . .</b>	<b>122</b>
<b>APPENDICES . . . . .</b>	<b>130</b>
Appendix A: Letter of Information and IRB Approval . . . . .	131
Appendix B: Posttest . . . . .	135
Appendix C: Pretest . . . . .	147
Appendix D: Online Surveys . . . . .	153
Appendix E: Allocation of Teaching Time . . . . .	161

Appendix F: The Normal Distribution Unit for the Activity Classes . . . . .	169
Appendix G: The Sampling Distribution for Sample Means Unit for the Activity Classes	195
Appendix H: Applets Used In Classes . . . . .	213
Appendix I: Examples of Revisions Made to Instructional Materials . . . . .	217
Appendix J: Examples of Teacher Logs . . . . .	238
<b>CURRICULUM VITAE . . . . .</b>	<b>255</b>

# **LIST OF TABLES**

Table	Page
1 Statistics courses offered by Mathematics and Statistics Department. . . . .	40
2 Colleges in which STAT 2000 students for the study were enrolled. . . . .	41
3 Items on the posttest that were copied, or slightly adapted, from existing resources. . . . .	50
4 Items on the posttest that were adapted from existing resources. . . . .	51
5 Items on the posttest that were created for this study. . . . .	51
6 Percentage of class time spent in teacher-centered activities. . . . .	63
7 Number of students eligible for the study in each class. . . . .	75
8 Descriptive statistics for the exam points by lecture and activity. . . . .	77
9 P-values for comparing exam scores by teaching method. . . . .	77
10 P-values for comparing exam scores by method and by semester. . . . .	79
11 Results from multiple regression for the base model. . . . .	80
12 Results for stepwise selection model. . . . .	81
13 Correlation of predictor variables with total exam points. . . . .	82
14 P-values for comparing scores for topics by teaching method. . . . .	84
15 P-values for comparing individual exam questions by teaching method. . . . .	85
16 P-values for the two-sided Wilcoxon rank-sum test for ATS questions. . . . .	86
17 P-values for the two-sided Wilcoxon rank-sum test for self-efficacy questions. . . . .	87
18 P-values for the two-sided Wilcoxon rank-sum test for end of semester survey ques- tions. . . . .	88
19 Cost of activities. . . . .	91
20 Results for how much students felt they learned with activities. . . . .	102
21 Results for how much students felt they enjoyed the activities. . . . .	102
22 Results for student desired group size. . . . .	102



23	Students' choices for preferred teaching method. . . . .	103
24	Amount of time spent in teacher-centered activities during the spring semester. . .	162
25	Amount of time spent in teacher-centered activities during the fall semester. . . .	163
26	Amount of time that students worked on their own during the lecture class. . . .	168

## LIST OF FIGURES

Figure	Page
1    Examples of original symmetric figures. . . . .	73
2    Examples of new symmetric figures. . . . .	73
3    Distribution of exam points by teaching method. . . . .	78
4    Distribution of final exam points and total points by class. . . . .	79
5    Scatterplots of predictor variables against total exam points. . . . .	82
6    Bar chart for self-efficacy item five: . . . . .	87

## INTRODUCTION

Data and statistics pervade all aspects of society. Statistics can be easily found on economics, health, politics, living quality, education, sports, and almost every facet of life. As the need for statistical knowledge, reasoning, literacy, and thinking has become apparent, an emphasis has been placed on statistics education. Unfortunately, statistics has long been viewed as a challenging subject by students. In an attempt to improve the effectiveness of statistics education, educators and researchers have proposed a variety of reform methods including fewer lectures, more active learning, more cooperative learning, using technology, and focusing more on concepts than procedures (Aliaga et al., 2010).

Notwithstanding the amount of literature on reforming statistics education, very few studies have been conducted to determine the effects of using the new techniques on students learning outcomes and attitudes. Regardless of the number of advocates of reform methods, the prior studies have given conflicting results. This research study compares the use of traditional lecture and activities in an introductory statistics course. The purpose of this study was to determine which teaching method resulted in higher student outcomes on both conceptual and procedural topics.

### Background of the Problem

Statistics are essential in science, medicine, psychology, and many other fields. The media commonly quotes numbers and statistics as evidence of their claims. It has become necessary for every citizen to be statistically literate. Every person should be able to interpret the statistics they encounter in their personal or professional lives and be able to make decisions based on the data (Aliaga et al., 2010; Batanero, Godino, & Roa, 2004a; Pfannkuch, 2008).

Since the 1980s, interest in statistics education has grown, resulting in an increase of statistical education research and literature. Statistics topics have become more prevalent in elementary, secondary, and university settings (Aliaga et al., 2010; Garfield, 2002). As the number of students taking introductory statistics courses has risen, the need for an improved statistics course has become apparent (Moore, 1997). Since, the reform movement started in the early 1990s (delMas, Garfield, Ooms, & Chance, 2007), a multitude of ideas on how to change the introductory statistics

course have been presented (Cobb, 1992; Garfield, 2002; Moore, 1997). The more prevalent suggestions include more focus on concepts, more active learning, and more technology (Aliaga et al., 2010; Cobb, 1992; Moore, 1997).

The committees and educators have proposed that statistics should be taught with fewer lectures and more active learning methods. In fact, literature for many educational fields discuss the use of active learning strategies. Nevertheless, while many teachers are trying to introduce innovative methods, a substantial number of teachers and professors find themselves resorting to the old standby of a lecture (Meletiou-Mavrotheris, Lee, & Fouladi, 2007; Meletiou-Mavrotheris, 2003).

Many established teachers either have no desire to revamp their curriculum, no time to do so, or no idea how to actually undertake such a monumental task. Incoming teachers are presented with new ideas of how to teach, but few examples. Many teachers have never sat in a class that was not primarily taught using lectures; given this fact, it is not surprising that it is difficult to find classes that are actually being taught with fewer lectures (Bonwell, 1995).

Nevertheless, as the literature on statistics reform has expanded, more instructors have started trying to revise their courses. Anecdotal evidence suggests that the majority of these instructors have enjoyed their new courses (Garfield, 2002) and some studies have suggested improved student attitudes and learning outcomes (Batanero, Tauber, & Sánchez, 2004b; Carlson & Winqvist, 2011; Keeler & Steinhorst, 1995). However, few comparative studies have been done to actually validate the theory of improved student outcomes with reform methods.

### **Lack of Existing Research**

The literature is unclear on whether using the reform methods increases students' conceptual understanding (delMas et al., 2007). The majority of evidence supporting the use of reform methods in statistics is anecdotal (Weltman & Whiteside, 2010). Garfield postulated,

It is one thing to state that statistical thinking and reasoning should be the focus of a course, or should be the desired course outcomes. It is another matter entirely to achieve this, and to determine how well students are able to think and reason using statistical information upon completion of their first course, or later in their lives. Many researchers believe that appropriate content, a focus on data analysis and real problems, and careful use of high quality technological tools will help students better achieve the suggested course goals and outcomes. However no one has yet demonstrated that a

particular set of teaching techniques or materials will lead to the desired outcomes.  
(2002, p. 10)

Research on active learning techniques for statistics has shown mixed results. Some studies have shown higher attitudes or test scores with active learning techniques (Keeler & Steinhorst, 1995), while other studies have shown little or no effect on student learning outcomes (Meletiou-Mavrotheris et al., 2007; Pfaff & Weinberg, 2009). Finally, some studies have shown a detrimental effect when using active learning methods (Brandsma, 2000; Weltman & Whiteside, 2010). One possibility for the conflicting results is the variety of active learning methods that have been used in the studies (Carlson & Winqvist, 2011). In fact, Garfield (2002) proposed that it is difficult to find instructors who agree just how to implement the reform suggestions.

The most obvious flaw in the existing research is that few comparative studies have been completed in which a statistics course has been taught using lecture in some sections and reform methods in other sections. Even fewer are studies in which the sections were taught for the same course, in the same semester, and by the same teacher (Giraud, 1997). Variation in exam scores, student comprehension of concepts, and student attitudes is expected to occur from instructor to instructor. The ability and background of students also varies from section to section and semester to semester. A study is needed that controls for these sources of variability.

The majority of the existing studies on statistics education examine the results of integrating a few activities into the existing lecture course (Batanero et al., 2004b; Brandsma, 2000). Other studies have been conducted by teaching a lecture course and implementing cooperative learning for part of the class time (Giraud, 1997; Keeler & Steinhorst, 1995). Studies that actually teach the entire statistics course with reform methods are very rare.

### **Purpose of the Study**

As there is little data comparing student results in lecture versus activity-based statistics courses, it is important to conduct further research in this context. The defining purpose of this study was to establish if there is a difference in students' mastery of statistical concepts between the teaching methods, and if so, to determine which method resulted in higher comprehension and mastery of statistical concepts among students.

Another purpose of this study was to determine if it is feasible to teach an entire course without lectures, or with minimal lecture as the case may be. Would the preparations be too time consuming or expensive? Would the students resist the change? Non-lecture methods tend to require more time to cover the required material than lecture does; would it be possible to cover the entire curriculum using the active learning methods?

There are a few studies in which the professors have completely changed the curriculum or taught using innovative textbooks designed to be active learning friendly. However, one of the goals of this study was to teach an activity-based course without changing the textbook or typical curriculum established in the department. In other words, could the new teaching method be used with the established curriculum and textbook?

The final objective of this study was to develop enough activities to teach the entire course using active learning methods, with minimal lecture. In addition, student comments and the instructor's informal assessments would be used to improve the materials developed for each course.

In order to gather information about whether there is a difference in the results of the learning methods, and if so, to determine which method produces higher comprehension and mastery of statistical concepts among students, a developmental or design research study was implemented. One section of an university level introductory statistics was taught using each method. The first section was taught entirely with lectures. The second section was taught using minimal lecture; the methods included hands on activities, discovery-based units, and group work.

Students' results on their midterm and final exams were analyzed to determine which method produced higher comprehension and mastery of statistical concepts, or if there is no difference in student outcomes between the methods. And yet, looking at the overall scores might not show the entire picture. The lecture and activity sections were also compared to determine which method produces higher results on conceptual questions, and which method facilitates higher results on procedural questions. Students' attitudes towards statistics were assessed to determine which method resulted in more positive attitudes.

When strengths and weaknesses of each approach are understood, changes can be implemented to improve statistics education.

## **Research Questions**

The following questions guided this research study.

1. Which teaching method resulted in higher levels of student comprehension of statistical concepts and ability to apply statistical procedures?
2. Which teaching method resulted in higher student outcomes for conceptual level questions?
3. Which teaching method resulted in higher student outcomes for procedural level questions?
4. What was learned about implementing an activity-based statistics course?

As the study was conducted, the following questions were also considered, though informally.

5. Which teaching method produced higher student learning outcomes for specified statistical content?
6. Were there differences in student attitudes towards statistics or the statistics course, based on which teaching method was used?

## **Defining Terms**

The following terms are defined for use throughout this study.

- Active learning: Any teaching method in which students are directly involved in the learning process, e.g. through collaborative learning, projects, or discussion (Bonwell, 1995).
- Activities: The actions, teaching methods, and instructional material used during the active learning class.
- Applet: An interactive technology application designed for a specific purpose; usually internet-based.
- Assessment: A process of gathering information about a student's progress on the learning goals of a course (Garfield, 1994). Assessment can also refer to the instrument used to gather information.

- Conceptual: Related to ideas and concepts; often used in terms such as conceptual knowledge, conceptual understanding, or conceptual level questions.
- Constructivist Learning: A teaching method in which students are encouraged and enabled to construct their own knowledge and understanding, rather than simply receiving information (Garfield, 1995).
- Content: Related to what the students are supposed to learn (Cobb & Moore, 1997).
- Cooperative Learning: A teaching method in which students work together to facilitate their own and each other's learning (Giraud, 1997).
- Group Work: The colloquial term for cooperative learning.
- Pedagogy: Related to the teaching methods used to help the students learn (Cobb & Moore, 1997).
- Procedural: Related to the specific steps and knowledge needed to solve a problem; often used in terms such as procedural knowledge, procedural understanding, or procedural level questions.
- Reform Methods: The methods that have been proposed for improving statistics education; e.g. fewer lectures, more active learning, cooperative learning, technology, and assessment (Aliaga et al., 2010).
- Self-efficacy: A student's view of his or her own knowledge and ability (Hall & Vance, 2010).



## **LITERATURE REVIEW**

The literature review that follows describes the need for statistics education reform, the benefits and implementation of active learning techniques, and specific reform oriented suggestions for teaching statistics. Finally, the existing statistics education literature is examined for previous studies that compare the effects of using active learning methods versus lecture.

The literature was searched using terms such as statistics, teaching statistics, statistics education reform, active learning, activities versus lectures, group work, statistics without lecture, workshop statistics, teaching statistics with activities, statistical reasoning, and GAISE. The reference lists of articles were used to discover more articles. The archives for the last five years of the Journal of Statistics Education were searched manually.

### **Why Are Reform Efforts Needed?**

Introductory statistics courses are often considered difficult and confusing by students (e.g. see Batanero et al., 2004a; Cobb & Moore, 1997; delMas, Garfield, & Chance, 1999; delMas et al., 2007; Gal, 2004; Garfield & Ben-Zvi, 2007). Many students do not understand the basic concepts, and their limited understanding may be shallow and fragmented (Chance, delMas, & Garfield, 2004; delMas et al., 1999; Garfield & Ben-Zvi, 2007; Moore, 1997). Students are often able to come to a correct solution without understanding the statistical concepts behind the procedure (Garfield, 1994; Meletiou-Mavrotheris et al., 2007). Reforms in statistics education are intended to enable students to emerge from a statistics course statistically literate, with reasoning abilities they can accurately apply to statistical situations, and the ability to apply statistical methods correctly (Garfield & Chance, 2000).

### **Students Need to be Actively Engaged in the Learning Process**

Traditionally, teachers are the authority in the classroom, delivering the knowledge to their students and controlling what the students learned. The reform movement focuses on the student. Learning comes by the student's activities, with the teacher acting as a guide and facilitator (Garfield, 1995; Moore, 1997, 2005). As Moore affirmed, "Students are not empty vessels to be

filled with knowledge poured in by teachers; they inevitably construct their own knowledge by combining their present experiences with their existing conceptions” (1997, pp. 124-125).

As Garfield (1995) asserted, learning is more than remembering information. Explaining to students the correct method to solve a problem does not always facilitate learning. Moore described the result of traditional teaching with lectures:

The result is often a formal knowledge of facts and procedures divorced from intuition and from the student’s knowledge of other subjects. Formal knowledge is fragile—students cannot solve problems formulated in unfamiliar ways and cannot apply the facts and procedures they have learned to higher-order tasks such as analyzing open-ended situations and solving problems that require several steps and selection from a wide body of available procedures. That is, teaching as information transfer tends to leave students with an algorithmic rather than a conceptual understanding.  
(2005, p. 1)

### **Many Students Find Statistics Challenging**

Statistics has historically been difficult for many students, especially at the college and university level. Students are often confused by statistical ideas (delMas et al., 1999). Results in statistics and probability are often counterintuitive (Batanero et al., 2004a). Two of the main topics of statistics, probability and statistical inference, are notoriously challenging for students. Cobb and Moore (1997) attested that probability is one of the hardest mathematical concepts to grasp (Gal, 2004). Students tend to have difficulties comprehending statistical inference, along with the underlying concept of sampling distributions (delMas et al., 1999, 2007; Gal, 2004).

Much of the material covered in an introductory statistics course is new to students. Even when a math prerequisite is established for the course, students come from a variety of mathematical backgrounds with a assortment of mathematical skills and attitudes. Commonly, students dislike or even fear mathematics. Students in math classes often express a dislike for story problems, and statistics by nature involves many story problems.

In addition, difficulties arise from the state of statistics education at the post secondary level. Many instructors teaching statistics at the college and university level are not statisticians. The majority are mathematicians, but many are from other departments (Garfield, 2002). The few instructors who are actually statisticians often teach large classes of up to 500 students (Garfield, 2002).

### *Difference Between Mathematics and Statistics*

A common viewpoint is that statistics is part of the mathematics discipline. Indeed, many statistics courses are taught by mathematicians. Nevertheless, many statisticians and educators hold the conviction that statistics is not merely a sub-branch of mathematics, but is more akin to physics or engineering (Cobb & Moore, 1997; Garfield, 2003) in that statistics is its own field that uses mathematics. A statistics problem may require mathematics, but it also involves generating data, exploring data, choosing an appropriate statistical model or test, and analyzing data. In addition, statistics requires a different type of reasoning than mathematics does (Garfield, 2003).

An obvious difference between mathematics and statistics is that statistics is based in context (delMas, 2004). Mathematics can deal with numbers stripped of any context. In statistics, the context of the numbers can be crucial; solutions to statistical problems need to be communicated in context (Cobb & Moore, 1997; Garfield, 2003).

Perhaps the most distinct difference between mathematics and statistics is in the nature of each field. Mathematics is deterministic; questions ordinarily have a single correct solution. While on the other hand, statistics is based on variability and uncertainty (Garfield, 2003). This difference commonly leads to confusion among students who enter a statistics course after a long mathematical career.

Because mathematics commonly has only one correct answer to a problem, statistics teachers trained in mathematics often focus on a procedural approach to statistics (Nicholson & Darnton, 2003). As a result, students can often give a correct answer to a statistical question without understanding the underlying statistical concepts (Garfield, 1994; Meletiou-Mavrotheris et al., 2007). Even when students are asked to write a conclusion or interpret their findings, it is possible to do so in one sentence, parroting back memorized and formulaic interpretations, without statistical understanding.

### *Statistical Concepts can be Abstract*

Many concepts introduced in statistics are unfamiliar and abstract in their mathematical nature (delMas, 2004). The abstract nature of statistical concepts commonly proves challenging for students. For example, standard deviation is often introduced as a formula without an explicit connection to a concrete idea; students often ask, “What can I do with it?” or “What does it really

mean?” As another example, consider a p-value. Students typically desire a concrete interpretation of the p-value such as the common, but false, notion that a p-value is the probability that the null hypothesis is true (Schneider, 2008). The theoretically correct, but very abstract, interpretation may feel like circular logic.

### *Statistics Deals with Variability and Uncertainty*

Statistical content is based on variability and uncertainty (Cobb & Moore, 1997; Garfield, 2003). Students are used to deterministic answers in mathematics, making it frustrating when they can not get a firm answer in statistics. In statistics, there is always a possibility of an error and everything is related to probabilities. Students often express the ideas, “How can I be sure?” or “We can not tell if our answer is correct, so what is the point?”

### *Difficulties with Probability*

Probability is often considered one of the most difficult topics in not only statistics, but basic mathematics (Cobb & Moore, 1997; Moore, 1997; Garfield, 1995, 2003). One unique element of probability is the counterintuitive results. As an example, consider tossing a coin five times, each toss resulting in a tail. Intuitively, people expect that one of the next few tosses must be a head. This is seen in the common statement, “my luck has to change soon.” In truth, each coin flip is independent. The results of the previous coin tosses have no bearing on the next coin toss (Batanero et al., 2004a). Bayes’ rule is also counterintuitive; students are asked to consider the difference between the probability that a person with a disease tests positive for the disease and the probability that a person who tests positive actually has the disease. One reason that statistics and probability can be confusing is that rather than using logical reasoning where a proposition must be true or false, in probability there is no certainty about whether an event will occur or which event will occur (Batanero et al., 2004a).

Research has shown that students have many misconceptions about probability. As Garfield validated, “research suggests that even people who can correctly compute probabilities tend to apply faulty reasoning when asked to make an inference or judgment about an uncertain event, relying on incorrect intuition” (2003, p. 26). Students tend to treat all events as equally likely. In addition, students commonly fixate on outcomes of single events when viewing probability instead

of using a long run frequency approach (Garfield, 2003). For instance, a common method of reasoning is as follows: “the probability of developing a rash is 10%, and that is a small probability, so I will not get a rash.” The correct interpretation is that 1 out of 10 people will get a rash.

### *Difficulties with Randomness*

Another confusing aspect of statistics is randomness. In everyday situations, randomness implies a chaotic or arbitrary process while the term random has a much different meaning in statistics. Many students have an incorrect intuition about which events are random (Garfield, 2003). For example, consider tossing a coin eight times. Students tend to think that {HHTHTHTH} is a more likely outcome than {HHHHHTTT}. Finally, random sampling is also misunderstood (Garfield, 1995).

### *Difficulties with Formal Inference*

Formal inference involves drawing conclusions about a population from a sample. Inference is typically taught with confidence intervals and significance tests. Both types of procedures rest on the concept of sampling distributions; understanding sampling distributions has been shown to be difficult for students even at the university level (Cobb & Moore, 1997; Cobb & McClain, 2004). Chance et al. described students’ difficulty with inferential statistics.

While many students may be able to carry out the necessary calculations, they are often unable to understand the underlying process or properly interpret the results of these calculations. This stems from the notoriously difficult, abstract topic of sampling distributions that requires students to combine earlier course topics such as sample, population, distribution, variability, and sampling. Students are then asked to build on these ideas to make new statements about confidence and significance. However, student understanding of these earlier topics is often shallow and isolated, and many students complete their introductory statistics course without the ability to integrate and apply these ideas. Our experience as teachers of statistics suggests that the statistical inference calculations that students perform later in a course tend to become rote manipulation, with little if any conceptual understanding of the underlying process. This prevents students from being able to properly interpret research studies. (2004, pp. 295-296)

The logic of confidence intervals is based on the question, “how often would the confidence interval contain the population parameter if many, many samples were gathered and the corresponding

confidence intervals computed.” There is no way to determine if the confidence interval computed is actually one of the many intervals that would contain the population parameter. This concept is typically very frustrating to students who wish to know the probability that the population parameter is in their confidence interval.

Basic significance tests suppose that there is no effect and then look for contradictory evidence in the data. If evidence in the data suggests that the assumption of no effect was false, then it is supposed that an effect exists. delMas (2004) asserted that, in general, people have difficulty understanding arguments that involve double negation. Hence, it should come as no surprise that many students have difficulty with the logic of significance tests.

### *Difficulties with the Language of Statistics*

Many words in statistics appear in everyday English, but their colloquial and statistical meanings are different (Gal, 2004). A few such words include random, independent, significant, power, and normal (Aliaga et al., 2010). For example, the “normal” distribution is very important in statistics and has a very specific meaning. Yet as commonly used in the English language, “normal” means ordinary or typical. Granted, the normal distribution is very common and often found in nature, but it is not the only ordinary or typical distribution.

### *More Common Misconceptions and Difficulties*

- There are several measures of central tendency including the mean, median and mode. Some students commonly use the mean and median interchangeably. On the other hand, other students believe that the mean is always the best measure of central tendency regardless of the distribution of the data set (Gal, 2004; Garfield & Chance, 2000; Garfield & Ben-Zvi, 2007).
- Students tend to have difficulty designing and understanding representative samples (Garfield, 2003).
- Students tend to focus on individual data points instead of the distribution of a data set (Bakker & Gravemeijer, 2004; Cobb & McClain, 2004; Garfield & Ben-Zvi, 2007).

- Students are confused when the expected value for a random variable is not one of the possible values. This is especially true when the possible values for the random variable are whole numbers. delMas (2004) proposed that a similar phenomenon occurs when students find a value for the sample mean which was not included in the original data set.
- Students enter a classroom with prior beliefs, misconceptions, and incorrect intuition. Unfortunately, changing beliefs and intuition can be a lengthy and arduous process (Garfield, 1995; Garfield & Ben-Zvi, 2007; Tempelaar, 2004).

### **Proposed Methods for Teaching Statistics**

While many resources and suggestions to reform statistics education can be found, the prominent source is the Guidelines for Assessment and Instruction in Statistics Education: College Report (GAISE). This report was created under the GAISE project funded by the American Statistical Association. The GAISE college report gives six recommendations for teaching statistics which are quoted here (Aliaga et al., 2010, p. 4).

1. “Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding, rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing conceptual understanding and analyzing data.
6. Use assessments to improve and evaluate student learning.”

The GAISE recommendations and suggestions from other sources are discussed below.

### **Cultivate Statistical Literacy, Thinking, and Reasoning**

At the core of every introductory statistics course should be statistical thinking and reasoning (Garfield, 2002). It is unreasonable to expect students leaving their first course in statistics to remember all the methods for analyzing data they have learned. It is even more unreasonable

to expect students to be ready to analyze real world messy data sets after one statistics course. Nonetheless, a reasonable expectation is that students can develop their statistical thinking and reasoning abilities. Hence one of the major goals in a statistics course should be to cultivate the students' statistical literacy, thinking, and reasoning.

Statistical thinking, statistical reasoning, and statistical literacy are common terms in the literature. The line between statistical reasoning and statistical thinking is very thin (delMas, 2004; Garfield, 2002). The definitions vary slightly from researcher to researcher and often overlap (Garfield & Ben-Zvi, 2007).

### *Statistical Thinking*

Statistical thinking involves the ability to apply statistical knowledge and methods (delMas, 2004). Statistical thinking includes recognizing the need for data and understanding the role of data production. Statistical thinking is the ability to recognize the variability present in almost all aspects of life and to be able to explain and account for that variability (Aliaga et al., 2010; Garfield, 2002).

### *Statistical Reasoning*

Statistical reasoning describes how people deal with statistical ideas and how they analyze and interpret statistical information (Garfield, 2003). Statistical reasoning appears when students are asked to justify their answers, draw conclusions, and make inferences. Statistical reasoning involves understanding why and how results occur; why statistical models were chosen, how sampling distributions underlie statistical inference, and more (delMas, 2004).

Statistical reasoning is very different from the more familiar mathematical reasoning (Garfield, 2003). Mathematical reasoning is based on formal logic. Mathematical thought often involves mental processes with no context or link to the physical world (delMas, 2004), whereas the context of a problem is very important to statistical reasoning.

### *Statistical Literacy*

Statistical literacy can be viewed as having the necessary statistical skills, knowledge, and beliefs to properly assimilate and use the readily available data in an increasingly data driven



world (Gal, 2004). Statistical literacy requires that one understands the basic language of statistics (Aliaga et al., 2010). One possible definition is given by Gal, “statistical literacy is the ability to understand and critically evaluate statistical results that permeate daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions” (2004, p. 48). Not only must people be able to correctly analyze and interpret their own data or statistical results they find, but they also need to be able to communicate their results and interpretations (Gal, 2004).

Statistical literacy is rapidly becoming viewed as essential not only for students, but for all adults. Data can be found in all walks of life and chance is always present (Gal, 2004). Research has shown that the abilities to reason correctly and make good judgments based on data are imperative in sensible decisions and solving problems that appear in everyday life. In fact, statistical training has been shown to be necessary for people in many various careers (Gal, 2004).

Statistical literacy can serve individuals and their communities in many ways. It is needed if adults are to be fully aware of trends and phenomena of social and personal importance: crime rates, population growth, spread of diseases, industrial production, educational achievement, or employment trends. It can contribute to people’s ability to make choices when confronted with chance-based situations (e.g., buying lottery tickets or insurance policies, and comprehending medical advice). It can support informed participation in public debate or community action. The need for statistical literacy also arises in many workplaces.... (Gal, 2004, p. 49)

### **Use Real Data**

Statistics is based on numbers with context (Cobb & Moore, 1997). The GAISE report, and many other researchers, have theorized that instructors should use real data whenever possible (e.g. see Aliaga et al., 2010; Cobb & Moore, 1997; Moore, 1997). Although, on occasion, realistic data needs to be created to illustrate a concept. Students are more involved and interested when real, not just realistic, data is used. Real data will be even more effective if students are interested in the subject or are involved in the generation of the data (Aliaga et al., 2010).

### **More Focus on Conceptual Understanding**

The typical introductory statistics course covers too much material for students to learn and truly understand. If the majority of the time is spent on procedures and skills, students often

have a fragmented, shallow understanding of statistics. However, if an emphasis is placed on understanding the concepts, students will retain more after the class. In addition, if the conceptual understanding is present, specific procedures are easier to learn in the future (Aliaga et al., 2010; Moore, 1997).

Statistics instructors should focus more on concepts instead of the algorithms and formulas (Aliaga et al., 2010; Keeler & Steinhorst, 1995). As the GAISE report remarked, instructors should “view the primary goal as not to cover methods, but to discover concepts” (Aliaga et al., 2010, p. 17). When courses focus on procedural knowledge, students are often able to successfully solve the problem or give a formulaic interpretation, without ever understanding the ideas and concepts behind their work (Garfield, 1994). Bakker & Gravemeijer (2004) suggested that students should develop their intuition about a concept, perhaps using informal and possibly imprecise language, before the formal definitions and vocabulary is introduced.

### **Fewer Formulas and Computations**

In an introductory statistics class, mathematical theory should be downplayed, with more focus on the conceptual understanding of the statistical processes. In fact, too much mathematical theory in the beginning can deter students’ development of statistical ideas (Gal, 2004; Moore, 1997). Many of the formal derivations of the formulas and techniques should be saved for a later class.

Many formulas have no place in the curriculum anymore. Any formulas with no pedagogical value should be removed from the curriculum (Aliaga et al., 2010). With the availability and easy access to calculators and software, there is no need to teach students’ to be number crunchers (Garfield, 2002; Keeler & Steinhorst, 1995; Meletiou-Mavrotheris et al., 2007).

Some formulas should be included in the curriculum because the formulas give insight into the statistical concept (Aliaga et al., 2010; Gal, 2004). One such example is standard deviation. The formula for sample standard deviation,

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

helps students understand the process of finding a standard deviation. The formula sheds light on the concept of finding the “average” deviations from the mean when students realize they have to

subtract the mean from each data value. On the other hand, while the formula

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n y_i)^2}{n-1}}$$

is computationally easier, the formula does not lend itself to conceptual understanding of standard deviation. Since calculators can be used to compute standard deviation quickly and efficiently, there is no need for the computationally easy formula that has no pedagogical value (Aliaga et al., 2010; Moore, 1997).

### **Use More Active Learning**

Active learning is used as an umbrella term for any methods that involve the students more directly in their learning. Using active learning can be considered the most important reform method for teaching statistics. In fact, many of the the suggestions given here could be classified under active learning. Active learning involves students directly, enabling students to discover and understand concepts at a deeper level; students are also more engaged (Aliaga et al., 2010).

#### *Active Learning in the General Educational Literature*

Active learning requires students to take a more active, rather than passive, approach to learning. Students are required to be more than a sponge ready to receive information from the professor. Active learning requires more than listening to the professor; students must involved in the learning process through discussion, writing, reading, brainstorming, exploring, and more. The goal of an active learning class is to develop students' skills and understanding, not to transfer knowledge to the students. Active learning requires and facilitates higher order thinking such as evaluation and analysis (Bonwell, 1995).

Students need to be actively engaged in their learning. Students do not learn as much by listening to a teacher (Chickering & Gamson, 1987). Researchers have demonstrated that students have a short attention span for lectures. According to Thomas (1972) the amount of information retained by students in a lecture dramatically declines after 10 minutes.

Perhaps most importantly, even though an instructor might be able to cover more material through a lecture, students typically retain less of the learned material because they were not

engaged in the learning (Moore, 1997). As the ancient Greek Sophocles wrote, “One must learn by doing the thing, for though you think you know it-you have no certainty until you try.” An old proverb expresses the same sentiment, “I hear, I forget; I see, I remember; I do, I understand” (Rosenthal, 1995, p. 223). Studies have shown that active learning methods give better results than teacher-centered methods in the following areas: “application of concepts, problem solving, attitude, motivation, group membership and leadership skills...” (McKeachie, Pintrich, Lin and Smith as quoted in Rubin & Hebert, 1998, p. 26).

### *Active Learning in the Statistics Education Literature*

As Moore (1997) asserted, the views of teaching statistic have moved from information transfer to a constructivist approach. Instead of listening or reading, students should be active participants in their learning experience, furnishing their own ideas (Bakker & Gravemeijer, 2004; Garfield, 1995).

Scheaffer, Watkins, Gnanadesikan, and Witmer synthesized the need for active learning in catching the students’ attention.

Their fast-paced world of action movies, rapid-fire TV commercials, and video games does not prepare today’s students to sit and absorb a lecture, especially on a supposedly dull subject like statistics. To capture the interest of these students, teaching must move away from a lecture-and-listen approach toward innovative activities that engage students in the learning process. (1996, p. vii)

Specific suggestions for active learning in statistics include group work, projects, student presentations, laboratory activities, and group discussions (Garfield, 2002).

### **Use Technology**

An important teaching tool is technology (Aliaga et al., 2010; Garfield & Ben-Zvi, 2007). Teachers must understand how technology can be used effectively in the classroom. Technology can be used to automate calculations. The reduced time in number crunching allows students more time to understand concepts and practice problem solving (Meletiou-Mavrotheris et al., 2007). The ease of doing calculations allows time spent in discovery. Yet the benefits of using technology go far beyond saving time. Exploring data is possible in ways never before experienced. Students can

see the results of hundreds or thousands of experiments in seconds. Applets can be used to help students visualize concepts and abstract ideas (Garfield, 1995; Garfield & Ben-Zvi, 2007; Schneider, 2008). Students can easily see several different representations of the same data. Computer applets and simulation allow students to discover relationships and gain a better understanding of random phenomena (Aliaga et al., 2010; Garfield, 1995; Moore, 1997).

Simulations have been suggested as a means to illustrate abstract ideas. For instance, sampling distributions are considered one of the hardest concepts in introductory statistics. Simulations allow students to explore the properties of sampling distributions and develop their understanding of the process of sampling distributions (Chance et al., 2004).

Rossman posed several ideas on how to integrate technology into a statistics course. He used these methods in his workshop approach to teaching statistics.

First, technology is used to perform the calculations and present the visual displays necessary to analyze real datasets, which are often large and cumbersome. Freeing students from these computational chores also empowers the instructor to focus attention on the understanding of concepts and interpretation of results. Second, technology is used to conduct simulations, which allow students to visualize and explore the long-term behavior of sample statistics under repeated random sampling. Whereas these two uses of technology are fairly standard, the most distinctive use of technology within the workshop approach is to enable students to explore statistical phenomena. Students make predictions about a statistical property and then use the computer to investigate their predictions, revising their predictions and iterating the process as necessary. (1995, p. 1)

### *Cautions on Using Technology*

The majority of instructors would agree that using technology in the classroom is beneficial (Aliaga et al., 2010; Garfield, 2002). Nonetheless, as Meletiou-Mavrotheris et al. (2007) and delMas et al. (1999) acquiesced, little empirical research has actually been conducted to determine the effect of using technology on students' statistical conceptual understanding. Anecdotal evidence suggests that students are more engaged and interested while using technology. But the few experiments that have been conducted show only small gains in student learning (Chance et al., 2004).

In spite of the many recommendations to use technology, using technology does not guarantee statistical understanding. Indeed, the technology must be chosen or designed with care (Aliaga et al., 2010; delMas et al., 1999; Garfield & Ben-Zvi, 2007). Meletiou-Mavrotheris et al. (2007)

cautioned that software designed for educational use with conceptual understanding should be used instead of software created for professional data analysis. However, delMas et al. (1999) discovered that using even excellent software was not enough to ensure students understanding of statistical ideas.

### **Use Appropriate Assessments**

Formative and summative assessments need to be aligned to the course objectives in both content and learning levels (Aliaga et al., 2010). If instructors focus on conceptual understanding during the lectures, but only test on procedural skills, students will develop their procedural skills at the cost of conceptual understanding (Aliaga et al., 2010; Garfield, 1994). Because students are most often tested on computational skills, students can often give a correct solution without understanding the problem or solution (Garfield, 1994).

Many researchers endorse a variety of assessments including projects, presentations, and activities (Garfield, 1994; Garfield & Chance, 2000). However, the most common type of assessment is still exams, quizzes, and homework (Garfield, 2002).

### **Use Less Lecture**

As Cobb attested, “Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures do not work nearly as well as many of us would like to think” (1992, p. 9). Research shows that the traditional lecture typically is not very effective. Students often do not remember or understand all they are taught during a lecture (Garfield, 1995; Moore, 2005). The traditional lecture should no longer be the primary method of instruction, but it should be supplemented as much as possible with active learning methods (Aliaga et al., 2010).

### *When Is Lecture Appropriate?*

While Moore (1997) encouraged a variety of activities, he cautioned that lecture still has its place. Lectures are useful to help students see the big pictures, or learn the technical statistical language. Cangelosi (2003) indicated that direct instruction is appropriate when teaching conven-

tions and new procedural skills. Nonetheless, while lecture still needs to be used in the classroom, it should be used less (Moore, 1997).

### *Use Interactive Lectures*

One possible method to make lectures more effective is to use interactive lectures (Macdonald & Teed, 2012). The lecture can be broken up with short segments of active learning methods such as demonstrations, discussions, or activities. Methods should be used that allow all students to participate. Weltman and Whiteside (2010) also postulated that having short periods of active learning throughout the lecture can be helpful. They suggested stopping the lecture every 10-15 minutes for a short time. This break in the lecture can be used to have students work on a problem, discuss with their groups, or just digest the material.

## **Cooperative Learning**

Group work, or more formally, cooperative learning is defined as having students work together in small groups to facilitate their own and the other group members' learning (Giraud, 1997). It is generally acknowledged that group work can be an effective learning technique (e.g. see Garfield, 1993; Giraud, 1997; Meyers, 1997; Rosenthal, 1995). Cooperative learning can lead to increased student understanding, positive student attitudes, and greater self-efficacy (Garfield, 1995; Giraud, 1997; Keeler & Steinhorst, 1995; Towns, 1998).

Benefits of group work include: students teach each other, at least one student in a group will probably be able to understand a given topic, students discuss ideas, and students can give each other feedback (Giraud, 1997; Towns, 1998). As students answer each other's simpler questions, the instructor has more time to help students develop a deeper conceptual understanding (Keeler & Steinhorst, 1995). Outside the academic arena, a benefit to cooperative learning is that working in groups can help students develop the collaborative skills they need in the workforce (Towns, 1998).

As students learn in groups, they have the opportunity to learn new information and then share it, thus enabling them to partake in the active learning process. Students teaching students is a very effective method (Rubin & Hebert, 1998). Research shows that students at a higher level

can help students at the lower level. Indeed, cooperative groups form a natural setting for peer tutoring (Giraud, 1997).

Anecdotal evidence suggest that students in cooperative learning structures ask more questions than students in a lecture. Instructors are able to gauge the understanding of students in the class by observing comments and questions in the groups (Giraud, 1997). Hence, instructors in a collaborative learning environment should be in a better position to assess students' understanding than instructors in a traditional lecture course.

### *Suggestions for Implementing Cooperative Learning*

Methods for forming groups are varied. Garfield (1993) noted that allowing students to self select groups is a common method. For elementary and secondary students, teacher involvement in group formation is generally recommended, but self selection is more accepted for college students (Garfield, 1993; Keeler & Steinhorst, 1995). However, for activities involving critical thinking, Cuseo (1997) recommends that instructors carefully choose the groups. Cuseo explained that there is not a specific recommended group size, but most groups range from 3-6 students (Cuseo, 1997). Depending on the situation, homogenous or heterogeneous group formations might be desired. Students can be placed into groups based on gender, ethnicity, or statistical ability (Cuseo, 1997; Keeler & Steinhorst, 1995).

The instructor should walk among the students as they work in groups. This will help students stay on task. The instructor can also encourage critical thinking and evaluate the students' progress and comprehension; the instructor should ask guiding questions to spark students' ideas (Cuseo, 1997).

### **Use Physical and Concrete Activities**

Cobb and Moore (1997) speculated that statistics instructors should employ physical, concrete activities; this will help students develop their intuition and reasoning (delMas, 2004). Educators have theorized that having students explore ideas with physical activities before using computer simulations will stimulate conceptual understanding (Aliaga et al., 2010).



### **Encourage Students To Make Predictions**

Students should make predictions about outcomes before collecting and analyzing data. Having students make predictions will help them realize the importance of having statistics and also improve their understanding of likelihood. Making previous predictions also engages the students in the problem and students will be more interested in the results of the study (Aliaga et al., 2010; Garfield, 1995).

### **Ensure Students Participate in All Stages of the Experiment**

Burgess (2001) proposes five stages in the process of statistics: explore data, formulate questions, use appropriate methods to answer questions, analyze and interpret results, and communicate the results. Students learn statistics best when they observe and participate in all steps of data analysis, from formulating the question, to reporting their findings (Aliaga et al., 2010). Thus lessons and units should cover all aspects in the cycle.

### **Develop Positive Student Attitudes Toward Statistics**

A common goal for any course would be to cultivate positive student attitudes towards the field of study. Students who feel positively towards statistics will be more likely to continue their study of statistics. They will also be more likely to use statistics in their lives and careers. Hence, the need to cultivate positive student dispositions towards statistics is prominent. Student attitudes can be also used to determine a course's effectiveness. Positive attitudes are a desired outcome in and of themselves, but attitudes have also been shown to correlate with learning outcomes (Carlson & Winqvist, 2011).

Students' attitudes towards statistics do not always improve throughout a course (Carlson & Winqvist, 2011; Garfield & Ben-Zvi, 2007). Instructors can not assume that students will grow to like and appreciate statistics just by taking the course. However, some researchers have theorized that using reform methods will create more positive student attitudes towards statistics (Giraud, 1997; Keeler & Steinhorst, 1995; Rubin & Hebert, 1998).

Self-efficacy is a subset of attitudes and beliefs dealing specifically with students' perceptions of their own abilities (Hall & Vance, 2010). Naturally, most instructors desire students to have

high self-efficacy at the end of a course. Self-efficacy has also been shown to correlate with student understanding and performance in a course (Finney & Schraw, 2003; Hall & Vance, 2010).

### **Some Existing Models for Teaching with Reform Statistics**

#### *The PACE Model*

A common approach to teaching statistics that incorporates several reform methods is the PACE approach. The PACE model is centered on projects, activities, cooperative learning with technology, and exercises to reinforce concepts (Garfield, 2002).

#### *The Workbook Approach*

Students in courses using a workbook approach typically work individually or in groups to fill out a workbook during class. This enables the instructor to spend more time with the students answering individual questions. The workbooks might be supplemented by short lectures.

Rossman and Chance (2012) developed a curriculum called “Workshop Statistics” with an accompanying textbook. Workshop Statistics uses a constructivist approach for learning statistics. The textbook is comprised of activities intended to lead students to discover the statistics concepts and learn to apply the statistical techniques.

### **Summary**

The constructivist approach to learning argues that students bring their own knowledge to the table and build upon past knowledge and experiences. Knowledge is constructed by each individual around their own framework. Thus each student needs to be actively engaged in learning rather than acting as a passive vessel receiving the teacher’s knowledge. Students need the opportunity to develop their own meaning for a concept (Garfield, 1995; Moore, 2005).

As a consequence, it is not sufficient for a teacher to lecture at the whiteboard for sixty minutes. The students need to be involved with active learning. They need to be interacting with the teacher and other students at a deeper level than providing the next step for the current problem. Methods that are an alternative to lecturing provide less time to cover concepts; however, students will learn the material at a deeper level. This will translate to more knowledge that students can actually use outside of the classroom (Aliaga et al., 2010; Garfield, 1995; Moore, 1997).

Students can learn through open-ended problems, appropriate feedback, group work, discussion, application, problem solving, explaining their thinking, and listening to other students explain their thinking. Students should be involved in laboratory exercises and projects (Aliaga et al., 2010; Garfield, 1995; Moore, 1997, 2005; Vithal, 2002). There should be a variety of learning activities in a classroom. Direct instruction still has a place in learning, but it should be used less often.

The student learning objectives have also changed; there is less focus on procedural knowledge and more importance placed on critical thinking, problem solving, and higher-order thinking (Aliaga et al., 2010). If students are to use their skills in the workplace, they need to be able to apply their knowledge to new situations (Garfield, 1995; Moore, 1997). That requires that they learn statistics at a higher conceptual level. Typical statistics courses used to, and many still do, focus on probability-based inference and formulas. However, the research suggests that students should be focusing on understanding concepts, seeing the big picture, data exploration, drawing conclusions from statistical tests, and learning how to apply their skills in a useful manner (Aliaga et al., 2010; Moore, 1997).

### **Obstacles to Using the Reform Oriented Methods**

Many instructors are worried about implementing active learning techniques in the classroom. Obstacles to switching from lecture oriented to active learning courses include concerns about preparation time, content coverage, resistance from students and colleagues, and having less control.

While many ideas exist for various activities involving the research-based “new” techniques, such as active learning, or changing the curriculum of a course to focus on explanatory data analysis and gathering data, it is difficult to fit the new ideas into existing prescribed curriculums. Active learning takes more time to cover the material than lecture does; teachers are concerned that they will not be able to cover all the required material (Carlson & Winqvist, 2011). For instance, when Cobb, McClain, & Gravemeijer (2003) wished to conduct a 14-week course focusing on bivariate data with eighth graders, they were forced to have students volunteer to give up their activity hour in order to find time for their course; the current mathematics teacher could not fit the new course into her curriculum with end-of-year tests looming.

Garfield (2002) discovered that instructors using more active learning required more time to develop their materials and plan for class. Developing new activity-based materials is very time

extensive. Instructors commonly find that it is easier and faster to use previously prepared materials than to reform their curriculum or teaching methods. In addition, using new strategies might require more resources than lecturing (Bonwell, 1995).

Unfortunately, students are used to lectures and will resist new methods. As Klionsky attested, “I found that students do not respond well to the complete elimination of lectures” (1998, p. 337). In fact, implementing active learning in the classroom often leads to lower student evaluations (Garfield, 2002). Students complain that they should not have to “teach themselves.” As Keeley, Shemberg, Cowell, and Zinnbauer described a typical student’s resistance to critical thinking, “I want you (the expert) to give me answers to the questions; I want to know the right answer” (1995, p. 140). The lecture format is also easier for most students because they merely need to be efficient at taking notes and memorizing, not at understanding and analyzing.

Using the new reform methods can make an instructor feel out of control. The content becomes more student driven instead of teacher driven. Instructors will spend more time answering student questions which can become increasingly more diverse. In addition, if students are working individually or in groups, the instructor will have to answer the same question multiple times (Carlson & Winquist, 2011). A major obstacle to overcome with student-centered learning is that students work at different paces, and they all must be accommodated (Carlson & Winquist, 2011). Also, students often lack the skills to work effectively in groups (Rosenthal, 1995).

Students are used to teacher-centered activities and lectures. If student-centered methods are to be used, new norms must be established. Students must be trained to think critically, discuss with others, explain their thought processes, and give feedback. For example, Bakker and Gravemeijer (2004) found in their study of seventh graders that students were confused when asked to make predictions because doing such was contrary to the common norms and practices. Cobb (1999) realized in his study of second graders, that based on their past experience in school, the students believed that they were supposed to guess the answer that the teacher had in mind, instead of coming up with their own ideas.

Unfortunately, instructors seeking to reform their classroom are often met with resistance from colleges and departments. Reasons for this can include a set departmental curriculum, traditions, or the desire for more mathematical rigor (Garfield, 2002).

Finally, most instructors have not seen anything but lectures, and they do not know how to successfully implement active learning strategies. Bonwell formalized that “because lecture

classes have been the prevailing instructional approach seen most often by faculty when they were undergraduate and graduate students, many faculty have had limited personal experience with, and few role models for, active learning alternatives” (1995, p. 8). In addition, if the instructor feels that he is a good lecturer, there is no impetus for change.

### **Previous Research Studies**

The literature was searched extensively for previous studies involving teaching statistics classes with lectures versus active learning methods. The researcher found many papers with anecdotal evidence, surveys, or results from teaching a class with active learning techniques. Some of the results were promising, while other studies showed lower student outcomes with active learning techniques.

Nevertheless, there are few papers detailing actual comparative experiments (Garfield, 2002; Towns, 1998). Such studies were found for other academic fields (e.g. by Chappell & Killpatrick, 2007; Hake, 1998; Klionsky, 1998), but there are very few comparative experiments that have been conducted in statistics classes. The articles that were discovered are documented and discussed below.

### **Traditional Lecture Versus a Randomization Course**

Tintle, Vanderstoep, Holmes, Quisenberry, and Swanson (2011) developed an introductory statistics course based on randomization instead of asymptotic techniques. They changed the content to focus on randomization, and minimized the role of descriptive statistics and probability. In 2007, they taught 195 students using the traditional curriculum with lectures and computer based laboratory exercises. In 2009, they taught 202 students using their new randomization based curriculum. With this course, they attempted to use GAISE based methods of activities, group based discussion, and self-discovery; yet they also included lecture when necessary. They “transitioned from a more traditional mix of lecture and laboratory exercises, to a focus on tactile, self-discovery learning experiences supported by a mix of lecture and concept review” (Tintle et al., 2011, p. 6). Students were taught in multiple sections of approximately 25 students each.

Students in both courses were given the Comprehensive Assessment of Outcomes in a First Statistics course (CAOS) as a pretest and posttest. Students in the randomization course scored significantly better than the traditional course on five CAOS items. These items included questions about p-values and the interpretation of p-values. The traditional course showed significantly better results on only one out of the forty CAOS items: estimating standard deviation from a boxplot (Tintle et al., 2011). Yet when the overall scores for the entire CAOS test are considered, the students in the traditional course have slightly higher average scores (Tintle, Topliff, Vanderstoep, Holmes, & Swanson, 2012).

Overall, the randomization course showed promise to improve student comprehension of statistical inference, but there is not enough evidence to draw firm conclusions. Also, the authors changed not only their content, but the pedagogy. As a result, differences in students' learning can not be directly ascribed to either pedagogy or content (Tintle et al., 2011).

As a follow up study, Tintle et al. (2011) compared the retention of statistical knowledge of students in the traditional and randomization courses. Students were given the chance to take the CAOS test again four months after the course ended. Roughly 40% of students agreed to take the test. Students in the randomization course retained significantly more knowledge. While students in the traditional course achieved higher average total scores on the original posttest, students in the randomization course had higher average total scores four months later. Again, the difference in retention can not be attributed to either pedagogy or content; although the authors suspect that the new content had a large impact due to the fact that the increased retention mainly appeared on items that the new randomization curriculum concentrated upon, specifically inference (Tintle et al., 2012).

### **Traditional Lecture Versus a Workshop Approach**

The “Workshop Statistics” program is very prominent in statistics education. Beth Chance and Allan Rossman created a textbook comprised of student-centered activities designed to develop conceptual understanding. As Beth Chance and Allan Rossman are both leading experts in the field of statistics education, as well as the authors of “Workshop Statistics,” they were contacted to determine if they knew of existing comparative studies involving active learning methods and lecture. They both established that there was little research in this arena, and expressed the opinion that further research was worthwhile (Chance, 2012; Rossman, 2012).

As of 2012, Rossman had not conducted any experiments to see if his workshop statistics approach actually resulted in higher student outcomes (Rossman, 1995, 2012). In personal correspondence Rossman indicated that he only knew of one research study that compared Workshop Statistics to a control group (Rossman, 2012). John Zhang and Charles Bertness conducted the study and presented the results at the 5th International Conference on Teaching Statistics (ICOTS). The presentation was titled, “Two ways to teach an elementary statistics course: the workshop approach vs. the traditional approach. Which one is the winner?” (Zhang, Bertness, & Pan, 1998). Regrettably, no published paper was found. John Zhang was contacted to request a copy of the paper or the results; no response was received. The other two authors have left Indiana University and no contact information was available.

Nonetheless, the abstract of the presentation was available. The study involved teaching four sections of statistics using a workshop approach, although the abstract does not state if the Workshop Statistics materials were used. The next semester, four sections of the course were taught using a traditional lecture approach. The workshop approach involved class activities that were designed to lead students to explore and discover the statistical concepts. Minitab, a statistical software package, was used regularly. The activities were very student-centered with little teacher involvement. The study compared the students learning outcomes as well as attitudes towards the teacher methods. However, no indication was given in the abstract as to whether the learning outcomes and attitudes were better or worse for the workshop approach (Zhang et al., 1998).

### **Traditional Lecture Versus Constructive Hands on Activities**

As part of a PhD research study, Jane Brandsma (2000) taught 38 students in two sections of an introductory statistics course using a lecture approach. A year later, she taught 40 students in two sections of the same course using the same basic lecture approach, but she inserted “ten disjoint, constructive hands-on activities” (Brandsma, 2001, p. 9). Students in the class with activities scored higher on only one exam and students in the traditional class actually showed higher retention in interviews conducted eight weeks after the end of the course. Overall, students in the activities class performed better initially on a few select topics, but worse on most of the topics. Brandsma concluded that more research should be done, specifically with using activities throughout the entire semester instead of just ten disjoint activities.

### **Traditional Lecture Versus Computer Lab Activities**

Batanero et al. (2004b) conducted a research study on students' reasoning on the normal distribution. The experiment involved 117 students enrolled in an elective introductory statistics course. Two sections of the course were taught in a traditional lecture format. The teacher presented examples, guided student discussion, and asked guiding questions designed to lead students to develop their knowledge of the normal distribution. Two sections of the course were taught in a computer lab. Students worked in pairs on the computer on data analysis activities for the normal distribution.

Batanero et al. judged that the computer lab activities led to greater understanding of the graphical and abstract properties of the normal distribution. However, no numerical data or p-values comparing the sections were presented. Additionally, they only analyzed the results from a nine hour activity based unit inserted into the traditional lecture course.

### **Traditional Lecture Versus the PACE Method**

Meletiou-Mavrotheris et al. (2007) surveyed twenty-two student volunteers from various statistics courses within three months after the courses ended to determine the role of technology. Students were classified as having taken a non-technology class, taught primarily with lectures, or having been enrolled in a PACE course in which students conducted actual statistical investigations using technology. Six of the seven students from the PACE course reported that their feelings towards statistics had improved during the course; hence suggesting that the PACE method improves student attitudes. Nevertheless, almost all of the students from both teaching methods demonstrated a lack of conceptual understanding. Meletiou-Mavrotheris et al. (2007) speculated that the PACE course did not actually increase students' conceptual understanding; while students could use the technology to report answers, they did not see the purpose or statistical concepts underlying the technological activities. Finally, the conclusions from this study should be considered while remembering that it was a small sample of volunteers and the instruction and teaching materials were not controlled as part of the experiment.



### **Comparison of Traditional Lecture, Hybrid, and Active Learning Formats**

For this research study, over three hundred students were enrolled in seven business statistics courses classified into three formats. The lecture format consisted of the instructor lecturing for the entire period. The hybrid format incorporated short pauses for brainstorming, discussing, or solving a problem, every ten to fifteen minutes during the lecture. The active learning method involved the students working in small groups on packets while the instructor was available to answer questions (Weltman & Whiteside, 2010).

Assessments consisted of twenty multiple choice questions per topic. The cognitive domain of the questions ranged from simple knowledge to application. The study focused on three topics: the binomial distribution, the central limit theorem and standard error, and the p-value approach to hypothesis testing (Weltman & Whiteside, 2010).

Interestingly, the highest overall results came from the lecture method, then the hybrid, with the active learning approach showing the lowest overall results. Yet, when the data were analyzed in more depth, the format, the GPA, and the interaction between format and GPA were all significant. Students with high GPA achieved lower scores when in the hybrid and active learning courses. On the other hand, students with low GPA achieved the highest scores when in the active learning format. Weltman and Whiteside (2010) concluded that active learning is not effective for high GPA students, however the active learning and hybrid methods tend to bring the low GPA students to the same level as everyone else in the class.

### **Control Group Versus a Workbook Approach**

Carlson and Winqvist (2011) conducted a study comparing a workbook approach to a control group. They administered the Survey of Attitudes Toward Statistics (SATS-36) to 59 students on the first and last day of the semester. Four sections were taught using a workbook approach. Students could work individually or in pairs to complete their workbooks. The workbooks were designed to help students construct the statistical ideas. Every day was started with a 15-20 minute lecture, out of a 75 minute class. The content coverage was the same as for a traditional class.

The results were compared to data on 235 students in 20 sections of statistics courses gathered from the developer of SATS-36. Students in the workbook approach had significantly more positive attitudes towards statistics and higher self-efficacy. Interestingly, at the end of the course, students

in the workbook approach believed that statistics was more difficult than did the control group. However, it is important to note that there is no information about how the courses were taught for the comparison group.

After examining the results from the Weltman and Whiteside (2010) study, Carlson and Winqvist (2011) were concerned that the workbook approach would be detrimental to the high GPA students so they examined the results from the final exam. The high GPA students still performed better than the medium GPA students.

### **Traditional Lecture Versus Small Group Cooperative Learning**

A traditional lecture section was taught in the spring of 1990. Two cooperative learning sections, which were slightly different, were taught in the fall of 1990 and fall of 1991. All sections used the same textbook and similar exams. Students formed pairs themselves and then the instructor combined the pairs into groups of four to create as heterogeneous groups as possible (Keeler & Steinhorst, 1995).

The cooperative learning sections were still primarily lecture-based. During the first semester, the cooperative learning came when students would work in pairs on a question for a few minutes at 10-15 minute intervals throughout the lecture. Once a week, students were given fifteen minutes to work on their homework with their groups of four. Students were encouraged to study with their groups outside of class, but Keeler and Steinhorst admitted that such study groups were rare.

After reviewing the results from fall of 1990, the structure of the course was adapted. Students self selected pairs, but no groups of four were formed. Material was presented in a lecture for 25-30 minutes, and then students worked in pairs for 20-25 minutes (Keeler & Steinhorst, 1995).

Keeler and Steinhorst (1995) concluded that more students in the cooperative learning sections passed the course. Also, students in the cooperative learning sections had higher overall scores in the courses. No p-values or significance results were reported. Anecdotally, students in the cooperative learning sections appeared to be more engaged during class.

### **Traditional Lecture Versus Cooperative Learning Format**

This research study involved students in a psychology introductory applied statistics course. Forty-four students were enrolled in a cooperative learning section and fifty-one students were in

a traditional lecture section. Both sections were taught by the same instructor, during the same semester, using the same textbook, examples, and homework. Results for the two sections were compared for one of the midterm exams as well as the comprehensive final exam (Giraud, 1997).

Students were randomly assigned to groups of five members with varied statistical ability. No formal training on working in groups was presented, nor were formal group structures established. The majority of the time spent in the cooperative learning section was still lecture based. Out of the 2.5 hours of class time per week, groups were usually given thirty minutes to an hour to work together; occasionally, the entire seventy-five minute class period was allotted to group work (Giraud, 1997). Students in the lecture class worked on the homework outside of class. The cooperative learning section was given class time to work on the homework in groups.

A pretest measuring algebra skills was given at the beginning of the semester. Analysis of scores on the second exam show that students who performed poorly on the pretest scored higher in the cooperative learning section than in the lecture section. For the final exam, students scored significantly higher in the cooperative learning section; mathematical ability as measured by the pretest score had no affect (Giraud, 1997). The majority of students responded favorably when asked about their cooperative learning experience. In addition, Giraud felt that he was better able to gauge the students' understanding in the cooperative learning section.

## Summary

Tintle's study with the randomization course had mixed results. In addition, the two methods were used in classes taught two years apart. The randomization course changed not only the teaching method, but the content. Brandsma's research with the constructive hands on activities resulted in a negative effect from the activities, but the study only involved adding ten disjoint activities to a lecture course a year after the first course.

The experiment conducted by Batanero consisted of only nine hours of computer lab activities inserted into a lecture course; in addition, no significance tests were reported. Meletiou-Mavrotheris's research for the PACE method involved results from a voluntary survey and the students in both types of courses showed a lack of conceptual understanding. Carlson and Winquist reported that students in their workbook approach had more positive attitudes and higher self-efficacy than the control group; however, there is no information on the teaching methods used for the control group.

Weltman and Whiteside's research on the lecture, hybrid, and active learning classes showed that the highest results came from the traditional lecture classes, but that the active learning classes helped the lower GPA students achieve higher results. Keeler and Steinhurst conducted research on lecture versus small group cooperative learning over three semesters, however, the cooperative learning sections were still primarily lecture based. Giraud's research was very similar with cooperative learning used in a primarily lecture based course, but both classes were taught the same semester by the same teacher.

In conclusion, the eight studies that were uncovered during the literature search gave mixed results. In addition, desired elements of the research design were found in some of the research studies, but no study had all the desired elements. None of the studies taught one course with lectures and the other course primarily with active learning methods, in the same semester, by the same instructor, and without changing the content.

### **Possible Assessment Instruments**

The assessment instrument for this research study needed to be carefully chosen. Many classroom exams focus on procedural skills and simple knowledge; these exams would not work to assess conceptual understanding (delMas et al., 2007; Garfield & Chance, 2000). For this study, questions that focused on conceptual understanding were vital, although procedural questions were also desired. Therefore, a search was made for statistics assessments. Special attention was paid to assessments that had been validated. Several possible assessments were found.

#### **Statistical Reasoning Assessment (SRA)**

The Statistical Reasoning Assessment was developed and then validated under an NSF grant. Since many existing statistics tests primarily consisted of procedural questions, Konold and Garfield designed an assessment instrument to assess statistical reasoning. Garfield proposed that at the time of development, no instruments were available to evaluate the statistical understanding and reasoning of high school students (Garfield, 2003; Tempelaar, 2004).

The assessment is comprised of 20 multiple choice items and consequently is an objective instrument. Since the SRA is a conceptual based reasoning test, students are asked to choose the response

that is closest to their thinking, instead of being asked to choose the correct answer (Garfield, 2003). Many of the questions were designed to assess the students for common misconceptions and errors (Tempelaar, 2004).

The SRA contains items to diagnose students' reasoning about data, representations of data, statistical measures, uncertainty, samples, and association (Garfield, 2003). The assessment is designed to not only evaluate students' correct understanding, but to also assess the common mistakes students make while reasoning with statistics (Garfield, 2003).

The items on the SRA went through a process including content validation by an expert panel, revisions, administering the instrument as a pilot test, and more revisions (Garfield, 2003). In an initial study, the correlation was computed between the SRA score and various exam scores from an introductory statistics course. The results indicated that the assessment of students' statistical reasoning was unrelated to their learning outcomes in the course. The correlation between items was low suggesting that the reliability of the instrument was not very high (Sundre, 2003; Tempelaar, 2004).

The items on the SRA are interesting, but purely conceptual and quite limited in content; the majority of SRA items are focused on probability. As the goal of this research study was to assess students' ability on both conceptual and procedural questions, and we desired to cover more statistical concepts, the SRA would not work as the primary assessment device.

### **Quantitative Reasoning Quotient (QRQ)**

Sundre (2003) modified the SRA to create a new assessment. Some of the SRA items that presented multiple choices were split into new items that each had one idea and students had the choice to agree or disagree. This enabled the instrument to assess more misconceptions individually. Other than splitting the SRA items, the QRQ and SRA instruments contain the same items (Sundre, 2003; Tempelaar, 2004).

One reason to split the SRA items was the hope that it would help with the low internal consistency. Splitting the SRA items also meant that the instrument would gather more information about the misconceptions and correct reasoning. The final reason to split the SRA items was to enable the QRQ to be scored by a computer (Sundre, 2003).

### **Comprehensive Assessment of Outcomes in Statistics (CAOS)**

delMas et al. (2007) set out to design an reliable assessment instrument that would cover a broad range of concepts and be suitable for the general introductory statistics course. The Comprehensive Assessment of Outcomes in Statistics instrument was designed as part of the Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project.

The CAOS instrument went through three years of revision including feedback from experts and content validity assessments. Experts from the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) were consulted on the content relevance for the final version of CAOS 4. They agreed that CAOS measures the majority of important statistical concepts (delMas et al., 2007).

The CAOS 4 was administered as a posttest in 2005-2006 to 1470 students in thirty-five introductory statistics courses at post secondary institutions across the United States. The majority of courses required a high school algebra or college algebra background. The Cronbach's alpha coefficient is .82; this is an acceptable reliability level (delMas et al., 2007).

The CAOS was administered to only 763 students as both a pretest and posttest at the beginning and end of the semesters. The results were surprising. The average score increased significantly from 44.9% to 54% over the semester, that 9% difference correlates to a an improvement of 3.3 to 4 of the forty items. The most interesting fact is that after taking a statistics course, on average, students could only answer 54% of the questions correctly (delMas et al., 2007).

The CAOS test items are all conceptual in nature, as opposed to procedural (delMas, 2004; Tintle et al., 2012). Close analysis of the items showed that approximately 30% of the items were comprised of creating or interpreting graphs. As this is more emphasis than is usually placed on graphs in STAT 2000 at Utah State University, it was determined that not all of the CAOS items would be necessary. In addition, some procedural questions would need to be added. Nevertheless, the CAOS instrument served as a basis for several items on the posttest for this study.

### **ARTIST Topic Scales**

As part of the ARTIST project, the ARTIST Topic Scales were developed. The scales are short tests each designed to assess one of eleven topics. Each scale contains seven to fifteen multiple

choice items (delMas, 2006). According to an email correspondence with delMas, no systematic analysis or validation study has been conducted on the ARTIST Topic Scales (delMas, 2012).

### **Reasoning about P-values and Statistical Significance (RPASS )**

The RPASS-4 instrument was designed by Lane-Getaz (2007a) as part of her doctoral dissertation. The RPASS is comprised of twenty seven multiple choice questions dealing with p-values and statistical significance at the conceptual understanding level. The instrument had low reliability as measured by internal consistency (Lane-Getaz, 2007b).

### **Assessment of Inferential Reasoning in Statistics (AIRS)**

Park (2012) created the AIRS assessment instrument for her dissertation. The AIRS-3 instrument consists of thirty-four multiple choice conceptual questions. Ten of the questions were adapted from items in the SRA, ARTIST topic scales, CAOS, and RPASS assessments. The instrument entails informal and formal inference in relation to sampling distributions and hypothesis testing. The reliability coefficient, measured by internal consistency, was at an acceptable level of .81.

### **Statistics Concepts Inventory (SCI)**

The Statistics Concepts Inventory was designed for engineering students (delMas et al., 2007). As such, the SCI was not a prime candidate to assess students in a introductory algebra-based statistics course. The main topic for the SCI appeared to be probability. In addition, the authors declined to respond to a request to use the item, so the SCI was eliminated from the list of potential assessment instruments.

### **Picking an Assessment to Use**

Interesting, many items showed up on multiple assessments or were slightly adapted. For example, items 9-11 on AIRS-3 are copied from items 11-13 on CAOS-4. Most of the similar problems are exact duplicates, but some have slight word changes, while others have shortened or lengthened prompts, or split or combined items.

None of the existing assessments were judged to be a close fit for this research study. Most of the assessments had a limited selection of topics. In addition most of the questions were conceptual.

While conceptual questions were desired for this study, procedural questions needed to be included as well. Therefore, a new assessment instrument was needed for a posttest. Questions were copied or adapted from the existing assessments and other questions were developed as needed. The format and composition of the posttest are further discussed in the methods chapter, see page 49.



## METHODS

Two sections of STAT 2000: Statistical Methods were taught in the spring semester 2013. One section was taught with traditional lecture methods. The other section was taught using active learning methods. That section will henceforth be referred to as the activity class. The activity class was taught with as little lecture as possible; most days, less than five minutes was spent lecturing, if any time was used at all. Teaching methods for the activity class included group work, group activities, class activities, and interactions with applets. The emphasis in the activity class was on discovering and understanding concepts. The experiment was repeated during the fall semester 2013.

### Research Design

A design research or developmental research methodology was used for this study. The main characteristics of many design research studies are to create instructional materials, conduct experiments using the teaching materials, and then analyze the results and refine the instructional materials (Bakker & Gravemeijer, 2004; Cobb et al., 2003; Edelson, 2002). Perhaps the simplest definition of developmental research comes from Cobb, “it involves both instructional *development* and classroom-based *research*” (1999, p. 6). Edelson (2002) asserted that design research has the additional benefit of researchers being intimately involved in improving the practice of education. In addition, design research by nature commonly contributes improved instructional materials or ideas that educators can directly implement.

A key component of the design research methodology is that not all decisions are made before the study begins. The research study is instead thought of as an iterative process of designing materials, implementing the materials, analyzing the results, and then enhancing the materials (Cobb et al., 2003; Edelson, 2002). Indeed, even the conjectures and proposed student learning trajectories are revised throughout the semester (Cobb, 1999). This research study was designed in a way to allow the instructor to use informal and formal assessments to guide the development of the future activities throughout the semester.

## Participants and Setting

### Setting

At Utah State University (USU), there are four introductory level statistics courses taught by the Mathematics and Statistics Department as summarized in Table 1. STAT 1040 is tailored to students with a lower mathematical background; the course is typically taught without mathematical symbols and traditional formulas. While STAT 3000 requires a calculus level background, STAT 2000 and 2300 are algebra-based statistics courses. STAT 2000 students come from a variety of colleges, see Table 2. However, STAT 2300 is specifically for business students. Hence, it was decided that STAT 2000 was the most general introductory level statistics course, and the most appropriate setting for this research study.

The Utah State University course catalog describes the course objective for STAT 2000 as “Introduction to statistical concepts, graphical techniques, probability, distributions, estimation, one and two sample testing, chi-square tests, and simple linear regression, one-way ANOVA” (Utah State University, 2012).

For Spring 2013, the lecture class was taught MTWF 9:30-10:20 and the activity class was taught MTRF 12:00-12:50. During the Fall 2013 semester, the lecture class was taught MTWF 10:30-11:20 while the activity class was MTWF 8:30-9:20. The lecture sections were scheduled by the department and the activity sections were scheduled in the only available time slot for the mathematics computer lab.

Table 1: Statistics courses offered by Mathematics and Statistics Department.

Course	Prerequisite	Credit hours	Average number of students
STAT 1040 Introduction to Statistics	MATH 1010 - Intermediate Algebra	3	500
STAT 2000 Statistical Methods	MATH 1050 - College Algebra	4	80
STAT 2300 Business Statistics	MATH 1050 - College Algebra	4	300
STAT 3000 Statistics for Scientists	MATH 1100 - Calculus Techniques or MATH 1210 - Calculus I	3	200

Table 2: Colleges in which STAT 2000 students for the study were enrolled.

College	Percentage of Students
Agriculture	34.2
Education	28.1
Natural Resources	24.0
Science	5.5
Social Science	3.4
Undeclared	4.8

### Participants

The target research population was university students enrolled in an introductory, algebra-based, statistics course. Students should be enrolled in a small statistics course with between thirty and fifty students; this decision was based on the typical small class size, not on any power calculations. Neither gender or age was a consideration. Since participants must be university students, their ages were typically be between 18 and 24 years. However, there was no upper age limit. Minors were not invited to participate.

STAT 2000 is taught at USU every fall and spring semester. Approximately 70-80 students sign up for the course each semester. The majority of the students take the class as a major requirement.

Students who needed STAT 2000 for their major self selected one of the two sections, presumably based on time constraints. Participants were not told in advance that the two sections would be taught differently. The default small class size for the department is 40 students, although class sizes often fluctuate between 30-50 students due to demand and seating availability. The activity class was taught in a computer classroom with a capacity of 32 students. The lecture class had an enrollment cap of 40 students, but more students were allowed to join the course as needed to accommodate the demand for the course.

Every student in the course was invited to participate in the study. All students were given a letter of information, see Appendix A, on the first day of class. The letter of information described the study and what participation in the study would entail. All students were given the opportunity

to decline to participate. Students who declined still participated fully in the course, but data from those subjects was not used in the research.

Initially, the study was blind. Students were not told that the two classes were different, just that they were participating in a study. In addition, students were not told that one class was an activity class and that one class was lecture-based. However, students were not asked to keep anything about the class secret, and they might have talked to students from other sections after the semester started. After the first month, the instructor would occasionally bring up the fact that the two classes were being taught differently when it related to the day's lesson, such as how to set up an experiment with two samples. Even then, students were not given much detail on the difference in the classes.

### **Procedures**

The subjects participated in one of two sections of STAT 2000. Students were expected to do all the course work and assessments typical in a college course. The required coursework included participation during class, homework, surveys, exams, the pretest, and the posttest. Students who choose not to participate in the research study were still required to complete all the coursework, but their data were not analyzed for the research study.

### **Course Content**

Cobb and Moore (1997) pronounced that statistics and data analysis is comprised of three aspects: data generation, exploratory data analysis, and statistical inference. The traditional STAT 2000 course at Utah State University does not focus heavily on data generation and exploratory data analysis; typically only 2-3 hours are spent on such topics. As one of the goals of this study was to try to create and implement an activity-based course that fit into the traditional curriculum, only a few days were spent on data generation and exploratory data analysis. The majority of the time was spent on probability, sampling distributions, and statistical inference.

The units covered during the semester include:

- Gathering and Describing Data
- Probability

- Random Variables and Expected Value
- Binomial Distribution
- Normal Distribution
- Sampling Distributions
- Confidence Intervals
- Hypothesis Tests
- Multiple Sample Hypothesis Tests
- Chi Square Tests
- Linear Regression

### **Class Structure**

Both classes were developed to incorporate the best teaching methods that were suited for that class's structure. Great effort was undertaken to ensure that the instructional materials for both classes were as clear as possible. The emphasis was placed on conceptual understanding over procedural knowledge for both classes. Students in both classes were strongly encouraged to always interpret their answers in the appropriate context rather than simply giving a probability or p-value.

Students in both classes were required to use the same formulas; formulas that had no pedagogical value were removed from the curriculum. Extensive time spent number crunching was avoided; instead, students were often given computer output and asked to interpret it or use it in their further computations. Applets and simulations were used in each class to develop student understanding of statistical concepts.

As it would be too time intensive to find enough real data sets to use for every example, this was not a focus for the study. However, many examples for both classes came from data presented in "Introduction to the Practice of Statistics, 7th edition" or "Essentials of Business Statistics, 3rd edition". Both textbooks claim to use real data, so, the students in this research study were presumably exposed to real data. No explicit measures were taken to enhance students statistical

thinking, reasoning, and literacy beyond designing materials to enhance their understanding of statistical concepts.

Informal and formal assessment was used for each class. During the lecture class, the instructor was able to assess understanding by listening to the students comments and questions. More feedback was gained from the activity class though because the instructor was able to talk to more students during the time period. Students in both classes were given the option to write comments on the exit slips every day, giving the instructor one more source of informal assessment; the exit slips were pieces of paper that the students were required to fill out every day for attendance purposes. The formal assessments were aligned to the learning goals and a large portion of the questions focused on conceptual rather than procedural understanding.

#### *Format of the Lecture Class*

For the lecture class, direct instruction was used for the entire course. However, techniques were implemented to ensure that the lectures were as effective as possible. The students were given guided notes which were designed to be as clear as possible, with sufficient examples. Nonetheless, since the goal of the study was to compare *traditional* lecture and activity approaches to teaching statistics, the methods used in the lecture class were kept similar to the instructor's past experiences as a student and teacher.

Attempts were made to give thirty second pauses several times throughout the lecture to allow students to catch up and assimilate the new information (Ruhl, Hughes, & Schloss, 1987). Sufficient wait time was given after an instructor's question and after a student response to allow students to process the information (Rowe, 1986).

Students were occasionally given two to four minutes to work on a problem on their own, but only when the instructor felt that sufficient examples had already been covered. Students did not work together.

During the lecture, an attempt was made to stress the concepts behind the procedures. The intention was to have as much of a constructivist lecture as possible. A few applets and simulations were used during the lectures. The applets and simulations chosen for the lecture class had been used previously by the instructor in other lecture-based statistics courses.

*Format of the Activity Class*

The activity class consisted primarily of student-centered activities. Students were involved in a variety of activities including: class or group activities to gather data, teacher led activities to discover statistical concepts, and group activities to discover and implement statistical concepts. The primary method of instruction was group activities. The best description of the course would be a workbook approach that incorporated physical and technological activities.

The main focus of the activity class was to discover and understand the statistical concepts. Even though conceptual understanding was stressed in the lecture class, the activity class structure led to even more emphasis on conceptual understanding. Additionally, students were often asked to make predictions about a phenomenon or concept.

Direct instruction in the form of lecture was used only when deemed absolutely necessary. The lecture segments were kept to a minimum and comprised a minority of the class time. On average, less than 10% of the class time was spent in direct instruction. The typical reasons for direct instruction were to discuss the results from the students activity or to introduce a new algorithm with an example.

An activity workbook was designed by the instructor for the course using a constructivist approach. Students were typically given pages from the workbook to work on each day. The instructor would circulate among the students asking and answering questions. The students were encouraged to work together, asking questions and explaining their thought processes to their partners. To facilitate immediate feedback, keys were available to the students as they worked with their groups. The immediate feedback often resulted in cognitive dissonance and a learning opportunity.

Pairs of students were formed naturally when each table had two chairs. Students were allowed to self select their partners. Partnerships varied from day to day, but held fairly constant after the first month. Students were always allowed to work in groups of four if they desired, but usually chose to work in pairs. For some activities, they were required to work in groups of four. The groups of four were formed by having one pair turn around and work with the pair behind them.

The activity structure of the class allowed for more time to gather data. Hence the activity class actually spent time gathering data unlike the lecture class. This led to the opportunity for more real data sets, particularly data in which the students had a vested interest. The activity students also had the opportunity to use physical and concrete activities. They completed several

activities in which they gathered data using physical objects. The gathering of data meant that the activity students were engaged in more of the stages of experimentation.

In addition, the activity class was exposed to much more technology. The activity class used the same applets as the lecture class and more. Of particular potential benefit was the chance that the students had to explore the applets themselves, rather than watching the instructor use the applet.

### **Instructional Materials**

Guided notes were created for the lecture class. The notes were adapted from guided notes used by the instructor in a previous lecture-based course. However, the previous notes were for a different textbook, and extensive changes and additional writing was required. The guided notes were prepared and compiled in advance and submitted to the USU bookstore as a course reader for the spring semester of 2013. Supplementary material was given to the students as the need arose.

The workbook for the activity class took much more time to design. Hence, each activity for the spring semester was created only a day or two in advance. Each day the students were given a packet with a portion of the workbook. This did have the benefit of allowing the instructor to adapt the design of future materials based on the previous activities, specifically the timing, informal assessments, and comments by students. Inspiration for some of the activities was found in online sources, and the activities were adapted. However, the majority of the activity outlines were developed by the instructor. The data used in examples for the activity workbook came from textbooks, online activity ideas, and student data collection.

After teaching the course during the spring semester, both the activity and lecture materials were evaluated and refined. As we desired to compare the results between the spring and fall semesters, an effort was made to keep the materials similar between semesters. Nonetheless, any major errors or flaws were fixed. Some units were rearranged to facilitate a better flow of material. The material was reworded in any instances where the instructor noticed confusion during the spring semester. Many of the directions for the activity class were revised to enhance clarity. During the fall semester, the materials were again revised for future use.

For the spring semester, students in the lecture course bought the guided notes as a course reader from the USU bookstore. Students in the activity class bought “reference notes”, a short book of reference materials, definitions, and common formulas created for the course, from the



USU bookstore. During the fall semester, students in the lecture course and the activity course bought the guided notes or the activity book as appropriate from the bookstore.

Although many examples were adapted from textbooks, an attempt was made to avoid any infringement of copyright material. The avoidance of the copyright material was required to submit the material as a course reader through the USU bookstore, and it allows the future publication of the instructional materials.

### **Instructor**

Since teachers vary a great deal in personality, lecture styles, and more, it seemed ideal to compare courses taught by the same instructor; instructor behavior can influence student attitudes and learning outcomes (Carlson & Winqvist, 2011). In addition, it is wise to compare courses taught during the same semester. Ideally teachers learn and grow each time they teach a class, so students taught in later semesters might have higher results due to the teacher's new knowledge base. Hence, Jennifer Loveland taught both sections in the spring semester. The experiment was repeated in the fall semester with Jennifer Loveland teaching two sections again.

During the spring semester, Jennifer Loveland missed one day of classes due to a conference. Another statistics graduate student taught both sections of the class. The materials were prepared in advance. She lectured for the lecture class and the students worked in groups for the activity class. The topic for both classes was the sampling distribution for sample proportions.

### **Textbook**

The textbook "Introduction to the Practice of Statistics, 7th edition" by Moore, McCabe, and Craig (2012) was chosen for a variety of reasons. There are many textbooks that claim to be conceptually intensive and even a few textbooks that are focused on activities, two examples are "Workshop Statistics" (Rossman & Chance, 2012) and "Statistics in Action" (Watkins, Scheaffer, & Cobb, 2004). However, the "Introduction to the Practice of Statistics" textbook is traditionally used for STAT 2000 at Utah State University. For the spring semester, the textbooks had already been ordered through the bookstore, and as a graduate student, the instructor did not have the authority to change the textbook.

However, even if the textbook could have been changed, one of the goals of the research study was to revise the teaching methods while still fitting in with the existing curriculum; this included using the existing textbook. Yet, according to Garfield (2002), the “Introduction to the Practice of Statistics” is a “reform” textbook and one of the most common textbooks used in introductory statistics. Hence, using “Introduction to the Practice of Statistics” as the textbook would have been an acceptable choice under any circumstances.

### **Exams**

Exams were all multiple-choice to facilitate objective grading and eliminate any subconscious effort of the teacher to grade the activity exams differently. The first day of the course, students were given a pretest, which was a shortened version of the final exam. Due to the time constraint of fifty minute class periods, there were four midterm exams. The midterm exams were given to both classes on the same day to minimize potential cheating. The final exam was administered as a common final for both sections during a two-hour period. All exam questions were unique, not appearing in other exams or homework, with the exception of the final. Fifteen questions from the final were on the pretest. However, students had no access to the pretest or their answers after taking the pretest.

Extensive reviews were created for each test; the reviews were the same for each teaching method and each semester. Students were encouraged to work on the reviews outside of class and bring questions to class on the review days. For the lecture class, the students asked questions which the teacher answered in a lecture format. The students in the activity class worked in pairs or groups on the review and asked the instructor questions as the need arose.

### **Homework**

Homework assignments were the same each semester. Homework was assigned per unit and due the day after the unit was finished. This resulted in homework being due once or twice a week. Students were encouraged to work on their homework each day for a more even workload. In an attempt to minimize confounding factors in the study design, the same homework was used for both teaching methods. Consequently, the homework should affect students in both classes the same way.

The majority of the homework came from the textbook, though some problems were added by the instructor. Reasons to add additional problems consisted of giving students practice with multiple choice problems, changing the directions to be more consistent with the lecture and activities, to give more practice with conceptual problems, or to cover topics that were not presented in the textbook.

### **Confidentiality**

Because students had to take the exams and complete assignments as part of the course, anonymity was not maintained. Grades were assigned based on students' performance on the assessments. However, the surveys were partially anonymous. Students took the surveys in Canvas and their names were not looked at until after the end of the semester. The survey results were not linked to their assessment results until after the end of the semester. This way, the instructor would not associate the survey results with the individual students. Their grades for the surveys depended only on completion of the survey, not the actual survey results. Hopefully, this encouraged honesty among the students.

## **Data Sources and Instruments**

Data from the pretest, midterm exams, and posttest were analyzed to decide which teaching method resulted in higher student achievement on the learning goals, assuming there was a difference between the classes. All exams were multiple choice for improved objectivity. Surveys were analyzed to determine which teaching method resulted in more positive students attitudes towards statistics.

### **Pretest and Posttest**

The final exam was a comprehensive exam that concentrated on the main ideas from the semester that were judged to be most important. The objective was to create an exam that covered the statistical ideas that we judged were the most important to have a student actually retain at the end of the course.

The first step in creating the final exam was to create a list of objectives. A panel of statistics professors at Utah State University, comprised of Kady Schneider, John Stevens, and Jurgen Symanzik, reviewed the list of objectives and revisions were made. Next, questions were selected from the CAOS, RPASS, AIRS, SRA, and QRQ instruments; some items were copied exactly while others were adapted, see Tables 3 and 4. When a topic did not appear in any of the aforementioned assessments, an item was created. Items for procedural knowledge were also created; the created items are detailed in Table 5. After several revisions, the tentative questions were sent to the panel of instructors to be judged for relevance to objectives and the conceptual versus procedural learning levels. As the panel made suggestions, the posttest was revised again.

Table 3: Items on the posttest that were copied, or slightly adapted, from existing resources.

Item on Posttest	Existing Instrument	Item Number
1-2	CAOS	3, 5
7	CAOS	17
8	ARTIST Probability	4
9	SRA	2
10	ARTIST Probability	1
11	ARTIST Probability	6
20	CAOS	16
26-28	CAOS	25-27
29	CAOS	19
	ARTIST Tests of Significance	3

After the posttest was finalized, the pretest was designed. Originally, the intent was to give the same assessment for the pretest and posttest. However, final exams are allotted 110 minutes at Utah State University. Yet if the pretest was to be taken during class time, only 50 minutes would be available. In addition, since procedural questions were to be included in the final, the majority of the students would not be able to do those questions at the time of the pretest. It was decided that the pretest should be a shorter version of the final. Questions for the pretest were taken directly from the final. The pretest was used as an indicator of existing statistical ability and reasoning.

Table 4: Items on the posttest that were adapted from existing resources.

Item on posttest	Existing Instrument	Item Number
3-4	ARTIST Data Representation	10-11
5	CAOS	14
6	ARTIST Data Collection	4
12	CAOS	11-13
	AIRS	9-11
13-14	ARTIST Normal Distribution	3-4
17-19	CAOS	34-35
	ARTIST Sampling Variability	9-14
22	ARTIST Confidence Intervals	5
23-25	CAOS	28-31
31	RPASS	5
32	ARTIST Tests of Significance	9
33	CAOS	23-24
	AIRS	19-20
34	ARTIST Tests of Significance	1
35	ARTIST Tests of Significance	2
41	CAOS	22
42	ARTIST Quantitative Bivariate Data	10
	CAOS	21
43	CAOS	20

Table 5: Items on the posttest that were created for this study.

Item on posttest	Topic
15	Find the probability for a sample proportion.
16	Find the shape, mean, and standard deviation for the sampling distribution of the sample mean.
21	Compute a confidence interval for the population mean.
30	Meaning of the level of significance.
36-38	Conduct a hypothesis test for the mean.
39	When to reject the null hypothesis.
40	Find a p-value.

The final was comprised of 43 multiple choice items. A formula sheet was given to the students. The pretest contained 15 multiple choice items and a condensed formula sheet. The same pretest and posttest were used each semester. See Appendices B and C for the pretest and posttest.

### **Midterm Exams**

There were four midterm exams throughout the semester. Each exam contained 20 to 28 multiple choice questions. The questions were designed by the instructor or adapted from various textbook question banks. A variety of conceptual and procedural questions were included. Students were given a formula sheet and the necessary probability tables for each exam. In order to facilitate comparing results across the semesters, the same midterms were used each semester; however, the order of the answers was changed for some of the questions to help eliminate cheating.

### **Exit Slips**

The students' attendance needed to be tracked for research purposes. This was facilitated by having students turn in an exit slip with their name at the end of every class. The exit slips were also used to judge student comprehension and understanding. The exit slips are a variation of common feedback devices, such as minute papers and "muddiest point in the lecture" questions (Mosteller, 1989). Students were encouraged, but not required to respond to the questions on the exit slip.

The questions were the same each day. The questions were:

- What did you find particularly helpful about today's lesson?
- What do you still feel confused about?
- Suggestions to improve today's lesson:

The exit slips were reviewed after every class and the comments considered. If several students expressed a similar question, the question was addressed during the next class. On occasion, the instructor responded directly to a student's question via email. The comments were often used to revise the instructional materials.

### **Informal Assessment**

During the activity class, as the students worked, the instructor circulated among the students, answering and asking questions. This enabled the instructor to gauge the students' understanding of the topic. When it became obvious that there were common questions or misconceptions, these were addressed as a class discussion, usually during a recap.

### **Instructor Logs**

The instructor kept detailed logs throughout each semester. The timing of the activities and lectures was recorded. Interesting or pertinent student comments were noted. The instructor listed possible changes and improvements to be made to the instructional materials.

### **Online Surveys**

Participants were asked to complete online surveys for class credit; see Appendix D. The surveys were designed to take less than 10 minutes, but students typically took between one and three minutes to complete the surveys. The surveys were used to assess students attitudes towards statistics, their attitudes towards the teaching methods, and their self-efficacy. Some of the surveys were created by the instructor, while others were found in the literature.

The Attitudes Toward Statistics (ATS) survey developed by Wise (1985) was used for this research study. The ATS scale is available online at <http://bit.ly/Tl3ATj> . We also used a modified version of the Current Statistics Self-Efficacy (CSSE) survey created by Finney and Schraw (2003). The CSSE survey is available online at <https://www.stat.auckland.ac.nz/~iase/cblumberg/finney2.pdf>. We modified the CSSE to have a five point scale instead of six possible responses. The first seven items on the modified self-efficacy survey came from the CSSE. The remaining five items were added by the instructor to reflect the goals of the course. The Attitudes Toward Statistics and modified self-efficacy surveys were administered to students at the end each semester.

## Data Analysis Methods

The results from assessments and surveys were summarized with numerical and graphical methods. Data from the pretest, midterm assessments, and the posttest were analyzed to decide which teaching method resulted in higher student achievement toward the learning goals. Surveys were analyzed to determine which teaching method resulted in the higher improvement of students attitudes towards statistics.

The SAS software system was used for data analysis. Data analysis techniques included descriptive statistics and the nonparametric Wilcoxon rank-sum test. Multiple regressions were conducted to control for other variables. The research questions, hypotheses, and analysis techniques are discussed below.

This study entailed exploring the effects of the two teaching methods. As such, it was important to consider whether the students in the lecture classes were similar to the students in the activity classes. This question was explored for gender, field of study, previous GPA, and pretest scores. The gender ratio was analyzed using a chi-square test of homogeneity and the student field of study was examined with Fisher's exact test. The previous GPA and pretest scores for the two methods were explored using nonparametric Wilcoxon rank-sum tests.

### Research Question One

The primary research question asked, *"Which teaching method resulted in higher levels of student comprehension of statistical concepts and ability to apply statistical procedures?"* The current literature proposes that using active learning techniques will increase student learning outcomes. Therefore, the hypothesis was that the activity class would show higher overall results on all the exams.

First, the final exam scores were compared for the lecture and activity sections. The final exam was chosen for the first analysis because many of the exam questions were drawn from previously validated instruments; the final exam was also reviewed by a panel of experts and went through an extensive revision process. The data analysis was completed with the Wilcoxon rank-sum test.

Second, the total exam scores for all the tests were compared for the lecture and activity sections. However, it is important to consider that other variables could have an effect on the total



exam scores. Hence, multiple regression was conducted using the possible predictor variables. The possible predictor variables were teaching method, fall or spring semester, morning or afternoon, pretest score, attendance, homework scores, previous GPA, gender, and whether the student has taken a statistics class previously. Stepwise variable selection was used to choose a reasonable regression model.

### **Research Question Two**

The second research question asked, “*Which teaching method resulted in higher student outcomes for conceptual level questions?*” The hypothesis was that the activity class would show higher results on the conceptual level questions on the exams. The total exam scores for all the conceptual level questions was compared between the two teaching methods using the Wilcoxon rank-sum test.

### **Research Question Three**

The third research question asked, “*Which teaching method resulted in higher student outcomes for procedural level questions?*” The hypothesis was that the activity class would show higher results on the procedural level questions on the exams. A reasonable hypothesis could be that the lecture class would outperform the activity class on the procedural questions since lectures typically focus more on procedures. However, we hypothesized that since the activity class students had more practice doing the procedures themselves instead of watching the teacher, that the activity class would still outperform the lecture class. The total exam scores for all the conceptual level questions were compared between the two teaching methods using the Wilcoxon rank-sum test.

### **Research Question Four**

This research question was, “*What was learned about implementing an activity-based statistics course?*” This question was answered in two parts. The teacher’s logs and observations were examined to determine what the researcher learned as the instructor. The students’ responses to exit slips and surveys were analyzed to ascertain the students’ responses to an activity-based statistics course.

## Informal Research Questions

### *Research Question Five*

This question was “*Which teaching method produced higher student learning outcomes for specified statistical content?*” The exam scores were broken down into topics and the topics were analyzed for differences between the two teaching methods using the Wilcoxon rank-sum test. Data on each student’s response to every exam question were collected. If there were significant differences between classes based on the individual questions, the questions could be analyzed to look for patterns. Because the data for individual questions was binary for each student, the only options were right or wrong, Fisher’s exact test was used. For this question, we did not hypothesize which teaching method would produce higher results for each topic or question; hence, two-sided tests were used.

### *Research Question Six*

This research question was, “*Were there differences in student attitudes towards statistics or the statistics course, based on which teaching method was used?*” Data on each student’s response to the end of semester surveys, attitude surveys, and self-efficacy surveys were analyzed. As suggested by Boone and Boone (2012), the questions on the Attitudes Toward Statistics and Current Self-Efficacy surveys were Likert type items and as such, the results were ordinal data. Due to the ordinal nature of the data, parametric tests such as the t test were not appropriate. The Wilcoxon rank-sum test was used instead. Again, no hypotheses about which teaching method would result in more positive attitudes or self-efficacy was made. So two-sided tests were used.

## Significance Levels and Adjusting for Multiplicity

Multiple significance tests were conducted for this study. Conducting multiple tests necessitates controlling the familywise error rate. The familywise error rate is the probability of rejecting at least one true null hypothesis when multiple hypothesis tests are conducted (Abdi, 2010). For example, if 100 tests are performed at the  $\alpha = .05$  level, then even if the null hypothesis was true, we would expect to see approximately five significant results, or false positives which are more formally known as Type I errors, just by chance.

The significance tests conducted for this study are divided into families as described henceforth. A family of tests is defined as a set of tests conducted jointly on a set of data (Abdi, 2010; Lehmann & Romano, 2012). Raftery, Abelly, and Braseltony (2002, p. 261) defines a family as “a collection of inferences for which it is meaningful to take into account some overall measure of errors.” Which tests should be considered a family can be an ambiguous choice and depends on the situation (Lehmann & Romano, 2012).

Each family of tests was allotted a total significance level of .05. Within the families, the Bonferroni-Holm method was used to control the familywise error rate. The Bonferroni-Holm method, which analyzes p-values sequentially from smallest to largest, is preferable to the Bonferroni method because it does not assume independence of the hypotheses (Abdi, 2010). The Bonferroni-Holm method is also more powerful than the Bonferroni method. For some families, the Bonferroni method was used first to divide the significance between subfamilies and then the Bonferroni-Holm method was used within the subfamilies. The tests in each family and the methods used for controlling the familywise error rate are detailed below.

*Family 1: Compare Exam Scores*

Within this family, the Bonferroni-Holm method was used to control for the familywise error rate.

1. Compare final exam points by method.
2. Compare total points by method.
3. Compare conceptual points by method.
4. Compare procedural points by method.

*Family 2: Multiple Regression*

1. Conduct a multiple regression on total points with predictors such as pretest points, gender, method, semester, etc.

*Family 3: Compare Topics and Individual Questions*

1. Compare the exam scores for each topic by method.

- Use  $\alpha = .025$  and Bonferroni-Holm.
2. Compare the proportion of students who answered the question correctly for the 143 individual exam items by method.
    - Use  $\alpha = .025$  and Bonferroni-Holm.

*Family 4: End of Class Survey*

Within this family, the Bonferroni-Holm method was used to control for the familywise error rate.

1. Compare the students' answers to each of the following questions by method.
  - (a) How useful do you think statistics will be to you in the future, either in your career or your personal life?
  - (b) How do you feel about statistics?
  - (c) How much did you enjoy this statistics class?
  - (d) How engaged were you during class?
  - (e) How motivated were you to come to class?

*Family 5: Attitudes Toward Statistics*

1. Compare the total score for all the attitude questions by method.
  - (a) Use  $\alpha = .025$ .
2. Compare the scores for the 29 individual attitude questions by method.
  - (a) Use  $\alpha = .025$  and Bonferroni-Holm.

*Family 6: Self-Efficacy*

1. Compare the total score for all the self-efficacy questions by method.
  - (a) Use  $\alpha = .025$ .

2. Compare the scores for the 12 individual self-efficacy questions by method.

(a) Use  $\alpha = .025$  and Bonferroni-Holm.

#### *Family 7: Learning and Enjoyment of Activities*

Within this family, the Bonferroni-Holm method was used to control for the familywise error rate.

1. Test if feelings towards activities improved significantly throughout semester, with regard to *learning*?
2. Test if feelings towards activities improved significantly throughout the semester, with regard to *enjoyment*?

### **Sample Size Versus Experimental Unit**

An experimental unit is considered to be the smallest unit that can have a treatment applied to it (Federer & King, 2007; Lazic, 2010; Stoker, King, & Foster, 1981). The sample size, usually reported as  $n$ , is commonly defined as the number of experimental units. However, the definition of sample size can be ambiguous. Many times, researchers consider the number of observations or the number of subjects as the sample size. The debate of the true sample size often arises when data has a hierarchical structure (Lazic, 2010).

In this study, we anticipated 140-150 participants. So the sample size could be the number of students, or approximately 140. However, the teaching method is the treatment. Because the teaching method is used for an entire class at a time, and not applied to the students individually, the four classes are the experimental units. Hence the sample size could be considered four.

Because the treatments were applied to classes and not individual students, the data has a hierarchical structure. The concern is that students within a class might not be independent observations, and that there might be a clustering effect with the classes (Lazic, 2010; Stoker et al., 1981). This concern can be addressed by using the average value for each experimental unit instead of the value from each participant; unfortunately, this method loses much of the data (Festing, 2012). In addition, it is impractical for situations with small numbers of experimental units. Another method is to use hierarchical models to analyze the data (Stoker et al., 1981).

Hierarchical structures often arise in educational research. Many times, it is impractical to apply a teaching method or treatment to individual students. Since teaching methods are usually applied to an entire class, we could expect to see many studies with hierarchical structure and small sample sizes. Yet, if the sample size is considered to be the number of experimental units, the time and cost to conduct a study with a sample size of, for example, forty, can be prohibitive.

In an attempt to determine what the common practice in education research entails, a search was conducted for any research studies that compared two or more teaching methods by applying different methods to two or more sections of a course. Studies from any subject, not just statistics, were considered. Search terms such as “compare teaching methods” were used. The articles were analyzed to determine if the sample size was considered to be the number of students or the number of classes, and if the experimenters adjusted for the clustering effect. The search continued until twenty articles were found that had enough detail to determine their statistical methods. Nineteen of the twenty articles used the number of participants as the sample size and did not account for any hierarchical structure (e.g. Tintle et al., 2011; Walsh & Hewson, 2012; Stanton, 1997; Thomas, 2008; Chappell & Killpatrick, 2007; Carlson & Winqvist, 2011; Giraud, 1997). The articles described studies with 2 to 19 classes. Only one article acknowledged the fact that students within a class might not be independent and that there could be a class effect (Reder, Cummings, & Quan, 2006). The researchers used a hierarchical general linear mixed regression model to account for the clustering effect. However, their study included 36 classes; it is less common to see studies with so many classes.

Based on the literature search, the sample size for this research study was determined to be the number of participants. Although we acknowledge that students within the classes might not have been independent, we did not account for the hierarchical structure of the data. The decision is largely due to the fact that there were only four classes.

Because only four sections were taught, resulting in only four experimental units, we realize that this study is not large enough to confirm the effectiveness of either teaching method for future sections of students. Therefore, this study should be considered more of a pilot study (Van Teijlingen & Hundley, 2001) than a confirmatory study.

## RESULTS

In this chapter, the results are detailed. The actual implementation of the teaching methods is discussed. Examples of instructional methods and activities are examined. The development of the instructional material is also discussed. Next, exam scores are compared by teaching method. Then student attitudes are compared by teaching method. Finally, the teacher observations and student comments are analyzed.

### Implementation of Teaching Methods and Development of Instructional Material

#### Time Allocation

##### *Activity Classes*

One of the goals of this study was to teach the activity classes with as little lecture as possible. This was largely accomplished by having students spend much of their time working in groups. There were also days when the entire class was involved in an activity to gather data or discover concepts. However, occasionally, the instructor could think of no way to introduce a topic or new procedure without lecturing; there are some things that students can not discover for themselves, usually set conventions or traditions. The time that was spent in teacher-centered methods during the activity classes can be broken into four categories.

- **Lecture:** This category consists of the times when the instructor did all of the talking. Students could ask questions, but the instructor did not rely on students to give answers or provide the next step. The lecture category was not interactive. Lecture was mainly used when introducing definitions and names set by convention. Some examples include introducing the definition of Type I and II errors, defining  $R^2$ , and outlining the requirements for a distribution to be called Binomial. Appendix E has a list of topics that lecture was used to introduce.
- **Recap:** As a class, the instructor reviewed what the students had already learned, or addressed issues and questions that arose during the activity. The instructor also summarized the results of an activity during a recap. Sometimes the recap was fairly interactive as the students described what they learned, but the recap could also be non-interactive when

the teacher cleared up questions and addressed misconceptions. The recap usually occurred during the last 2-4 minutes of class, but was not done every day. Examples of recap include reviewing all the probability rules, discussing the difference between the assumption  $n \geq 30$  and the assumption that both  $np \geq 10$  and  $n(1-p) \geq 10$ , and discussing the correct interpretation of  $R^2$ .

- **Introduce Algorithms & Work Examples:** For this category, the teacher would introduce new procedures and formulas, as well as show students how to do a new type of problem. This was more interactive than the “lecture” category as students were asked to provide the next step, or to justify the next step. However, since the material was new, the instructor was still the focus and provided many of the ideas.
- **Teacher Led Discovery:** The teacher led the class through activities designed to discover new concepts. These were very interactive sessions in which the students were supposed to supply the answers and the instructor’s role was to be a facilitator or guide. This category also includes activities in which the entire class was involved in gathering data or other hands-on activities. Examples include using an excel simulation to motivate expected value, having a root beer taste test to discover the properties of the Binomial distribution, and discovering the formulas and properties of confidence intervals.

In my opinion, only the lecture, introduce algorithms and work examples, and recap categories can be counted as traditional lecture. So 9.7% and 11.1% of the available teaching time was spent in traditional lecture during the spring and fall semesters respectively. Only 20.9-22.1% of the time was spent in any teacher-centered activity. Table 6 describes the percentage of the class time that was spent in each category of teacher-centered activities throughout the semester. Tables 24 and 25 in Appendix E itemize the time allocation in more detail. Appendix E also contains a list of what days involved teacher-centered activities. If the class was not involved in a teacher-centered activity, students were working in groups; this was approximately 80% of the activity class time.

### *Lecture Classes*

The default teaching method for the lecture classes was for the instructor to lecture for the entire period. Occasionally, when the instructor felt that enough examples had been covered by lecturing, the students were given a few minutes to try a problem on their own. They did not



Table 6: Percentage of class time spent in teacher-centered activities.

Category	Spring Semester Percentage	Fall Semester Percentage
Lecture	2.6	2.5
Recap	1.6	3.3
Introduce Algorithms & Work Examples	5.5	5.4
Teacher Led Discovery	11.2	11.0
Total Time in Traditional Lecture	9.7	11.1
Total Time in Teacher-Centered Activities	20.9	22.1

work with any other students. Afterward, the instructor would review the results and students could ask any questions that arose when they tried the questions. Table 26 in Appendix E itemizes the time that students were given to work on their own for both semesters. During the spring semester, students spent a total of 31 minutes throughout the entire semester working on their own. Students in the fall semester spent 63 minutes working on their own.

The following is an example of a typical lecture session. The topic was confidence intervals. The teacher introduced the formula for the  $z$  confidence interval for a population mean. The formula was given to the students, they did not discover the formula. The teacher led the students through three examples of finding and interpreting confidence intervals. The teacher stopped every few minutes to ask for questions. The teacher often waited for students to supply the next step. The students were led through two examples which demonstrates the result of changing the sample size and confidence level. The findings were discussed.

### Examples of Activities

The following are a few summarized examples of activities that students completed in the activity course. The actual activities for the Normal Distribution Unit and the Sampling Distribution Unit are in Appendices F and G respectively. Students in the activity class spent time working in groups to discover concepts, gather data, and practice procedures. The entire class worked together, brainstorming and discussing to discover concepts and gather data. Occasionally, the teacher would lecture to introduce a topic, or show students an example of a new procedure.

### *Sampling Variability*

On the first teaching day, a jar of M&Ms was provided. Students were told that they needed to estimate the percentage of red M&Ms in the jar. Because it would take too long to count all the M&Ms in the jar, students were going to take samples of red M&Ms. This led naturally to a discussion about populations and samples and the need for samples. Each pair in the class took their own sample and counted the percentage of red M&Ms. They then recorded their results on the board for the class to see. A histogram of the sample percentages was made. Students were then led in a discussion about how different students found different sample percentages, thus introducing the idea of sampling variability.

The true percentage of red M&Ms in the entire jar was revealed. Students were led to the realization that even though the pairs estimates vary, most of the estimates were close to the true percentage for the entire jar. Students were told at the beginning that whoever had the closest estimate could take the remained or the M&Ms home. After revealing the true percentage of red M&Ms, students pointed out that it was not fair. The students who were the closest did not do anything better, they were just lucky. This led to a discussion on how we never know how close our sample is to the true parameter, but we hope it is close. Because this activity took place on the first day, the ideas and vocabulary were not necessarily technical. The activity served as an interesting initiation to statistics while introducing students to the very important idea of sampling variability.

### *Normal Distribution*

Students worked in groups for this entire unit. The first day, students worked in pairs because they would be using computers. Students used real data from approximately normal distributions to count the percentage of observations within one, two, and three standard deviations of the mean. This led to their discovery of the empirical rule. Then students read a half page introduction about the normal distribution. Next, students used an applet to discover how the mean and standard deviation affect the shape of the normal distribution. After the normal distribution has been introduced, the empirical rule was formally stated and students practiced applying it. Finally, students used an applet to find probabilities for specific  $z$  values and vice versa for the standard

normal distribution. Using an applet first to find the probabilities helped develop students' intuition about using the normal curve and later helped them use the standard normal table.

The second day, students were encouraged to work in groups of four to facilitate brainstorming. Students started by reading a half page introduction for the standard normal distribution. Then, students used the standard normal table to find probabilities and  $z$  values. Even though they had not been instructed on how to use the standard normal table, between brainstorming in groups, asking the instructor questions, and looking at the key, students were able to figure out how to use the table. Students were encouraged to draw pictures and use logic instead of memorizing procedures. Finally, students discovered that the  $z$ -score is equal to the number of standard deviation the value is from the mean. Students then practiced standardizing values from various distributions.

The third day, students were allowed to choose between groups of four and working in pairs. Students typically chose to work in pairs when given the choice. No instruction was given by the teacher, but the students' worksheets included steps on how to use the table to find probabilities for any normal distribution. Students were typically able to figure out the procedure with little difficulty. Next, students practiced finding values from any normal distributions based on the probability. Finally, students were given the normal quantile plots activity. The instructor first gave directions for about thirty seconds. In this activity, students sorted histograms with attached normal quantile plots into three categories: normal, kind of normal, and definitely not normal. Then students discovered that a normal quantile plot looks like a straight line for a normal distribution. The instructor spent the last two minutes in class going over the results for the normal quantile plot activity.

### *Expected Value*

Students were led in a discussion with the entire class about average values. First, students were given a table with ratings and the percentage of people who gave each rating to a song. Students were asked how they could find the average of the 100 ratings using the percentages. This led to a brainstorming session and the discovery of the formula for expected value.

Next, students were motivated to use expected value with the example of a game. A wheel is presented and students were told that they could win or lose money based on which area the spinner lands on. The percentages of the area and the amount they gain was given. Students

were first asked their feelings about whether they thought they should play the game, and if the carnival would win or lose money if lots of people play their game. This was just based on their gut feeling. Then an excel simulation was used to spin the wheel many times. Students realized that they could not predict whether the next spin would result in winning or losing money, but that they could start to make long run predictions about the average. The formal definition and formula for expected value were then introduced.

Finally, students worked in groups on the remainder of the packet. This consisted of finding the expected value for several hypothetical carnival games. Students were told that they were helping to set up the carnival and their job was to determine if the carnival would win or lose money in the long run for each proposed game. Students were also asked to find the expected net gain for the carnival for each game for 2000 people, and the total net gain for all the games combined. This introduced students to the rules of expected value at an intuitive level. Students could answer these questions without any coaching or realizing that they were using formal rules. As time permitted, students worked on more expected value problems involving blackjack, insurance, lotteries, and more.

### *Binomial Distribution*

Students were told that the class was going to participate in a taste test for three varieties of root beer. Since most students claimed they would not buy the generic brand of root beer, they were told that the goal of the taste test was to identify the generic root beer. At the beginning of class, ten volunteers tasted each of the three varieties of root beer. The root beer was poured into numbered cups before students entered the room to make the experiment blind. Their results were recorded and saved for later in the discussion.

As a class, the teacher led the students in a discussion about the probability distribution of how many students could blindly guess correctly if one student took the test, and if three students took the test. As students filled out the probability distribution for three students, the instructor helped students see patterns in the probabilities. This was used later to develop the binomial probability formula. As an introduction to future hypothesis testing, students were asked, as a gut feeling, “so if all three students are blindly guessing, how many students would have to guess correctly for you to believe that they can tell the difference”.

Next, the instructor mentioned that having ten students take the test would be more accurate. Students were asked how many possible outcomes there were in the sample space for ten students and if they really wanted to fill out a table for all the possible outcomes and probabilities. Since there are 1,024 possible outcomes, the students naturally did not want to fill out the table for all the possible outcomes. The instructor reminded the students of the patterns they saw for three students and led the students in a brainstorming session to develop the formula for binomial probabilities. Since students had not learned about combinations, the number of ways to get  $x$  correct guesses from ten students was supplied for the students, and they could complete the rest of the probability table.

The instructor revealed how many students correctly identified the generic brand. The students calculated the probability of having that many students or more guess correctly. To further build the reasoning needed for hypothesis tests, the students were asked if they believe that the taste testers were blindly guessing based on the calculated probability.

Finally, the instructor introduced the formal definition of sample counts and the binomial distribution, and went through one example of finding binomial probabilities. The previously described activity took one 50 minute period. The next day, students worked in groups to identify situations that could be described with the binomial setting, find probabilities with the formula, find probabilities with the table, and find binomial means and standard deviations.

### *Confidence Intervals*

Students were first led in a class discussion about the motivation and theory behind confidence intervals. The next part of the activity was designed to introduce the meaning of the actual confidence level. Students were told that they needed to estimate the percentage of the earth that is land. Working in groups, they tossed a globe, keeping track of the proportion of tosses for which their finger landed on land. At this point, they had not learned about formulas for confidence intervals or the critical values. Instead, they were given a formula with the critical value and told to plug in their value for the sample proportion and write down the resulting confidence interval. After each group did this twice, they wrote their results on the board for the class to see. Similarly, they found confidence intervals for the mean heights and weights of NBA basketball players in the 2001-2002 season, and the average sum if they rolled two dice five times. These examples were chosen because the true population values were known. After students recorded all their results,

the class was led in a discussion. Students realized that they all found different confidence intervals, but the intervals seemed to be centered around the true population values. However, most, but not all of the confidence intervals actually contained the true population value. Students were shown simulated results and discovered that the percentage of intervals that actually contain the true population value is approximately equal to the confidence level.

The next day, the students were led in a class brainstorming session in which they discovered the three confidence interval formulas: z confidence interval for a mean, t confidence interval for a mean, and z confidence interval for a proportion. The t distribution could be introduced at this point or earlier in the semester. The instructor then worked through an example for each of the formulas.

The next two days, students worked in groups. They practiced finding and interpreting confidence intervals. The workbook led them to discover what happens to the margin of error if they change the confidence level or sample size. Students also practiced finding the necessary sample size for a desired margin of error, but only after they discovered the formula they needed.

### **Technology Used in Classes**

Six applets were used for the lecture class. The instructor led the students through the applets. The applets were chosen because they had been used by the instructor in previous lecture based classes. A detailed list of the applets is available in Appendix H.

Eighteen applets and five excel simulations were utilized during the activity class. The activity class also used MegaStat, a plug in for excel, to run a simple ANOVA analysis. Students usually explored the applets in pairs. There were four applets for which the teacher led the exploration of the applet as part of a class discussion. During the fall semester, the classroom with the computers was not available on Tuesdays. There were several times when the class met in a campus computer lab on Tuesdays to access the applets. But there were two applets that required a short enough amount of time, that the instructor led the exploration of the applets rather than meeting in the computer lab for the entire period.

The instructor created two applets, categorical versus quantitative data and exploring the shape of the normal distribution, for the course. Two excel simulations were created for the ANOVA unit, and three excel simulations were adapted from other sources for other units.

### *Technological Issues*

While using technology has great benefits, such as allowing students to visualize data and see the results from many experiments, complications occasionally arose. The classroom had a closet full of laptops for the students' use. Five to ten minutes was typically required to get the computers out and load the applet or simulation, as the computers were fairly slow.

The links to the applets were posted in advance on Canvas, the learning management system used at Utah State University. Once, when the links were copied into Canvas, the “~” symbol disappeared from the link. Students were unable to access the applet which slowed down the class. The instructor re-posted the link and emailed it to students directly. Many of the applets required JAVA. Due to settings on the laptops, the students were occasionally unable to access the applets. This annoyed the students enough for them to complain on the exit slips.

The instructor created an excel simulation for ANOVA and planned to have the students explore it in pairs. The simulation had been created in Excel 2010, but the students' computers had Excel 2007, so they could not use the simulation. The instructor led the students through the simulation as a class, but ten minutes had already been wasted trying to get the simulation to work on the student laptops. In addition, the simulation would have been much more effective if students were exploring the simulation in pairs, rather than watching the teacher.

### **True Difference Between Teaching Methods**

The biggest difference between the classes was the teaching method. The activity-based classes were involved in group work, class activities, and direct instruction while the lecture classes only had direct instruction. Sometimes the lessons for both classes were almost identical, most commonly during hypothesis tests when more practice was involved than discovery. However, the instructor would lecture through all the material in the lecture classes, while in the activity classes, the students would do the work in groups, with at most a short teacher introduction to a new procedure.

Even though the instructor tried to incorporate some discovery type lessons into the lecture classes, the activity classes were much more discovery-based. Usually, the instructor tried to explain the reasoning behind a new procedure in the lecture classes to make sure that the students understood the concepts behind the procedures, but in the activity classes, the students discovered many of the procedures and concepts themselves.

For example, when introducing the formula  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  in the lecture classes, the instructor drew Venn diagrams to illustrate the need to subtract the probability of the intersection after double counting. For the activity class, students were asked to find  $P(A \cup B)$  which they were able to easily do using logic due to the setup of the problem. Then they were asked to find  $P(A) + P(B)$  and students realized the answers did not match. Through some hints, students were able to realize that they had double counted the intersection. Finally, students were asked to come up with the formula.

As another example, when introducing the binomial distribution, the instructor told the students the probability formula, and then explained the logic behind each piece of the formula. In the activity class, the teacher guided students as a class through discovering patterns in probabilities and generating the probability formula.

For confidence intervals, the teacher guided all the classes for both teaching methods through the theory underlying the basic confidence interval formula (estimate  $\pm$  critical value  $\times$  standard error). But the activity class then discovered the meaning of the confidence level through several interactive activities, while the lecture class watched the teacher demonstrate the meaning with an applet. In addition, the teacher guided the activity class through discovering three confidence interval formulas. The lecture students were just told each new formula. The lecture class had a very short segment where after calculating several margins of error, they realized the effect of changing the sample size or confidence level. They were then told that if they rearranged the margin of error formula, they could find the formula for the necessary sample size. The activity class had a much longer segment on margin of error. They discovered for themselves the effects of changing sample size or confidence level. They also came up with the formula for finding the necessary sample size.

Some of the topics were motivated differently between the teaching methods. For example, during the lecture classes, the teacher simply introduced the formula for expected value and told students that the expected value was a long run average. For the activity classes, the instructor helped students see the need for a long run average by talking about carnival type games.

The activity classes collected real data throughout the semester, but they also used many data sets that were found in textbooks; they used more data sets from the textbook than data sets they collected. The lecture classes only used data sets from textbooks.

The notes for the activity classes usually had more steps and clearer explanations than the guided notes for the lecture classes. This resulted because activity students were expected to



be able to go through the notes without the instructor's assistance. In the lecture classes, the instructor was able to expound on the notes, and so the pre-written notes did not have to be as detailed.

### **Revision of the Instructional Materials**

After both the lecture and activity classes each day, the student comments and questions during class, comments on exit slips, and instructor observations were used to revise and improve the instructional materials. This process was undertaken during the spring semester, and the materials were further revised during the fall semester.

Typographical errors were fixed in the notes for both classes. In the activity classes, errors from copying and pasting were commonly found in the keys. Oftentimes, an example needed to be reworded, or a few words needed to be added for clarification. Sometimes, the directions for the activities were revised to lessen student confusion. In the activities, I often needed to add or change steps to facilitate the discovery process. Sometimes, I decided the wording I used in one class was better than the wording in the other class, so I changed the notes so that both classes were similar. Occasionally, the activities needed to be changed or revised to help students better develop their understanding.

Several times the units or examples were reordered. This helped with the flow of the material. By working on a specific example before another example, an important point or distinction could be made. Occasionally, I moved an example forward in the notes because I felt it was more important than other examples in the section. This way I could ensure that I covered the most important examples first. I also found a few instances where a difficult question was used to introduce the topic and was then followed by an easier question; it seemed logical to switch the order of the questions in that situation.

Below are some examples of the revisions and reasoning behind the revisions. A partial list of the revisions can be found in Appendix I. The revisions in the appendix are the ones that I documented as I made the revisions, but they do not account for all the revisions I made. In particular, I did not document as many revisions during the spring semester, partly because I was so busy developing the new activities. Minor changes such as fixing typos are not detailed in the appendix. Overall, I felt pleased with the revisions I made between the two semesters and

optimistic about the changes after the fall semester. However, no specific data were collected that would allow for the evaluation of the effectiveness of specific revisions.

**Example.** I completely changed the flow of the probability section for the lecture class. When I initially made the guided notes, I followed the structure of the textbook. In the textbook, the basic probability rules and methods are covered in section 4.2, and then covered again in section 4.5 in a more advanced manner. Partly because of this, and partly because of the way I organized the notes, we covered the some topics several times. I felt that the examples were jumping from topic to topic and that the entire unit was very disorganized. When I created the activities I felt that the organization of the probability unit was more effective than the unit in the guided notes. For the second semester, I reorganized the unit for the lecture class. All of the addition rules for disjoint and non disjoint events were covered together. The multiplication rule for independence was taught once and then revisited after the general multiplication rule, but it was clear that the section was a review. The conditional probability section was moved up in the unit. I felt much better about the probability unit the second time I taught it. The examples and topics seemed to flow better.

**Example.** The first semester for both classes, I followed the structure of the textbook and taught the binomial distribution after the sampling distribution for the mean, and before the sampling distribution of the proportion. I decided this was not logical. When we were working with the binomial distribution we needed the skills we learned in the random variables unit. In fact, the binomial distribution felt like an extension of the random variables unit; that made it feel out of place in the sampling distribution unit, leading to extra confusion. For the second semester, I moved the binomial distribution to right after the random variables. The revised order of topics is random variables, binomial distribution, and then the normal distribution.

**Example.** The first semester, we did an activity where students sorted histograms into three categories: left skewed, symmetric, and right skewed. Students were given no prompting as to what the categories should be, and they developed their own names. Unfortunately, all the symmetric distributions happened to be mound shaped, see Figure 1, and the students called the category the normal or bell category instead of focusing on symmetry. To encourage students to focus on the symmetry aspect, I added some new symmetric distributions, see Figure 2.

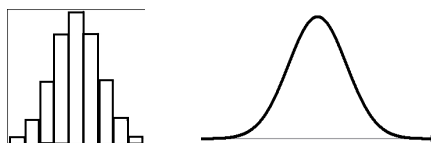


Figure 1: Examples of original symmetric figures.

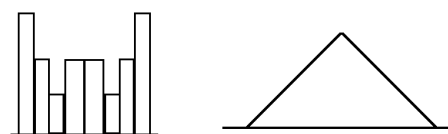


Figure 2: Examples of new symmetric figures.

**Example.** For the linear regression unit, there was an example in the notes for both classes about using home value to predict upkeep. During the second semester, I had the students try to do a prediction that was extrapolation. Then they found and interpreted  $r^2$ , and then they filled out a table with various home values to decide whether or not the predictions would be extrapolation. I decided that the order of the problem did not make sense. I moved the table for extrapolation before the  $r^2$  problem to right after the first extrapolation question.

In addition, the units for home worth were in thousands of dollars. I added an extra column to the table to force students to compute the  $x$  value for home worth before deciding whether they should predict the upkeep value or if it would be extrapolation. I also changed one of the home values from \$300,000 to \$320,000 so that the  $x$  value was close to the existing data range, but not in it. In the future, this should offer the chance to have a debate over whether a prediction is extrapolation if the  $x$  value is really close to the existing data range.

**Example.** After the second semester, for the beginning of the advanced regression section when population parameters are introduced, I replaced the guided notes section with what we used in the activity section. The notes were similar, but the activity section was a little better. I think that the activity notes were set up better because I designed the notes for students to read themselves. Because I planned on students reading and learning the material without my help, I had to put much more effort into making the material clear and understandable. In addition, anything I planned on saying to students in the lecture class had to be included in the activity notes since I was not going to be lecturing to the activity students.

### *Extra Handouts for the Lecture Class*

During the first semester, I created all the guided notes before the semester began. Throughout the semester, as I created material for the activity course, I occasionally came up with an idea for teaching a topic that I thought would work better than what I had created for the lecture class. I created a few extra handouts for the lecture class to supplement or replace their existing guided notes. The extra handouts were on using the complement rule for “at least one” situations, Bayes Theorem, extra problems for expected value and variance, extra problems for hypothesis testing, interpreting p-values and significance levels, and extra problems for linear regression. The extra handouts were still presented using lecture. For the second semester, the extra handouts were integrated into the guided notes.

## **Compare Exam Scores by Teaching Method**

In this section, the student data from exam scores are analyzed to answer research questions one, two, three, and five. The points from the final exam, the points from all the exams combined, the points from conceptual questions, and the points from procedural level questions are compared by teaching method. A multiple regression is conducted to determine which predictors are most useful for predicting total exam scores. In addition, exam items are examined individually and by topic for differences in student scores between the teaching methods. First though, the activity and lecture classes are compared to determine if the samples were roughly equivalent with regard to gender, previous GPA, pretest scores, and area of study.

### **Final Set of Participants**

At the university level, students commonly add and drop classes during the first few weeks. Hence enrollment was counted after three weeks instead of the beginning of the semester. One student declined to participate in the study. Three students added the class late and were ineligible for the study. A total of six students dropped the course, one student from the lecture class and two students from the activity class each semester. The overall participation in the study rate was 93.5%.

Table 7: Number of students eligible for the study in each class.

	Spring Semester	Fall Semester	Total
Lecture	43	44	87
Activity	31	28	59
Total	74	72	146

### Were the Classes Similar?

#### *Previous GPA*

Due to an outlier, the Wilcoxon rank-sum test with a normal approximation was conducted to compare the previous GPA for students in the lecture and activity courses. There was no significant difference in previous GPA between students enrolled in the activity and the lecture classes; the two-sided p-value was 0.4278.

#### *Pretest Points*

Due to potential outliers, the Wilcoxon rank-sum test with a normal approximation was conducted to compare the pretest points. The resulting two-sided p-value of 0.9089 shows that there was no significant difference in the pretest points between the participants in the two different types of classes.

#### *Gender*

To assess if the two sections had similar gender ratios, a chi-square test of homogeneity was carried out. With a p-value of 0.9686, there is no evidence that the activity and lecture courses differ significantly in regard to gender.

#### *Area of Study*

To assess if the two teaching methods had similar students with respect to their chosen major, a chi-square test for homogeneity was conducted on which college a student was enrolled in. Fisher's exact test was used because many expected counts were less than five. The p-value was .4933.

There is no reason to believe that there was a significant difference in which college a student was enrolled in between the two teaching methods.

### *Summary*

In conclusion, there is no evidence of significant differences in the two classes with regard to the pretest score, previous GPA, gender or field of study. Students enrolled in the activity courses were fairly similar to the students enrolled in the lecture courses for these four variables.

### **Compare The Exam Scores**

The total points for all the exams, the points from the final exam, the points from all conceptual level questions, and the points from all procedural level questions were analyzed. The descriptive statistics are in Table 8. Box plots for the total points, final exam points, conceptual points, and procedural points by teaching method are in Figure 9. The box plots show that the data for each category is skewed with potential outliers.

The hypotheses were that the activity class would have higher scores for the total points, final exam points, conceptual points, and procedural points. The points for each category were skewed, but a t test would still have been appropriate due to the large sample sizes. However, due to outliers, the one-sided non parametric Wilcoxon rank-sum test with a normal approximation was used to compare the scores between the lecture and activity sections. See Table 9 for the results. The Bonferroni-Holm method was used to adjust the p-values which were then compared to a significance level of  $\alpha = .05$ .

The evidence did not suggest that there were any significant differences between the scores for the activity and lecture classes for the total points, final exam points, conceptual points, or procedural points. As can be seen in Table 8, even if the differences between the activity and lecture classes had been statistically significant, the averages never differed by more than two points between the classes; so the differences would not have been practically significant either.

Table 8: Descriptive statistics for the exam points by lecture and activity.

	Available Points for Category	Lecture Classes $n = 87$	Activity Classes $n = 59$
Total Points	143	$\bar{x} = 118.9$ $s = 13.73$	$\bar{x} = 121.6$ $s = 12.47$
Final Exam Points	43	$\bar{x} = 35.6$ $s = 4.69$	$\bar{x} = 36.86$ $s = 4.29$
Conceptual Points	94	$\bar{x} = 68.2$ $s = 7.86$	$\bar{x} = 70.10$ $s = 7.36$
Procedural Points	53	$\bar{x} = 42.2$ $s = 6.27$	$\bar{x} = 42.9$ $s = 5.73$

Table 9: P-values for comparing exam scores by teaching method.

	Standardized Test Statistic, $Z$	Raw P-value	Bonferroni-Holm Adjusted P-value
Total Points	1.1650	0.1220	0.2440
Final Exam Points	1.8249	0.0340	0.1360
Conceptual Points	1.4914	0.0679	0.2038
Procedural Points	0.4956	0.3101	0.3101

*Research Question One: Which Teaching Method Resulted in Higher Levels of Student Comprehension of Statistical Concepts and Ability to Apply Statistical Procedures?*

This question was partially answered by comparing the final exam scores by teaching method. The final exam scores were chosen because many items came from existing statistical assessment instruments, and the final exam was reviewed by an expert panel. The hypothesis was that the activity classes would outperform the lecture classes. The adjusted p-value for comparing the final exam scores was .1360. Hence, there is no evidence that the activity teaching method resulted in significantly higher levels of student comprehension of statistical concepts and ability to apply statistical procedures.

However, there is the possibility that other covariates are affecting exam scores. Perhaps, if those covariates were accounted for, the teaching method might be shown to have a significant relationship with the student comprehension of statistical concepts and ability to apply statistical procedures. A multiple regression to explore this possibility is conducted below.

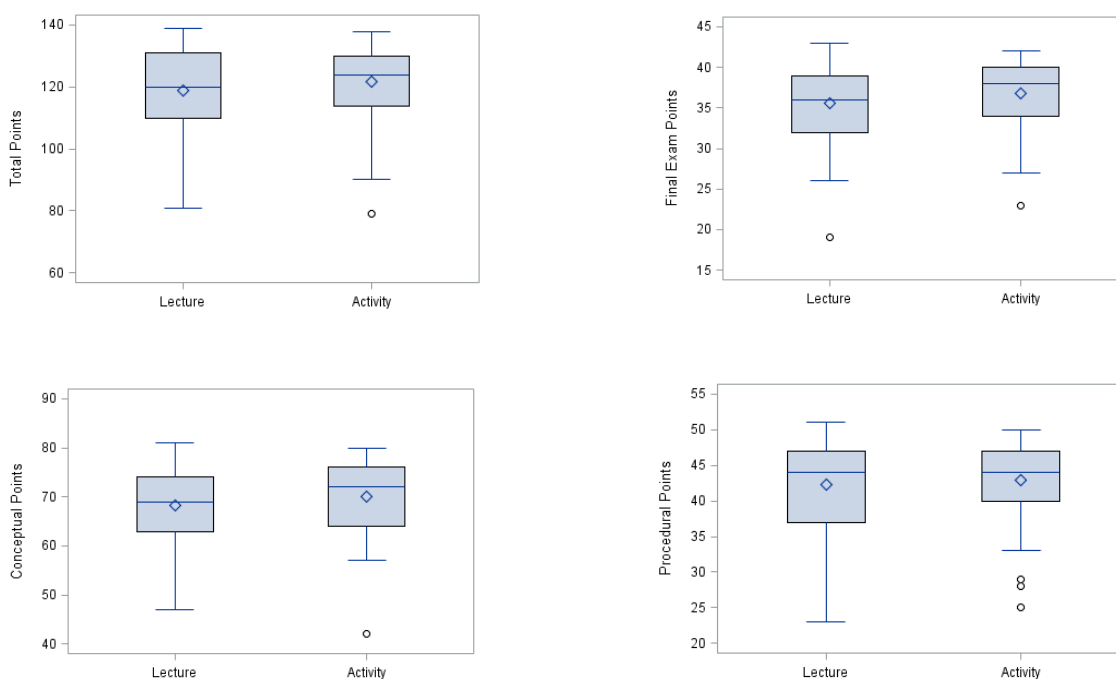


Figure 3: Distribution of exam points by teaching method.

*Research Question Two: Which Teaching Method Resulted in Higher Student Outcomes for Conceptual Level Questions?*

The points from all exams for conceptual level questions were compared by teaching method. The hypothesis was that the activity class would perform higher on conceptual level questions. The adjusted p-value is .2038, leading to the conclusion that the activity class did not perform significantly higher on conceptual level questions.

*Research Question Three: Which Teaching Method Resulted in Higher Student Outcomes for Procedural Level Questions?*

The points from all procedural level questions were compared by teaching method. Again, the hypothesis was that the activity class would do better than the lecture class on procedural level questions. However, the adjusted p-value is .3101. Hence, there is no evidence that the activity class performed significantly higher on procedural level questions.



### Compare Exam Scores Within Semesters

Although it is not a formal research question, comparing the results by semester led to interesting results. The Wilcoxon rank-sum tests were repeated, but the tests were conducted for the spring and fall semesters separately, see Table 10 for the results. Even after adjusting for multiplicity using the Bonferroni-Holm method, the activity class was significantly higher in total points, final exam points, conceptual points, and procedural points at the  $\alpha = .05$  level during the spring semester. However, the activity class did not outperform the lecture class during the fall semester; the adjusted p-values are all one. Since the activity class did better during the first semester, but not during the second semester, when we compared the exam points for both semesters combined, we did not find any significant differences.

Table 10: P-values for comparing exam scores by method and by semester.

	Spring Semester	Fall Semester	Both Semesters Combined
Total Points	.0033 (.0100)	.8614 (1.000)	.1220 (.2440)
Final Exam Points	.0006 (.0026)	.5484 (1.000)	.0340 (.1360)
Conceptual Points	.0036 (.0100)	.7255 (1.000)	.0679 (.2038)
Procedural Points	.0173 (.0173)	.9185 (1.000)	.3101 (.3101)

Note 1: The raw p-values are given first. The adjusted Bonferroni-Holm p-values for that semester are given second in parentheses.

Note 2: The hypotheses are that the activity classes scored higher than the lecture classes. So all p-values are one-sided. In the spring semester, the activity class scored higher in each category. In the fall semester, the lecture class scored higher in each category.

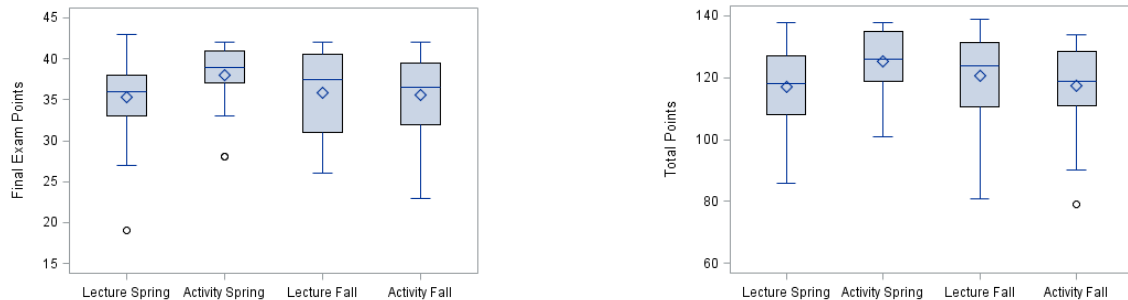


Figure 4: Distribution of final exam points and total points by class.

## Conduct a Multiple Regression

### *Base Model*

A multiple regression was computed for the total points on all exams. The possible explanatory variables were pretest scores, fall or spring semester, time of day the class was held, teaching method, previous GPA, percentage of attendance, homework scores, gender, whether the student took statistics previously, and the interaction between teaching method and previous GPA. The assumptions for multiple regression were met; the residuals looked normal and there were only slight signs of heteroscedasticity.

The entire model had a test statistic of  $F = 6.53$  and a p-value less than 0.0001. So at least one of the explanatory variables has a significant relationship with the total exam points. However,  $R^2 = 0.3276$ . So the model only explains 32.76% of the variation in total exam scores. The less biased adjusted  $R^2$  value is 0.2774. The teaching method does not appear to be a significant predictor with a p-value of 0.7944. The interaction between teaching method and previous GPA is not significant either with a p-value of 0.7129.

Table 11: Results from multiple regression for the base model.

Variable	Parameter Estimate	Standard Error	t Value	Pr >  t	Standardized Estimate
Intercept	55.6505	10.9232	5.0900	<.0001	
Pretest Points	0.5956	0.2867	2.0800	0.0397	0.1592
Semester	-2.9533	1.9490	-1.5200	0.1321	-0.1119
Time of Day	2.6414	2.0059	1.3200	0.1901	0.1001
Teaching Method	-3.9526	15.1354	-0.2600	0.7944	-0.1472
Previous GPA	12.1329	2.8023	4.3300	<.0001	0.4082
Attendance Percentage	0.1827	0.0941	1.9400	0.0542	0.1696
Homework Percentage	0.0663	0.1020	0.6500	0.5169	0.0572
Gender	-0.6875	1.9506	-0.3500	0.7250	-0.0260
Taken Statistics Previously	-0.2876	2.7825	-0.1000	0.9178	-0.0078
Interaction between GPA and Method	1.6675	4.5213	0.3700	0.7129	0.2109

### Stepwise Selection Model

Stepwise selection was used to determine a model with less predictor variables. The significance level for a variable to enter or leave the model was set at 0.10. The stepwise selection model has a test statistics of  $F = 20.08$ , and a p-value less than 0.0001. The coefficient of determination is  $R^2 = .2993$  with an adjusted  $R^2$  value of .2844. With an adjusted  $R^2$  value of only .2844, this model does not explain much of the variation in total exam scores. The regression model would not be useful for making predictions.

The variables left in the model are students' previous GPA, scores on the pretest, and percentage of attendance. The explanatory variables in the model are all significant at the  $\alpha = .05$  level, see Table 12. The previous GPA appears to be the most significant predictor. A one standard deviation increase in a student's GPA will result in a predicted increase of 0.44488 standard deviations for total points, assuming pretest points and attendance remain constant.

Table 12: Results for stepwise selection model.

Variable	Parameter Estimate	Standard Error	t Value	Pr >  t	Standardized Estimate
Intercept	55.55512	8.79489	6.32	<.0001	0
Previous GPA	13.22246	2.16608	6.1	<.0001	0.44488
Pretest Points	0.62007	0.26538	2.34	0.0209	0.16572
Attendance Percentage	0.21651	0.07892	2.74	0.0069	0.20105

The correlations of the predictor variables with the total exam scores are not very high, see Table 13. The scatterplots of each predictor variable and the total exam points are in Figure 5. The pretest points and total exam scores only have a correlation of 0.1240; considering the correlation and the scatterplot, the pretest points do not really seem to have a relationship with total exam scores. However, with a correlation of 0.4908, the students' previous GPA does have a relationship, although not a very strong relationship, with total exam scores. As would be expected, students who have prior high GPA scores did better in the class. Overall, as students attended class more, their total exam scores increased; the correlation was 0.3009. The relationship between attendance and total exam scores is not very strong either. There is a lot of individual variation in total exam scores from student to student that has not been accounted for by any of the covariates.

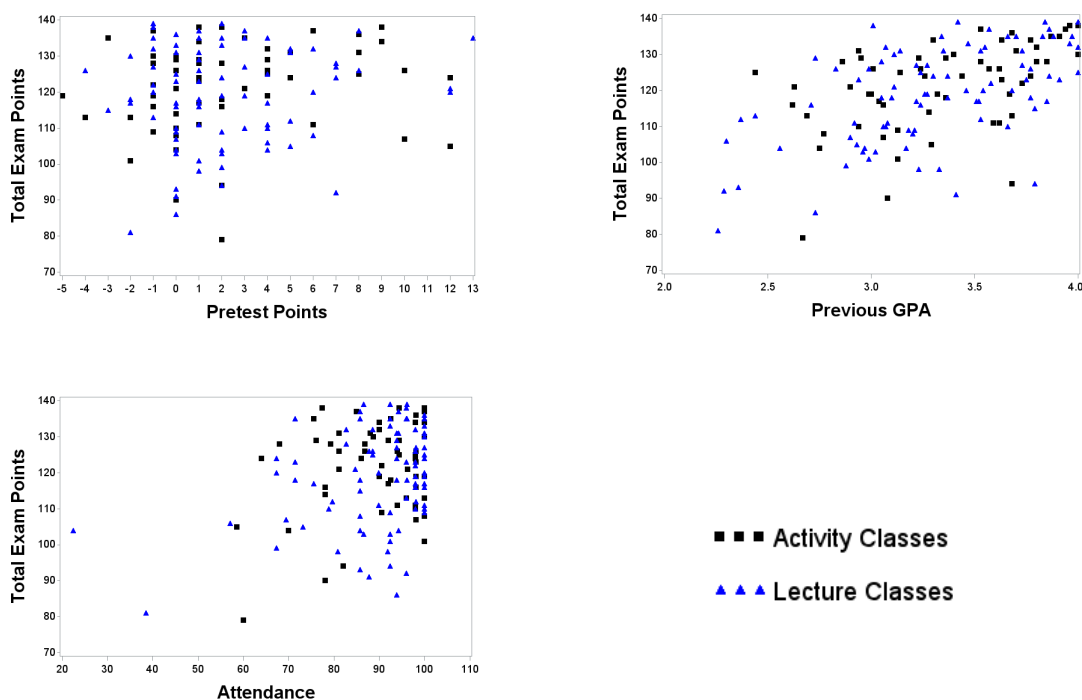


Figure 5: Scatterplots of predictor variables against total exam points.

Table 13: Correlation of predictor variables with total exam points.

Intercept	Correlation
Pretest Points	0.1240
Previous GPA	0.4908
Attendance Percentage	0.3009

*Research Question One Continued: Which Teaching Method Resulted in Higher Levels of Student Comprehension of Statistical Concepts and Ability to Apply Statistical Procedures?*

The total points were used for the multiple regression because more information was contained in the scores from all combined exams than would be available from just the final exam points. The teaching method has a p-value of 0.7944 in the base model, and does not appear as an explanatory variable in the stepwise selection model. This led to the conclusion that teaching method does not have a significant relationship with the total exam scores. Combining the results from the multiple regression and the earlier comparison of scores from the final exam, it is clear that there

is no evidence that the activity teaching method leads to higher levels of student comprehension of statistical concepts and ability to apply statistical procedures.

### **Compare Scores for Topics and Individual Questions by Teaching Method**

The exam questions were organized into topics. The topics were analyzed to determine if one teaching method resulted in higher scores. The individual questions were also analyzed. If significant differences existed between teaching methods for any topics or individual questions, the data could be analyzed to look for patterns. Since no hypothesis was made as to which teaching method would be better for specific topics or questions, two-sided hypothesis tests were used.

#### *Compare Scores for Topics*

The exam questions were categorized into sixteen topics that were compared with the Wilcoxon rank-sum test. The results were then analyzed at the  $\alpha = .025$  significance level, after the p-values were adjusted using the Bonferroni-Holm method. The p-values are in Table 14. Since the smallest adjusted p-value is .0350, the conclusion is that none of the topics have significant differences between scores for the activity and lecture classes.

#### *Compare Scores for Individual Exam Questions*

There were 143 questions from all the exams. So the Bonferroni-Holm method was used to adjust for the multiplicity to control the familywise error rate. The data for each question was binary for each student, so the chi-square test of independence was the appropriate choice to compare the proportions of correct answers by teaching method. However, many of the expected counts were less than five. So Fisher's exact test was used to compare the results for individual questions by teaching method at the  $\alpha = .025$  level. The four individual questions with the smallest p-values are shown in Table 15. So the smallest adjusted p-value is 0.6028, and the rest of the adjusted p-values are 1.000. Clearly, there are no significant differences in the proportion of students who answered the individual questions correctly for the activity and lecture classes, for any of the individual exam questions.

Table 14: P-values for comparing scores for topics by teaching method.

Topic	Standardized Test Statistic, $Z$	Raw P-values	Adjusted P-values
Descriptive Statistics	3.0634	0.0022	0.0350
Confidence Intervals	2.1043	0.0354	0.5303
Binomial Distribution	1.8994	0.0575	0.8051
Statistical Reasoning	1.7123	0.0868	1.0000
ANOVA	-1.6357	0.1019	1.0000
Linear Regression	1.0831	0.2787	1.0000
P-values	0.9463	0.3440	1.0000
Probability	-0.9053	0.3653	1.0000
Sample Proportions	0.8922	0.3723	1.0000
Sampling Distributions	0.6958	0.4866	1.0000
Hypothesis Testing	0.6717	0.5018	1.0000
Sample Means	0.2633	0.7923	1.0000
Chi-Square Tests	0.1371	0.8910	1.0000
Normal Distribution	0.0691	0.9449	1.0000
Causation vs Association	-0.0660	0.9474	1.0000
Expected Value	0.0127	0.9899	1.0000

*Research Question Five: Which Teaching Method Produced Higher Student Learning Outcomes for Specified Statistical Content?*

There were no significant differences for any of the topics or individual questions based on teaching method. Hence, we concluded that the two teaching methods did not produce different levels of student learning for any of the specified statistical topics.

### **Compare Student Attitudes by Teaching Method**

In this section, data from the Attitudes Toward Statistics survey, modified Current Self-Efficacy survey, and the End of the Semester survey is analyzed to answer research question six. Since no hypothesis was made about which teaching method would result in higher, more positive student attitudes, two-sided tests were used.

Table 15: P-values for comparing individual exam questions by teaching method.

Exam	Question Number	Raw P-value	Adjusted P-value
Exam 4	26	.0044	.6028
Exam 3	13	.0079	1.000
Exam 4	5	.0180	1.000
Exam 1	19	.0261	1.000
All other questions		>.0261	1.000

### Analyze Attitudes Toward Statistics

The Attitudes Toward Statistics (ATS) scale had five possible choices for each item: strongly disagree, disagree, neutral, agree, and strongly agree. Some of the items reflect positive attitudes towards statistics, while other items represent negative attitudes towards statistics. The student responses to each item were converted to numerical scores from one to five, with five representing the highest, most positive attitude towards statistics. Hence on items with a positive slant, the option “strongly agree” was given a score of five, whereas on negative items, the score of one was given to the option “strongly agree”.

The data were ordinal, and definitely not normal, so the two-sided Wilcoxon rank-sum test was used for comparisons. The numeric scores on all 29 questions were added together for a total attitudes score. The two-sided p-value for the Wilcoxon rank-sum test was 0.8104. After comparing the p-value to the significance level of  $\alpha = .025$ , we have no reason to believe that the participants in the activity and lecture classes have different overall attitudes towards statistics.

Next, we compared the results on each of the 29 items on the survey by teaching method. Since we were doing 29 tests, we would expect to see some small p-values just by chance. The familywise error rate was controlled by using the Bonferroni-Holm method. All of the p-values were adjusted, see Table 16. The adjusted p-values were then compared to the significance level of  $\alpha = .025$ . None of the tests for the 29 items were statistically significant. Once again, there is no evidence of a difference in students’ attitudes towards statistics between the two teaching methods.

Table 16: P-values for the two-sided Wilcoxon rank-sum test for ATS questions.

Question	P-value	Adjusted P-value	Question	P-value	Adjusted P-value
2	0.5731	1.000	1	0.2155	1.000
5	0.7650	1.000	3	0.1380	1.000
8	0.6326	1.000	4	0.9057	1.000
10	0.2220	1.000	6	0.4609	1.000
12	0.2328	1.000	7	0.6546	1.000
14	0.5069	1.000	9	0.7511	1.000
15	0.1041	1.000	11	0.2983	1.000
16	0.2733	1.000	13	0.3606	1.000
17	0.5808	1.000	18	0.6160	1.000
20	0.2421	1.000	19	0.9344	1.000
21	0.6206	1.000	22	0.6546	1.000
23	0.4493	1.000			
24	0.4732	1.000			
25	0.7465	1.000			
26	0.0950	1.000			
27	0.3384	1.000			
28	0.7369	1.000			
29	0.2178	1.000			

Note: The left table consists of questions for which the lecture classes had higher, more positive attitudes towards statistics. The right table includes the questions for which the activity classes had higher scores.

### Analyze Self-Efficacy Survey

Similarly, due to the ordinal nature of the data, the two-sided Wilcoxon rank-sum test was used to compare results for the self-efficacy survey. The p-value for the total self-efficacy score is 0.1424, with the lecture class having slightly higher results, but not significantly so since the significance level was  $\alpha = .025$ . The 12 individual items were also tested and the p-values were adjusted using the Bonferroni-Holm method, see Table 17, and then compared to  $\alpha = .025$ . None of the 12 individual items were significantly different. Hence, there is no reason to believe that there was a difference in self-efficacy between the students who experienced the two teaching methods. Each



Table 17: P-values for the two-sided Wilcoxon rank-sum test for self-efficacy questions.

Question	P-value	Adjusted P-value	Question	P-value	Adjusted P-value
1	0.9098	1.0000	7	0.7714	1.0000
2	0.1689	1.0000	8	0.8397	1.0000
3	0.0637	0.7004	9	0.2346	1.0000
4	0.2029	1.0000	10	0.1011	1.0000
5	0.0284	0.3403	11	0.4023	1.0000
6	0.2328	1.0000	12	0.7771	1.0000

Note: Item 8 was the only question on which students in the activity classes had higher self-efficacy than students in the lecture classes.

self-efficacy item was measured on a scale from one to five. For most items on the survey, the two teaching method's self-efficacy averages differed by less than 0.1.

Figure 6 has the bar charts for item 5 on the self-efficacy scale, "Distinguish between a Type I error and a Type II error in hypothesis testing." It is the item with the smallest p-value from the significance tests comparing the two methods. The figure reinforces the conclusion that there was not a big difference in the distribution of self-efficacy between the two classes.

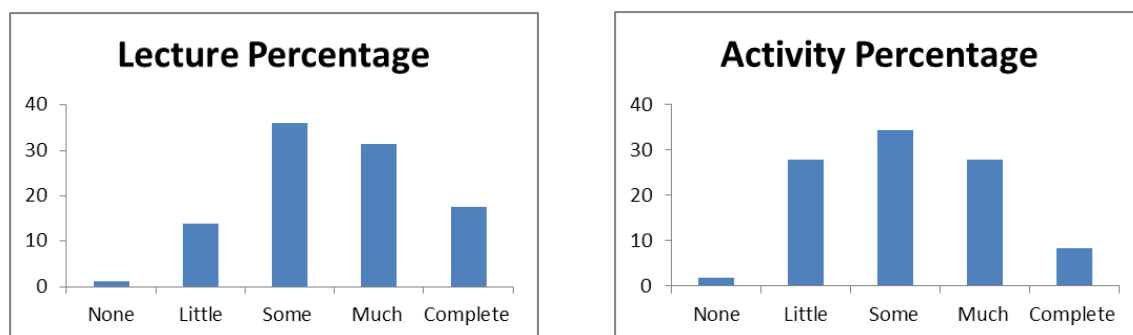


Figure 6: Bar chart for self-efficacy item five: Distinguish between a Type I error and a Type II error in hypothesis testing.

### Analyze End-of-Semester Survey

Student results on the end of semester survey were analyzed for the attitude related questions. Student answers were converted to a numerical scale giving ordinal data. The two-sided Wilcoxon rank-sum test was used to compare the results to the following questions.

Table 18: P-values for the two-sided Wilcoxon rank-sum test for end of semester survey questions.

Question	P-value	Adjusted P-value
useful	0.3405	1.0000
feel	0.7513	1.0000
enjoy	0.6364	1.0000
engaged	0.2119	0.8476
motivated	0.1098	0.5489

- How useful do you think statistics will be to you in the future, either in your career or your personal life?
- How do you feel about statistics?
- How much did you enjoy this statistics class?
- How engaged were you during class?
- How motivated were you to come to class?

Because there were five items, the Bonferroni-Holm method was needed to adjust the p-values, see Table 18. Afterward, the adjusted p-values were compared to the significance level  $\alpha = .05$ . There were no significant differences between the activity and lecture teaching methods for any of the five questions.

**Research Question Six: Were There Differences in Student Attitudes Towards Statistics or the Statistics Course, Based on Which Teaching Method Was Used?**

There were no significant differences between the classes based on which teaching method was used for the end of semester questions, the attitudes towards statistics questions, or the self-efficacy questions. The conclusion is that there were no differences in student attitudes towards the statistics course or statistics in general, based on which teaching method was used.

## Results from Teacher Observations

The following consists of the researcher's reflections as the instructor. The sources for the reflections include the teacher logs, the teacher's observations during the class, student comments that the instructor overheard, and insights from reading the students' comments on the exit slips throughout the semester. Examples of the teacher logs can be found in Appendix J.

### Activities

Students seemed to enjoy having activities. The first few weeks of class, having activities and group work was a novelty. Students were excited to be more engaged than in a traditional lecture class. However, as the semester progressed, students became less excited about the activities. In particular, just as students get bored being in a lecture class every day, students complained that working on packets in groups every day became monotonous.

Although the initial enchantment with activities wore off, students still seemed to enjoy activities. They especially liked the more interactive activities. Students also enjoyed days when a variety of methods were used including class activities, teacher-centered activities, and group work.

I was occasionally surprised at how much students could do without much teacher guidance. For the symmetry activity, students were told to sort images into three categories. The students did not seem intimidated by having to sort things into categories without more specific instructions. I was amazed at how fast they separated the graphs into the "correct" categories of left skewed, right skewed, and symmetric.

I felt that students in the activity classes were able to learn many concepts without lecture. For example, I never showed them how to find probabilities using the normal distribution or how to standardize. Yet, they were able learn the procedure by following the steps outlined in their notes. The spring semester had an additional day, so we covered the chi-square and F tests for variance(s). I let the activity class work in groups with no teacher introduction and they seemed to learn the material without any significant problems.

### *Food-Based Activities*

Many students seemed to particularly enjoy food-related activities. Not only did they enjoy the food, but the food seemed to make the activity more concrete and easier for the students to relate to. The food-based activities seemed to be engaging. However, the allure of the food seemed to wear off toward the end of each semester, and more and more food ended up in the trash. Students were not as excited to get M&Ms after the second time. During the fall semester, class was at 8:30 in the morning, and the instructor overheard several students complain that it was too early for chocolate or root beer. When asked what the most memorable activity was, 69% of the students mentioned an activity dealing with food.

### *Cost of Activities*

The total cost of the activities classes for both semesters was \$98.43. I bought various items for hands on experiments such as M&Ms and root beer. We also needed plates, cups, bags, and other such items for the experiments. I did borrow a set of inflatable globes, worth \$20-\$30, and a set of dice, worth \$10-\$20, for the activities as well. See Table 19 for a detailed list of the costs. I also made copies of the keys and printed supplements to activities; this was an added cost to the department, but not a cost that I tallied. No cost was incurred for lecture classes beyond the typical copying of the syllabus, exams, and a few occasional handouts.

### *Student Resistance to Activities*

I knew before I started the research study that students might be resistant to the non-lecture teaching methods. I did not receive many written complaints about the new methods, but there were a few. From the few written complaints and overheard comments, I felt that students did not think that I was teaching them when they were working in groups. As one person said, "I could do worksheets at home." Another student said, "Giving us a packet and letting us do it ourselves seems like a huge waste of tuition dollars." So there was resistance to the new teaching methods, but not as much as I expected.

The biggest source of student resistance could be seen in the fact that the majority of students wanted more teacher guided activities and teacher introductions. I think it was difficult for students to come from years of lecture-based classes, especially math classes, to a statistics class where they

Table 19: Cost of activities.

Item	Cost (dollars)
M&Ms	49.76
Root beer	9.04
Cookies	9.57
Beads	4.97
Plates	7.38
Plastic bags	5.79
Cardstock	2.63
Draw bags	3.33
Gloves	3.99
Cups	1.98
Total	98.43

were largely responsible for their learning. They wanted the teacher to show them how to do everything.

### Teacher-Centered Activities

Whenever I lectured, or guided the class through activities, I received positive comments. Students particularly liked it when I gave an introduction to the topic before they started working. Students usually mentioned that it was easier to understand the material when I played a more active role. However, while students seemed to feel that they were not understanding the material when they worked in groups, as the instructor, I felt that they were understanding the material at least as well as the lecture students.

While I still believe that using non-lecture methods can be effective, I plan on using more teacher-centered methods the next time I teach. Until students become used to non-lecture-based classes, I do not think it will be as effective to try to teach with as little lecture as possible. At the very least, I would give a 5-10 minute introduction of the new material each day.

Most of the teacher-centered time occurred during the hypothesis tests section. I found that there were many concepts and formulas that were harder for students to discover, and many that were set by convention, and hence not discoverable. The only alternative that I could see to introducing these topics by lecture was to have students read the material at home. However, I

believe that students learn more from a lecture and the chance to ask questions, than they do from reading.

The goal was to teach the course with as little lecture as possible. I now believe that introductory statistics can be taught with a minimal amount of lecture, in my case about 10% of teaching time. However, I do see the benefit of still using lecture occasionally as a teaching device.

### **Group Work in the Activity Class**

As the instructor, I saw several benefits of students working in groups. I've discovered over the years as both a student and a teacher, that in a lecture format, it is possible to follow along with the teacher and feel okay with the material. But when it comes time to do the homework alone, the material suddenly seems harder. My favorite aspect of having students work in groups during class was that students were required to work on the material themselves. By doing so, they could address issues and questions as they arose because the students had access to the instructor to ask questions. The students realized that they did have questions and what their difficulties were. By doing group work, students had the chance to try the procedures themselves during class, not when they were home alone with no access to the instructor.

Another benefit to group work is that students can share ideas. Oftentimes, if one student can not figure out how to do a problem, or why something needs to be done, someone else in the group can. In this way, students become peer tutors. The student needing help receives it. The student giving help has the opportunity to further understand the topic by teaching it.

### *Students Working at Different Rates*

Students tended to work at different rates. This was very obvious between groups or pairs, but it was also sometimes an issue for people within the groups. Several students complained throughout the semester that their partner either worked too slow or too fast. The common complaint was that if a partner worked too fast, the student had to scramble to keep up and so the student did not comprehend and understand the material completely. In each class, there were a few of the highest scoring students that did not like to work in groups, and often refused, because they felt that the other students slowed them down.

One of my biggest problems as an instructor was to create an activity packet that would take the correct amount of time. The slower people needed to be able to finish the critical information, but the fastest people needed enough material to be engaged the entire time.

Whenever a student left in the last 10 minutes, it caused a chain reaction and many students would start packing up and leaving, although some might stay and finish. If people left in the last five minutes, then typically everyone would leave regardless if they were finished. It was very frustrating to me.

I would often bring some copies of their homework, or something else for students to work on if they had extra time, but they would not start it unless they had 10 minutes left. Otherwise they just left. Most days, I just tried to add enough extra problems to the packet to occupy everyone the entire time. The extra problems did not contain new or essential information, but were extra practice. Students were better at finishing extra problems that were already attached to the packet, than they were at working on an extra handout.

### *Group Size*

At the beginning of the first semester, I tried to have students work in groups of four. After a few weeks, and looking at the results of a student survey, I let students work in groups of two. The majority of the students said that it was too difficult to work with a group of four because they worked at such different paces. They believed it was easier to only have to worry about keeping pace with a partner. Throughout the semester, I occasionally asked them to work in groups of four, but the groups usually devolved into partnerships quickly. It might have also been a matter of convenience since, with the classroom setup, working in groups of four required two students to turn around and share the desk of the students behind them. In addition, I learned not to have students work in groups of four if they needed to use a computer, because it was too difficult for everyone to see the screen.

As the instructor, I could see the benefit of both group sizes. I appreciated the groups of four, because having four people in a group meant that there was a bigger pool for brainstorming ideas. I felt like students had the opportunity to learn more in the group of four. Throughout the semester, I occasionally asked them to work in groups of four on the days when they would be doing more brainstorming.

Perhaps a more effective method would be to alternate between the group sizes often enough that students were comfortable with both sizes. On days when the students would mostly be practicing, they could work in pairs. On days when the activity would focus on discovering concepts, the students could work in groups of four. However, I think that specific steps would be needed to ensure that the students know how to effectively work and communicate in groups. In addition, perhaps their group work would need to be counted as part of their grade to give more motivation.

### *Activity Keys*

The students had access to the keys to the activities. Usually there was a key per pair of students, although occasionally they had to share the key among a group of four students.

Because students had access to the key, they often tried a problem and then realized they had done something wrong. This brought an opportunity for cognitive dissonance. Many great teaching moments happened when they discussed with their group or with the instructor why their method was wrong.

On the other hand, some students would look at the key before trying the problem. Instead of using the key to check their work, they copied the key and just tried to make sure they understood each step. A few students complained that the work seemed really easy in class because they were following the key, but then they did not know what to do on the exam.

A balance needs to be struck on the accessibility of the keys. I believe that students need to be able to check their answer after each question. If students work for an extended period of time without checking the key, they often end up ingraining incorrect procedures in their mind. In addition, students do not necessarily go back and check all their work carefully if they have done several problems without checking the key. On the other hand, I think that students need to try the problem first before checking the key. As the instructor, I do not believe I struck that balance. I was also limited by the classroom setup because students could not easily leave their chairs. So the key had to be close enough for them to access it without getting up.

### *Students Asking and Answering Questions*

As an instructor, I really liked having students work in groups with the option to ask the instructor questions. I felt like I was able to talk to more students and answer their individual



questions. Because I answered more questions from more students than I typically would in a lecture class, I felt better able to gauge the class's understanding of the material. I used this increased knowledge of the students' understanding to help plan for both future activity and lecture lesson.

In a lecture situation, usually a minority of the students actually ask questions. Having the instructor come around the room and talk to groups or individual students allows students to ask questions without the pressure of the entire class watching and listening. I felt that many more students were able to gain help. In addition, the questions and answers are tailored to the specific students, not to the entire course.

Many semesters, there were one or two students who ask many questions, some relevant, many not. In a lecture situation, this can be distracting and, for the irrelevant questions, a waste of time. If students work in groups and ask the instructor questions individually, one student's questions do not derail the entire class. The instructor might have to answer the same common question from different groups, but that takes the teacher's time, not all the students' time.

There was not enough room in the classroom to walk easily between the desks and walking between the desks usually entailed requiring the students to move their chairs. The process could be distracting. So I usually waited and walked around every 3-5 minutes unless someone had a question. However, I did find that students were more likely to ask a question if I was already walking by.

As a caveat, I discovered that there are some classes when the students do not ask many questions in the group work setting. Some days I only received one question in the entire 50-minute period. Those days were boring as the instructor.

As the instructor, I believe that having students work in groups with the opportunity to ask questions can be a very effective teaching method. However, steps might need to be taken to train the students to ask questions. Classroom norms might need to be specifically established, instead of hoping they develop naturally.

### *Instructor Asking and Answering Questions*

I was not very good at asking students questions as they worked. I would usually ask more questions at the beginning of the semester and ask fewer and fewer questions throughout the semester. It was partly because the room was fairly small and it was difficult to walk around the students. It was partly because I have not trained myself to do it well enough yet. It was mostly

because I disliked interrupting the students' concentration. It did not feel worth it to stop them while they were working. But I asked questions when I saw that they were writing something incorrectly, when I heard them say something wrong, when I heard them say something interesting that I could talk about, or when I heard them ask a question to their partner. I believe that the group work method could be even more effective if I developed the personal trait of asking more questions. However, I am still wrestling with the dilemma of whether it is better to ask questions or not distract the students.

### *Amount of Examples Covered*

Interestingly, students in the activity classes covered more examples than the lecture classes. Granted, the discovery lessons took longer for the activity class to complete than it took the instructor to cover the same information in the lecture classes. However, when students in the activity classes worked in groups, they could work through examples much faster than the instructor could lead the lecture class through the same examples. So overall, the activity classes had the advantage of discovering new material, and completing more examples.

### **Gathering Data**

Students in the activity class occasionally gathered data to analyze. However, I learned that any activity to gather data takes quite a bit of time. Class time is wasted when students need to record their results on a sheet for the entire class. In addition, it takes time to compute all the descriptive statistics for any data set. So to save time, whenever the students gathered data, I tried to compute all the descriptive statistics for them. Usually they gathered data the day in advance. This allowed me to compute the statistics overnight and have the summary data ready for students to analyze the next day. I believe that such time saving measures, and not having students gather data on a regular basis, allowed me to use activity-based methods and still be able to cover the entire existing curriculum.

I learned that students need fairly specific instructions when gathering data. The students can be involved in developing the specific instructions, but the instructions need to be very clear and probably repeated a few times. For instance, at the beginning of the semester, I asked students to tell me their height in *inches*. The direction to measure their height in inches was specific, but

apparently not stressed enough. Several students, both semesters, reported their height as five inches. However, that did lead to an opportunity to discuss how to deal with outliers.

Later in the semester, I wanted to have each student count the number of chocolate chips in a Western Family Chocolate Chip Chewy Cookie. I set out the cookies and paper plates before class so that the students could count the chocolate chips before class to save time. Then the students could eat the cookies. I had not thought of giving specific instructions on counting chocolate chips, or even realized that there were multiple ways to count the chocolate chips. Someone asked if the students were supposed to count the chips they could see from the top, or if they were supposed to tear the cookie apart. As a class, we decided they should tear the cookie apart, but several students had already eaten their cookie. So we ended up with an interesting data set. That experience really emphasized the need for clear, specific, and repeated instructions when gathering data.

## **Applets**

The applets were very useful tools. The applets seemed to be effective in helping students discover new concepts and visualize the topics. The students seemed to find the applets helpful as well as engaging and enjoyable.

I think that the chance for students in the activity classes to explore the applets themselves, instead of watching the teacher demonstrate the applet, was very beneficial. I think students had more of a chance to discover concepts at a personal level and also to realize what questions they had.

On the other hand, I occasionally realized that students in the activity classes did not understand how the applet worked, or what they were supposed to be learning. In the lecture class, the instructor could discuss what the applet was simulating. For example, during the second semester, several students in the activity class seemed really confused with the sampling distribution applet because they did not know what the applet was simulating. They did not truly understand that the applet was showing them the distribution of all the possible values of the sample mean. The directions to the activity did explain the purpose of the applet, but students had a tendency to start trying to use an applet without reading the directions carefully. Perhaps the more complicated applets need to be explained by the instructor before the students start exploring the applet.

While I liked having students explore applets themselves, sometimes it was not very time efficient. For example, I created an applet for learning the difference between categorical and

quantitative variables. I wanted students to use the applet so that they would experience cognitive dissonance when they realized zip code and drivers license numbers are not quantitative. This cognitive dissonance would arise because students are not allowed to drag those variables to the quantitative category. My applet worked and I had a lot of students ask me why the two variables were not quantitative. However, it felt inefficient to get the computers out for a three-minute activity. In the future, if the computers would be needed for less than five minutes, I think it would be more efficient for the teacher to demonstrate the applet to the entire class.

### **Individual Attention for Students**

I was able to get to know students much better in the activity classes. Granted, the activity classes were smaller than the lecture classes in this study, but the main reason was that I was able to spend much more time helping students at the individual level. I also had a lot of time to just observe the students while they were working. This enabled me to learn the names of more students from the activity classes.

### **Time Required to Prepare for Class**

The first semester, developing the guided notes for the lecture class took 2-3 hours per day. The time was spent reading the textbooks, typing the notes, and searching for examples. Developing the activities and workbook for the activity class took much longer, usually 5-10 hours per day. Much of the time was spent brainstorming ways to teach a concept without lecturing. Once a way to teach the concept was chosen, examples needed to be found and the activity needed to be typed. Creating the keys also took a significant amount of time. In addition, applets and simulations needed to be found, or made for the activities. If the activities needed any props such as flashcards or food, those needed to be prepared as well.

The second semester, much less time was required to prepare for both classes, but the activity class still required more time. For the lecture class, I could quickly read over the guided notes for the day to refresh my memory. That usually took ten minutes. Because we would cover as much as possible each day and then pick up where we left off with the lecture class, I did not have to worry very much about the timing. For the activity class, I needed to figure out which activities the students could cover in the 50-minute period every day. In addition, any props needed to

be prepared. If students needed to use applets, the links needed to be posted on Canvas. More time was also needed to plan what I would say as the teacher to help the students brainstorm and develop ideas on their own. Usually 30-60 minutes was spent preparing for the activity class each day.

### **Difference Between Sections of the Same Course**

I've heard teachers state that every section of students is different. I observed that phenomenon during this study when I taught 4 sections of the same statistics course during the year. I did notice slight differences between the two lecture classes in regard to the amount of questions, types of questions, and types of responses from students. However, because I was able to interact with students at a much more personal level in the activity class, and I had a lot of time to just observe the students, I saw a very big difference between my two activity classes.

Even on the second day of class, I wrote in my notes that the students in the fall activity class did not seem as willing to brainstorm as students in the spring activity class. Throughout the semester, students in the fall activity class were not as willing to work in groups, often working quietly by themselves. Only a few partnerships even asked each other questions on a regular basis. The fall activity class did not seem to be willing to brainstorm in class discussions, or offer ideas. This could have been due to the time difference; the spring class met at noon, but the fall class met at 8:30 am. Perhaps the students just were not alert yet so early in the morning, or the differences might have been due to which students happened to be in the section.

### **Attendance**

For all the classes, attendance would steadily drop throughout the semester. Not only would students not come to class, but many students would come late. This happened each semester and for each teaching method. (No data is available on student tardiness because attendance was tracked by exit slips turned in at the end of class.) Tardiness was a particular concern for the activity classes. If the teacher spent the first 5-10 minutes introducing a topic, but many students came 5-10 minutes late, then they completely missed the introduction.

### Nuances for Lecture Classes

One of the benefits of the lecture format was the opportunity to point out more subtle nuances to students. I found that when I was teaching the lecture classes, there were more opportunities to point out important things I really wanted students to know or even just subtle nuances.

For example, when reviewing for the first exam, students in both classes had trouble knowing whether to compare  $P(A | B)$  to  $P(A)$  or to  $P(B)$  to establish independence. I was able to tell students in the lecture classes which comparison to make and why several times throughout the review. For hypothesis tests, I frequently pointed out that we find evidence of our conclusion, we do not prove our conclusion. Whenever we needed to use a probability table, I made the class tell me which table we needed and why. During linear regression, I was able to point out many times that regression does not establish causation. The following is a quote from one of my teacher logs.

I think students in the activity class are having a harder time differentiating between Z and T tests because they have not been listening to me state the difference over and over like my lecture class has. I have had people in both activity classes look at the words “population” or “sample” and choose Z or T instead of looking for “population standard deviation” and “sample standard deviation.”

Granted, I tried to impart the same information to the activity classes, but I think there is a difference in expecting students to read a note while they are working and having the instructor point it out several times a day.

### Exam Reviews

We spent 1-2 days in class reviewing for each exam. The activity students worked in groups, and I went through problems for the lecture classes. I really liked having students review in groups because they could really figure out what questions they had and ask the instructor. When the review is covered in a lecture format, it is easy for students to follow along and think they understand everything, but when it is time for the exam, they are not necessarily ready. I believe that there is great benefit in allowing students to try the problems themselves. Nonetheless, it was nice to do the reviews as lecture because I could emphasize and point out things that I thought were important.

## Results from Students' Comments and Surveys

### Analyze Feelings About Activities Survey

The Feelings about Activities survey was administered online to the activity classes at approximately 3-5 weeks and 11-12 weeks into each semester. Students were required to take the surveys for participation points, but not all students responded. See Appendix D for more details about the survey.

#### *Learning with Activities*

Students were asked, "How would you rate your learning with activities versus lectures? (Do not consider how how entertaining or enjoyable activities are, just how well you learn with the activities.)" The results are summarized in Table 20. In regard to how much they learned with activities, at 3-4 weeks, 62.7% of students responded that they learned better with activities than lectures. At 11-12 weeks, 76.8% of students thought they learned better with activities.

The Wilcoxon signed-rank test was used to compare the first and second survey results since the data are ordinal and students took both surveys, resulting in paired data. The one-sided p-value of 0.0122 (Bonferroni-Holm adjusted p-value of 0.0244) led to the conclusion that students' opinions toward how much they learned with the activities became significantly more positive from about 4 to 12 weeks.

#### *Enjoyment of Activities*

Students were asked, "How would you rate your enjoyment of the activities versus lectures? (Do not consider how you learn, just how much you enjoy being in the class.)" The results are summarized in Table 21. The majority of students, 76.3% and then 67.9%, of the students either liked or loved the activities when compared to lectures, with regard to how much the enjoyed the course.

The Wilcoxon signed-rank test was used to compare the first and second survey results since the data are ordinal and students took both surveys. The unadjusted and adjusted p-values were both 1.000. Hence, there was no evidence that students opinions towards their enjoyment of activities improved significantly from four to twelve weeks.

Table 20: Results for how much students felt they learned with activities.

Preference	First Percentage	Second Percentage
Hate activities: learn much worse with activities than lectures	5.1	0.0
Don't like activities: learn a bit worse with activities than lectures	20.3	10.7
Don't care: learn about the same with activities than lectures	11.9	12.5
Kind of like activities: learn a little bit better with activities than lectures	37.3	55.4
Love activities: learn much better with activities than lectures	25.4	21.4

Table 21: Results for how much students felt they enjoyed the activities.

Preference	First Percentage	Second Percentage
Hate activities	6.8	3.6
Don't like activities	15.3	8.9
Don't care	1.7	19.6
Kind of like activities	37.3	37.5
Love activities	39.0	30.4

*Desired Group Size*

Students were asked if they would rather work in pairs or groups of four. The results are summarized in Table 22. The majority of students, 44.1% for the first survey and 62.5% for the second survey, preferred to work in pairs instead of groups of four.

Table 22: Results for student desired group size.

Preference	First Percentage	Second Percentage
Pairs	44.1	62.5
Groups of 4	33.9	12.5
I don't care	22.0	25.0



### Analyze Choice of Teaching Method Question

Students were given a survey at the end of each semester. Students in the activity classes were asked what would be their first, and second, choices for the teaching method if they took another statistics class. The results are in Table 23. The majority of students, 63.93%, chose having the teacher introduce a topic and then spending the remainder of the time on group work as their first choice. The majority second choice at 54.10% was teacher guided class activities. This suggests that students desired the teacher to have a larger part in the instruction for at least part of the period.

Table 23: Students' choices for preferred teaching method.

Preferred method	First choice percentage	Second choice percentage
Group work	3.28	13.11
Teacher introduce topic then group work	63.93	21.31
Teacher guided class activities	18.03	54.10
Traditional lecture	14.75	11.48

### Analyze Student Comments from Exit Slips

Students had the opportunity to make comments every day. This means that they could repeat the same comment on different days. In addition, students answered three questions, so students made similar comments under different question headings. The questions were:

- What did you find particularly helpful about today's lesson?
- What do you still feel confused about?
- Suggestions to improve today's lesson:

When analyzing students' comments, loose analysis was used (Miles, Huberman, & Saldana, 2013). In loose analysis, qualitative data are examined without preexisting categories. Instead, the student comments were read and then tagged with categories. Through an iterative process, the student comments were read and reread several times and the categories were refined. An attempt was

made to find the overall themes in the student comments. Those themes are discussed below. For each theme, supporting quotations from students are included. The grammar has been fixed for several comments, but otherwise the quotations are word for word. Since many comments are only fragments of sentences, the quotes have been organized into the categories helpful, confused, and suggestions to correspond to the three questions.

The questions were open ended, so student comments vary greatly. Students were not asked specifically about the themes below, so the number of students making a comment does not necessarily reflect the percentage of students that share that viewpoint. In addition, comments were not required and students tended to fill out the exit slips at the end of class. Many students either did not write anything, or wrote something very quickly in order to leave. Students tended to write more comments at the beginning of the semester, and then the number of comments decreased steadily throughout each semester.

### **Comments from Lecture Classes**

#### *Lecture*

Four people found the lecture easy to follow or understand. Eight students said the pacing of the lecture was good. However, four students commented that the pace was too fast. Of the comments on the pacing of the lecture, there was one day in which one student said the pace was too slow and another student said the pace was too fast. That demonstrates that students learn at different paces and it would be impossible to please all students. There were 25 comments from 16 students throughout the semester about specific statistical techniques that they were taught that were very useful, or a specific teaching tool that was effective.

#### *Helpful Quotes:*

Everything was easy to understand.

Everything was pretty straightforward.

You go at a good speed. I can follow you and I don't feel behind or lost.

We go through the material step by step instead of rushing through it.

You teach at a very good pace. You are very thorough. I actually like this class.

I like the chart for standard deviation and the way it organized my work.

I like how you showed us when to reject or fail to reject based on the [decision rule] picture. It was very helpful!

That you used the guilty vs not guilty and criminal system [when learning about hypothesis tests].

*Suggestion Quotes:*

We go very fast sometimes. I think a little more time to process information would help it sink in.

I felt like I was writing a lot and had a hard time understanding what we were doing due to the pace of the lecture. I feel I'm getting it though. Just takes time for me.

*Good Overall Lesson or Teacher*

There were 38 comments from 20 unique people about the overall lesson or the teacher. These were comments that did not mention specific qualities of the lesson or teacher. Students commented that the day's lesson was good. Some said that no improvement was needed or commented on the fact that the lesson was easy to understand. Some said that the teacher was good.

*Helpful Quotes:*

It was all marvelous.

Good job, no improvement necessary.

Clear and helpful lecture.

You're doing great! I really enjoy this class.

You are a great teacher. Seems like the material is simple right now.

*Guided Notes*

Students made 21 comments, although there were only 18 unique people that made the comments, about appreciating the guided notes. Students found that having the guided notes made following the lecture easier and they could pay more attention due to having to write less. They liked that their notes were the same, and the same format, as the notes that I wrote on at the document camera. They found it convenient to have all the notes pre-printed and bound. Although a few students complained throughout the semester about how big the book of guided notes was and that the book was not double sided.

*Helpful Quotes:*

The notebook is great and helpful for being able to keep up with lecture.

I find the guided notes very helpful. I use them to study and help with homework. It was worth the money to me.

The notes that are provided for us. So I don't have to hurry and write things down.

Guided notes make listening in lectures a lot easier.

### *Examples*

There were 91 comments from 30 unique people about the examples being helpful. Student comments included such ideas as there were many examples and that there were a variety of examples. Students also appreciated that the examples were clear and easy to understand, that the examples included real life applications, and that the examples were broken down into steps.

#### *Helpful Quotes:*

Examples were very well thought out and supported the material properly.

I appreciate real life examples like the salary question.

I could understand better when we used real life examples.

It was explained very clearly, and a lot of examples were given.

### *Questions and Teacher Explanations*

Seven students commented on the teacher's willingness to let students ask questions and then answer the question. In addition, one student commented on liking the fact that the teacher asked a lot of questions to the students. There were 16 comments from 13 students saying that the teacher explanations and answers to questions were helpful.

#### *Helpful Quotes:*

Thanks for letting us ask so many questions today.

Thank you for explaining simply how to get the expected value.

I like that you ask a lot of questions, even though we don't always answer them. It makes me think more.

I think you did a good job at explaining the material. It seems easy and makes sense.

### *Exam Reviews*

Thirty-one people made 49 comments throughout the semester about the exam reviews. They said that the exam reviews were very useful in preparing for exams and that it was helpful to go over them in class.

#### *Helpful Quotes:*

Going through the test review step by step in class.

Thank you for always taking the time to review. This has been very helpful.

I like reviewing a whole day before the exam. Very helpful.

### *How to Set Up a Problem*

There were 25 comments, from 8 unique students, about having difficulty setting up a problem and knowing which formula or procedure to use. Some students mentioned that they could apply the procedure correctly, they just did not know which procedure to use. However, there were six students that commented that my explanations helped them know which formula to use.

#### *Helpful Quotes:*

Going over multiple problems that are similar & writing all of the formulas on the board.

#### *Confused Quotes:*

Trying to distinguish the types of problems, and set them up, is difficult.

Knowing how to set up the equations and which one to use.

I'm good once I get the formula correct, but sometimes I don't get how to set the formula up.

### *Time for Students to Work on Their Own*

Six students commented that having time to try problems on their own was helpful. In addition, one student expressed the desire for more time to work on their own.

#### *Helpful Quotes:*

Getting to work that problem on our own and then go through it with you.

Doing problems on our own is super helpful! Thank you!

#### *Suggestion Quotes:*

I would love to have time to try and work some of the problems before the answer is given.

### *Extra Handouts*

During the spring semester, there were several extra handouts given to the students. For the fall semester, those extra handouts were incorporated into the guided notes. There were 14 comments from 9 students about the extra handouts being useful.

#### *Helpful Quotes:*

The explanations in words and practical examples of hypothesis and testing.

Thank you for the clarification with the worksheet.

Thanks for going over more of the random variables. It helped me understand better.

### *Miscellaneous*

- There were three comments when students wanted to know why we care and what we use it for.
- There were two comments on wanting more application and real life context.
- Two students mentioned that it was helpful that I wrote down everything I was saying.
- Four students found the applets helpful.

## **Comments from Activity Classes**

### *Activities*

There were 12 comments, from 8 unique students about liking the activities in general. The students liked being actively involved or moving around. Not only was class more engaging and enjoyable, but students thought that the material was easier to learn and relate to. There were 35 comments from 27 students comments about appreciating the hands on and interactive side of the activities. Nine students said they liked activities that helped them visualize the topic. In addition, there were 61 comments, from 35 unique students, on liking specific activities. For instance, there were ten comments about how useful the standard deviation flashcards were for visualizing and understanding standard deviation. Students particularly liked activities involving food.

*General Helpful Quotes:*

Activities are fun.

Doing activities in class make students more interested.

I really like the activities and group work. So far very helpful!

It was an active lesson, making it easier to pay attention.

The activities really help to learn the subject.

*Hands-On Helpful Quotes:*

I like hands on activities & class discussions!

Getting out of my chair.

Being hands on with the M&M sampling.

Hands on activities are very fun and helpful to learn the topic.

*Visual Helpful Quotes:*

Visual with M&Ms helped.

The computer program was good to visualize the data.

Being able to visually and actively understand these graphs and variables was helpful.

I liked the visual learning.

*Specific Activity Helpful Quotes:*

Putting the different graphs into categories—it's a memorable way to learn the different types.

The flash cards helped as a visual aid for better understanding.

The applet, graphing things out to understand standard deviation. The flashcards really solidified my understanding better than the applet.

I enjoyed the cookie experiment and am excited to see the results.

You used living examples!

I loved the human example for histograms.

I liked the interactive root beer activity in class and the walk through in class.

### *Good Overall Lesson*

There were 14 comments, from 11 unique students, that the lesson was good or needed no improvement. These were comments that did not mention specific qualities of the lesson. Students commented that the day's lesson was good. Some said that no improvement was needed or commented on the fact that the lesson was easy to understand. Interestingly, no one mentioned the instructor, probably because the class did not involve many teacher-centered activities.

#### *Helpful Quotes:*

It was all helpful.

It was great and I had lots of fun.

### *Activity Book or Packets*

There were 18 comments, from 12 unique students, that the activity book was helpful. They mentioned that the activity book or packets helped them learn the material. Six students commented that the examples in the activity book were helpful. One student particularly liked the variety of examples, and another mentioned the real life examples.

#### *Helpful Quotes:*

The pink book does teach well.

The book was helpful in learning how to solve probability.

Step by step directions & examples.

The handouts in class really help.

I feel the packet was very helpful with good examples to understand the material.

### *Amount of Practice*

Fifteen students made a total of 38 comments about the amount of practice being helpful. They said that the repetition was good. Two people said that they needed more examples, but one student said there were too many examples.

#### *Helpful Quotes:*

I do like the amount of practice.

Being able to do problems the whole time.



Lots of practice. I understand it.

*Suggestion Quotes:*

Fewer questions that are repeats.

I really enjoy the class worksheets. You could add a few more problems to it just to help ingrain it in us even more.

*Applets*

Thirty students made 55 positive comments about the applets. They mentioned that the applets were effective in helping them visualize and learn the concepts. Students also found the applets engaging and even fun. However, there were four complaints about the technological difficulty of getting the applets to work. One student felt that it was not worth getting the computers out for the day with only a five minute applet.

**Quotes:**

*Helpful:*

The applet helped to see how it works.

The applets were like games. They were fun and I learned a lot.

The applets are awesome for visuals. They help a ton.

I love using applets and these felt like games.

Playing with the graphs to see the changes in SD.

*Suggestion:*

A lot of the applets were unnecessary to have pairs or groups break out a laptop just to see what you could have easily shown us quickly on the smart board.

Applets sometimes don't work

I hate how slow the applets are.

The first 5 minutes with the computers (ours didn't work); our computer didn't let us log in. By the time we got another, it was time to move on.

*Teacher-Guided Activities*

There were 144 comments from 39 unique people that expressed an appreciation or desire for teacher-guided activities. Some students expressed appreciation on the days that we had teacher-

guided activities, while other days, they suggested that we have more teacher-guided activities. Thirty-eight of the total 144 comments were a desire for more teacher-guided activities. The rest for the comments were from students mentioning that the teacher-guided activities were helpful that day.

The comments were from days with various types of teacher-guided activities. Some days entailed the teacher lecturing, some days the teacher-guided students through discovery-based lessons, sometimes the teacher gave a recap at the beginning or end of class, and other days the entire class was involved in an activity to gather and analyze information. Some days consisted of teacher-guided activities for the entire period, but most of the comments were from days when the teacher-guided activities were at the beginning of class and then students worked in groups. In fact, 26 of the 144 comments specifically mentioned that it was helpful to have the teacher introduce the topic, and then have the students work in groups.

*Helpful Quotes:*

I liked that we discussed this as a class and not just on our own.

Working through problems as a class is super helpful, mostly because I can see where the numbers we use come from.

I like when you lecture. It helps to know if I am doing it right. I really benefit from learning new material by you teaching it.

Going through the packet with us and explaining it really helped me understand every concept.

I liked doing some problems together. I understand the “why” much better and learned better.

I really liked learning as a class. Teaching ourselves with group work and packets is a lot harder for me than learning through lecture. I feel like I am more confident.

You walked us through the steps and then turned us loose to work in groups.

Very helpful that you went over the packet first and then let us go do it in groups.

*Suggestion Quotes:*

Go over one example before hand, I was so confused!

I learn better if you show us how to do a problem before I do it.

Go over the new material first before throwing us in.

Maybe explain and teach more than us just trying to figure it out.

Maybe teach a bit before-just to review and teach what the big picture is.

Wish you would help more and teach us, not just hand us a handout and make us do it ourselves.

### *Group Work*

There were 98 positive comments about group work from 47 unique students. Students mentioned that group work was helpful because they could discuss ideas with their groups. They could also ask questions in their groups. However, there were twenty-two comments about the difficulty of working with groups. Students mentioned that it was difficult to work with groups or partners if they worked at different speeds. In addition, when students talked instead of working, it was distracting to themselves and other students. Some of the students commented that it would be easier to work in pairs than groups of four; other students wanted to work on their own instead of with a partner. One student mentioned that he had difficulty because he was shy in groups.

#### *Helpful Quotes:*

It was nice having time to sit down and just work with others you could ask questions.

Working in our groups to bounce ideas around.

I like working on things on my own with my partner. It makes me really pay attention to what we are learning.

Working in pairs with a key is very helpful and effective.

Working with partners allowed me to ask questions and have another perspective.

Group work helped us to learn the material for ourselves.

I liked the pair/group work. It helped to have more than one brain thinking about one problem.

#### *Suggestion Quotes:*

Working in groups is hard when they don't stay together.

Studying in a group sometimes hindered my learning because some people were quicker at understanding a certain concept than others.

Having to work in groups pressured me into having to rush through work as fast as the others were without actually understanding it first.

Sometimes I feel rushed to understand as quickly as my partner. What if we worked independently, but collaborated with partners when we need help or to collaborate?

Some of the group work hindered my learning because not everyone participated, or none of us knew what to do, or it was awkward to get into groups.

The group work was a two edge sword. It really helped me a lot, but some days it was a lot easier to get distracted. With the groups, on the days I didn't feel like doing math it was easy to start talking about something else instead of doing my work.

Allow people to work on their own but with a partner they can ask for help. I really prefer going at my own pace.

### *Teacher Helping Groups*

Nine students made 14 comments in which they deemed it helpful when the instructor went around answering questions during group work time. They thought that it was nice to be able to try the problems on their own, and with their groups, but to still have the chance to ask the instructor questions.

#### *Helpful Quotes:*

You were walking around talking to each group.

It was helpful to work with a partner and have the teacher go around and help.

I enjoyed having time to really think things through on my own and with the group, but still be able to have you be there to answer questions.

### *Variety of Teaching Methods*

There were 11 comments from 7 students about liking the variety of teaching methods. Having hands-on activities and group work appeared to be a nice break from lecture classes. Students seemed to enjoy the days in which there were class activities and group activities.

#### *Helpful Quotes:*

Class discussion & trying it for ourselves.

I loved the variety and using different ways to learn.

I like the different activities mixed with your lecture.

I like both lecture and individual time.

I like days when you start us, we do a little on our own, and then we synthesize what we did. I also like changing the pattern of class frequently (i.e. groups of 4, class, pairs, etc...).

Working on activities, talking with others, and still having instruction.

I LOVE having class and individual time as well as group time. It helped me learn.

I really liked multiple ways we learned. The computer, excel, worksheets, etc. . .

### *Exam Reviews*

There were 64 comments from 21 students about the exam reviews being helpful. Five of those 64 comments specifically mentioned the appreciation of class time for the review. However, three students wanted me to lecture for the reviews.

#### *Helpful Quotes:*

I love that we get class time to work on the review. That way we get the opportunity to ask questions.

The review packet was very helpful with the formulas and concepts.

#### *Suggestion Quotes:*

Maybe the day before the test, go through with us and explain, so we know we're doing it right.

Do example problems with each formula under your direction.

### *How to Set Up a Problem*

There were 15 comments from 12 students about having difficulty setting up a problem and determining which procedure or formula to use.

#### *Confused Quotes:*

When to use equations, especially with word problems.

Just knowing how to set up each problem from the question.

Remembering which formulas to use and when.

### *Miscellaneous*

- Four students liked having the answer key available as they worked in groups. However, one student said that the answer key shouldn't be available because it was too easy to just follow the key and feel over-confident.
- Four students said the pace was too fast when I did lecture.

- Two students said they weren't sure how everything fits together.
- There were three comments when students wanted to know why we care and what we use it for.
- There were four comments about a specific teaching technique being helpful.
- Three students said the teacher explanations were helpful.

### **Summary**

For the lecture class, students seemed to find the lecture effective. They appreciated having the guided notes as the notes made it easier for students to follow the lesson. The variety and multitude of examples was judged helpful by the students. Students found the teacher's explanations helpful and liked being able to ask questions. They liked being able to try a problem on their own and then go over the problem as a class. The biggest difficulty seemed to be in knowing how to set up a problem.

The activity class seemed to enjoy the activities. They liked the variety of teaching methods and being more actively involved. Students liked activities that helped them visualize the topic. The applets were deemed helpful by many students. While the majority of students found group work helpful, especially when the teacher was available for questions, they did not like learning primarily by group work. Many students wanted to have more teacher-centered activities; students wanted to have the teacher at least introduce the material. The instructor got the impression that students were okay with working in groups, as long as they had some teacher-centered activity each day. The main difficulties with group work were when students worked at different speeds or when students talked in their groups instead of working.

## DISCUSSION

The previous chapters detailed the literature on teaching statistics with active learning methods, the methods of the research study, and the analysis of the data. This chapter addresses the limitations of the study and future research possibilities. Finally, the conclusions are discussed.

### Limitations and Future Research Possibilities

One of the limitations of the course was the lack of explicit development of necessary sociomathematical norms such as requiring students to justify their answers, having students ask other students to clarify their answers, and having students evaluate other students arguments (Cobb, 1999; Cobb et al., 2003). In particular, most of the class time consisted of students working in groups. No instruction was given to students about how to work effectively in groups. The group work was not graded in any way. In addition, students were also required to brainstorm both in their groups and as a class to discover concepts. An effort should have been made to develop students' abilities to brainstorm and supply ideas to a discussion at the beginning of each semester. Perhaps if the class was taught again, but sociomathematical norms were developed first, the activity-based methods would be more effective.

One of the goals of the study was to use activity-based methods within the existing curriculum. The pedagogy was changed, but not the curriculum. If the curriculum of the statistics course was more flexible, more time could be spent gathering data. Students could choose their own data sets and collect more data. More projects could be incorporated into the course. In addition, the focus of the class could be changed from understanding and being able to apply the major statistical concepts and procedures to developing statistical literacy and reasoning. After all, one statistics course is not enough to train students to be effective analyzers of data. The first course should focus on developing statistical reasoning and then future classes can develop their technical statistical abilities (Garfield, 2002).

The Attitudes Towards Statistics (ATS) (Wise, 1985) survey was used for this study. However, the Survey of Attitudes Toward Statistics (SATS) (Schau, Stevens, Dauphinee, & Vecchio, 1995) scale might be a better choice for a future study. The SATS survey has a specific pretest and

posttest version. This would allow a more precise measurement of the change in student attitudes towards statistics.

Students in both the activity and lecture sections completed the same homework and exam reviews. This resulted in students from both classes having the same practice on the concepts. In particular, having the same exam reviews and spending one or two class periods reviewing for each class, could have leveled the playing field. Because students often relearn, or refresh their memory during an exam review, perhaps true differences between the teaching methods were masked by students having so much practice that was identical between the teaching methods. In addition, because students could learn, and fill in any missing gaps in their understanding while doing homework and reviews, it is difficult to link any results to specific activities. For a future experiment, students could be given short quizzes at the end of each activity to measure understanding. This would help distinguish the effects of the two teaching methods and also enable the evaluation of individual activities.

The students' comments from the exit slips were analyzed for major themes. However, because the questions were open ended, students wrote comments on whatever thought was foremost in their minds that day. So even though only four students in the lecture classes said that applets were helpful, perhaps 80% of students found the applets helpful. Or maybe only 20% of students found the applets helpful. In order to truly judge how effective various teaching techniques were, students should have been asked to rate the effectiveness of the specific techniques. For example, students in the activity class could have rated the effectiveness of group work, teacher guided activities, applets, hands-on activities, etc. This would have led to more accurate and useful data, especially when trying to compare the lecture and activity-based teaching methods.

An effort was made to make the lecture classes as effective as possible. Pre-printed guided notes were available which let students spend less time writing and more time trying to understand the material. The instructor tried to explain the reasoning and help students understand concepts as much as possible, instead of just focusing on procedures. The instructor attempted to teach at a slow pace, and to make the explanations as clear as possible. Students were guided through discovering some concepts. Perhaps there was no apparent difference between the two teaching methods because the instructor was a particularly good lecturer, not a traditional lecturer.

There are so many potential confounding variables in any study dealing with students. Students' scores in the statistics class could have been affected by their level of natural aptitude towards



mathematics, the amount of time they studied, how much they slept the night before the test, if they were sick during the test, or how much homework they completed. Their scores could also be determined by how much effort they put into the class. Did students in the activity classes put their full effort into learning the material, or did they just copy answers from the keys? Did students in both classes try the homework problems and review problems before looking at the keys? Perhaps a student would have performed better with a different partner or group. Without being able to control for all the potential confounding variables, it is difficult to truly measure the effect of the two teaching methods.

Perhaps the biggest limitation of this research study was the small sample size. Only 146 students were in the study; 87 students were in the lecture classes and 59 students in the activity classes. More importantly, there were only four sections taught for the study. When exam scores were compared for the lecture and activity sections during just the spring semester, students in the activity class did score significantly higher than students in the lecture class. However, when the results were compared during just the fall semester, students in the activity class did worse, but not significantly so, on the exams than students in the lecture class. The instructor observed differences between student characteristics between the two activity sections. Overall, the students in the activity section in the spring appeared to be more alert, more willing to brainstorm, and more willing to work in groups than the students in the fall activity section. This difference was apparent from the beginning of the fall semester. (This difference was based on the instructor's informal observations, not any specific formal data.) At this point, with only four sections, we can not separate the effect of the two teaching methods from the natural variation that occurs from section to section in any course. In addition, some researchers would argue that since the experimental unit can be considered to be the entire section, the sample size of this study was only four.

This study needs to be repeated with many more sections. Because class characteristics can vary based on which students are enrolled, more sections are needed to truly isolate the effect of the teaching method. In addition, the sections could be taught by different instructors. If each instructor taught several sections, it would be easier to isolate the effect of the teaching method from any instructor effect. Until this study is repeated with more sections and instructors, any research on comparing the teaching methods should be considered a pilot study, rather than a confirmatory study.

Finally, this study measured the outcomes of lecture and activity-based teaching methods as defined by one instructor. Other instructors could lecture differently with different results. In addition, there are many ways that active learning techniques could be implemented in a statistics course. Perhaps another instructor's version of activity-based teaching methods would prove to be more effective.

The limitations discussed above limit the ability to generalize the results of this study. The small number of sections is one of the biggest obstacles to being able to generalize the results. The other main obstacle is the fact that the instructor might have applied lecture and activity-based methods different than other instructors.

### **Final Conclusions**

This study did not demonstrate that activity-based teaching methods are more effective than traditional lecture for teaching introductory statistics. Students in the activity classes did not do significantly better on the exams; the activity classes did not score significantly higher on the procedural or conceptual questions, or for any specific topics or statistical questions. Neither teaching method resulted in higher, more positive attitudes towards statistics.

This study did show that it was possible to teach an entire course with activity-based methods and minimal lecture. In addition, it was possible to teach the activity class within the bounds of the existing curriculum and textbook. Sufficient activities to teach the entire course were developed. However, developing the activities was more time consuming than preparing for lectures, as well as more costly. Even the second semester, after the materials were developed for both the activity and lecture classes, the activity class still required more preparation time.

One of the biggest contributions of this study was the development of the instructional materials. Materials were designed and revised for both lecture and activity formats. These materials could be used by instructors in any algebra-based, introductory statistics course. In particular, because developing activities is so time intensive, the materials developed for the activity classes will be a valuable resource for statistics instructors. In this study, the activity materials were at least as effective as the traditional lecture materials.

Students in the activity classes did appreciate certain aspects of the teaching method. They enjoyed the variety of the teaching methods to relieve monotony. The hands on activities appeared

to be helpful in learning. Some students appreciated the chance to work in groups, especially when the teacher was available for asking questions.

However, the majority of students in the activity classes wanted more teacher-centered activities. Some students just wanted more lecture, but many students wanted to have the teacher introduce the topic each day before students worked in groups. The students' second choice for a teaching method was for the teacher to lead the entire class through the activities.

I believe that students would be much more accommodating to having an activity-based statistics course if more teacher-centered methods were used. Just like students in the lecture class wanted more variety, students in the activity class got bored of working in groups. They wanted the group work to be interspersed with teacher-centered activities.

Because students are still unused to activity-based courses, initial reluctance can be expected when implementing activity-based methods. Students used to lecture can feel that they are not being taught when the teacher steps out of the limelight and allows students to drive their own learning. Until active learning is actually practiced on a widespread scale, instructors might need to slowly introduce activity-based methods. The other option is to specifically develop the necessary sociomathematical norms and ensure that students realize that the new active learning methods require them to take responsibility for their own learning. Students would need to understand that the new methods were being used because they are effective, not because the teacher does not want to teach.

Finally, individual student characteristics can vary greatly. What is effective for one student, might be detrimental for another student. Some students might perform better in lecture situations, whereas other students could excel with activity-based methods. With this in mind, it can be difficult to define the "best" teaching method. Perhaps, the "best" teaching method would include a variety of techniques. Courses could use both activity and lecture-based methods. If the activity and lecture-based methods were interspersed on a regular basis, the needs of more students could be met. I now imagine an ideal class would consist of 30-50% teacher-centered activities, but not necessarily lecture, and the remainder of the time would be spent in group work.

## REFERENCES

- Abdi, H. (2010). Holm's sequential bonferroni procedure. In *Encyclopedia of Research Design*, edited by N. J. Salkind. Thousand Oaks, California: Sage.
- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., et al. (2010). *Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report*. Alexandria, VA: American Statistical Association.
- Bakker, A. & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 147–168.
- Batanero, C., Godino, J. D., & Roa, R. (2004a). Training teachers to teach probability. *Journal of Statistics Education*, 12(1).
- Batanero, C., Tauber, L. M., & Sánchez, V. (2004b). Students' reasoning about the normal distribution. In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 257–276.
- Bonwell, C. C. (1995). *Active Learning: Creating Excitement in the Classroom*. Center for Teaching and Learning, St. Louis College of Pharmacy, Missouri.
- Boone, H. N. & Boone, D. A. (2012). Analyzing likert data. *Journal of Extension*, 50(2).
- Brandsma, J. A. (2000). *Data Collection and Analysis: Examining Community College Students' Understanding of Elementary Statistics through Laboratory Activities*. (Doctoral dissertation). North Carolina State University.
- Brandsma, J. A. (2001). Recent dissertations. *Statistical Education Research Newsletter*, 2(3), 9.
- Burgess, T. A. (2001). Assessing the statistics knowledge of preservice teachers. In *24th Annual Conference of the Mathematics Education Research Group of Australasia*, edited by J. Bobis, B. Perry, & M. Mitchelmore. Turramurra, NSW, pages 114–121.
- Cangelosi, J. (2003). *Teaching Mathematics in Secondary and Middle School: An Interactive Approach*. 3rd ed. Upper Saddle River, NJ: Pearson Education.

- Carlson, K. A. & Winkist, J. R. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education*, 19(1).
- Chance, B. (2012). Email interview.
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 295–323.
- Chappell, K. K. & Killpatrick, K. (2007). Effects of concept-based instruction on students' conceptual understanding and procedural knowledge of calculus. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 13(1), 13–37.
- Chickering, A. W. & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin*, 39(7), 3–7.
- Cobb, G. W. (1992). Teaching statistics. In *Heading the Call for Change: Suggestions for Curricular Action*, edited by L. A. Steen. No. 22 in MAA Notes, Washington DC: Mathematical Association of America, pages 3–43.
- Cobb, G. W. & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801–823.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Cobb, P. & McClain, K. (2004). Principles of instructional design for supporting the development of students' statistical reasoning. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 375–395.
- Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, 21, 1–78.
- Cuseo, J. B. (1997). Guidelines for group work. *Cooperative Learning and College Teaching*, 7(3), 11–16.

- delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 79–95.
- delMas, R. C. (2006). ARTIST topic scales. Retrieved from <https://apps3.cehd.umn.edu/artist/index.html>.
- delMas, R. C. (2012). Email interview.
- delMas, R. C., Garfield, J., & Chance, B. L. (1999). A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3).
- delMas, R. C., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.
- Edelson, D. C. (2002). Design research: What we learn when we engage in design. *The Journal of the Learning Sciences*, 11(1), 105–121.
- Federer, W. T. & King, F. (2007). *Variations on Split Plot and Split Block Experiment Designs*. Hoboken, New Jersey: John Wiley and Sons.
- Festing, M. (2012). Experimental unit. Retrieved from [http://isogenic.info/html/6\\_\\_experimental\\_unit.html](http://isogenic.info/html/6__experimental_unit.html).
- Finney, S. J. & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, 28, 161–186.
- Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*, edited by D. Ben-Zvi & J. Garfield. Dordrecht: Kluwer Academic Publishers, pages 47–78.
- Garfield, J. (1993). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, 1(1).
- Garfield, J. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education*, 2(1).

- Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25–34.
- Garfield, J. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, 10(2).
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22–38.
- Garfield, J. & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396.
- Garfield, J. & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning*, 2, 99–125.
- Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, 5(3).
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64–74.
- Hall, S. & Vance, E. A. (2010). Improving self-efficacy in statistics: Role of self-explanation and feedback. *Journal of Statistics Education*, 18(3).
- Keeler, C. M. & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, 3(2).
- Keeley, S. M., Shemberg, K. M., Cowell, B. S., & Zinnbauer, B. J. (1995). Coping with student resistance to critical thinking: What the psychotherapy literature can tell us. *College Teaching*, 43(4), 140–145.
- Klionsky, D. J. (1998). A cooperative learning approach to teaching introductory biology. *Journal of College Science Teaching*, 27(5), 334–338.
- Lane-Getaz, S. J. (2007a). *Development and Validation of a Research-based Assessment: Reasoning about P-values and Statistical Significance*. (Doctoral dissertation). University of Minnesota.
- Lane-Getaz, S. J. (2007b). Toward the development and validation of the reasoning about p-values and statistical significance scale. In *Proceedings of the International Statistical Institute (ISI)*

- / *International Association of Statistical Education (IASE) Satellite Conference on Assessing Student Learning in Statistics*, edited by B. Phillips & L. Weldon. Voorburg, The Netherlands: International Statistical Institute.
- Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BioMed Central Neuroscience*, 11(5).
- Lehmann, E. L. & Romano, J. P. (2012). Generalizations of the familywise error rate. In *Selected Works of E. L. Lehmann*, edited by J. Rojo. Selected Works in Probability and Statistics, Springer, pages 719–735.
- Macdonald, H. & Teed, R. (2012). Interactive lectures. Retrieved from <http://serc.carleton.edu/sp/library/interactive/index.html>.
- Meletiou-Mavrotheris, M. (2003). On the formalist view of mathematics: Impact on statistics instruction and learning. In *Proceedings of Third European Conference in Mathematics Education*, edited by A. Mariotti. Bellaria, Italy: European Research in Mathematics Education Society.
- Meletiou-Mavrotheris, M., Lee, C., & Fouladi, R. T. (2007). Introductory statistics, college student attitudes and knowledge - A qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology*, 38(1), 65–83.
- Meyers, S. A. (1997). Increasing student participation and productivity in small-group activities for psychology classes. *Teaching of Psychology*, 24(2), 105–115.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2013). *Qualitative Data Analysis: A Methods Sourcebook*. 3rd ed. Thousand Oaks, CA: Sage.
- Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65, 123–165.
- Moore, D. S. (2005). Preparing graduate students to teach statistics: Introduction. *The American Statistician*, 59(1), 1–3.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Introduction to the Practice of Statistics*. Seventh ed. New York: W.H. Freeman and Company.



- Mosteller, F. (1989). The "muddiest point in the lecture" as a feedback device. *Harvard-Danforth Teaching Center Newsletter*, 10–21.
- Nicholson, J. & Darnton, C. (2003). Mathematics teachers teaching statistics: What are the challenges for the classroom teacher? In *Proceedings of the 54th Session International Statistics Institute*. Berlin: International Statistical Institute.
- Park, J. (2012). *Developing and Validating an Instrument to Measure College Students' Inferential Reasoning in Statistics: An Argument-Based Approach to Validation*. (Doctoral dissertation). University of Minnesota.
- Pfaff, T. J. & Weinberg, A. (2009). Do hands-on activities increase student understanding?: A case study. *Journal of Statistics Education*, 17(3).
- Pfannkuch, M. (2008). Training teachers to develop statistical thinking. In *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics: Challenges for Teaching and Teacher Education*, edited by C. Batanero, G. Burrill, C. Reading, & A. Rossman. Monterrey, Mexico: Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.
- Raftery, J. A., Abelly, M. L., & Braselton, J. P. (2002). Multiple comparison methods for means. *SIAM Review*, 44(2), 259–278.
- Reder, S., Cummings, P., & Quan, L. (2006). Comparison of three instructional methods for teaching cardiopulmonary resuscitation and use of an automatic external defibrillator to high school students. *Resuscitation*, 69(3), 443–453.
- Rosenthal, J. S. (1995). Active learning strategies in advanced mathematics classes. *Studies in Higher Education*, 20(2), 223–228.
- Rossman, A. J. (1995). Workshop statistics: Using technology to promote learning by self-discovery. In *Research on the Role of Technology in Teaching and Learning Statistics*, edited by J. Garfield & G. Burrill. International Statistical Institute.
- Rossman, A. J. (2012). Email interview.
- Rossman, A. J. & Chance, B. L. (2012). *Workshop Statistics: Discovery with Data*. 4th ed. Wiley.

- Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education*, 37, 43–50.
- Rubin, L. & Hebert, C. (1998). Model for active learning: Collaborative peer teaching. *College Teaching*, 46(1), 26–30.
- Ruhl, K. L., Hughes, C. A., & Schloss, P. J. (1987). Using the pause procedure to enhance lecture recall. *Teacher Education and Special Education*, 10(1), 14–18.
- Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55(5), 868–875.
- Scheaffer, R. L., Watkins, A., Gnanadesikan, M., & Witmer, J. A. (1996). *Overview of Activity-Based Statistics: Student Guide*. New York: Springer-Verlag.
- Schneider, K. (2008). Two applets for teaching hypothesis testing. *Journal of Statistics Education*, 16(3).
- Stanton, S. (1997). *Effectiveness Comparison for Two Methods of Teaching Process Simulation to Industrial Technology Students*. (Doctoral dissertation). Purdue University.
- Stoker, H. W., King, F., & Foster, B. F. (1981). The class as an experimental unit. *Evaluation Review*, 5(6), 851–859.
- Sundre, D. L. (2003). Assessment of quantitative reasoning to enhance educational quality. In *Annual American Educational Research Association Meeting*. Chicago, Illinois.
- Tempelaar, D. (2004). Statistical reasoning assessment: an analysis of the SRA instrument. In *ARTIST Roundtable Conference on Assessment in Statistics*. Lawrence University.
- Thomas, C. N. (2008). *A Comparison of Three Instructional Methods—Teacher-Directed Lecture and Text-Based Instruction, Analog Video Instruction, and Multimedia Anchored Instruction—on the Knowledge, Beliefs, and Skills of Preservice Teachers*. (Doctoral dissertation). University of Texas.
- Thomas, J. (1972). The variation of memory with time for information appearing during a lecture. *Studies in Adult Education*, 4, 57–62.

- Tintle, N., Topliff, K., Vanderstoep, J., Holmes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal*, 11(1), 21–40.
- Tintle, N., Vanderstoep, J., Holmes, V.-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education*, 19(1).
- Towns, M. H. (1998). How do I get my students to work together? Getting cooperative learning started. *Journal of Chemical Education*, 75, 67–69.
- Utah State University (2012). General catalog. Retrieved from <http://catalog.usu.edu/content.php?catoid=4&navoid=678>.
- Van Teijlingen, E. R. & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, 35.
- Vithal, R. (2002). Learning to teach statistics through project work. In *Proceedings of the Sixth International Conference on Teaching Statistics*, edited by B. Phillips. Cape Town, South Africa: International Statistical Institute and International Association for Statistical Education.
- Walsh, C. A. & Hewson, J. (2012). A comparison of two methods of teaching research to master of social work students. *International Journal of Higher Education*, 1(1), 14–21.
- Watkins, A. E., Scheaffer, R. L., & Cobb, G. W. (2004). *Statistics in Action: Understanding a World of Data*. Emeryville, CA: Key Curriculum Press.
- Weltman, D. & Whiteside, M. (2010). Comparing the effectiveness of traditional and active learning methods in business statistics: Convergence to the mean. *Journal of Statistics Education*, 18(1).
- Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, 45, 401–405.
- Zhang, J., Bertness, C., & Pan, K. (1998). Two ways to teach an elementary statistics course: The workshop approach vs. the traditional approach. Which one is the winner? In *Abstract from the 5th International Conference on Teaching Statistics*. International Association for Statistical Education.

## APPENDICES

**Appendix A: Letter of Information & IRB Approval**



## Institutional Review Board

USU Assurance: FWA#00003308



### Exemption #1

## Certificate of Exemption

**FROM:**

Melanie Domenech Rodriguez, IRB Chair

True M. Rubal, IRB Administrator

**To:** Kady Schneiter, Jennifer Loveland

**Date:** October 24, 2012

**Protocol #:** 4692

**Title:** Traditional Lecture Versus Gaise Statistics Courses: A Comparison Of Student Learning Outcomes

The Institutional Review Board has determined that the above-referenced study is exempt from review under federal guidelines 45 CFR Part 46.101(b) category #1:

Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

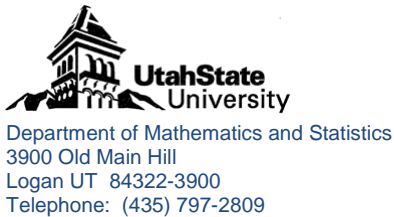
This exemption is valid for three years from the date of this correspondence, after which the study will be closed. If the research will extend beyond three years, it is your responsibility as the Principal Investigator to notify the IRB before the study's expiration date and submit a new application to continue the research. Research activities that continue beyond the expiration date without new certification of exempt status will be in violation of those federal guidelines which permit the exempt status.

As part of the IRB's quality assurance procedures, this research may be randomly selected for continuing review during the three year period of exemption. If so, you will receive a request for completion of a Protocol Status Report during the month of the anniversary date of this certification.

In all cases, it is your responsibility to notify the IRB prior to making any changes to the study by submitting an Amendment/Modification request. This will document whether or not the study still meets the requirements for exempt status under federal regulations.

Upon receipt of this memo, you may begin your research. If you have questions, please call the IRB office at (435) 797-1821 or email to [irb@usu.edu](mailto:irb@usu.edu).

The IRB wishes you success with your research.



Page 1 of 2

## LETTER OF INFORMATION

### *Traditional versus GAISE Statistics Courses: A Comparison of Student Learning Outcomes*

**Introduction/ Purpose** Dr. Kady Schneider and Jennifer Loveland in the Department of Mathematics and Statistics at Utah State University are conducting a research study to find out more about teaching lecture based courses versus activity based courses. You have been asked to take part because you are a student registered for STAT 2000, an introductory statistics course. There will be approximately 80 total participants in this research.

**Procedures** If you agree to be in this research study, you will be expected to complete all the required coursework including homework, projects, quizzes, surveys, exams, pre-test, and post-test. The exams, pre-test, and post-test will be administered in class. The surveys will be administered in Canvas. During class, you may be asked to comment on how any activities could be improved.

If you decide to do so, you may also be asked to participate in an interview to describe what aspects of the class helped you the most. This interview will be conducted only one time and will last approximately 15 minutes. Collected data will be analyzed to help improve the teaching of introductory statistics courses.

If you choose not to participate in the research study, you will still be expected to complete all the required coursework including homework, projects, quizzes, surveys, exams, pre-test, and post-test. However, your data will not be used for analysis.

**Risks** Participation in this research study may involve some added risks or discomforts. There is a small risk of loss of confidentiality but we will take steps to reduce this risk.

**Benefits** Information gained through this study will benefit future introductory statistics students by improving the teaching of introductory statistics courses and improving available materials. There will be no direct benefits to you as a participant.

**Explanation & offer to answer questions** Jennifer Loveland has explained this research study to you and answered your questions. If you have other questions or research-related problems, you may reach her at [jennifer.ellsworth@aggiemail.usu.edu](mailto:jennifer.ellsworth@aggiemail.usu.edu).

**Voluntary nature of participation and right to withdraw without consequence** Participation in research is entirely voluntary. You may refuse to participate or withdraw at any time without consequence or loss of benefits. You may be withdrawn from this study without your consent by the investigator. **If you do not want to participate, please complete the section at the end of this document and return it to Jennifer Loveland.**

**Confidentiality** Research records will be kept confidential, consistent with federal and state regulations. Only the investigator will have access to the data which will be kept in a locked file cabinet or on a password protected computer in a locked room. To protect your privacy, personal, identifiable information will be removed from study documents and replaced with a study identifier. Identifying



Page 2 of 2

### LETTER OF INFORMATION

#### *Traditional versus GAISE Statistics Courses: A Comparison of Student Learning Outcomes*

information will be stored separately from data. The information will be kept for two years after the end of the study.

**IRB Approval Statement** The Institutional Review Board for the protection of human participants at Utah State University has approved this research study. If you have any questions or concerns about your rights or a research-related injury and would like to contact someone other than the research team, you may contact the IRB Administrator at (435) 797-0567 or email [irb@usu.edu](mailto:irb@usu.edu) to obtain information or to offer input.

**Investigator Statement** "I certify that the research study has been explained to the individual, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered."

#### **Signature of Researcher(s)**

\_\_\_\_\_  
*Dr. Kady Schneider*  
Principal Investigator  
[kady.schneider@usu.edu](mailto:kady.schneider@usu.edu)

\_\_\_\_\_  
*Jennifer Loveland*  
Student Researcher  
[jennifer.ellsworth@aggiemail.usu.edu](mailto:jennifer.ellsworth@aggiemail.usu.edu)

#### **Decline Participation**

I do not wish to participate in this research study. I understand that I am expected to complete the course requirements, however you will not use my class materials as part of the data analysis.

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
Name (Please Print)

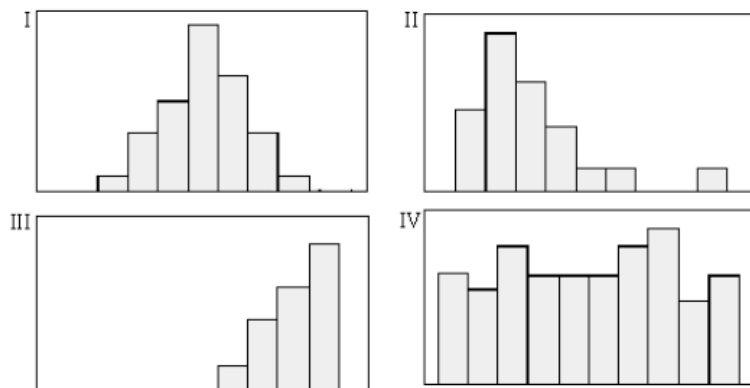


## **Appendix B: Posttest**

## Final Exam

**Problems 1 to 2 refer to the following situation:**

Four histograms are displayed below. For each item, match the description to the appropriate histogram.



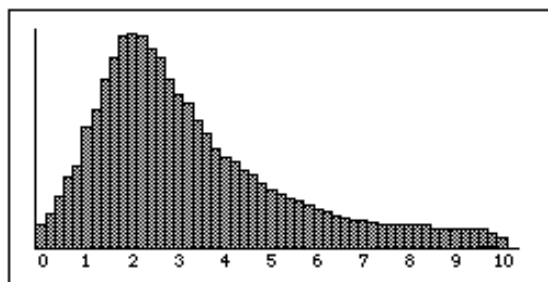
**Problem 1.** A distribution for a set of quiz scores where the quiz was very easy is represented by:

- Histogram I.
- Histogram II.
- Histogram III.
- Histogram IV.

**Problem 2.** A distribution for the last digit of phone numbers sampled from a phone book (i.e., for the phone number 968-9667, the last digit, 7, would be selected) is represented by:

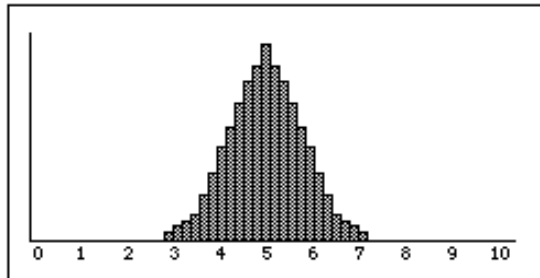
- Histogram I.
- Histogram II.
- Histogram III.
- Histogram IV.

**Problem 3.** Select the description that best represents the shape of the following distribution.



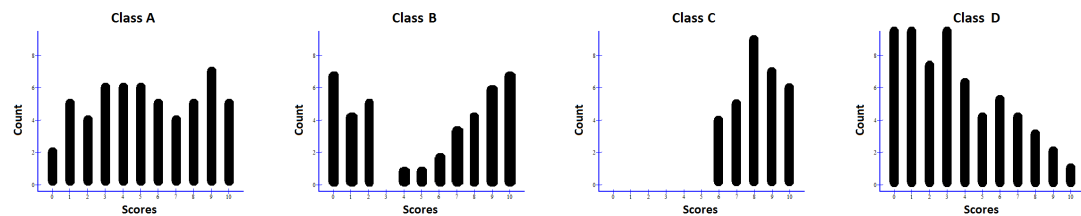
- Left skewed
- Right skewed
- Normal
- Bimodal

**Problem 4.** Select the description that best represents the shape of the following distribution.



- a. Left skewed
- b. Right skewed
- c. Normal
- d. Bimodal

**Problem 5.** Four histograms are presented below. Each histogram displays test scores for one of four different statistics classes. Which of the classes would you expect to have the lowest standard deviation? (They are all drawn on the same scale.)



- a. Class A
- b. Class B
- c. Class C
- d. Class D

**Problem 6.** A researcher is studying the relationship between a vitamin supplement and cholesterol level. What type of study needs to be done in order to establish that the amount of vitamin supplement *causes* a change in cholesterol level?

- a. Observational study
- b. Controlled experiment
- c. Time Series study
- d. Survey

**Problem 7.** Imagine you have a barrel that contains thousands of candies with several different colors. We know that the manufacturer produces 35% yellow candies. Five students each take a random sample of 20 candies, one at a time, and record the percentage of yellow candies in their sample. Which sequence below is the most plausible for the percent of yellow candies obtained in these five samples?

- a. 30%, 35%, 15%, 40%, 50%.
- b. 35%, 35%, 35%, 35%, 35%.
- c. 5%, 60%, 10%, 50%, 95%.
- d. Any of the above.

**Problem 8.** Suppose you read on the back of a lottery ticket that the chances of winning a prize are 1 out of 10. Select the best interpretation.

- You will win at least once out of the next 10 times you buy a ticket.
- You will win exactly once out of the next 10 times you buy a ticket.
- You might win once out of the next 10 times but it is not for sure.

**Problem 9.** The following message is printed on a bottle of prescription medication:

WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.

Which of the following is the best interpretation of this warning?

- You should use the medication, there is hardly a chance of getting a rash.
- For application to the skin, apply only 15% of the recommended dose.
- If a rash develops, it will probably involve only 15% of the skin.
- About 15 of 100 people who use this medication develop a rash.

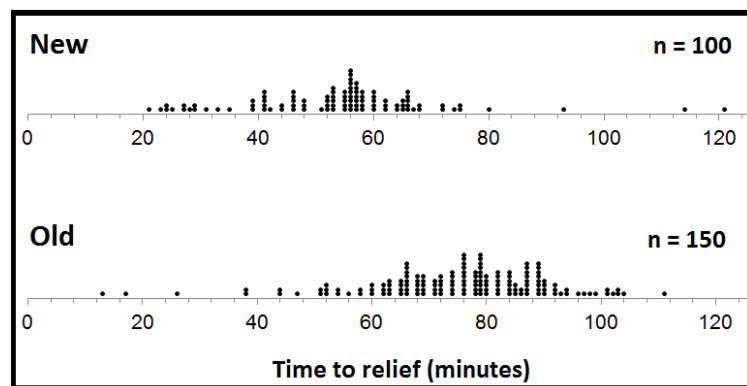
**Problem 10.** Bob and Jacob each bought one ticket for the same lottery each week for the past year. Bob has not won a single prize yet. Jacob just won a \$20 prize last week. Who is more likely to win a prize this coming week if they each buy only one ticket?

- Bob.
- Jacob.
- They have an equal chance of winning.

**Problem 11.** Colin is flipping a fair coin. Heads has just come up 5 times in a row! The chance of getting heads on the next throw is

- less than the chance of getting tails since we are due for a tails.
- equal to the chance of getting tails since the flips are independent and the coin is fair.
- greater than the chance of getting tails since heads seem to be coming up.

**Problem 12.** A drug company developed a new formula for their headache medication. To test the effectiveness of this new formula, 250 people were randomly selected from a larger population of patients with headaches. 100 of these people were randomly assigned to receive the new formula medication when they had a headache, and the other 150 people received the old formula medication. The time it took, in minutes, for each patient to no longer have a headache was recorded. The results from both of these clinical trials are shown below. Which of the following conclusions is most accurate?



- The old formula works better. Two people who took the old formula felt relief in less than 20 minutes, compared to none who took the new formula. Also, the worst result - near 120 minutes - was with the new formula.
- The average time for the new formula to relieve a headache is lower than the average time for the old formula. I would conclude that people taking the new formula will tend to feel relief about 20 minutes sooner than those taking the old formula.
- I would not conclude anything from these data. The number of patients in the two groups is not the same so there is no fair way to compare the two formulas.

**Problems 13 and 14 refer to the following situation:**

The scores on a test of motivation are normally distributed, with a mean of 35 and a standard deviation of 14. Higher scores correspond to more motivation.

**Problem 13.** Gato scored 49 on this test. Gato scored higher than what proportion of the population?

- a. 0.32
- b. 0.49
- c. 0.68
- d. 0.84

**Problem 14.** 2.5% of the students scored higher than Shamu. What was her motivation score?

- a. 25
- b. 49
- c. 63
- d. 95

**Problem 15.** In the 2012 presidential election, Obama received 25% of the popular votes in Utah. If Holly takes a random sample of 100 people from Utah, what is the probability that the proportion of people who voted for Obama is greater than .35?

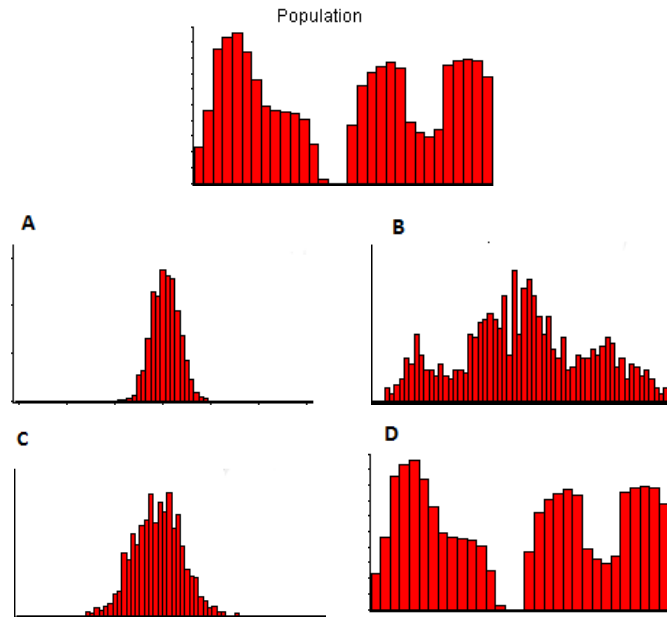
- a. 0.0104
- b. 0.0179
- c. 0.9896
- d. Can't be answered with the given information.

**Problem 16.** A certain right skewed distribution has a population mean of 32 and a population standard deviation of 2.3. What is the sampling distribution of the sample mean for  $n = 100$ ?

- a. Skewed right with a mean of 32 and a standard deviation of .23.
- b. Skewed right with a mean of 4.52 and a standard deviation of 2.3.
- c. Approximately normal with a mean of 32 and a standard deviation of 2.3.
- d. Approximately normal with a mean of 32 and a standard deviation of 0.23.

**Problems 17 to 19 refer to the following situation:**

A distribution for a population of test scores is displayed below. Possible sampling distributions of the sample mean were found for sample sizes 1, 2, 15, and 50. The graphs are all drawn to the same scale.



**Problem 17.** Which graph best represents a distribution of sample means for a thousand samples of size 15?

- Graph A
- Graph B
- Graph C
- Graph D

**Problem 18.** What do you expect for the shape of the sampling distribution of the sample mean for all possible samples of size  $n = 50$ ?

- Shaped more like a normal distribution than like the population distribution.
- Shaped more like the population distribution than like a normal distribution.
- Shaped like neither the population or the normal distribution.

**Problem 19.** What do you expect for the variability (spread) of the sampling distribution of the sample mean for  $n = 2$ ?

- Same as the population.
- Less variability than the population (a narrower distribution).
- More variability than the population (a wider distribution).

**Problem 20.** A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?

- a. Sam, because there are more candies, so his bag can have more brown candies.
- b. Sam, because there is more variability in the proportion of browns among larger samples.
- c. Kerry, because there is more variability in the proportion of browns among smaller samples.
- d. Both have the same chance because they are both random samples.

**Problem 21.** Jim wants to find a 98% confidence interval for the mean price of all houses in Salt Lake. He takes a sample of 24 houses and finds a sample mean of \$160,000 and a sample standard deviation of \$15,000. Find the 98% confidence interval. Assume normality.

- a. \$122,500 to \$197,500
- b. \$125,110 to \$194,890
- c. \$152,345 to \$167,654
- d. \$152,878 to \$167,121

**Problem 22.** Justin and Hayley conducted a mission to a new planet, Vulcan, to study arm length. They took a random sample of Vulcan residents and calculated a 95% confidence interval for the mean arm length of all Vulcans. What does a 95% confidence interval for arm length tell us in this case? Select the best answer:

- a. I am 95% confident that this interval includes the sample mean arm length.
- b. I am confident that about 95% of all Vulcan residents will have an arm length within this interval.
- c. I am 95% confident that most Vulcan residents will have arm lengths within this interval.
- d. I am 95% confident that this interval includes the population mean arm length.

**Problems 23 to 25 refer to the following situation:**

A high school statistics class wants to estimate the average number of chocolate chips in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a 95% confidence interval for the average number of chips per cookie (18.6 to 21.3). Below are several possible interpretations of the 95% confidence level. Indicate if each interpretation is valid or invalid.

**Problem 23.** We are 95% certain that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.

- a. Valid.
- b. Invalid.

**Problem 24.** There is a 95% probability that the population mean is between 18.6 and 21.3.

- a. Valid.
- b. Invalid.

**Problem 25.** Of all the possible samples the class could have chosen, about 95% of the resulting confidence intervals would contain the population mean.

- a. Valid.
- b. Invalid.

**Problems 26 to 28 refer to the following situation:**

A research article reports the results of a new drug test. The drug is to be used to decrease vision loss in people with Macular Degeneration. The article gives a  $p$ -value of .04 in the analysis section. Problems 26, 27, 28 present three different interpretations of this  $p$ -value. Indicate if each interpretation is valid or invalid.

**Problem 26.** The probability of getting results as extreme as or more extreme than the ones in this study if the drug is actually not effective is 0.04.

- a. Valid.
- b. Invalid.

**Problem 27.** The probability that the drug is not effective is 0.04.

- a. Valid.
- b. Invalid.

**Problem 28.** The probability that the drug is effective is 0.04.

- a. Valid.
- b. Invalid.

**Problem 29.** A graduate student is designing a research study. She is hoping to show that the results of an experiment are statistically significant. What type of  $p$ -value would she want to obtain?

- a. A large  $p$ -value.
- b. A small  $p$ -value.
- c.  $p$ -values are not related to statistical significance

**Problem 30.** Nik is conducting a hypothesis test with a level of significance of  $\alpha = .05$ . Which of the following is **NOT** a correct meaning of the level of significance?

- a. The level of significance tells us how much evidence against the null hypothesis Nik needs to reject the null hypothesis.
- b. The level of significance is the probability that the null hypothesis is true.
- c. The level of significance is the probability of rejecting a null hypothesis that is in fact true.
- d. Nik will only reject the null hypothesis if he observes a result that is so extreme that it would only happen 5% of the time if the null hypothesis is true.



**Problem 31.** Joe and Ashley are two graduate students at different colleges conducting different experiments to gauge the effect of a new medicine. They both chose a significance level of  $\alpha = .05$ . Ashley finds a  $p$ -value of .043 and Joe finds a  $p$ -value of .032. Which of the following statements is correct?

- a. Ashley found stronger evidence of an effect.
- b. Joe found stronger evidence of an effect.
- c. They both found the same amount of evidence of an effect.

**Problem 32.** A researcher conducts an experiment on human memory and recruits 15 people to participate in her study. She performs the experiment and analyzes the results. She obtains a  $p$ -value of .17. Which of the following is a reasonable interpretation of her results?

- a. This proves that her experimental treatment has no effect on memory.
- b. She should reject the null hypothesis.
- c. She failed to find evidence that her experimental treatment has an effect on memory.
- d. She found evidence that her experimental treatment has an effect on memory.

**Problem 33.** A researcher in environmental science is conducting a study to investigate the impact of a particular herbicide on fish. He has 60 healthy fish and randomly assigns each fish to either a treatment or a control group. The fish in the treatment group showed higher levels of the indicator enzyme.

Suppose a test of significance was correctly conducted and showed a statistically significant difference in average enzyme level between the fish that were exposed to the herbicide and those that were not. What conclusion can the graduate student draw from these results?

- a. The sample size is too small to draw a valid conclusion.
- b. He has proven that the herbicide causes higher levels of the enzyme.
- c. There is no evidence that the herbicide causes higher levels of the enzyme for these fish.
- d. There is evidence that the herbicide causes higher levels of the enzyme for these fish.

**Problem 34.** The makers of Mini-Oats cereal have an automated packaging machine that is set to fill boxes with 24 ounces of cereal. At various times in the packaging process, a random sample of 100 boxes is taken to see if the machine is filling the boxes with an average of 24 ounces of cereal. Which of the following is a statement of the null hypothesis being tested?

- a. The machine is filling the boxes with the proper amount of cereal.
- b. The machine is not filling the boxes with the proper amount of cereal.
- c. The machine is not putting enough cereal in the boxes.
- d. The machine is putting too much cereal in the boxes.

**Problem 35.** Which definition of a  $p$ -value is the most accurate?

- a. the probability that the observed outcome will occur again.
- b. the probability of observing an outcome as extreme or more extreme than the one observed if the null hypothesis is true.
- c. the value that an observed outcome must reach in order to be considered significant under the null hypothesis.
- d. the probability that the null hypothesis is true.

**Problems 36 to 38 refer to the following situation:**

It is known that the Honda Odyssey minivan has an average 25 miles per gallon on the highway. Chelsey is one of the engineers at Honda. She has developed a new part that she believes will increase the mean gas mileage of the minivans, but she needs to conduct a hypothesis test before approaching her boss. She takes a sample of 20 vans with her new part and finds a sample mean of 26 miles per gallon. Assume the population standard deviation is 1.3 miles per gallon and that the gas mileages are normally distributed.

**Problem 36.** Which hypothesis test should Chelsey use?

- a. One Sample z Test for Means
- b. One Sample t Test for Means
- c. One Sample z Test for Proportions
- d. The assumptions are not met. Chelsey shouldn't perform any of these tests.

**Problem 37.** What are the appropriate null and alternative hypotheses?

- a.  $H_0 : \mu = 25$  versus  $H_a : \mu < 25$
- b.  $H_0 : \mu = 25$  versus  $H_a : \mu \neq 25$
- c.  $H_0 : \mu = 25$  versus  $H_a : \mu > 25$
- d.  $H_0 : \bar{x} = 26$  versus  $H_a : \bar{x} > 26$

**Problem 38.** What is the test statistic?

- a. 0.77
- b. 3.44
- c. 4.47
- d. 15.38

**Problem 39.** KC needs to conduct a hypothesis test. He is unsure of how to use the p-value to decide whether to reject or fail to reject the null hypothesis. Which of the following should you tell him?

- a. He should always reject the null hypothesis if the sample size is at least 1000.
- b. He should reject the null hypothesis if he gets a small p-value.
- c. He should reject the null hypothesis if he gets a large p-value.
- d. He shouldn't use the p-value at all to decide whether to reject the null hypothesis.

**Problem 40.** McKenna is conducting a hypothesis test on 40 patients to determine if a new treatment increases hair growth. She is using a z test and she finds a test statistic of 2.54. What is her  $P$ -value?

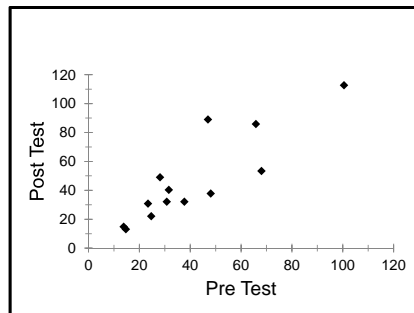
- a. 0.0055
- b. 0.0110
- c. 0.3641
- d. 0.9945

**Problem 41.** Researchers surveyed 1,000 randomly selected adults in the U.S.

A statistically significant, strong positive correlation was found between income level and amount of recycling. Please select the best interpretation of this result.

- This result indicates that there is an association between income level and the amount of recycling, but we don't know if a higher income level causes people to recycle more than people who earn less money.
- This sample is too small to draw any conclusions about the relationship between income level and amount of recycling for adults in the U.S.
- This result indicates that earning more money causes people to recycle more than people who earn less money.
- This result proves that earning more money causes people to recycle more than people who earn less money.

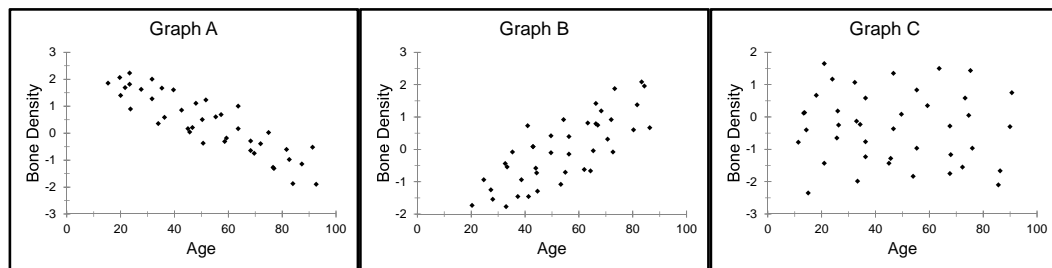
**Problem 42.** Dr. Jones gave students in her class a pretest about statistical concepts. After teaching about hypotheses tests, she then gave them a posttest about statistical concepts. Dr. Jones is interested in determining if there is a relationship between pretest and posttest scores, so she constructed the following scatterplot and calculated the correlation coefficient.



Which of the following is the best interpretation of the scatterplot?

- There is a moderate positive linear relationship between pretest and posttest scores.
- There is a strong negative linear relationship between pretest and posttest scores.
- There is little, if any, linear relationship between pretest and posttest scores.
- All of the students' scores increased from pretest to posttest.

**Problem 43.** Bone density was measured for a group of women. Lower scores correspond to lower bone density. Which of the following graphs shows that as women grow older they tend to have lower bone density?



- Graph A.
- Graph B.
- Graph C.

Final Exam Formula Sheet

Sampling Distributions

$\mu_{\bar{x}} = \mu$	$\mu_{\hat{p}} = p$
$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Confidence Intervals

Formula	Degrees of Freedom	Assumption
$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$		normal or $n \geq 30$
$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$	$df = n - 1$	normal or $n \geq 30$
$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$		$np \geq 10$ and $n(1-\hat{p}) \geq 10$
$b_1 \pm t^* SE_{b_1}$	$df = n - 2$	

Probabilities

Formula	Assumption
$z = \frac{x - \mu}{\sigma}$	normal
$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$	normal or $n \geq 30$
$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$	$np \geq 10$ and $n(1-p) \geq 10$

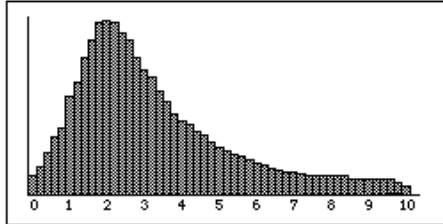
Hypothesis Tests

Formula	Degrees of Freedom	Assumption
$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$		normal or $n \geq 30$
$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$df = n - 1$	normal or $n \geq 30$
$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$		$np_0 \geq 10$ and $n(1-p_0) \geq 10$

## Appendix C: Pretest

## Pretest

**Problem 1.** Select the description that best represents the shape of the following distribution.



- a. Left skewed
- b. Right skewed
- c. Normal
- d. Bimodal
- e. I don't know.

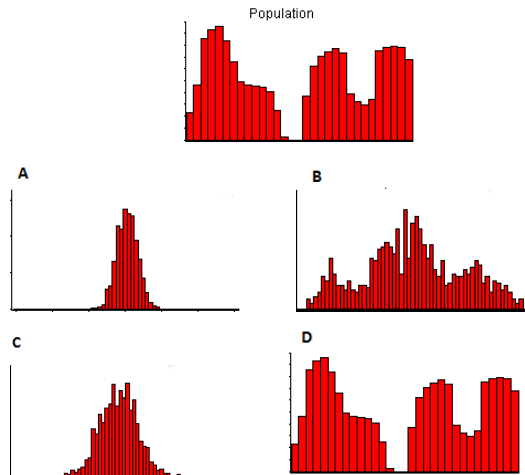
**Problem 2.** Bob and Jacob each bought one ticket for the same lottery each week for the past year. Bob has not won a single prize yet. Jacob just won a \$20 prize last week. Who is more likely to win a prize this coming week if they each buy only one ticket?

- a. Bob.
- b. Jacob.
- c. They have an equal chance of winning.
- e. I don't know.

**Problem 3.** The scores on a test of motivation are normally distributed, with a mean of 35 and a standard deviation of 14. Higher scores correspond to more motivation. Gato scored 49 on this test. Gato scored higher than what proportion of the population?

- a. 0.32
- b. 0.49
- c. 0.68
- d. 0.84
- e. I don't know.

**Problem 4.** A distribution for a population of test scores is displayed below. Possible sampling distributions of the sample mean were found for sample sizes 1, 2, 15, and 50. The graphs are drawn to scale. Which graph best represents a distribution of sample means for a thousand samples of size 15?



- Graph A
- Graph B
- Graph C
- Graph D
- I don't know.

**Problem 5.** A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?

- Sam, because there are more candies, so his bag can have more brown candies.
- Sam, because there is more variability in the proportion of browns among larger samples.
- Kerry, because there is more variability in the proportion of browns among smaller samples.
- Both have the same chance because they are both random samples.
- I don't know.

**Problem 6.** Jim wants to find a 98% confidence interval for the mean price of all houses in Salt Lake. He takes a sample of 24 houses and finds a sample mean of \$160,000 and a sample standard deviation of \$15,000. Find the 98% confidence interval. Assume normality.

- \$122,500 to \$197,500
- \$125,110 to \$194,890
- \$152,345 to \$167,654
- \$152,878 to \$167,121
- I don't know.

---

**Problems 7 to 9 refer to the following situation:**

A high school statistics class wants to estimate the average number of chocolate chips in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a 95% confidence interval for the average number of chips per cookie (18.6 to 21.3). Below are several possible interpretations of the 95% confidence level. Indicate if each interpretation is valid or invalid.

**Problem 7.** We are 95% certain that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.

- a. Valid.
- b. Invalid.
- e. I don't know.

**Problem 8.** There is a 95% probability that the population mean is between 18.6 and 21.3.

- a. Valid.
- b. Invalid.
- e. I don't know.

**Problem 9.** Of all the possible samples the class could have chosen, about 95% of the resulting confidence intervals would contain the population mean.

- a. Valid.
- b. Invalid.
- e. I don't know.

---

**Problems 10 to 12 refer to the following situation:**

A research article reports the results of a new drug test. The drug is to be used to decrease vision loss in people with Macular Degeneration. The article gives a  $p$ -value of .04 in the analysis section. Problems 10, 11, 12 present three different interpretations of this  $p$ -value. Indicate if each interpretation is valid or invalid.

**Problem 10.** The probability of getting results as extreme as or more extreme than the ones in this study if the drug is actually not effective is 0.04.

- a. Valid.
- b. Invalid.
- e. I don't know.

**Problem 11.** The probability that the drug is not effective is 0.04.

- a. Valid.
- b. Invalid.
- e. I don't know.

**Problem 12.** The probability that the drug is effective is 0.04.

- a. Valid.
  - b. Invalid.
  - e. I don't know.
-



**Problem 13.** A graduate student is designing a research study. She is hoping to show that the results of an experiment are statistically significant. What type of  $p$ -value would she want to obtain?

- a. A large  $p$ -value.
- b. A small  $p$ -value.
- c.  $p$ -values are not related to statistical significance.
- e. I don't know.

**Problem 14.** A researcher conducts an experiment on human memory and recruits 15 people to participate in her study. She performs the experiment and analyzes the results. She obtains a  $p$ -value of .17. Which of the following is a reasonable interpretation of her results?

- a. This proves that her experimental treatment has no effect on memory.
- b. She should reject the null hypothesis.
- c. She failed to find evidence that her experimental treatment has an effect on memory.
- d. She found evidence that her experimental treatment has an effect on memory.
- e. I don't know.

**Problem 15.** It is known that the Honda Odyssey minivan has an average 25 miles per gallon on the highway. Chelsey is one of the engineers at Honda. She has developed a new part that she believes will increase the mean gas mileage of the minivans, but she needs to conduct a hypothesis test before approaching her boss. She takes a sample of 17 vans with her new part and finds a sample mean of 26 miles per gallon. Assume the population standard deviation is 1.3 miles per gallon and that the gas mileages are normally distributed. Which hypothesis test should Chelsey use?

- a. One Sample  $z$  Test for Means
- b. One Sample  $t$  Test for Means
- c. One Sample  $z$  Test for Proportions
- d. The assumptions are not met. Chelsey shouldn't perform any of these tests.
- e. I don't know.

### Formulas that you might need

(You will only need a few of these formulas.)

1.  $\mu_{\bar{x}} = \mu$

2.  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

3.  $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$

4.  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

5.  $z = \frac{x - \mu}{\sigma}$

6.  $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$

7.  $\left[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \right]$

8.  $\left[ \bar{x} \pm t^* \frac{s}{\sqrt{n}} \right]$  with  $df = n - 1$

9.  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

10.  $b_1 \pm t^* SE_{b_1}$  with  $df = n - 2$

11.  $n = \frac{1}{4} \left( \frac{z^*}{m} \right)^2$

## **Appendix D: Online Surveys**

## Your Statistical Background Survey

This survey was given during the first week of all four classes.

**Student Directions:** Please take this survey so that I can get to know you and your statistical background. I will not look at the names when I review the results. Please answer honestly as your answers will not affect your grade in anyway.

### Questions:

1. Have you ever taken another statistics course, and if so which one?
  - AP statistics
  - USU STAT 1040
  - USU STAT 2000
  - USU STAT 2300
  - Another statistics course
  - This is my first statistics course
2. How do you feel about taking this statistics course?
  - I really, really, don't want to be here.
  - I am nervous.
  - I don't care either way.
  - I am kind of excited.
  - I am really excited.
3. How useful do you think statistics will be to you in the future, either in your career or your personal life?
  - I will never use it again.
  - I will use it once or twice.
  - I will use it occasionally.
  - I will use it often.
4. How do you feel about statistics?
  - I hate statistics.
  - I dislike statistics.
  - I don't care either way.
  - I like statistics.
  - I love statistics.
  - Since I have never taken statistics, I can't judge yet.

### Feelings About Activities Survey

This survey was given to the spring activity class at 3-4 weeks and again at 11-12 weeks.

The survey was given to the fall activity class at 4-5 weeks and again at 11-12 weeks.

**Student Directions:** I am curious about how the class feels about working with activities. I want to know how you feel about your learning and also your enjoyment of the class.

Please answer honestly. I won't look at any names until the semester is over. So your answers won't affect your grades or my feelings toward you.

#### Questions:

1. How would you rate your learning with activities versus lectures?  
(Do not consider how entertaining or enjoyable activities are, just how well you learn with the activities.)
  - I love the activities. I learn much better with activities than lectures.
  - I kind of like the activities. I learn a little bit better with activities than lectures.
  - I don't care either way. I learn about the same with activities or lectures.
  - I don't really like the activities. I learn a bit worse with activities than lectures.
  - I hate the activities. I learn so much worse with activities than lectures.
2. How would you rate your enjoyment of the activities versus lectures?  
(Do not consider how you learn, just how much you enjoy being in the class.)
  - I love the activities. Activities are so much better than lectures.
  - I kind of like the activities. Activities are a little better than lectures.
  - I don't care either way.
  - I don't really like the activities. Activities are a little worse than lectures.
  - I hate the activities. Activities are so much worse than lectures.
3. Would you rather work in pairs or groups of 4?
  - Pairs
  - Groups of 4
  - I don't care

### End of Class Survey

This survey was given during the last week of the course to students in all four classes.

**Student Directions:** Please answer each question honestly. I won't check the results until after the grades are submitted.

#### Questions:

1. How useful do you think statistics will be to you in the future, either in your career or your personal life?
  - I will never use it again.
  - I will use it once or twice.
  - I will use it occasionally.
  - I will use it often.
2. How do you feel about statistics?
  - I hate statistics.
  - I dislike statistics.
  - I don't care either way.
  - I like statistics.
  - I love statistics.
3. How much did you enjoy this statistics class?
  - I hated it.
  - I disliked it.
  - Neutral.
  - I liked it.
  - I loved it.
4. How engaged were you during class?
  - I often slept.
  - I was very bored.
  - I was kind of bored.
  - I was kind of engaged.
  - I was very engaged.

5. How motivated were you to come to class?
  - I often found reasons to skip class.
  - I only came for the participation points.
  - I wanted to come to class so that I didn't miss anything.
6. **Lecture Classes Only:** How did you feel about the guided notes that we used in class?
  - They were very useful; definitely worth the price.
  - They were okay.
  - They weren't worth the price.
7. **Fall Activity Class Only:** How did you feel about the activity book that we used in class?
  - They were very useful; definitely worth the price.
  - They were okay.
  - They weren't worth the price.
8. **Activity Classes Only:** If you were to take another statistics course, what teaching method would be your **first** choice?
  - traditional lecture
  - teacher introduce topic for 10 minutes, and then group work
  - teacher guided activities as a class
  - group work
9. **Activity Classes Only:** If you were to take another statistics course, what teaching method would be your **second** choice?
  - traditional lecture
  - teacher introduce topic for 10 minutes, and then group work
  - teacher guided activities as a class
  - group work

### Modified Current Statistics Self-Efficacy

This survey was administered to students in both classes at the end of each semester.

**Student Directions:** Please rate your confidence in your current ability to successfully complete the following tasks. The item scale has 5 possible responses:

- (1) no confidence at all,
- (2) a little confidence,
- (3) some confidence,
- (4) much confidence
- (5) complete confidence.

For each task, please mark the one response that represents your confidence in your current ability to successfully complete the task.

**Questions:**

1. Interpret the p-value from a statistical procedure.
2. Select the correct statistical procedure to be used to answer a research question.
3. Interpret the results of a statistical procedure in terms of the research question.
4. Explain what the value of the standard deviation means in terms of the variable being measured.
5. Distinguish between a Type I error and a Type II error in hypothesis testing.
6. Distinguish between a population parameter and a sample statistic.
7. Explain the difference between a sampling distribution and a population distribution.
8. Find the probability of an event.
9. Conduct a hypothesis test.
10. Calculate a confidence interval.
11. Interpret a confidence interval.
12. Explain what the 95% confidence **level** means.



### Attitudes Toward Statistics Survey (ATS)

This survey was given during the last week of class for both classes each semester.

**Student Directions:** For each of the following statements mark the rating category that most indicates how you currently feel about the statement. Please respond to all of the items.

**Possible Responses:**

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

**Questions:**

1. I feel that statistics will be useful to me in my profession.
2. The thought of being enrolled in a statistics course makes me nervous.
3. A good researcher must have training in statistics.
4. Statistics seems very mysterious to me.
5. Most people would benefit from taking a statistics course.
6. I have difficulty seeing how statistics relates to my field of study.
7. I see being enrolled in a statistics course as a very unpleasant experience.
8. I would like to continue my statistical training in an advanced course.
9. Statistics will be useful to me in comparing the relative merits of different objects, methods, programs, etc.
10. Statistics is not really very useful because it tells us what we already know anyway.
11. Statistical training is relevant to my performance in my field of study.
12. I wish that I could have avoided taking my statistics course.
13. Statistics is a worthwhile part of my professional training.
14. Statistics is too math oriented to be of much use to me in the future.
15. I get upset at the thought of enrolling in another statistics course.
16. Statistical analysis is best left to the “experts” and should not be part of a lay professional’s job.

17. Statistics is an inseparable aspect of scientific research.
18. I feel intimidated when I have to deal with mathematical formulas.
19. I am excited at the prospect of actually using statistics in my job.
20. Studying statistics is a waste of time.
21. My statistical training will help me better understand the research being done in my field of study.
22. One becomes a more effective “consumer” of research findings if one has some training in statistics.
23. Training in statistics makes for a more well-rounded professional experience.
24. Statistical thinking can play a useful role in everyday life.
25. Dealing with numbers makes me uneasy.
26. I feel that statistics should be required early in one’s professional training.
27. Statistics is too complicated for me to use effectively.
28. Statistical training is not really useful for most professionals.
29. Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

## Appendix E: Allocation of Teaching Time

### Amount of Time Spent in Teacher-Centered Activities

During the spring semester, there were 58 days in the semester. Students took exams on five of those days, leaving 53 teaching days at 50 minutes apiece. Table 24 details the amount of teaching time that was spent in teacher-centered activities during the spring semester for the activity class.

During the fall semester, there were only 57 days in the semester, leading to 52 teaching days at 50 minutes apiece. Table 25 analyzes the amount of class time that was spent in teacher-centered activities during the fall semester.

Table 24: Amount of time spent in teacher-centered activities during the spring semester.

Topic	Introduce Algorithms & Work Examples	Lecture	Recap	Teacher Led Discovery
Gathering Data				13
Describing Data	5	4	2	22
Random Variables	6		3	
Expected Value and Variance	6		7	12
Binomial Distribution	13	6		35
Normal Distribution	3.5		2.5	
Sampling Distributions			2	9
Confidence Intervals	31	8	12	34
Intro to Hypothesis Tests		10		26
Z Test, T Test, Test for Proportions	38	10	10	43
Two Sample Tests	8	5		19
ANOVA	5	10	2	18
Drug Activity for P-values				27
Chi Square Goodness of Fit	10	3		11
Chi Square Test of Independence	9	5		6
Linear Regression		3		6
Critical Values	9	5		11
Total in Minutes	143.5	69	40.5	292
Total in Hours	2.4	1.2	.68	4.9

Table 25: Amount of time spent in teacher-centered activities during the fall semester.

Topic	Introduce Algorithms & Work Examples	Lecture	Recap	Teacher Led Discovery
Gathering Data			2	18
Describing Data	7	9	18	8
Probability			13.5	
Random Variables			4	
Expected Value and Variance	12		11	15
Binomial Distribution	13.5	4		28
Normal Distribution			2	
Sampling Distributions		2	5	9
Confidence Intervals	29		16.5	33
Intro to Hypothesis Tests		11		24
Z Test, T Test, Test for Proportions	40	8	3.5	62
Two Sample Tests	7	9		13
ANOVA	9	5	2	16
Drug Activity for P-values				24
Chi Square Goodness of Fit	7	7		10
Chi Square Test of Independence	9	6		7
Linear Regression		2	4	8
Critical Values	6	2	3	12
Total in Minutes	139.5	65	84.5	287
Total in Hours	2.3	1.1	1.4	4.8

### **List of Topics That Were Lectured For The Activity Class**

The following is a list of topics for which the instructor lectured during the activity class. Lecture was usually used to introduce a new topic or concept.

- Picking classes for histograms
- Dealing with outliers
- Sample counts and requirements for a binomial distribution
- Results for sampling distributions
- T distribution
- Margin of error
- Significance levels, type I and II errors, etc.
- Official concepts for hypothesis testing
- Theory behind comparing two samples
- ANOVA, discuss  $R^2$
- Chi-square goodness of fit test
- Chi-square test of independence
- Least squares method
- Critical values

### **List of Days With and Without Teacher-Centered Activities**

These results are for the spring semester activity class. The fall semester activity class was similar.

#### **Days Without Any Teacher-Centered Activities**

1. Tuesday, Jan 15–Probability
2. Thursday, Jan 17–Probability
3. Friday, Jan 18–Probability
4. Tuesday, Jan 22–Probability
5. Thursday, Jan 24–Bayes Theorem and Review for Exam 1
6. Friday, Jan 25–Review for Exam 1
7. Monday, Feb 4–Normal Distribution
8. Friday, Feb 8–Sampling Distribution
9. Friday, Feb 15–Sampling Distribution for Proportions
10. Friday, Feb 22–Confidence Intervals
11. Monday, Feb 25–Confidence Intervals and Review for Exam 2
12. Tuesday, Feb 26–Review for Exam 2
13. Tuesday, Mar 5–Z tests
14. Thursday, Mar 7–Significance level and Errors
15. Friday, Mar 22–Two Sample Tests
16. Monday, Mar 25–Chi-Square and F Tests for Variance(s)
17. Tuesday, Mar 26–Review for Exam 3
18. Thursday, Mar 28–Review for Exam 3
19. Tuesday, Apr 2–ANOVA
20. Tuesday, Apr 9–Correlation and Scatterplots
21. Friday, Apr 12–Linear Regression
22. Monday, Apr 15–Linear Regression
23. Tuesday, Apr 16–Review for Linear Regression
24. Thursday, Apr 18–Review for ANOVA and Chi Square
25. Tuesday, April 23–Complete Problem
26. Thursday, Apr 25–Review for Final
27. Friday, Apr 26–Review for Final

**Days With Less Than Five Minutes of Teacher-Centered Activities**

1. Friday, Jan 11–Describing Data
2. Tuesday, Jan 29–Random Variables
3. Tuesday, Feb 5–Normal Distribution
4. Thursday, Feb 7–Normal Distribution
5. Thursday, Feb 14–Binomial

**Days With More Than Five Minutes of Teacher-Centered Activities**

1. Tuesday, Jan 8–Gathering Data
2. Thursday, Jan 10–Describing Data
3. Monday, Jan 14–Standard Deviation
4. Thursday, Jan 31–Expected Value
5. Friday, Feb 1–Variance and Rules for Expected Value and Variance
6. Monday, Feb 11–Sampling Distributions
7. Tuesday, Feb 12–Binomial
8. Tuesday, Feb 19–Confidence Intervals
9. Thursday, Feb 21–Confidence Intervals
10. Thursday, Feb 28–Intro to Hypothesis Testing
11. Monday, Mar 4–Intro to Hypothesis Testing
12. Friday, Mar 8–T test
13. Monday, Mar 18–Matched Pairs
14. Tuesday, Mar 19–Z test for proportions
15. Thursday, Mar 21–Two Sample Tests
16. Monday, Apr 1–ANOVA
17. Thursday, Apr 4–Chi Square Goodness of Fit
18. Friday, Apr 5–Drug Activity
19. Monday, Apr 8–Chi Square Test of Independence
20. Thursday, April 11–Linear Regression
21. Monday, Apr 22–Critical values



**Days With Exams**

1. Monday, Jan 7–Pretest
2. Monday, Jan 28–Exam 1
3. Friday, Mar 1–Exam 2
4. Friday, Mar 29–Exam 3
5. Friday, Apr 19–Exam 4

**Time That Students Worked On Their Own  
During The Lecture Class**

Students were occasionally allowed to work on their own during the lecture class. This occurred when the instructor felt that enough examples had already been covered as a class.

Table 26: Amount of time that students worked on their own during the lecture class.

Topic	Time in Spring	Time in Fall
Probability		7
Random Variables		4
Normal Distribution	5	11
Sampling Distribution		3
Confidence Intervals	9	4
Hypothesis Tests	17	12
Two Sample Tests		7
ANOVA		8
Chi Square Test of Independence		7
Total	31	63

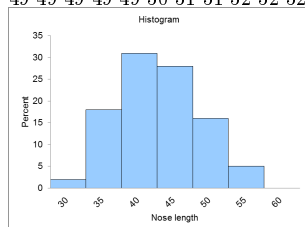
Note: The time is in minutes.

## **Appendix F: The Normal Distribution Unit for the Activity Classes**

## Normal Distribution—Discover the Empirical Rule

1. Here are the lengths of 100 adult noses. The mean is 45.07 and the standard deviation is 5.91.

31 34 35 35 36 36 36 37 37 38 38 38 39 39 39 39 39 39 39 40 40 40 40 41 41 41 41 41 42 42 42 42 42 42 43  
43 43 43 44 44 44 44 44 44 44 44 44 44 44 44 45 45 45 45 45 46 46 46 46 47 47 47 47 47 48 48 48 48 48 48 49 49  
49 49 49 49 49 50 51 51 52 52 52 52 53 53 53 53 53 54 54 54 55 56 56 59 59

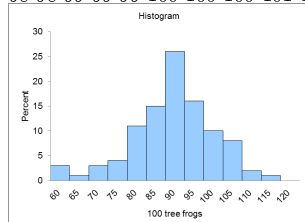


- How would you describe the shape of the distribution?
- What is the interval  $\text{mean} \pm \text{one standard deviation}$ ?
- What percentage of the data is within one standard deviation of the mean?
- What is the interval  $\text{mean} \pm \text{two standard deviations}$ ?
- What percentage of the data is within two standard deviations of the mean?
- What is the interval  $\text{mean} \pm \text{three standard deviations}$ ?
- What percentage of the data is within three standard deviations of the mean?
- Let's summarize our results.

	interval	percentage of the data
within one standard deviation of the mean		
within two standard deviations of the mean		
within three standard deviations of the mean		

2. Here are the lengths for 100 frogs. The mean is 91.54 and the standard deviation is 10.63.

63 63 64 66 70 74 74 78 78 78 79 80 80 80 82 82 82 82 83 84 84 84 85 85 85 85 86 86 87 87 87 87 87 88 88 88 88  
90 90 90 90 91 91 91 91 92 92 92 92 92 92 93 93 93 93 94 94 94 94 94 94 94 95 96 96 96 96 96 97 97 98 98 98  
98 98 99 99 99 100 100 100 101 102 102 102 102 103 103 105 106 107 107 107 108 109 109 110 112 116

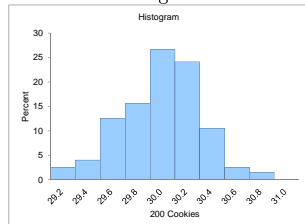


- (a) How would you describe the distribution?

- (b) Fill out the table.

	interval	percentage of the data
within one standard deviation of the mean		
within two standard deviations of the mean		
within three standard deviations of the mean		

3. Here are the weights for 200 cookies. The mean is 30.039 and the standard deviation is 0.316.

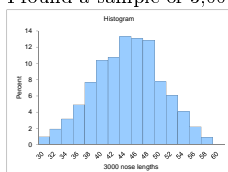


- (a) How would you describe the distribution?

- (b) To save time, I've filled out the table for you.

	interval	percentage of the data
within one standard deviation of the mean	(29.999,30.625)	65%
within two standard deviations of the mean	(29.677,30.941)	96%
within three standard deviations of the mean	(29.361,31.257)	99%

4. I found a sample of 3,000 nose lengths. The mean is 44.76 and the standard deviation is 5.76 .



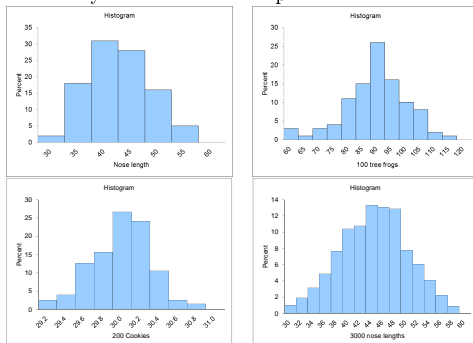
- (a) How would you describe the shape of the distribution?

- (b) To save time, I filled out the table for you.

	interval	percentage of the data
within one standard deviation of the mean	(39.00, 50.52)	68.76%
within two standard deviations of the mean	(33.24, 56.28)	95.56%
within three standard deviations of the mean	(27.48, 62.04)	99.83%

### Summarize all our Results

5. How did you describe the shape of the distributions?



	nose lengths	frogs lengths	cookies	more nose lengths
Sample size	100	100	200	3,000
Percentage of data within <b>one</b> standard deviation of the mean	60%	73%	65%	68.76%
Percentage of data within <b>two</b> standard deviation of the mean	97%	94%	96%	95.56%
Percentage of data within <b>three</b> standard deviations of the mean	100%	100%	99%	99.83%

- (a) Do you notice any patterns in the percentages of the data within **one** standard deviation of the mean?
- (b) Do you notice any patterns in the percentages of the data within **two** standard deviations of the mean?
- (c) Do you notice any patterns in the percentages of the data within **three** standard deviations of the mean?

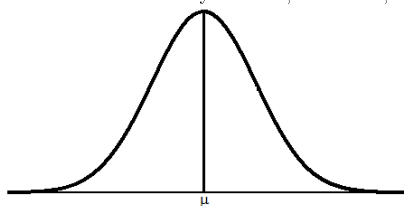
## Normal Distribution—Introduction

The normal distribution is very common. We find that the distribution of heights of people, weights of people, tree diameters, scores on tests, repeated measurements of the same quantity, lengths of baby pythons, yields of corn, and many more random variables are normally distributed.

Warning! Although many distributions are normal, there are still many random variables that do not have normal distributions.

We also use the normal distribution in many of our statistical procedures.

- The normal curve is symmetric, unimodal, and bell shaped. It is continuous.



- The shape of the normal curve is affected by the mean and by the standard deviation.
- The height of the normal curve at any point  $x$  is given by

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where  $\mu$  is the population mean and  $\sigma$  is the population standard deviation.

- Although there are many normal curves with different means and standard deviations, they all have common properties.
- Because we use the normal curve so much, we use the notation  $N(\mu, \sigma)$ .

## Normal Distribution—Shape

Use the applet at <http://mathnstats.com/applets/01-exploreNormalDistribution.html>

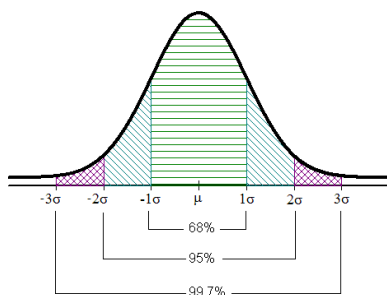
- Use the slider to change the mean.
  - What happens to the shape of the normal curve if the mean increases?
  - What happens to the shape of the normal curve if the mean is decreased?
- Use the slider to change the standard deviation.
  - What happens to the shape of the normal curve if you increase the standard deviation?
  - What happens to the shape of the normal curve if you decrease the standard deviation?
- Where is the center, or the middle, of the normal curve?

## Normal Distribution–Empirical Rule

### The 68-95-99.7 Rule

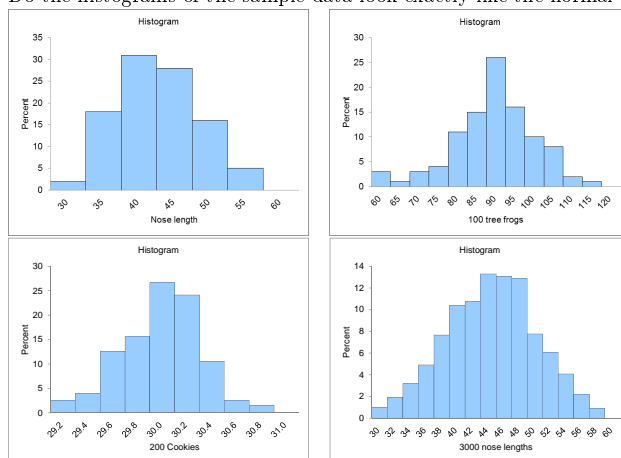
In the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ :

- Approximately 68% of the observations fall within 1 standard deviation of the mean.
- Approximately 95% of the observations fall within 2 standard deviations of the mean.
- Approximately 99.7% of the observations fall within 3 standard deviations of the mean.



10. Remember your previous results when we looked at some data sets.

(a) Do the histograms of the sample data look exactly like the normal curve?



(b) Do the histograms look kind of like the normal curve?

(When we use real data, our data won't follow the normal curve exactly, but there are many data sets that can be approximated by the normal curve.)



- (c) Here are the summary results for the sample data sets.

	nose lengths	frogs lengths	cookies	more nose lengths
Sample size	100	100	200	3,000
Percentage of data within <b>one</b> standard deviation of the mean	60%	73%	65%	68.76%
Percentage of data within <b>two</b> standard deviation of the mean	97%	94%	96%	95.56%
Percentage of data within <b>three</b> standard deviations of the mean	100%	100%	99%	99.83%

- i. Do the percentages of data within one, two, and three standard deviations from the mean match the 68-95-99.7 rule?
  - ii. Are the percentages in our actual data close to the 68-95-99.7 rule?
  - iii. Which data sets have percentages closest to the 68-95-99.7 rule?
  - iv. How do you think we can make sure our data sets follow the 68-95-99.7 rule better?
11. Use the 68-95-99.7 rule to describe each data set. I'll do the first one for you.
- (a) IQ test scores are supposed to be normally distributed with a mean of 100 and a standard deviation of 15. This means that
    - 68% of the people should have IQ scores between 85 and 115.
    - 95% of the people should have IQ scores between 70 and 130.
    - 99.7% of people should have IQ scores between 55 and 145.
  - (b) Suppose teenage boys in Utah are normally distributed with a mean of 66.5 inches and a standard deviation of 3.5 inches.
  - (c) Egg lengths are supposed to be normally distributed with a mean of 6 cm and a standard deviation of 1.4 cm.
12. Challenge: Egg lengths are supposed to be normally distributed with a mean of 6 cm and a standard deviation of 1.4 cm.
- (a) The longest 2.5% of the eggs should be at least how long?
  - (b) What is the cutoff length to be in shortest 50% of eggs?

## Normal Distribution–Standard Normal Distribution

- There is a special normal distribution.
- We call it the standard normal distribution.
- The mean is  $\mu = 0$ .
- Then standard deviation is  $\sigma = 1$ .
- We call the random variable  $Z$ .

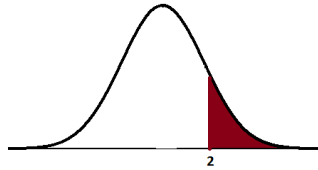
For this activity, you need the applet at <http://wise.cgu.edu/sdmmod/pzapplet.asp>

- Click on the graphic button.
- Set the mean to 0 and the standard deviation to 1.
- The  $z$  score is the  $z$  value for the standard normal distribution.
- You can drag the dividing line. This line divides the area under the normal curve into the area to the left and the area to the right.
- If you click the 2-tails button, then you can see the area in the middle and the area in the combined tails.
- You can enter the  $z$  score in the box and then press enter, or drag the line.
- You can also enter a probability in the box and then press enter, or drag the line.
- Remember that for a continuous distribution,  $P(X \geq 2)$  is the same as  $P(X > 2)$ .

13. Play with the applet for a few minutes. Click the buttons and drag the dividing line until you feel comfortable working with it.

14. Use the applet to find the following probabilities. Draw a picture for each problem.

- (a) Find the area under the curve to the right of  $z = 2$ . In symbols this is  $P(Z > 2)$ .



- (b) Find the area under the curve to the left of  $z = 2$ . In symbols this is  $P(Z < 2)$ .

- (c) What is the relationship between  $P(Z > 2)$  and  $P(Z < 2)$ ?

- (d) Find the area under the curve to the right of  $z = -2$ . In symbols this is  $P(Z > -2)$ .

(e)  $P(Z \geq -2)$

(f)  $P(Z < 1)$

(g) Find the area under the curve between  $z = -1$  and  $z = 1$ . In symbols this is  $P(-1 < Z < 1)$ .

(h)  $P(-1.2 < Z < 1.2)$

15. The empirical rule says that

- approximately 68% of the data is within 1 standard deviation of the mean.
- approximately 95% of the data is within 2 standard deviations of the mean.
- approximately 99.7% of the data is within 3 standard deviations of the mean.

We can find the exact percentages for a normal distribution using the applet.

- (a) To find the percentage of data within 1 standard deviation of the mean, we need to find  $P(-1 < Z < 1)$ .  
Draw a picture.
- (b) What is the exact percentage of data within 2 standard deviations of the mean? Find  $P(-2 < Z < 2)$ .  
Draw a picture.
- (c) What is the exact percentage of data within 3 standard deviations of the mean? Find  $P(-3 < Z < 3)$ .  
Draw a picture.

16. Sometimes we want to know the  $z$  value that divides the curve into certain probabilities. Draw a picture for each problem.

(a) What  $z$  value has 0.3 area to the left?

(b) What  $z$  value has 0.7 area to the right?

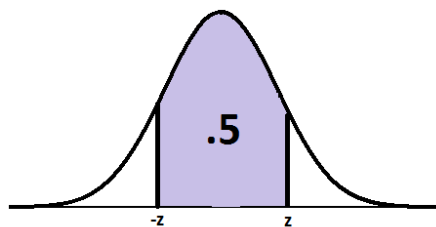
(c) What  $z$  value has 0.3 area to the right?

(d) Find the  $z$  value such that  $P(Z > z) = .4$ .

(This is really fancy notation to say find the  $z$  value that has .4 probability to the right.  
Some books say find the  $k$  value such that  $P(Z > k) = .4$ .)

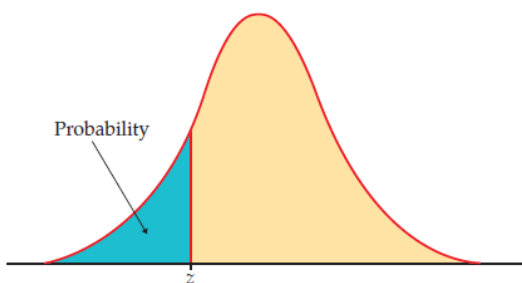
(e) What  $z$  value has 0.1 area to the right?

(f) What  $z$  value do we need so that the area in the middle between  $z$  and  $-z$  is 0.5?



## Normal Distribution–Standard Normal Table

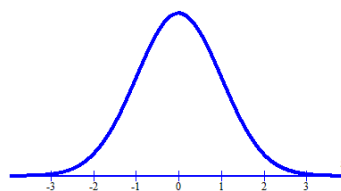
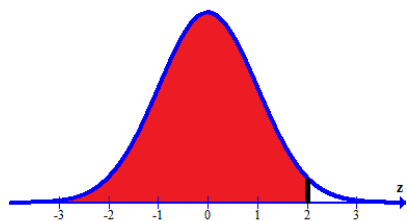
- There is a special normal distribution.
- We call it the standard normal distribution.
- The mean is  $\mu = 0$ .
- Then standard deviation is  $\sigma = 1$ .
- We call the random variable  $Z$ .
- We can use Table A at the back of our book to find probabilities for the standard normal curve.
  - **Notice:** the table gives the area under the standard normal curve to the *left* of  $z$ .
  - However, if you are creative, you can use the table to find any other area that you want.
    - \* Just remember the complement rule and that the normal distribution is symmetric.
  - The ones and tenths place of the  $z$  values are down the left column. The hundredths place of the  $z$  values are across the top row. The probabilities are in the middle of the table.



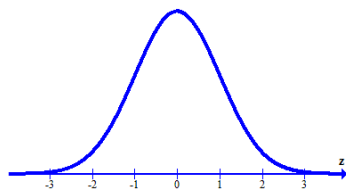
Use the table to find the probabilities. Make sure you shade the picture appropriately. I expect you to think creatively about how to find these probabilities. There are usually at least two ways to find the answers.

A.  $P(Z \leq 2)$

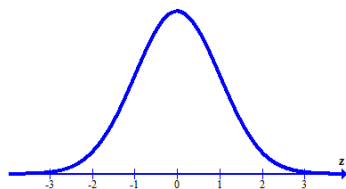
B.  $P(Z \geq 2)$



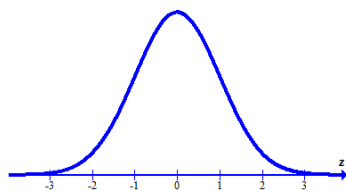
C.  $P(Z \leq -2)$



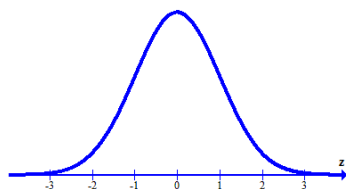
D.  $P(Z < -2)$



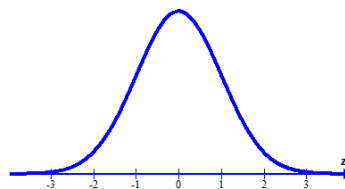
E.  $P(Z \geq -2)$



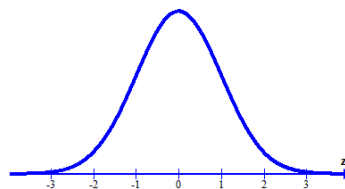
F.  $P(Z \leq 0.58)$



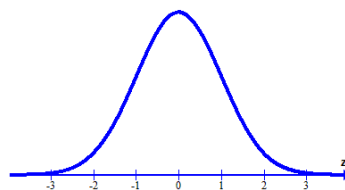
G.  $P(Z \leq 1.25)$



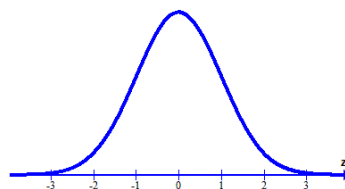
H.  $P(Z \geq 1.25)$



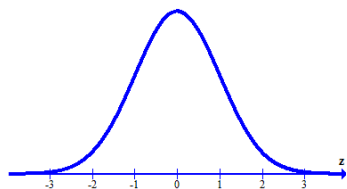
I.  $P(Z \leq -1.25)$



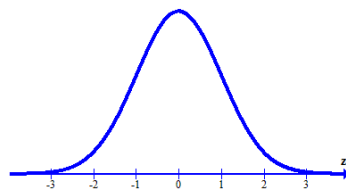
J.  $P(Z \geq -1.25)$



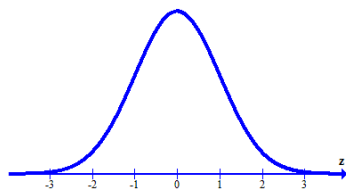
K.  $P(Z < -1.39)$



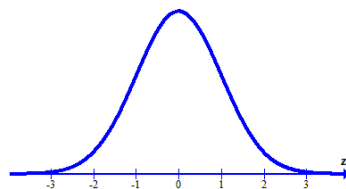
O.  $P(0 \leq Z \leq 2)$



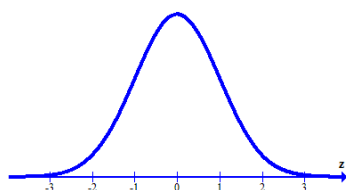
L.  $P(Z \leq 4.5)$



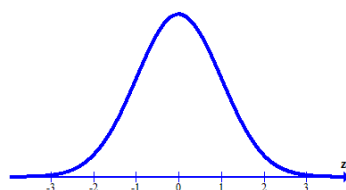
P.  $P(1.21 < Z \leq 2.47)$



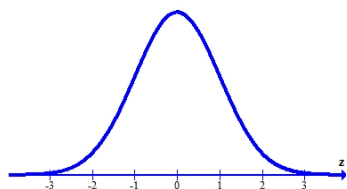
M.  $P(Z > 3.4)$



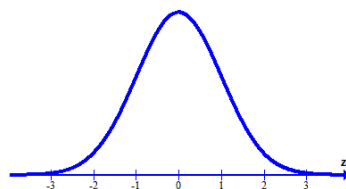
Q.  $P(-1.3 \leq Z < 2.7)$



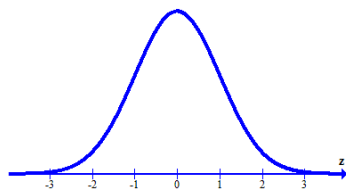
N.  $P(Z > 5)$



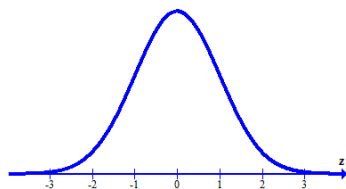
R.  $P(-3 \leq Z \leq 3)$



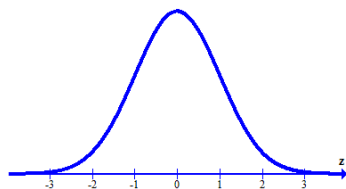
S.  $P(2 < Z < 3.18)$



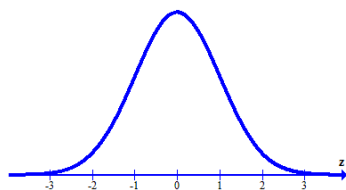
T.  $P(Z < -2 \text{ or } Z \geq 2)$



U.  $P(Z \leq -1.72 \text{ or } Z \geq 0.63)$



V.  $P(Z < -3 \text{ or } Z \geq 0)$



Use the table to find the value of  $z$  such that:

W. The area to the left of  $z$  is .3192.



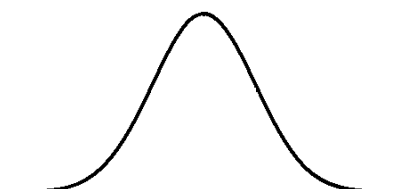
X. The area to the left of  $z$  is .6808.



Y. The area to the left of  $z$  is .0853.

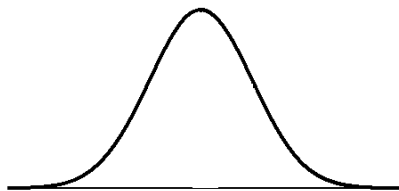


Z. The area to the right of  $z$  is .9147.

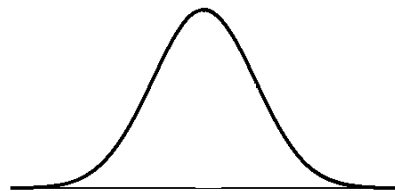




AA. The area to the left of  $z$  is .9382.



AE. The area to the right of  $z$  is .9319.



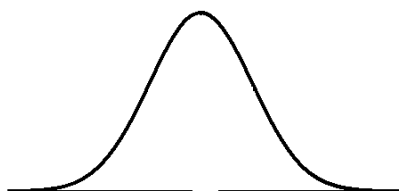
AB. The area to the left of  $z$  is .9265.



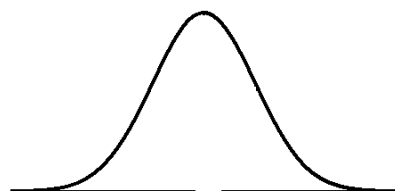
AF. The area between  $-z$  and  $z$  is .9956.



AC. The area to the right of  $z$  is 0.5.



AG. The area between  $-z$  and  $z$  is .383.



AD. The area to the right of  $z$  is .1170.



AH. The area in the tails has a combined probability of .10. (area to the right of  $z$  and area to the left of  $-z$ )



## Normal Distribution–Standardization

We have used the standard normal table to find probabilities, but what if we wanted to find probabilities for another normal distribution.

Because there are so many possible normal curves, it would be impossible to make a table of probabilities for all of them. Instead we have one table of probability for the standard normal curve.

In general we can find the z-score for any observation and any normal distribution:

$$z = \frac{\text{value} - \text{mean of value}}{\text{standard deviation of value}}$$

### Standardizing

If you have a random variable  $X$ , that is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ , you can standardize it with the formula

$$Z = \frac{X - \mu}{\sigma}$$

The new random variable,  $Z$ , has a standard normal distribution with mean 0 and standard deviation 1.

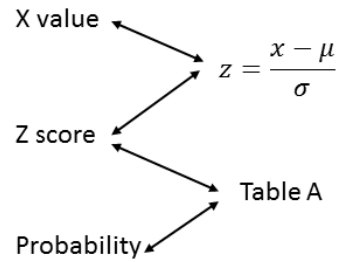
17. You are working with a normal distribution with mean 64.5 and standard deviation 2.5. Fill in the table. I'll do a few for you.

	original value	z-score
three standard deviations below the mean	57	$z = \frac{57 - 64.5}{2.5} = -3$
two standard deviations below the mean	59.5	$z = \frac{59.5 - 64.5}{2.5} = -2$
one standard deviation below the mean	62	
the mean	64.5	
one standard deviation above the mean		
two standard deviations above the mean		
three standard deviations above the mean		

18. What do you notice about the z-score and how many standard deviations the value is from the mean?



## Normal Distribution-Finding Probabilities



**You must always standardize (find the z-score) before you find the probability for a normal distribution!**

**Step 1:** You know the original  $x$  value.

**Step 2:** Find the z-score using the formula  $z = \frac{x - \mu}{\sigma}$ . (Draw a picture.)

**Step 3:** Use the standard normal table (Table A) to find the probability.

20. The test scores on a test are normally distributed with a mean of  $\mu = 290$  and a standard deviation of  $\sigma = 16$ . What is the probability that a randomly selected person has a score greater than 273?

21. The length of baby smurfs has an average of 10 inches and a standard deviation of 1.7 inches. Assume the lengths of baby smurfs is normally distributed. The baby smurf Bill is measured at 9.3 inches. What percentage of baby smurfs are shorter than Bill?

22. SAT scores are approximately Normal with mean 1026 and standard deviation 209.

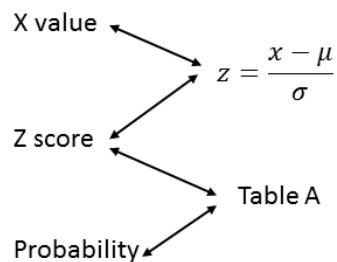
(a) What percentage of students have scores less than 1000?

(b) What proportion of all students have scores greater than or equal to 1300?

(c) The NCAA considers students to be partial qualifiers for athletic scholarships if their SAT score is between 720 and 820. What percentage of all students are partial qualifiers?

23. An automaker claims that the mileages of all its midsize cars are normally distributed with a mean mileage  $\mu = 33$  mpg and  $\sigma = .7$  mpg.
- (a) Calculate the probability that the mileage,  $X$ , of a randomly selected midsize car will be between 32 mpg and 35 mpg.
  - (b) Find the probability that  $X$ , the gas mileage of a randomly selected car, is less than or equal to 31.2.
  - (c) **Your feelings:** A testing agency randomly selects a car from the automaker and finds that it has a gas mileage of 31.2. (Do you think that the automaker is telling the truth?)
24. Marketing research indicates that hot chocolate tastes best if its temperature is between 153° F and 167°F. A restaurant claims that their hot chocolate has a mean of 160 degrees and a standard deviation of 5.37 degrees. Estimate the probability that  $X$ , the temperature of a randomly selected cup of hot chocolate, is *outside* the optimal temperature range.

## Normal Distribution-Knowing Probabilities and Finding X Values



**Step 1:** You know the probability. Draw a picture.

**Step 2:** Use table A to find the z-score.

**Step 3:** Work backwards to find the original value  $x$  with the formula  $z = \frac{x - \mu}{\sigma}$ .

25. Scores on the SAT Verbal test in recent years follow an approximate normal distribution with mean  $\mu = 505$  and  $\sigma = 110$ .

(a) How low must a student score to be in the bottom 15% of all students taking the SAT?

(b) How high must a student score in order to place in the top 10% of all students taking the SAT?

(c) How high must a student score to be in the top 30% of all students taking the SAT?

26. Extensive testing indicates that the lifetime of the Everlast battery is normally distributed with a mean of  $\mu = 60$  months and a standard deviation of  $\sigma = 6$  months.  
The Everlast's manufacturer has decided to offer a free replacement battery to any purchaser whose Everlast battery does not last at least as long as the minimum lifetime specified in its guarantee. How can the manufacturer establish the guarantee period so that only 1 percent of the batteries will need to be replaced free of charge.
27. Consider the ISTEP test scores which are approximately Normal with a mean of 572 and a standard deviation of 51.
- (a) How low a score is needed to be in the bottom 25% of students that take this exam?
- (b) We are often interested in knowing the middle 50% of test scores (with equal area in each tail). What score would you need to get to be in the middle 50% of test scores?



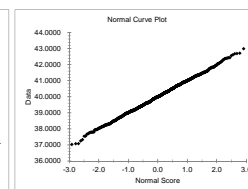
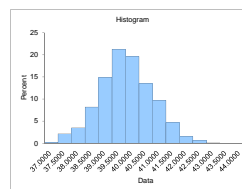
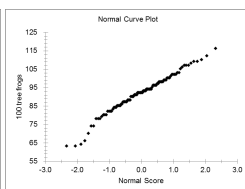
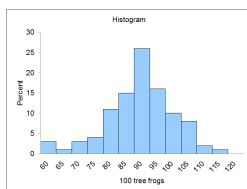
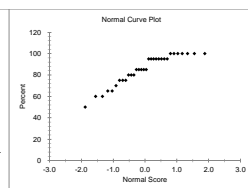
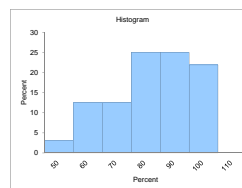
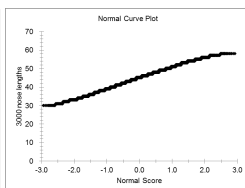
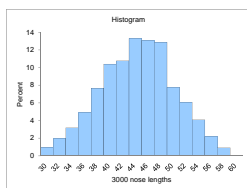
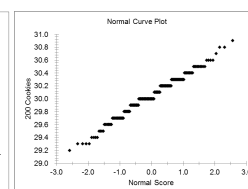
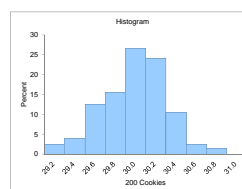
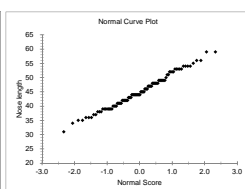
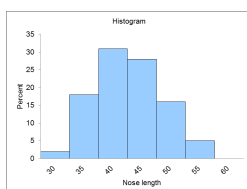
## Normal Quantile Plot Activity

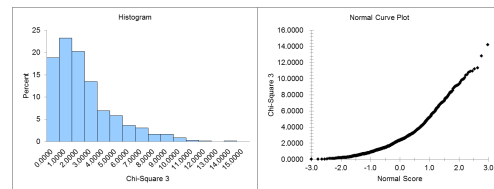
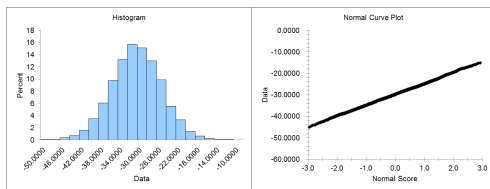
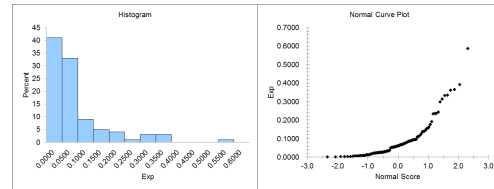
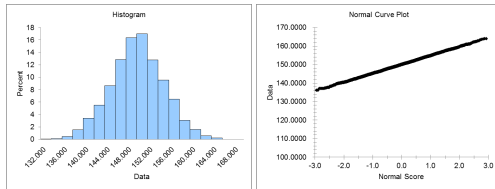
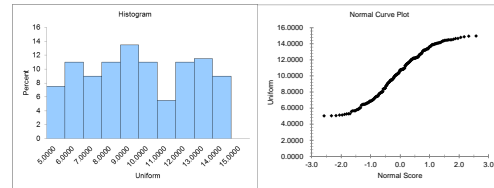
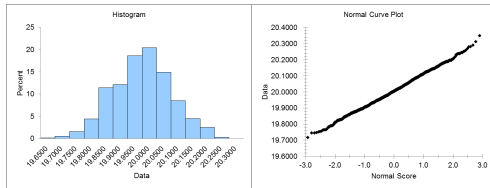
### Directions to Instructor:

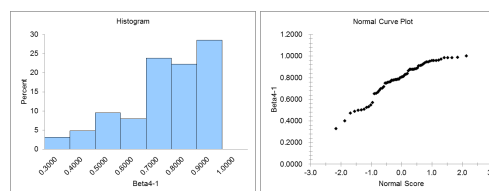
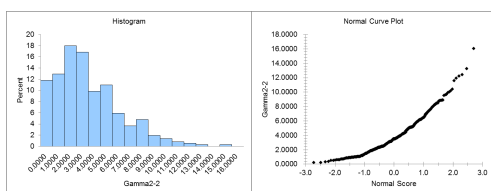
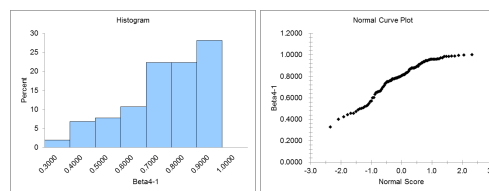
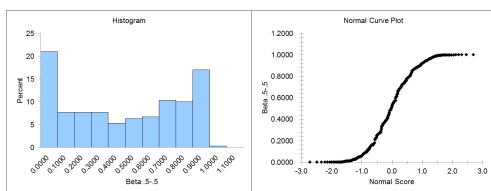
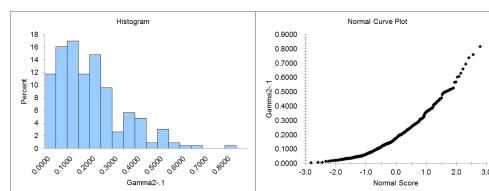
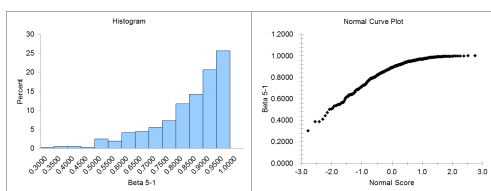
- Make a copy of the following images for each group.
- Cut the images out, but keep the histograms and the normal quantile plots for each data set attached together.
- Distribute the set of images to each group, along with the set of instructions below.
- Instruct the students to use the histograms to sort the data sets into three categories: very normally distributed, kind of normally distributed, and definitely not normally distributed.
- Tell the students to look at the normal quantile plots for the three categories and see if they can find a pattern.
- Students should realize that the normal distributions have normal quantile plots that are a straight line. The closer the data is to a normal distribution, the straighter the line in the normal quantile plot.

### Directions to Students:

1. Looking just at the histograms, sort the distributions into three categories: very normally distributed, kind of normally distributed, and definitely not normally distributed.
2. Pay attention to the shapes of the histograms and the shape of the normal quantile plots. Do you see any patterns?
3. If the histograms seem to be normally distributed, what does the normal quantile plot look like?







**Appendix G: The Sampling Distribution for Sample Means Unit for the Activity  
Classes**

## Sampling Distribution-Introduction

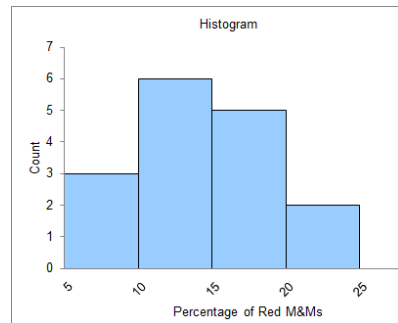
### Data from Class Experiment with M&M's:

- Remember when we did the M&M activity. This is the data from a previous year.
- I had a jar of M&M's and each pair in the class took a sample of about 70 M&M's and counted the percentage of red M&M's in their sample.
- The true percentage of all the M&M's in the jar was 14.5%.

Pair	Percentage of red M&M's in your sample
1	12.3
2	15.9
3	17
4	8.9
5	9.6
6	15.8

Pair	Percentage of red M&M's in your sample
7	14.1
8	14.1
9	15.6
10	18.8
11	12
12	22

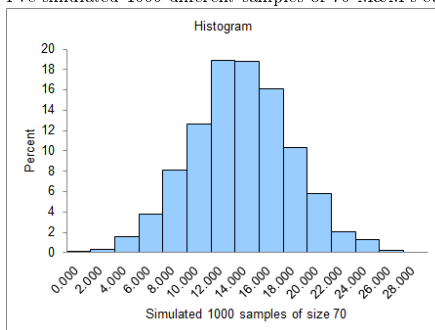
Pair	Percentage of red M&M's in your sample
13	24
14	14.7
15	11.8
16	8



- Did each pair get the exact same sample of red M&M's?
  - Did each pair get the same percentage of red M&M's?
  - What is the approximate center of the distribution for the percentage of red M&M's in the samples? Is it close to the true percentage 14.5%?
- Since each pair got different percentages of red M&M's in their sample, we say that the percentage of red M&M's is a random variable. This is because the value changes from sample to sample.
  - We often want to know what the distribution of the random variable would be if we took every possible sample. This is called the sampling distribution.

4. Now we only took 16 different samples. Imagine that we had a bigger jar with more M&Ms. We could potentially take thousands or millions of different samples.

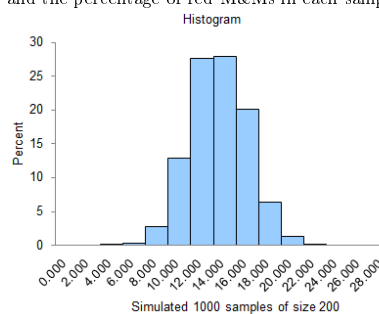
(a) I've simulated 1000 different samples of 70 M&M's each and the percentage of red M&Ms in each sample.



i. Does it look like the distribution is centered close to the true percentage of the population 14.5%?

ii. What is the shape of the distribution?

(b) Suppose we wanted to take a larger sample of 200 M&Ms each. I've simulated 1000 samples of 200 M&Ms each and the percentage of red M&Ms in each sample.



i. Does it look like the distribution is centered close to the true percentage of the population 14.5%?

ii. What is the shape of the distribution?

(c) Did the center of the distribution change when we went from 70 M&Ms in a sample to 200 M&Ms in a sample?

(d) Did the standard deviation change when we went from 70 M&Ms in a sample to 200 M&Ms in a sample?

## Sampling Distribution-Introduction

### All Possible Samples

From any population, there are many different samples of size  $n$  that can be chosen.

Let a population be  $\{0, 1, 2, 3, 4\}$ . The possible samples of size 2 that could be chosen without replacement and if we don't care about the order are:

$\{0,1\}, \{0,2\}, \{0,3\}, \{0,4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}$ , and  $\{3,4\}$ .

### Sampling Distribution

Since there are different samples that can be chosen, any sample statistic, (i.e. sample mean, sample mode, sample median, sample first quartile, sample standard deviation, sample range, sample maximum value, etc.), will be different based on which sample is chosen.

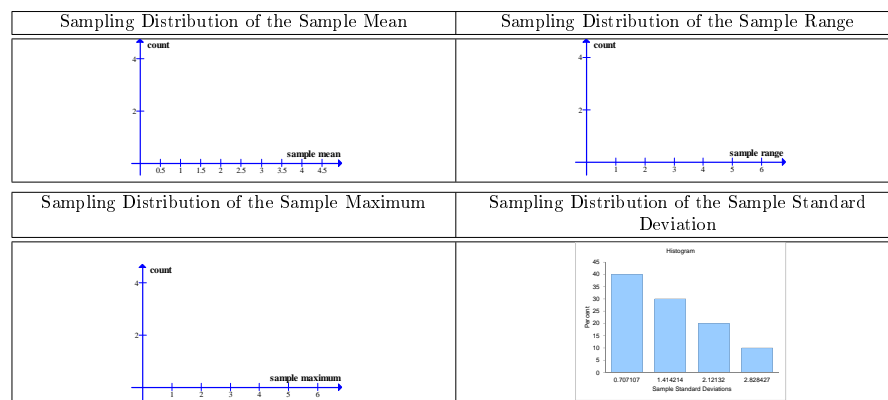
- Each statistic will be a random variable because its value changes from sample to sample.
- That means that each statistic also has its own distribution, called the sampling distribution.
- The idea that the value of the statistic changes from sample to sample is called **sampling variability**.

- Calculate the sample mean of each of the samples. Also calculate the range and maximum value for each sample. I've calculated the standard deviation for each sample for you.

Sample	Sample Mean	Sample Range	Sample Maximum	Sample Standard Deviation
$\{0,1\}$				0.707107
$\{0,2\}$				1.414214
$\{0,3\}$				2.12132
$\{0,4\}$				2.828427
$\{1,2\}$				0.707107
$\{1,3\}$				1.414214
$\{1,4\}$				2.12132
$\{2,3\}$				0.707107
$\{2,4\}$				1.414214
$\{3,4\}$				0.707107

- Notice that there are 10 different values for the sample mean depending on which sample is chosen. That means that the sample mean has its own distribution.
- We also see that the sample range, sample maximum, and sample standard deviation each have their own sampling distribution.
- Since each statistic has its own sampling distribution, we will want to know what the mean and standard deviation is of the sampling distribution.

- Draw the graph for the sampling distribution of each of the statistics.

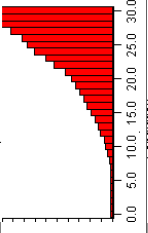
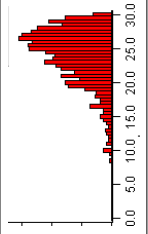
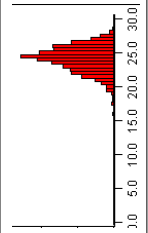
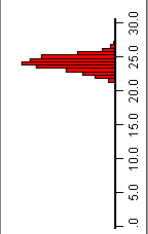
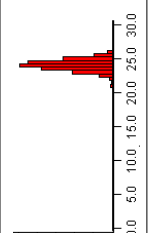
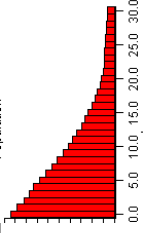
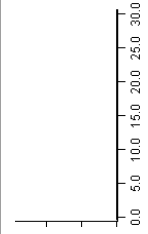
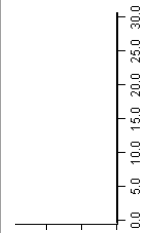
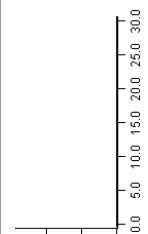
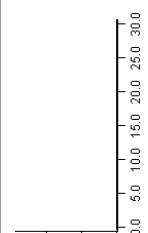
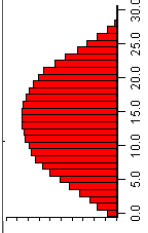
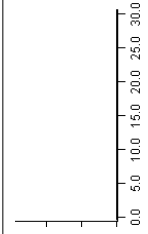
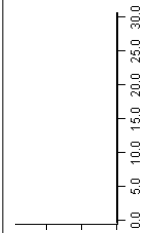
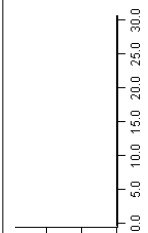
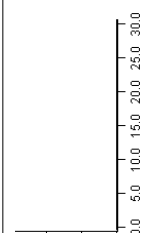


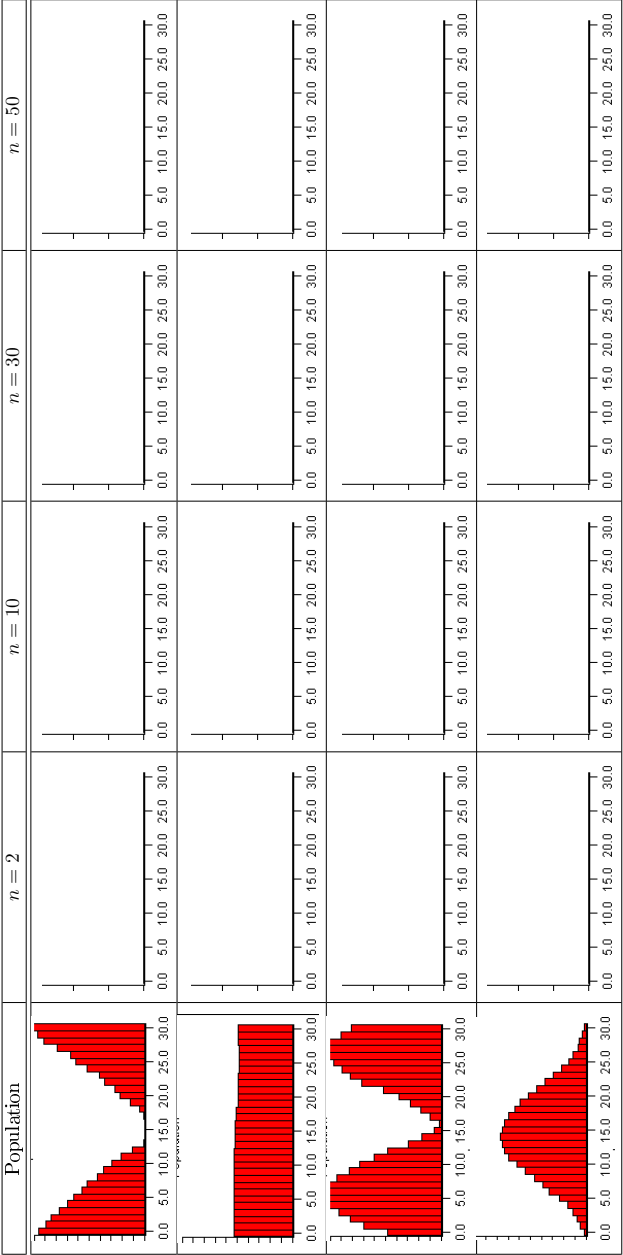


## Sampling Distribution-Shape

Use the applet <http://www-personal.umd.umich.edu/~pksmith/JavaStat/CentralLimitTheorem.html>

7. Go to the website; read the instructions on each screen, pressing continue until you get to the screen with the “resample” button.
8. Set the sample sizes to 2, 10, 30 and 50.
9. You can hold down the mouse button and drag the mouse to change the shape of the population. Then press resample.
10. This applet lets you find the sampling distribution of the sample mean.
11. Fill in the table by trying to create similar populations. Roughly sketch the sampling distribution for each of the sample sizes.

Population	$n = 2$	$n = 10$	$n = 30$	$n = 50$
				
				
				



12. What do you notice about the shape of the sampling distributions for sample size  $n = 2$ ?
13. What do you notice about the shape of the sampling distributions for sample size  $n = 50$ ?
14. How does the shape of the sampling distribution compare to the shape of the population distribution as the sample size increases?
15. What happens to the standard deviation of the sampling distribution as the sample size increases?
16. Try to create the weirdest populations you can think of. Is it possible to create a population that is so weird that the sampling distribution for sample size 50 doesn't look bell shaped?

## Sampling Distribution-Shape

Use the applet <http://www.intutor.com/statistics/CLAppClasses/CentLimApplet.htm>

17. Use the slider to make the number of samples as high as possible.
18. Use the slider at the bottom to adjust the sample size. Start with the slider all the way to the left for a sample size of 1 and slowly slide it to the right.
19. Repeat this for each of the four population types.
20. What happens to the shape of the sampling distribution for  $\bar{X}$  as the sample size increases?

## Final Conclusions

You can use both of the applets to answer these questions.

21. What does the sampling distribution of the sample mean look like?
22. What is the shape of the sampling distribution of the sample means if we start with a normal population?
23. What is the shape of the sampling distribution of the sample means if we start with a non-normal population?
24. What happens to the sampling distribution of the sample mean if we have a really weird shaped original population?

## Sampling Distribution-Mean and Standard Deviation

### Notation:

$\bar{x}$ : the sample mean of the individual observations in a single sample.

$\mu$ : the mean for the original population of individual values

$\sigma$ : the standard deviation for the original population of individual values

$\mu_{\bar{x}}$ : the theoretical mean for the population of *all possible sample means*

$\sigma_{\bar{x}}$ : the theoretical standard deviation for the population of *all possible sample means*

Use the applet [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html)

25. Go to the website and click the begin button.
26. You can click and drag the mouse to change the parent population.
27. You can change the sample size for the sampling distribution of means using the drop down box. You can also make the bottom graphs be the sampling distribution of the means for different sample sizes.
28. Notice that the mean and standard deviation are given for the parent population.
29. Also, next to each sampling distribution for the means, there is a mean and standard deviation.
30. For each parent population, click the button to get 10,000 samples. **Make sure you click "Clear lower 3" in between each sample or your results will be off.**
31. Use the default normal population. I've done this one for you.
  - (a) The mean of the original population is  $\mu = 16$ .
  - (b) The standard deviation of the original population is  $\sigma = 5$ .
  - (c) Now the sample mean has its own distribution. We are using the applet to find 10,000 different sample means.
    - i. This means that it is possible to find the mean of all the different sample means. We call this  $\mu_{\bar{x}}$ .
    - ii. We also want to find the standard deviation of all the different sample means. We call this  $\sigma_{\bar{x}}$ .
  - (d) Fill in the table with the data for the sampling distribution. You need to compute  $\sigma/\sqrt{n}$  for each sample size. (Remember that  $\sigma$  is the standard deviation of the original population and  $n$  is the sample size.)

Sample Size	Mean of all the sample means, $\mu_{\bar{x}}$	Standard deviation of all the sample means, $\sigma_{\bar{x}}$	$\sigma/\sqrt{n}$
2	16.01	3.52	$5/\sqrt{2} = 3.53$
5	15.99	2.26	$5/\sqrt{5} = 2.23$
16	16	1.23	$5/\sqrt{16} = 1.25$
25	16	1.01	$5/\sqrt{25} = 1.00$

32. Create a new population. (Don't forget to click "Clear lower 3" to reset the results.)
  - (a) What is the population mean  $\mu$ ?
  - (b) What is the population standard deviation  $\sigma$ ?
  - (c) Fill in the table with the data for the sampling distribution. You need to compute  $\sigma/\sqrt{n}$  for each sample size. (Remember that  $\sigma$  is the standard deviation of the original population and  $n$  is the sample size.)

Sample Size	Mean of all the sample means, $\mu_{\bar{x}}$	Standard deviation of all the sample means, $\sigma_{\bar{x}}$	$\sigma/\sqrt{n}$
2			
5			
16			
25			

33. Create a new population.

- What is the population mean  $\mu$ ?
- What is the population standard deviation  $\sigma$ ?
- Fill in the table with the data for the sampling distribution. You need to compute  $\sigma/\sqrt{n}$  for each sample size.

Sample Size	Mean of all the sample means, $\mu_{\bar{X}}$	Standard deviation of all the sample means, $\sigma_{\bar{X}}$	$\sigma/\sqrt{n}$
2			
5			
16			
25			

34. Create a new population.

- What is the population mean  $\mu$ ?
- What is the population standard deviation  $\sigma$ ?
- Fill in the table with the data for the sampling distribution. You need to compute  $\sigma/\sqrt{n}$  for each sample size.

Sample Size	Mean of all the sample means, $\mu_{\bar{X}}$	Standard deviation of all the sample means, $\sigma_{\bar{X}}$	$\sigma/\sqrt{n}$
2			
5			
16			
25			

35. What do you notice about the mean of the sampling distribution,  $\mu_{\bar{X}}$ , for the different sample sizes?

36. What do you notice about the standard deviation of the sampling distribution,  $\sigma_{\bar{X}}$ , for the different sample sizes?

37. How does  $\sigma_{\bar{X}}$  compare with  $\frac{\sigma}{\sqrt{n}}$ ?

- Why is the standard deviation,  $\sigma_{\bar{X}}$ , of the population of all possible sample means less than  $\sigma$ , the standard deviation of the original population?
  - Sample means are less variable than individual observations.
  - This is because each sample mean *averages out* high and low sample measurements.
  - So the sample mean can be expected to be closer to the population mean  $\mu$  than many of the individual population measurements would be.

## Sampling Distribution-Mean and Standard Deviation Review

### Notation:

$\bar{x}$ : the sample mean of the individual observations in a single sample.

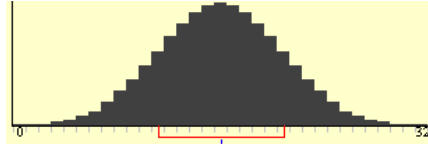
$\mu$ : the mean for the original population of individual values

$\sigma$ : the standard deviation for the original population of individual values

$\mu_{\bar{x}}$ : the theoretical mean for the population of *all possible sample means*

$\sigma_{\bar{x}}$ : the theoretical standard deviation for the population of *all possible sample means*

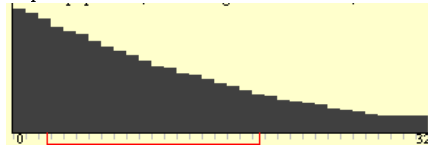
### 38. Population 1



- The mean of the original population is  $\mu = 16$ .
- The standard deviation of the original population is  $\sigma = 5$ .
- Now the sample mean has its own distribution. We are using the applet to find 10,000 different sample means.
  - This means that it is possible to find the mean of all the different sample means. We call this  $\mu_{\bar{x}}$ .
  - We also want to find the standard deviation of all the different sample means. We call this  $\sigma_{\bar{x}}$ .

Sample Size	Mean of all the sample means, $\mu_{\bar{x}}$	Standard deviation of all the sample means, $\sigma_{\bar{x}}$	$\sigma/\sqrt{n}$
2	16.01	3.52	$5/\sqrt{2} = 3.53$
5	15.99	2.26	$5/\sqrt{5} = 2.23$
16	16	1.23	$5/\sqrt{16} = 1.25$
25	16	1.01	$5/\sqrt{25} = 1.00$

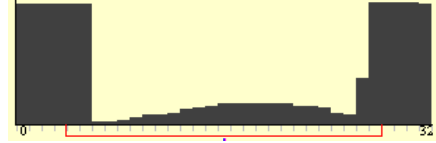
### 39. Population 2



- The mean of the original population is  $\mu = 10.56$ .
- The standard deviation of the original population is  $\sigma = 8.41$ .

Sample Size	Mean of all the sample means, $\mu_{\bar{x}}$	Standard deviation of all the sample means, $\sigma_{\bar{x}}$	$\sigma/\sqrt{n}$
2	10.51	5.95	$8.41/\sqrt{2} = 5.94$
5	10.52	3.73	$8.41/\sqrt{5} = 3.76$
16	10.56	2.09	$8.41/\sqrt{16} = 2.10$
25	10.58	1.70	$8.41/\sqrt{25} = 1.68$

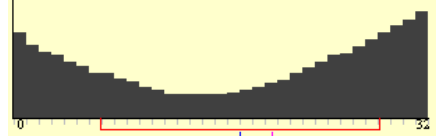
## 40. Population 3



- (a) The mean of the original population is  $\mu = 15.90$ .  
 (b) The standard deviation of the original population is  $\sigma = 12.61$ .

Sample Size	Mean of all the sample means, $\mu_{\bar{X}}$	Standard deviation of all the sample means, $\sigma_{\bar{X}}$	$\sigma/\sqrt{n}$
2	15.86	8.93	$12.61/\sqrt{2} = 8.92$
5	15.88	5.63	$12.61/\sqrt{5} = 5.64$
16	15.95	3.19	$12.61/\sqrt{16} = 3.15$
25	15.84	2.54	$12.61/\sqrt{25} = 2.52$

## 41. Population 4



- (a) The mean of the original population is  $\mu = 17.52$ .  
 (b) The standard deviation of the original population is  $\sigma = 11.08$ .

Sample Size	Mean of all the sample means, $\mu_{\bar{X}}$	Standard deviation of all the sample means, $\sigma_{\bar{X}}$	$\sigma/\sqrt{n}$
2	17.55	7.82	$11.08/\sqrt{2} = 7.83$
5	17.45	4.92	$11.08/\sqrt{5} = 4.95$
16	17.52	2.78	$11.08/\sqrt{16} = 2.77$
25	17.52	2.22	$11.08/\sqrt{25} = 2.22$

42. What do you notice about the mean of the sampling distribution,  $\mu_{\bar{X}}$ , for the different sample sizes?

43. What do you notice about the standard deviation of the sampling distribution,  $\sigma_{\bar{X}}$ , for the different sample sizes?

44. How does  $\sigma_{\bar{X}}$  compare with  $\frac{\sigma}{\sqrt{n}}$ ?

- Why is the standard deviation,  $\sigma_{\bar{X}}$ , of the population of all possible sample means less than  $\sigma$ , the standard deviation of the original population?
  - Sample means are less variable than individual observations.
  - This is because each sample mean *averages out* high and low sample measurements.
  - So the sample mean can be expected to be closer to the population mean  $\mu$  than many of the individual population measurements would be.

## Sampling Distribution-Central Limit Theorem

### The Sampling Distribution of $\bar{X}$

Assume the original population has mean  $\mu$  and standard deviation  $\sigma$  and we randomly select a sample of size  $n$ .

Then the population of all possible sample means has:

- Mean:  $\mu_{\bar{x}} = \mu$
- Standard deviation:  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- Shape:
  - It has an exactly normal distribution if the original population was normally distributed (for any size of sample).
  - It has an *approximate* normal distribution (for *any* original population) if the sample size is large enough.

### Central Limit Theorem

If the sample size  $n$  is large enough, then the population of all possible sample means is approximately normally distributed, for *any* original population.

As the sample size  $n$  increases, the normal curve becomes a better approximation for the sampling distribution of  $\bar{X}$ .

**Rule of Thumb:** If our sample size is at least 30, we can reasonably assume that the sampling distribution of  $\bar{X}$  is normal.

---

45. A trucking company knows that the population mean billing time is  $\mu = 19.5$  days. The population standard deviation is  $\sigma = 4.2$  days. They decide to take a sample of  $n = 65$  days and look at the sample mean.

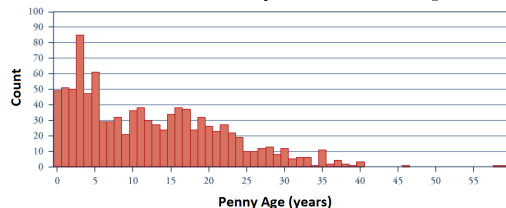
- (a) What is the mean of the sampling distribution?  $\mu_{\bar{X}}$
- (b) What is the standard deviation of the sampling distribution?  $\sigma_{\bar{X}}$
- (c) Can we use the normal curve as an approximation of the sampling distribution? (Hint: is the population normal or the sample size at least 30?)

46. Suppose the time  $X$  between text messages arriving on your cell phone has a normal distribution with a population mean  $\mu = 25$  minutes and standard deviation  $\sigma = 15$  minutes. You record the times between your next 10 messages and you are interested in the sample mean.

- (a) What is the mean of the sampling distribution?  $\mu_{\bar{X}}$
- (b) What is the standard deviation of the sampling distribution?  $\sigma_{\bar{X}}$
- (c) Can we use the normal curve as an approximation of the sampling distribution? (Hint: is the population normal or the sample size at least 30?)



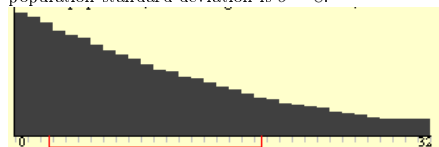
47. Someone counted thousands of pennies to find the ages of each pennies. This is the population.



The population mean is  $\mu = 12.3$  years and the standard deviation is  $\sigma = 9.6$ . You decide to take a sample of 50 pennies and look at the average age.

- Will the sampling distribution of the mean be normally distributed? (Hint: is the population normal or the sample size at least 30?)
- What is the mean of the sampling distribution?  $\mu_{\bar{X}}$
- What is the standard deviation of the sampling distribution?  $\sigma_{\bar{X}}$

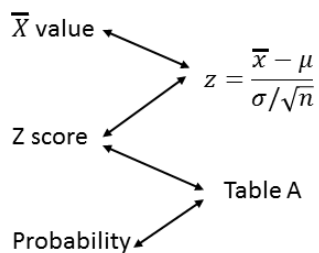
48. Here is a right skewed population. You want to take a sample of size  $n = 5$ . The population mean is  $\mu = 10$  and the population standard deviation is  $\sigma = 8$ .



- Will the sampling distribution of the mean be normally distributed?
  - What is the mean of the sampling distribution?  $\mu_{\bar{X}}$
  - What is the standard deviation of the sampling distribution?  $\sigma_{\bar{X}}$
49. Suppose that of all the adults in the U.S. the mean spending for Halloween is \$50 and the standard deviation is \$35. You decide to take a sample of 100 U.S. adults and look at the mean spending.
- What is the shape of the distribution of all the possible sample means?
  - What is the mean of all the possible sample means?  $\mu_{\bar{X}}$
  - What is the standard deviation of all the possible sample means?  $\sigma_{\bar{X}}$

### Sampling Distribution-Probabilities for Sample Means

If the original population is normal, or the sample size is large enough ( $n \geq 30$ ), you can use the formula  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  to find probabilities associated with the sample mean.



---

50. A trucking company knows that the population mean billing time is  $\mu = 19.5$  days. The population standard deviation is  $\sigma = 4.2$  days. They decide to take a sample of  $n = 65$  days and look at the sample mean.

(a) Can we use the normal distribution for this problem? (Hint: is the population normal or the sample size at least 30?)

(b) What is the probability that they get a sample mean that is less than 18 days?

(c) What is the probability that they get a sample mean that is greater than 21 days?

51. Suppose the time  $X$  between text messages arriving on your cell phone has a normal distribution with a population mean  $\mu = 25$  minutes and standard deviation  $\sigma = 15$  minutes. You record the times between your next 10 messages and you are interested in the sample mean.

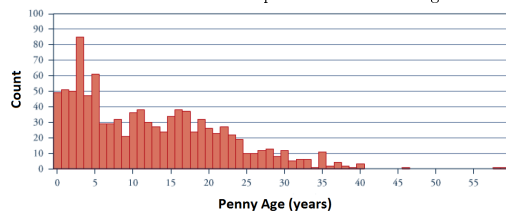
(a) Can we use the normal distribution for this problem? (Hint: is the population normal or the sample size at least 30?)

(b) What is the probability that the sample mean is less than 28 minutes?

(c) What is the probability that the sample mean is between 20 and 22 minutes?

(d) What  $\bar{x}$  value has 10% of the area to the right?

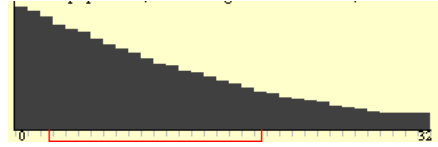
52. Someone counted thousands of pennies to find the ages of each pennies. This is the population.



The population mean is  $\mu = 12.3$  years and the standard deviation is  $\sigma = 9.6$ . You decide to take a sample of 50 pennies and look at the average age.

- (a) Can you use the normal distribution for this problem?
- (b) What is the probability that the sample mean is greater than 15 years?
- (c) What is the probability that the sample mean is between 10 and 15 years?
- (d) What is the probability that the sample mean is less than 9 years or greater than 15 years?

53. Here is a right skewed population. The population mean is  $\mu = 10$  and the population standard deviation is  $\sigma = 8$ . You want to take a sample of size  $n = 5$  and look at the sample mean.



- (a) Can you use the normal distribution for this problem?
  - (b) What is the probability that the sample mean is greater than 11?
54. Suppose that of all the adults in the U.S. the mean spending for Halloween is \$50 and the standard deviation is \$35. You decide to take a sample of 100 U.S. adults and look at the mean spending.
- (a) Can you use the normal distribution for this problem?
  - (b) Find the probability that the **total spending** for the 100 people is greater than \$5700.
  - (c) What is the cutoff to be in the lowest 20% of all the possible sample means?
55. Suppose that of all the adults in the U.S. the mean spending for Halloween is \$50 and the standard deviation is \$35. You decide to take a different sample of U.S. adults and look at the mean spending. However, you want to be able to control the standard deviation of the sampling distribution of the means.
- (a) You want  $\sigma_{\bar{x}}$  to be \$5 or less. How big of a sample do you need?
  - (b) You want  $\sigma_{\bar{x}}$  to be less than \$2. How many adults do you need to include in your sample?

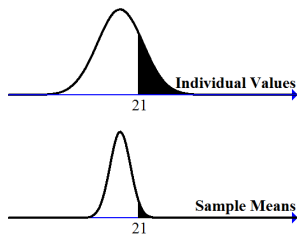
**Extra Problem: Sampling Distributions for Means**

56. The scores of all the high school students in the nation on the ACT exam were roughly *normally* distributed with a mean  $\mu = 19.2$  and a standard deviation  $\sigma = 5.1$ .

(a) What is the probability that a randomly chosen student scores at least 21 on the exam?

(b) What is the probability that a sample of 25 randomly chosen students has a sample mean of at least 21 on the exam?

(c) I have graphed the distributions of the individual values and the sample means for you on the same scale. Why do you think that  $P(X \geq 21)$  was so much bigger than  $P(\bar{X} \geq 21)$ ?



## **Appendix H: Applets Used In Classes**

### Applets Used in Lecture Classes

- Probability
  - Coin tossing applet
    - [nlvm.usu.edu](http://nlvm.usu.edu)
- Sampling Distributions
  - Central limit theorem applet
    - [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html)
  - Central limit theorem applet
    - <http://www-personal.umd.umich.edu/~pksmith/JavaStat/CentralLimitTheorem.html>
- Confidence intervals
  - Meaning of confidence level applet
    - [http://bcs.whfreeman.com/ips4e/cat\\_010/applets/confidenceinterval.html](http://bcs.whfreeman.com/ips4e/cat_010/applets/confidenceinterval.html)
- Introduction to Hypothesis Tests
  - Coin tossing applet
    - <http://www.math.usu.edu/~schneit/CTIS/PValue/>
  - Dice rolling applet
    - <http://www.math.usu.edu/schneit/CTIS/ChiSquare/>

### Applets Used in Activity Classes

\*Unless otherwise stated, the students explored the applets in their groups.

- Describing Data
  - Categorical versus quantitative applet
    - <http://mathnstats.com/applets/Categorical-Quantitative.html>
  - Standard deviation applet
    - <http://www.math.usu.edu/~schneit/CTIS/SD/>
- Probability
  - Coin flipping applet
    - [http://nlvm.usu.edu/en/nav/frames\\_asid\\_305\\_g\\_4\\_t\\_5.html?from=topic\\_t\\_5.html](http://nlvm.usu.edu/en/nav/frames_asid_305_g_4_t_5.html?from=topic_t_5.html)
  - Hospital Excel simulation
  - Pick a Number Excel simulation



- Expected Value
  - (Teacher led) Spinner Excel simulation
- Normal Distribution
  - Shape applet
    - <http://mathnstats.com/applets/01-exploreNormalDistribution.html>
  - Probabilities applet
    - <http://wise.cgu.edu/sdmmod/pzapplet.asp>
- Sampling Distributions
  - Shape of sampling distribution applets
    - <http://www-personal.umd.umich.edu/~pksmith/JavaStat/CentralLimitTheorem.html>
    - <http://www.intutor.com/statistics/CLAppClasses/CentLimApplet.htm>
  - Shape, mean, and standard deviation of sampling distribution applet
    - [http://onlinestatbook.com/stat\\_sim/sampling\\_dist/index.html](http://onlinestatbook.com/stat_sim/sampling_dist/index.html)
- Hypothesis Testing
  - (Teacher led) Coin flipping applet
    - <http://www.math.usu.edu/~schneit/CTIS/PValue/>
  - Significance level and error applet
    - <http://www.intutor.com/statistics/CurveApplet.html>
      - Student explored spring semester; teacher led fall semester
- ANOVA
  - (Teacher led) ANOVA Excel Simulation
  - (Teacher led) ANOVA Excel Simulation 2
  - ANOVA applets
    - <http://www.psych.utah.edu/stat/introstats/anovafash.html>
    - [http://bcs.whfreeman.com/ips4e/cat\\_010/applets/anova.html](http://bcs.whfreeman.com/ips4e/cat_010/applets/anova.html)

- Linear Regression

- Correlation applet

- <http://www.stat.uiuc.edu/courses/stat100/cuwu/Games.html>
      - Moved to <http://istics.net/stat/correlations/>
    - <http://illuminations.nctm.org/LessonDetail.aspx?ID=L456#first>
      - Moved to <http://illuminations.nctm.org/Activity.aspx?id=4187>
    - <http://www.math.usu.edu/~schneit/CTIS/scorrelation/>

- Line of best fit applet

- [http://www.ruf.rice.edu/~lane/stat\\_sim/reg\\_by\\_eye/](http://www.ruf.rice.edu/~lane/stat_sim/reg_by_eye/)
    - <http://hspm.sph.sc.edu/courses/J716/demos/LeastSquares/LeastSquaresDemo.html>
      - Moved to <http://sambaker.com/courses/J716/demos/LeastSquares/LeastSquaresDemo.html>

- Outliers and influential points applet

- <http://illuminations.nctm.org/LessonDetail.aspx?ID=L456#first>
      - Moved to <http://illuminations.nctm.org/Activity.aspx?id=4187>
      - Student explored spring semester; teacher led fall semester

**Appendix I: Examples of Revisions Made to Instructional Materials**

## Changes to Instructional Materials After the Spring Semester

### Changes to the Activity Class

#### *Gathering Data*

- On “Lurking Variables”, one group was trying to think of a reason that mattress prices would affect professor salaries instead of looking for a lurking variable. I addressed that with the group. I changed the directions so that they explicitly say that temperature is a lurking variable in the first example. That should help students realize they need to look for lurking variables in the subsequent problems.
- The only issue with directions was for the group survey. Students weren’t quite sure what they were supposed to do. I revised the directions.

#### *Descriptive Statistics*

- The first semester, we did an activity where students sorted histograms into three categories (left skew, symmetric, and right skew). Students were given no prompting on what the categories should be, and they developed their own names. Unfortunately, all the symmetric distributions were mound shaped, and the students called it the normal or bell category instead of focusing on symmetry. To fix the problem and encourage students to focus on the symmetry aspect, I added some new symmetric distributions.
- I added a section on dealing with outliers. The “Dealing with Outliers” section came about because two students told me their height was 5 inches on the survey. I thought their data entry error would be a good way to bring up outliers.
- I removed the comparing boxplots section because students were trying to read too much into what the center or what the spread was instead of just getting a general idea from the boxplot. I think the topic actually goes better with ANOVA because then the students aren’t in the descriptive statistics mindset. During the descriptive statistics section, students are trying to think very technically and specifically about mean versus median, standard deviation versus range, etc.

- For the activity on symmetry, I originally prepared the figures on white paper, but then I redid the figures so that the symmetry activity was one color and the mode activity was another color. This helped the students keep them separate.
- I added a section on finding the standard deviation by hand. I didn't plan on including it, but we had an extra five minutes.

### *Probability*

- I realized that I asked the students to record proportion of heads, but the applet tells them the percentage of heads. I changed the worksheet to ask for the percentage of heads.
- I realized that I didn't specifically say at the beginning that you only draw one piece of paper at a time for the color activity.

### *Normal Distribution*

- I added more examples of real data sets and their percentages so that students can see a pattern in the percentages of data within one, two , and three standard deviations of the mean.

### *Sampling Distribution*

- Students were very confused when they tried to fill out this table.

Sample Size	Mean of all the sample means, $\mu_{\bar{X}}$	Standard deviation of all the sample means, $\sigma_{\bar{X}}$	$\sigma/\sqrt{n}$
2			
5			
16			
25			

○

- They tried to put  $\mu$  and  $\sigma$  instead of  $\mu_{\bar{X}}$  and  $\sigma_{\bar{X}}$ . Other students also calculated  $\frac{\sigma_{\bar{X}}}{\sqrt{n}}$  instead of  $\frac{\sigma}{\sqrt{n}}$ .
- I added some clarifications to the directions to clear this up, as well as an example where I filled out a table for them.

- I changed “Find the probability that the **total spending** is greater than \$5700. “ to “Find the probability that the **total spending for the 100 people** is greater than \$5700.” This helped reduce confusion.

### *Confidence Intervals*

- I should have only had each group toss the globe once instead of twice. This would have saved some time. Also, I had each of the six data collection pages in a line at the front of the room so that students could move down the line and record results for the NBA and dice activities. But I didn’t need to have them actually write the  $\bar{x}$  and  $\hat{p}$  values, and confidence intervals. They could have just drawn the intervals. This would have saved time. I fixed this in the directions for next time.

### *Hypothesis Tests*

- Introduction to Hypothesis Testing Activity
  - I left a few more things blank so that the students could supply more of the steps and answers.
  - I changed the wording so that students have to realize it is the binomial distribution, instead of my saying it is the binomial distribution.
  - I reworded a few questions. The questions on example 2 were in a weird order.
  - I changed the order of a few questions.
  - I changed from “accept” to “fail to reject”. I only used “accept” in the first place because that is the term that the applet used.
- I had to add a statement, “Western Family claims that their cookies are comparable to Chips Ahoy cookies” to the cookie example after a student asked if it was a typo that the problem mentioned Chips Ahoy but we were dealing with Western Family cookies.
- There was a problem in the activity, that had two correct answers.
  - “In the cookie problem, we got a p-value of essentially 0 (software tells us it was  $1.95 \times 10^{-32}$ ). What is an appropriate conclusion?”

- a.) We have proved that the population mean number of chocolate chips is 33.
  - b.) We have proved that the population mean number of chocolate chips is less than 33.
  - c.) We didn't find evidence that the population mean number of chocolate chips is 33.
  - d.) We found extremely strong evidence that the population mean number of chocolate chips is less than 33."
- I changed option (c) to "We found extremely strong evidence that the population mean number of chocolate chips is 33." because (c) was technically correct. The correct answer was supposed to be (d).
- I didn't specify the level of significance on the Jillian diet problem, or the Matt and John problem. A few students asked questions about what their significance level should be even though any significance level from  $\alpha = .10$  to  $\alpha = .01$  would have given the same answers. I fixed this on the worksheet.
  - I used a computer to find the p-value for Matt and the students asked how they were supposed to find  $P(Z < -5.27) = .0000000682$ . I added a note to the worksheet that I found it using a computer.
  - On the question, "How likely do you think it will be to actually decide the dice with  $p = .0333$  is unfair? (What is the power?)" students were trying to find an exact value. I changed the directions to specify that they don't have to find the exact value.
  - Students seemed confused about example four on the matched pairs. I added an introduction to the example.
  - I added normal quantile plots to the confidence interval part. The same data was used for the hypothesis test part at the beginning of the day, so I didn't think I needed to add the normal quantile plots again. However, three pairs asked me how they were supposed to check the assumptions without the plots. They didn't realize that the data sets were the same.
  - After working on matched pairs, students were supposed to use the computers and MegaStat to do the one sample t test. However, due to various factors, most of the class didn't get to the computer analysis.

- It takes students 5+ minutes to get the computer ready and download the files.
- Since I told them that the computer activity was only if they had time, they didn't try very hard once they got to that point. I heard a lot of talking instead of working once they finished the matched pairs. (I even told them at the beginning that talking was not a good excuse for not getting to the computer activity.)
- Maybe half the students actually did the computer analysis on the first problem. But no one finished, and I really don't think anyone wrote down anything for the computer activity.
- The matched pairs part of the activity seemed like a good activity. For next time, I put in more matched pairs problems to replace the computer activity.

#### *Drug Activity for P-values*

- I didn't think the drug activity really helped with understanding the chi-square tests. I changed it to focus more on the relation to p-values and address proportions instead of the chi-square test of independence.
- I added an extra part to the last page: If the two treatments are equally effective, and the only difference in the success rates is due to who ended up in which sample, then there is only a \_\_\_\_\_ chance of getting 14 or more successes for Desipramine.

#### *ANOVA*

- \*\*I moved the computer problems after the non-computer problems. This semester, students didn't have time to get to the regular problems because the computer problems took too long. This way, they will have practice on problems similar to the test and homework and then any practice on the computer will just be icing on the cake.

#### **Changes to the Lecture Class**

##### *Gathering Data*

- I changed the flow and order of the gathering data unit.



### *Probability*

- I completely reordered the probability unit. The current organization felt like we were jumping from topic to topic and was confusing. This is a note from my teacher log, “I really think I need to reorganize the probability unit. I feel like we are jumping around in topics and it just hasn’t seemed to flow well. I think it should be probability, conditional probability, and then the multiplication rule for dependent and independent events.”

### *Random Variables*

- I added some more steps and practice problems to the expected value and variance section.

### *Normal Distribution*

- I redid the working with the z table” section. I needed to put the similar problems side by side instead of in a column.
- There was no section on finding z value for a specified probability. There is a section on finding X values, but not for finding z values. I had to do an impromptu short lesson with four problems for the topic. I added a section to the guided notes for next time.
- The first problems in “Finding Probabilities for any Normal Distribution” were in a bad order. The first problem had two X values which is more confusing. We needed to have a simple problem first.

### *Sampling Distribution*

- I redid the notes so that they discover  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ . I made it more like the activity class.
- I added more examples of the graphs of sampling distributions for different sizes so that they would have the graphs in their notes for future reference. Otherwise, the students would just watch me explore the applet, but they would have no visual reminders of the shapes of the graphs.
- I didn’t like the lecture on finding probabilities for sample means. I changed the section.

- I added questions about the sampling distribution before I asked questions about probabilities.
- I told them the new  $z$  formula,  $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ , before we started doing finding probabilities.
- I broke the problems down into more steps so that the lecture was closer to the activity.

### *Binomial Distribution*

- Once again, I didn't feel happy with the lecture. Since I decided to not make students find binomial probabilities by hand, we had to change the 69% to 70%. Also, all of the problems had probabilities higher than 50% so we had to use the table backwards. I did throw in an impromptu extra problem with a low probability. I changed the section so that we will do several problems with small probabilities then moved to the high probabilities. Also, I put a copy of the binomial probability table on each problem so that students could easily reference it.
- I fixed the wording on the example that asked if there was evidence to make it more intuitive.

### *Confidence Intervals*

- I didn't like the introduction to confidence intervals that I used for the lecture class. Among other things, it was confusing to bounce from  $n = 25$ , to  $n = 5$ , to  $n = 25$ .
- I used the introduction to confidence intervals that I created for the activity class as inspiration to completely revise the introduction for the lecture class.

### *Hypothesis Tests*

- I had accidentally labeled everything as coins not dice on introduction.
- The note on page 112 was confusing when compared to the bottom of page 109. The note in the reference pages was much better. I should tell students, "the level of significance is how much evidence we require to reject  $H_0$ . The p-value is how much evidence we actually found against  $H_0$ ." Students brought up several questions about this. I added a page to the notes about interpreting the amount of evidence and the p-values.

- I added an example where the decision changes based on the significance level after a student asked about it.

### *ANOVA*

- I changed the assumptions to match the assumptions from the activity notes.
- I removed the  $Data = Fit + Residual$  and finding residuals portion because it didn't fit well with the ANOVA lesson. It would have made more sense if we had already done the linear regression section. (The textbook does linear regression before ANOVA.)
- On the comparing boxplots, I changed the plots so that they show the mean because it gets confusing when I talk about comparing the means and students can't see the means.

### *Linear Regression*

- I reworked the advanced linear regression section. It took too long to get through the information to introduce the advanced regression topic and then the students forgot it all by the time we tried to apply it. I changed it so that we introduce a piece, try to apply it, then introduce another piece, etc.

## **Changes to Both Classes**

### *Probability*

- I decided to add an extra handout about using the complement rule for “at least one” situations after hearing the questions and comments in both classes.

### *Random Variables*

- I had planned to be done with random variables, but I didn't think the students had a good enough grasp on the rules for expected value and variance. So I made an extra handout with extra problems for rules of expected value and variance.

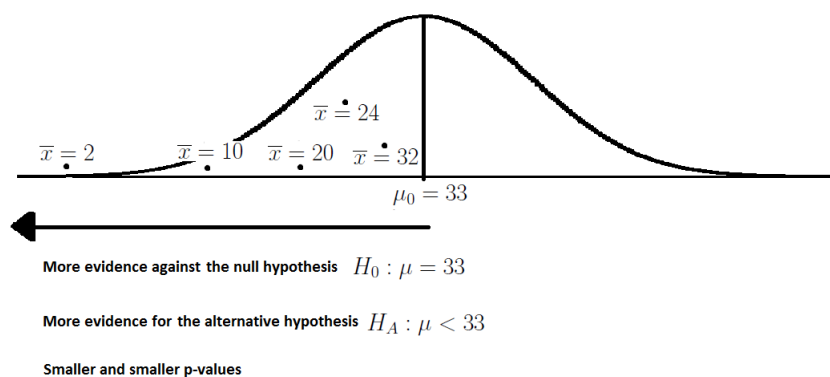
### *Binomial Distribution*

- Originally, I followed the structure of the course and taught the binomial distribution after the sampling distribution for the mean, and before the sampling distribution of the proportion. I decided that this was not logical. When we were working with the binomial distribution we needed the skills we learned in the random variables unit. In fact, it felt like an extension of the random variables unit; of course, that made it feel out of place in the sampling distribution unit, leading to extra confusion. For the second semester, I moved the binomial distribution to right after the random variables. The revised order of topics is random variables, binomial distribution, and then the normal distribution.

### *Hypothesis Tests*

- After class, a student asked me why small p-values mean stronger evidence. So I came up with this explanation. It worked well, and I used it later in both the activity and lecture classes and integrated it into the future notes.

#### **Sampling Distribution of $\bar{X}$**



- For two sample tests, a student wondered if both sample sizes have to be large or both have to be normal, or if it can be one of both. I changed the wording in the guided notes and activity packets to reflect that.
- I've decided that instead of saying  $\mu_1 < \mu_2$ , I should be in the habit of saying  $\mu_{CA} < \mu_{WF}$  and being more descriptive in my labels. I've been working on changing that for the notes and keys.

### *Chi-Square Tests*

- On the Chi-Square Goodness of Fit page, I added a note that it was a completely different test from the  $\chi^2$  test for variance. (One student asked me after class why we were relearning it.)
- On the official page for chi-square test of independence, I added a line that you use the test when “we want to know if the two variables affect each other (if there is a relationship).”

### *Linear Regression*

- I added “ $r^2$  is a positive number between 0 and 1. The higher the  $r^2$  value, the better our line fits the data.” to the activity and lecture notes.

### *Review for Exam Four*

- In the example for the leaning tower of Pisa, we coded the year 1974 as 74, etc. But in the world record 10k problem, we coded year 1974 as 1974. This difference caused a lot of confusion. I added a statement to the directions for both problems to hopefully address that.
- In example 37, the word “controlled” was confusing for students. I removed it.
- I had a couple students ask me questions about the conclusions. They thought that the conclusions “we didn’t find any evidence that the variables are independent” meant the same thing as “we found evidence that the variables are dependent”, etc. So I created the handout “evidence” for both classes. The handout also talks about interpreting the weight of the evidence for significance level versus p-value. I integrated the handout into the future notes in the hypothesis test section.

## Changes to Instructional Materials After The Fall Semester

### Changes to the Activity Class

#### *Descriptive Statistics*

- Since students seemed to have some issues with the histograms on the previous day, and I knew I would have extra time today, I made a extra handout with practice for histograms and boxplots. I included the fact that there is no rule on how to pick classes since that threw students last semester.

#### *Probability*

- There was a big typo on an extra problem packet where I had replaced the problem, but accidentally kept part C of the old problem.
- I added an example where I compare  $P(A | B)$  and  $P(A)$  to check independence before I ask them to try it.
- I changed the last question of the cola problem to be closer to the lecture. Instead of comparing  $P(C1 | S)$  and  $P(C1 | V)$ , I had them compare  $P(C1)$  and  $P(C1 | S)$ . This way, they won't be confused about what they have to compare when checking independence.

#### *Random Variables*

- I added an example for the discrete versus continuous section that is discrete but not whole numbers. Most students believe that the definition of discrete data is to be whole numbers.
- I saw many students today trying to find the expected value on the extra problems without making a table first of possible values and probabilities. So I have now changed the worksheets so that I added in a blank table for them to fill out for the first 3 extra problems.

#### *Binomial Distribution*

- I added a few more steps to the first problem with tables to try to make it more obvious what we are doing when we switch from  $p = .3$  to  $p = .7$ .

- For next time, I made parts (d) and (e) for the strikes problem optional, to be done if there is extra time. (Last year I had them skip it.) This will save probably 5 minutes allowing everyone to move on and finish the packet.

### *Sampling Distribution*

- I made minor changes to the directions for the sampling distribution, mean and standard deviation subsection, to try and clarify a few things.
  - Remember to “Clear Lower 3” to reset the results.
  - Create any new population. (There is no specific population.)
  - Emphasize that they are computing  $\sigma/\sqrt{n}$  where  $\sigma$  is the population standard deviation.
  - $n$  is the sample size.
- I added two examples of finding the necessary sample size for a desired  $\sigma_{\bar{X}}$  to the activity book after I realized there were no examples for that topic.
- I added an example where they find an individual probability and then a sample mean probability to the activity book as an extra problem.

### *Hypothesis Testing*

- I moved the extra handout for hypothesis testing to after the errors section in the activity book. It seemed out of place.
- On page 193, problem 345, I used a computer to get the exact p-value of .1807 which confused a couple people since it didn’t match the value on the table. I added a note about this.
- For problem 352 with the dice and errors, I added something to the key about the expected values for the two dice on the type II error question. This was a problem that the students seemed to struggle with.
- For matched pairs, I changed the garage example to be a one-sided alternative hypothesis. That way I can show them how to set up the one-sided hypothesis, which is harder for matched pairs, in the example that I do with them.

- I added more examples for matched pairs.
- I put the activity's introduction to the proportions hypothesis test in the guided notes for next time because I liked it so much. Specifically, "What is sampling distribution of  $\hat{p}$ ?" and "What is the test statistic following the usual guideline?"

### *ANOVA*

- I changed the directions for the first applet. The directions stated that students should use the sliders to change the standard deviation, but they actually needed to use the yellow buttons.
- I changed the directions for the second applet. A pair of students pointed out that no matter how far apart they moved the means, the F statistic and p-value didn't change. When we looked at the applet, the default setting has the means so far apart, that p-value is already minimized. I added a line telling students to start with the means close together and the standard error in the middle so that they can watch the F and p-value change.
- I realized that for the second applet I asked students to sketch the boxplots, but they weren't given boxplots. Instead they were given a type of dotplot.
- On the typical ANOVA table, I added "area in the right tail" for the p-value spot instead of leaving it blank.
- I added a spot for "Why do you think the ANOVA F test is always a right tail test?" at the beginning of the problems, after they have done the applets.
- \*\*I made an extra copy of the computer problems for ANOVA that does not require the computer. It always takes so long to get the computers out and working (usually 5-10 minutes), that it isn't always beneficial. Also, many students this semester only got them out in the last 5-10 minutes, so they didn't actually have time to work on the computer activity. This way, if students have enough time, they can do the computer activity. If students are slow, they can do the same practice problems, but they use the computer output given to them to save time. So I now have two potential activities depending on the time and computer availability.



- I added a couple more directions to the computer activity to hopefully help facilitate using MegaStat.

### *Chi-Square Tests*

- We didn't count M&Ms for the chi-square goodness of fit.
  - Last time, I noticed that students got tired of M&Ms part way through the semester.
  - Since so many people were throwing away food this semester, I just had students write down all the colors last time when we counted for the proportions and I saved the data.
- I told students to do the M&M problem with their data last since there was no key and I thought it would be easier to do a problem with the key first to make sure they knew how. So for next semester, I moved the M&M problem back to the end of the section.
- I changed if the “alternative hypothesis is true” to “if the null hypothesis is false” on the second page
- I added “k is the number of groups” right below chi-square test statistics.

### *Linear Regression*

- I really needed a new data set for introducing the line of best fit. When students draw their line of best fit, it looks the same for all the students, so the example doesn't make the point that I was hoping for.
- On the first problem that they have to interpret the components of the line, I added a hint to refer them back to the page that has the official definitions.
- I added a hint the first time they had to worry about units such as thousands of steps.

### **Changes to the Lecture Class**

#### *Probability*

- I redid the independence page because some of the examples were redundant.

- I added more space to the multiplication rule for independent events page so that students would have space to work the problems.

### *Random Variables*

- I moved the example for continuous versus discrete up so that it is right after the continuous introduction. This improved the flow of the lecture.
- I added a problem to find Linda's variance and then added the graphs for cars and trucks. This way, we could compare the variances and means.
- After a student's question, I made an extra graph with both mean and standard deviation equal to five. We can talk about the coincidence of having the same mean and standard deviation and how it is possible to not get negative values in that situation if you have a skewed distribution.

### *Binomial Distribution*

- I added some more space and explicit directions for going over the binomial probabilities tables backwards.

### *Normal Distribution*

- I added a statement about the  $z$  value being the standard deviations from the mean, just so I don't have to worry about forgetting it in future lectures.
- I changed  $P(z \geq 1.25)$  to  $P(Z \geq 1.25)$  to be more in line with conventional notation even though the book doesn't use the uppercase notation.
- I added a problem for the students to find the middle 90% of ISTEP scores on example 109 (page 74) so that they have to find two  $x$  values.
- I added some histograms to go with the normal quantile plots. Students always wonder why the normal curves have straight lines on the normal quantile plots when normal curves are "curved". I think the activity class understood this better in the past because they saw the histograms and normal quantile plots together.

### *Sampling Distribution*

- I changed the order of the applets so that we look at the simple one with just shapes first and then include means and standard deviations.
- There were no examples in the lecture about not being able to do a problem because the sample size was too small and the original distribution wasn't normal. I added one example.
- There were no problems in the lecture about finding the  $\bar{x}$  value for a specified probability. I added two examples.
- I also changed the wording slightly on the billing problem on the lecture so that hopefully it will be more intuitive as I try to make them start thinking about hypothesis tests.
- I reworded some of the bullet points in the lecture to match the activity better and removed the subsection about counts.

### *Confidence Intervals*

- I added a step to have students check the assumptions and broke the examples into more steps.

### *Hypothesis Testing*

- I added a step to interpret the p-value on one of the first few z test examples.
- I put the comparing two population means from the activity on page 227 into the guided notes.
- I added a couple more examples for two sample tests to the guided notes.

### *ANOVA*

- The assumptions for the ANOVA F test were stated better in the activity, so I copied them to the guided notes.
- I added a note that ANOVA is always a right tail test. I always say it, but it is nice to have it typed on the official test page.

- I added a spot for us to discuss why the ANOVA F test is always a right tail test.
- On the typical ANOVA table, I added “area in the right tail” for the p-value spot instead of leaving it blank.
- I moved the coefficient of determination,  $R^2$ , from the official page to after the first problem because I think there was a lot of theory before we got to any examples.
- I added checking the assumptions to the fruit fly example. This way they will have seen it once.

### *Chi-Square Tests*

- I added a snapshot of the binomial probabilities from the table to the gender problem so that students will have it in their notes.

### *Linear Regression*

- I added the plots that I had in the activity notes for extrapolation and impractical y-intercepts to the guided notes. I had drawn them in by hand each year, but I think it is nicer to have the plots already there and they look better when not hand drawn.
- On the first gas problem where we interpret all the components of the line, I didn’t have them find correlation, but I wrote it in by hand each semester. I added a step for finding correlation to the guided notes.
- On the home value and upkeep example, I had them try to do a prediction that was extrapolation. Then they found and interpreted  $r^2$ , and then they filled out a table with various home values to decide whether or not it should be extrapolation.
  - I added an extra column to force students to compute the  $x$  value for home worth before decided whether they should predict the upkeep value or if it would be extrapolation.
  - I moved this to after the extrapolation problem and before the  $r^2$  problem.
  - I changed \$300,000 to \$320,000 so that the  $x$  value is close to the existing data range, but not in it.

- On the outliers and influential observations, I added a plot that is an outlier, but not an influential point.
- For the beginning of the advanced regression section, with population parameters, I replaced the guided notes section with what we had in the activity section. They were similar, but the activity section was a little better.
  - I transferred everything over up to and including the first example since the activities was set up better.
  - The activity was set up better because students were reading it themselves with no help from me.
- In assumptions for test for slope, I added a note that instead of checking the formal assumptions, they are just looking at the three plots to make sure linear regression is appropriate.
- I added the page for estimating the regression parameter with the examples of four different samples from the same population. I used it during the lecture classes both semesters, but it wasn't in the guided notes.

## Changes to Both Classes

### *Probability*

- I realized that on examples for using the condition probability formula, that I asked them to find  $P(A | B)$  and I gave them  $P(A \cap B)$  and  $P(B)$ , but I never gave them  $P(A)$ . So it made it really easy to determine which probability to divide by in the formula. Students were just making sure that they divided by the bigger number. So I added in some extra probabilities to the examples.

### *Normal Distribution*

- I separated the problem for finding probabilities with gas mileages into more steps to make the reasoning behind the future hypothesis tests more clear. I now have students find the probability first, and then I say that they found a car with a gas mileage of 31.2, so do you think they are telling the truth?

### *Review Exam 3*

- The review has a problem where students conduct a T test on test scores that are given in percentages. No one said anything during the spring semester, but during the fall semester, several students in both classes expressed concerns about the percentages and whether they should change them to decimals. I explained that percentage in this case was just the units and didn't affect our calculation, but it still threw the students enough that I have changed it to points for the future.

### *ANOVA*

- I also added a spot to talk about the heights between men and women when introducing  $R^2$  because it seems like a very intuitive example.

### *Chi-Square Tests*

- In the development of the chi-square test of independence I switched from saying  $P(A \cap B)$  to  $P(A \text{ and } B)$  since our book doesn't always use the " $\cap$ " symbol, and I figured students might not remember what it means.
- I added the following comments to the degrees of freedom part on the official chi-square test for independence page so that I don't have to write it by hand.
  - $(\text{number of rows} - 1)(\text{number of columns} - 1)$
  - Don't count the totals.

### *Linear Regression*

- For the definition of a residual, I changed the order from predicted value minus observed value to observed value minus predicted value so that the order matches the formula.
- In the assumptions for the T test for slope, I changed errors to residuals since that is what they are more familiar with.
- In the extra problem for world record times, I added a sentence so that students could infer from the problem statement which were the explanatory and response variables.



## Appendix J: Examples of Teacher Logs



## Examples of Teacher Logs

### Lecture Class

#### *February 1—Expected Value and Variance*

- We started on rules for expected value and finished rules for standard deviation.
- We had 10 minutes left and I didn't want to start the normal distribution, so we did two problems from the homework. I figured it would be good practice because it would help firm up the concepts in the students' minds.
- I had two students make comments about how it seemed confusing, but they thought they just needed more practice. I think the difficulty is mostly due to the new notation.
- I had a few students say they liked doing the homework problems at the end, that it helped their understanding.
- I've decided that I really should have introduced the symbols  $\mu$  and  $\sigma$  before now.
- I think the guided notes need a little more reworking with more steps and some more practice problems.

#### *February 12—Central Limit Theorem and Binomial Distribution*

- Topics: pages 76 through example 108 on page 80.
  - Finish central limit theorem.
  - Sample counts.
  - Binomial distribution-identify.
  - Binomial distribution-probabilities.
- I decided on the fly to skip using the formula to find binomial probabilities and to just use the table since we were getting behind.
- It took 27 minutes to get through pages 76-78.
- This lecture itself was fine. Nothing really needed to be changed.

*October 1–Sampling Distributions*

- Pages: 76-79.
- Timing:
  - 6 minutes, Populations and samples, pg 76.
  - 4 minutes, All possible samples & sampling distribution, pg 76.
  - 23 minutes, Finding the sampling distributions, pg 77 and 78.
  - 8 minutes, Applets, pg 79.
  - 9 minutes, Results from applets, page 79.
- Next year, change the order of the applets so that we look at the simple one with just shapes first and then include means and standard deviations.
- Two students asked what  $n$  was and what the 10,000 reps were.
- Tomorrow, I need to go over the applet again quickly.
  - $N$  is the sample size.
  - We are drawing 10000 different samples of size 2 each and plotting each mean.

*November 20–T test for Slope with an Integrated Review of Linear Regression*

- Timing:
  - 6 minutes, extra handout, estimating the regression parameter: four samples from same population.
  - 9 minutes, t test for slope, pg 200.
  - 29 minutes, NEA and fat gain, first example, pg 201.
  - 6 minutes, start steps per day and body mass index, pg 202.
  - We couldn't finish the second example.
  - We are about the same point as last year, maybe just 10 minutes slower.

- Compare classes: the lecture class has only seen one test and one confidence interval so far and we aren't done. The activity class is much further along. They finished everything today and they had more examples to cover.
  - However, due to the lecture format, I can point out lots of things to my lecture class, but not my activity class. Examples include:
    - $R^2$  is useful to tell you how good your predictions will be.
    - Linear regression can't establish cause and effect! You need an experiment instead.
    - Linear regression is good for predictions.
    - Predicting a mean response is more accurate than predicting an individual response even though the process is the same.
    - We don't bother with the t test for the intercept because we usually don't care about the intercept, and forcing it to be zero usually ruins the fit of the line.
  - Students in the lecture class are also asking lots of questions, slowing us down. Some of the questions are irrelevant or truly silly.
    - For example, I told the students that they aren't required to use the different categories of strong, very strong, and extremely strong evidence based on the p-value. However, the categories are written in the guided notes. But one girl asks me what the categories are at least every other week. The same girl has also asked me questions such as "so what was  $t$ " when I have written on the paper  $t = 3.62$ .
  - The last few weeks, there are more questions in the lecture class than previously.

## Activity Class

### *January 18-Conditional Probability*

- Topics: Conditional Probability, and Conditional Probability & Independence.
- Timing:
  - Students worked in pairs for the entire period.
  - At 8 minutes, several groups were on the third page, "Conditional Probability and Independence".

- At 14 minutes, most groups were on page 4 or 5.
  - At 29 minutes, everyone was on page 6 with the cola problem.
  - At 40 minutes, three pairs were done, one pair was on page 6, but the rest were on page 7 or 8.
  - At 50 minutes, everyone at least got to the extra problems.
- I found some typos.
- Several groups called me over because they were trying to find conditional probabilities using a reduced sample space and they thought they were doing something wrong. I explained that using the reduced sample space is an easier method, but not required, so I didn't require the students to learn it to reduce confusion. I estimate that at least half the pairs started using the reduced sample space on their own.
  - On page 5, where they were supposed to find the probability of being faculty if you know you are female, I showed my work in many steps. The students used the reduced sample space and jumped to the final answer, but they compared their final answer to one of my intermediate steps and they kept wondering why they were wrong.
- One thing I have noticed is that the students aren't writing out the probability symbols for their problems. I think this is because they haven't watched me write out my work. I have shown my work on the keys, but they aren't incorporating the symbols in their work.
- I think today went well. In the future, it would be nice to somehow enforce the concept that the definition of independence requires the comparison of the conditional and unconditional probabilities. Several groups asked me why they were comparing the conditional probabilities. There was a point on the worksheet where I told them that the formal definition of independence required comparing  $P(A|B)$  and  $P(A)$ , but by the time they got to the problems, they had forgotten.
- One person asked for more keys.
- I think the activity was a good length, although I could add another extra problem for the fast students.

- I asked the class at the beginning if they preferred pairs or groups of four. The majority voted to work in pairs not groups.
  - No one raised their hands when I asked about working in groups and over half raised their hand when asked about working in pairs.
  - The pairs still seem to be working well.
- Based on comments in the lecture class and the activity class, students still aren't getting the at least one concept. I really need to make a good activity where they discover this concept.

*January 31-Expected Value*

- Timing:
  - Discovery: We spent 5 minutes on the DJ problem and informally discovering the formula for expected value.
  - Discovery & Simulation: We then spent 7 minutes on the spinner problem developing our intuition about the meaning of expected value.
  - Lecture-Algorithm: We spent 6 minutes going over the formal definition of expected value and my walking them through the formula to find the expected value of the spinner problem.
  - Students started working in pairs at 12:21.
  - At 12:27, most people were on page 4, but some were on pages 3 or 5.
  - At 12:42, people were on pages 6-8.
  - At the end, everyone was somewhere on the extra problems, but no one finished all the problems. But that is okay, since they are extra problems.
- The students responded very well and seemed to really understand the material. They were giving the "correct" suggestions and ideas very quickly.
- I really didn't like having to switch back and forth between the document camera and computer to simulate spinning the wheel. It wasted too much time.

- I could have let the students do the discovery part on their own, but I wanted to be able to point out several small nuances that they would probably miss. In addition, this is one of the first real algorithms and I thought students might need direct instruction to get started. If nothing else, the notation is complicated and confusing the first time students see it.
- I heard some interesting discussions on whether they would rather bet \$1 on 1000 bets or \$1000 on one bet.
- It was a good day.
- I had 5 people comment on the exit slips that going over something at the beginning together helped them understand the material better.
- The discovery lesson was very different from what I did with the lecture class even though the class was focused on me.

*Tuesday, February 5—Normal Distribution*

- Topics: Normal Distribution.
  - Finding probabilities from z scores and vice versa using the table.
  - Standardization.
- Timing:
  - Lecture-introduce table: I spent the first 3 minutes introducing the standard normal table and I did the first problem with them.
  - Then students worked in pairs.
    - At 12:10, most students were on pages 2-3.
    - At 12:22, most students were on pages 3-5.
    - At 12:31, most students were on pages 4-7. (There were only 7 pages.)
    - At 12:34, two pairs were done, but lots of people were still on pages 4-5.
  - Six pairs were done and left at 12:41.
  - A few pairs were on page 5 when class ended, but everyone else was done.

- I did a 30 second recap before they left.
  - The table gives areas to the left.
  - The table is in the back of the textbook. (I had forgotten to mention this earlier.)
  - The z-score is always the number of standard deviations from the mean.
- During group work, one student said “is it always true that  $z = 3$  is three standard deviations from the mean?” I said yes and he said, “cool”.
- Several people didn’t pay attention when I said there were two sides to the standard normal table, so as they started working, I heard a lot of comments of, “Oh, I didn’t realize it had two sides”, or “It helps if you look at the other side”.
- The students had a hard time trying to figure out how to find the in between probabilities and the z values for probabilities in between  $-z$  and  $z$ , but I think they understand it better now than the lecture class does.
- I am positive that they understand that the z-scores are the number of standard deviations from the mean better than the lecture class students.
- Today is the first time since the first week that all 33 students actually showed up. It is also the first time that I only made 32 copies.

### *February 12–Binomial Distribution*

- Topics: Introduce binomial distribution, how to identify it, and discover formula to find probabilities.
- The students were supposed to have time to work on identifying binomial distributions and using the formula, but we ran out of time. We only did the class activity.
- Activity: 9 minutes, root beer taste testing experiment.
  - When the students came into the room, I had three varieties of root beer poured into cups. The cups were labeled 1-3.
    - Walmart \$.84 (labeled #2)
    - A&W \$1.24 (labeled #3)

- Barq’s \$1.25 (labeled #1)
  - We had 10 taste testers try to identify which brand was the Walmart brand.
  - I introduced the activity by asking the class which brand they liked, if they thought they could tell which one was the cheap Walmart brand by taste, and if they would buy the Walmart brand.
  - We decided to see if the students could tell which brand was the Walmart brand.
  - 2 out of the first 3 students correctly identified the Walmart brand.
  - 6 out of 10 students correctly identified the Walmart brand.
  - The first 6 volunteers were very quick to volunteer and then I got 3 more when I asked again, but I had to ask a few times to get the last girl.
  - I told the students to write their choice on the piece of paper and not to talk to anyone so that their results would be independent.
- Lecture: Discovery (pages 1-2).
  - Then we spent 26 minutes going through the probabilities and probability distributions for  $Bin(3, .333)$  and  $Bin(10, .333)$ .
  - This was a good review of finding probabilities and probability distributions. We talked about the probability of having 6 out of 10 students guess correctly and whether they thought the students were blindly guessing.
  - This was a good introduction to the theory underlying hypothesis testing and it came very intuitively to my students as we did this activity. Unlike in my lecture class where they still stare at me blankly when I try to work in the intuitive hypothesis testing theory.
- Lecture: Introduce (6 minutes).
  - Introduce sample counts.
  - Introduce requirements to be binomial distribution.
    - One example to identify binomial distribution.
- Lecture: Introduce Algorithm (9 minutes).



- Introduce formula for binomial distribution.
- This requires talking about the combination and factorials.
- The students were very comfortable with everything but the combination. They were much better with this than the lecture class since we spent so much time discovering the formula. I kept trying to emphasize that they were able to come up with the formula, so they should just think through the problems instead of getting caught up in the complicated looking symbols in the formula.
- We then did the problems with milk cartons.
  - Is it binomial?
  - $P(X = 2)$
  - $P(X < 2)$
- Even though it took a lot longer than I had been expecting, it was a really good activity.

#### *August 28—Describing Data*

- Topics: categorical/quantitative variables applet, graphing categorical, living histogram and boxplot, and graphing quantitative variables.
- Timing:
  - 3 minutes, class business and introduction.
  - 17 minutes, applet and groups worked on graphing categorical variables.
    - This should have been much faster, but we had technical difficulties.
    - No one was logged in yet at 4 minutes. People started getting to the applet at 7 minutes.
  - 2 minutes, recap: what are variables, labels versus quantitative, category not like the others at the end.
  - 8 minutes, class activity: living histogram and boxplot.
  - 17 minutes, group work on graphing quantitative variables and finish categorical variables if necessary.

- 3 minutes, recap/clarify for issues on histograms, also discussed the difference between bar graphs and histograms.
- Not everyone finished the histograms, but they were all at least close.
- Some people finished the extra handout for graphing quantitative and categorical variables.
- The timing was okay, but we had trouble with the computers, so we should have had more time for graphing quantitative variables.
- Categorical/Quantitative Variables Applet
  - I warned students in advance that the applet would only take 5 minutes, so no one complained about having to get out the computers for a short exercise.
  - The applet worked well again; the cognitive dissonance worked well with the drivers license and zip code.
    - This year the students didn't ask me why the drivers license number wasn't quantitative like they did last semester, so I went around and asked them.
- Living Histogram and Boxplot
  - This went fine, but I ran out of tape.
  - I didn't get as many positive comments as last semester, but student attitudes were still positive.
- Graphing Quantitative Variables
  - The students had lots of trouble with this.
  - They were trying to find their counts without writing down the classes, and so they were messing up.
  - They didn't understand the directions for class width and lower boundary on the changing classes page. ( Students didn't have trouble with this last semester.)
  - So I stopped the class to do a recap during the last 3 minutes and pointed out a few things they had been messing up on.

- A few students were confused about how to find the 5 number summary. I need to clarify in advance that it was found using software and they won't be finding it themselves.
- In the future, remember to point out the difference between histograms with counts and percentages.

*September 3—Standard Deviation (Describing Data)*

- We went to the Ag Science computer lab since we were going to be working with the applets.
- We had trouble getting the applet to work.
  - Students were able to use the applet on Chrome, but Chrome wasn't on the teaching computer so I couldn't demonstrate the applet.
  - Also, Canvas was down, so I couldn't post the link to the applet and we had to search for it.
  - Since every student had a computer, it was harder to get them to work in pairs. Maybe 1/3 to 1/2 worked in pairs.
  - The applet took longer and was more difficult to do in the computer lab than in the regular classroom.
- Timing:
  - 44 minutes, activity and applet: students worked in pairs on the standard deviation applet and the standard deviation flashcards.
    - It took a lot longer this year for students to finish the applet part.
    - Almost all students had time for flashcards. The fastest students had about 10 minutes for the flashcards.
  - 3 minutes, recap: what does standard deviation mean and some of the interesting flashcards.
  - 7 minutes, lecture: how to find standard deviation by hand.
- Students were supposed to use the applet to compare the mean and standard deviation of the two data sets:

- Set 1: 4,4,4,4,5,5,5,5,6,6,6,6
- Set 2: 1,2,2,3,3,3,4,4,4,4,5,5,5,5,6,6,6,7,7,8,8,9.
- A student wanted to say that the second set had a higher standard deviation because there were more observations. I need to change the problem so that both sets have the same number of observations so that it isn't an issue.
- One student was using different scales when comparing the different data sets. I need to change the directions to make it clear that they need the same number of rows and columns on each window.
- I still like this activity because the students get an intuitive grasp of standard deviation.
  - I have the students try to develop an intuitive sense of standard deviation before we compute it.
  - I only have them actually compute standard deviation once for their homework. I think it is useful to see the formula once to help with understanding the meaning of standard deviation. However, there is no point in actually calculating standard deviation by hand in the future.
  - As one student commented, "I found it helpful that we first learned what the standard deviation represented and what affected its values before we actually jumped into how to find the standard deviation".
- There were favorable comments for both the applet and flashcards.

*October 7—Confidence Intervals intro*

- Timing:
  - Lecture-Discovery, 18 minutes: confidence intervals introduction pg 151-152.
  - Lecture-Introduce Algorithm, 3 minutes: earth example pg 153.
  - Group Activity, 13 minutes: students toss globe and compute confidence intervals, pg 153.
  - Lecture-Recap Results, 3 minutes: go over earth results.

- Student Activity, 11 minutes: collect NBA and dice results and record them.
- Lecture-Recap Results, 2 minutes: go over NBA and dice results.
- Students seemed concerned this semester about where the formulas and the critical numbers 1.645 and 1.96 came from.
- I think I could take out the dice example to save time.
- I had the students record their  $\hat{p}$  and  $n$  for the earth example and bring me the slip of paper. Then I used a spreadsheet to compute and record the confidence intervals while the students worked on computing the confidence intervals. This worked out well and saved time.
  - Last year, I had students compute the confidence intervals and then come to the board to record them.
- I told the students if they had extra time on the earth problem to start the NBA and dice problems, but I didn't introduce the problems until after we went over the earth results.
- Most people had lots of extra time while students were recording the NBA and dice results. I had students work in pairs for that.
- Once again, we ran out of time on the recording NBA and dice results and I had to stop it to go over the results. Even though I told them not to, students stood in a line and recorded each page one at a time which takes forever.
- Overall, I think it went well, but I felt like I was supplying all the discovery answers. The students aren't responding as well this semester.

#### *October 23-T Test*

- Timing:
  - Lecture-Thinking: 6 minutes, picture and explanation. I changed it to the true mean of 20.8.
  - Lecture-Discovery: 3 minutes, what do we do if we don't know  $\sigma$ ? We need to use a t test, pg 202.

- Lecture-Introduce: 3 minutes, introduce formal one sample t test for  $\mu$ , pg 202.
- Lecture-Example & Introduce: 11 minutes, do an example of a t test with the cookies, introduce using t table for p-values, pg 203.
- Work in Pairs: pg 204-209.
  - At 10 minutes left, most students were on pages 205-207.
  - Some students finished about 5 minutes early.
  - Lots of students were on at least page 207 at the end of class.
- We got a huge p-value for our cookie problem which really slowed us down on the example. That is why it took 11 minutes. But it did make for a good discussion.
  - Should I change this problem for the future to get a small p-value for a faster example?
  - Plus, the big p-value wasn't really helpful for learning how to use the T table.
- Lots of students came 5 minutes late, two girls came 10 minutes late.
- I told students how to tell the difference between a Z and T test at least 3 times during the discussion, but someone still asked me during group time.
- I took the extra cookie problem in case they have extra time. Everyone left early last year. But no one did it this semester even though some finished early.
- I told them to read “interpreting the results of hypothesis testing” on page 209 at home if they ran out of time.

#### *November 18-Linear Regression*

- Timing:
  - 3 minutes: Discovery, I gave the students a minute to draw their line of best fit and then had three students show their results, pg 292.
  - 5 minutes: the students worked in groups of four to determine their own criteria for deciding the best line to fit the data, pg 292.
    - I gave them rulers and the graph of the data so that they could develop their methods.

- I spent time walking around and talking to the groups.
  - 3 minutes: one student from each group shared their method with the class, pg 292.
  - 2 minutes: Lecture-Introduce, I introduced least squares method, pg 293.
  - Students spent the rest of the time on the packet, pg 293-300.
  - Most students didn't finish.
    - A couple finished, but the rest were spread out from pages 296-299.
    - I told them we would do page 300 together in class tomorrow.
- Two of the applets links no longer worked. Luckily I realized this last night and was able to find the new locations of the applets.
- Responses to brainstorming the line of best fit:
  - One group found average of many possible lines.
  - Most groups just put 4 points above the line and 4 points below the line.
  - One group did half the points above and below the line, but they wanted to somehow compensate for how far the points are from the line.
  - After they reported, I drew an outlier and asked if that should change the line, and then pointed out that it wouldn't according the half and half rule.
  - I also drew 2 points close to, but below the line, and 2 points far above the line and pointed out that they should affect the line, but they wouldn't according to the half and half rule.
  - Then I introduced least squares, emphasizing that that it wasn't the only method, just the most common.
  - They weren't as inventive this semester.
- They realized that they really had to pay attention to the directions for the applets today. Usually they try to just start working without reading the directions.
- I forgot to hand out the keys until 9:00 when students starting asking for them.
- Change: I really need a new data set. When students draw their line of best fit, it looks the same for all the students, so it doesn't make the point that I was hoping for.

- Change: on the first problem that they have to interpret, refer them back to the page that has the official definitions.
- Change: there was another new broken link and we reused one of the broken links from the previous activity too.



## CURRICULUM VITAE

Jennifer Loveland  
(February 2014)

jennifer.ellsworth@aggiemail.usu.edu

### Education

- |       |   |
|-------|---|
| Ph.D. | <p>May 2014<br/>Mathematical Sciences, Utah State University<br/>Specialization: College Teaching<br/>G.P.A. 4.0</p>  |
| M.S.  | <p>May 2011<br/>Mathematics, Utah State University<br/>Specialization: Mathematical Statistics and Statistics Education<br/>Thesis: <i>Mathematical Justification of Introductory Hypothesis Tests and Development of Reference Materials</i><br/>G.P.A. 3.96</p> |
| B.S.  | <p>May 2008<br/>Mathematics and Statistics Education, Utah State University<br/>Utah Teaching Certification Level 1<br/>G.P.A. 3.6</p>  |

### Teaching & Research Experience

#### Courses Taught

##### University Level

STAT 2000–Statistical Methods  
 STAT 2300–Business Statistics  
 STAT 3000–Statistics for Scientists  
 MATH 1010–Intermediate Algebra  
 MATH 1060–Trigonometry  
 MATH 1210–Calculus I  
 MATH 2020–Mathematics for Elementary Teachers  
 MATH 2210–Multivariable Calculus

##### Secondary Level

Prealgebra  
 Geometry  
 Algebra II

**Graduate Instructor**

Utah State University at Logan, Utah (2010-2013)

**Recitation Leader**

Utah State University at Logan, Utah (2012)

**Adjunct Instructor**

Utah State University for Distance Education (2012-2013)

Utah State University at Logan, Utah (2009-2010)

**Research Assistant**

Utah State University at Logan, Utah (2012)

**Secondary School Teacher, Mathematics**

Box Elder School District at Brigham City, Utah (2008-2009)

**Mathematics and Statistics Tutor**

Private Tutor (2004-2010)

Utah State University, Academic Resource Center at Logan, Utah (2006-2007)

**Teaching Honors**

Innovative Teacher of the Year, Mathematics and Statistics Department, *Utah State University*, Logan, Utah (2012-2013).

**Publications & Presentations**

Loveland, J., & Schneider, K. (2014, July). Teaching statistics with lectures or activities: A comparative study. Presentation at *The Ninth International Conference on Teaching Statistics*, Arizona, United States. International Statistical Institute and International Association for Statistical Education.

Sun, Y., & Loveland, J. (2014). Conditional Heteroskedasticity of Return Range Processes. Manuscript in referee process.