DOCUMENT RESUME

ED 461 396 JC 020 132

AUTHOR	Brandsma, Jane Ann
TITLE	Data Collection and Analysis: Examining Community College Students' Understanding of Elementary Statistics through Laboratory Activities.
PUB DATE	2000-00-00
NOTE	214p.; Doctor of Philosophy Dissertation, North Carolina State University.
PUB TYPE	Dissertations/Theses - Doctoral Dissertations (041) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	College Mathematics; Community Colleges; Conventional Instruction; *Educational Theories; Mathematics Activities; *Mathematics Instruction; Nontraditional Education; *Statistics; *Teaching Methods; *Two Year College Students; Two Year Colleges

ABSTRACT

This quantitative study compared two groups of community college students who were studying elementary statistics. The control group consisted of 38 students in two classes during the fall of 1998, and the treatment group contained 40 students in two classes during the spring of 1999. The treatment group participated in ten data collection and analysis activities in lieu of some teacher-centered instruction. The study determined if treatment students would: (1) show different levels of understanding; (2) write more accurate, detailed, and complete explanations on open-ended essay questions; (3) more readily see applications of statistics; and (4) develop different attitudes and beliefs about statistics. Results showed that students in the treatment group had better grades on the first of three tests (p <.0001), but on none of the selected final examination items. Administration of the Survey of Attitudes toward Statistics (SATS) and the STARC-CHANCE showed no statistically significant differences. The study concluded that the data do not indicate that including 10 constructive hands-on activities in an otherwise traditional course is sufficient to achieve broad gains in statistical understanding. It recommended that sequences of related activities be implemented to help students build on previous ideas and develop connections among concepts. (Contains 86 references.) (NB)

DATA COLLECTION AND ANALYSIS: EXAMINING COMMUNITY COLLEGE STUDENTS' UNDERSTANDING OF ELEMENTARY STATISTICS THROUGH LABORATORY ACTIVITIES

by

JANE ANN BRANDSMA

A dissertation submitted to the Graduate Faculty of

North Carolina State University

in partial fulfillment of the requirements for the Degree of

Doctor of Philosophy

MATHEMATICS EDUCATION

Raleigh

2000

APPROVED BY:

Glenda Carter

Draga Vidakovic

C. Jacqueline Dietz

Chair of Advisory Committee

ABSTRACT

BRANDSMA, JANE ANN. Data Collection and Analysis: Examining Community College Students' Understanding of Elementary Statistics Through Laboratory Activities. (Under the direction of Lee V. Stiff.)

Two groups of community college students studying elementary statistics were compared. The control group consisted of 38 students in two classes during the fall of 1998 and the treatment group consisted of 40 students in two classes during the spring of 1999. The treatment group participated in ten data collection and analysis activities in lieu of some teacher-centered instruction.

Quantitative results showed that students in the treatment group had significantly better grades on the first of three tests ($p < .0001$), but none of the selected final examination items. The treatment group also showed significantly greater understanding of one concept of the seven selected scales measured by the Statistical Reasoning Assessment (Garfield, 1998), the importance of large samples ($p = .0465$). Students encountered this concept many times throughout the semester as they collected data and pooled their results with those of their classmates. The Survey of Attitudes Toward Statistics (SATS) and the STARC-CHANCE Abbreviated Scale (SCAS) were administered to assess students' attitudes and beliefs. No statistically significant differences were determined.

Qualitative results indicated that while in some cases the writing of students in the treatment group showed greater depth of understanding, these students were also more likely to exhibit confusion among related topics such as correlation and regression, or confidence intervals and margin of error. Interviews conducted with ten students, eight

weeks after the course, indicated that the control group had greater retention of ideas than the treatment group.

The data do not indicate that including ten disjoint, constructive hands-on activities into an otherwise traditional course is sufficient to achieve the broad gains in statistical understanding advocated by the reform movements in mathematics and statistics education. The author recommends that sequences of related activities be implemented to help students build on previous ideas and develop connections among concepts. Additional research should be conducted with classes that use integrated, constructive, student-centered activities as the primary classroom instructional tool throughout the semester.

DEDICATION

This work is dedicated to my grandmother, Catherine,
who encouraged me from the time I was a child
to learn all I could.

It is through her influence and her prompting that
I developed my love for learning.

I hope to honor her through my life's work as an educator.

BIOGRAPHY

Jane Ann Brandsma was born and raised on Long Island, New York. She received the Bachelor of Science degree in Mathematics Education (1988) and the Master of Science degree in Mathematics (1990) from the University of New Orleans. She has taught mathematics and statistics at the post-secondary level in Louisiana, New York, and North Carolina over the past 12 years, beginning with a graduate teaching assistantship at the University of New Orleans. Currently, Jane teaches at Greensboro College in Greensboro, NC.

ACKNOWLEDGEMENTS

I am grateful to Lee Stiff for his guidance throughout my doctoral work, especially this dissertation. He has taught me many lessons about being a good educator, a good researcher, and a good person. I am also appreciative of the time and energy Draga Vidakovic has dedicated to helping me develop as a professional. Her encouragement and mentoring mean the world to me. Jackie Dietz and Glenda Carter have also provided guidance and direction throughout this process and I value their input and suggestions.

Special thanks go to Lynne Gregorio who took valuable time from her busy personal and professional lives to assist me with this study. She was willing to listen when I needed to talk, she was willing to talk when I needed to listen, and she gave generously of wisdom gained through the experience of having gone before. Denise Rowell and Brenda Cates were also wonderfully supportive colleagues throughout my program and the course of this study.

I thank my husband Terry most of all. I am deeply appreciative of both his personal and professional support. For more than fifteen years he has loved me, encouraged me, and supported me. As an academic librarian, Terry provided countless hours of research assistance, proofreading, style guidance, and computer-related consulting. This work could not have been completed without his understanding and his confidence in me.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

Present initiatives in statistics education call for a more data-driven curriculum with less emphasis on theory and formulas. As changes are beginning to be made in statistics courses at the undergraduate level, researchers have started to investigate the impact of such changes, particularly on students' understanding. "Compared with other pedagogies, . . . statistical education is in a relative early stage of its development. Even when relevant research has been conducted, statistical education specialists are only just beginning to be able to build on existing studies" (Hawkins, 1996, p.8). Much of the research conducted at the undergraduate level has focused on students at baccalaureate-granting colleges and universities. Yet, many students complete their mathematics course requirements at two-year community colleges before transferring to senior institutions.

Prichard (1995) suggests that community college students, often encountering mathematics as a barrier to their academic success, would benefit from pedagogical strategies currently being implemented as a result of the reform movement in statistics education and the *Curriculum and Evaluation Standards for School Mathematics* published by the National Council of Teachers of Mathematics (NCTM) in 1989. He also acknowledges that "the mathematics instruction and assessment that most community college students experience has emphasized knowledge-based, procedural learning, without significant regard to actual applications or problem-solving" (p. 30), and suggests that community colleges consider reforming their mathematics curricula. Grosof and Sardy (1993) remind us that "statistics courses *may* be, in fact for many are *most likely* to be, the [community college] students' only post-high school mathematics experience . . . . " (p. 251, emphasis in original). It follows that research investigating the effect of pedagogical change will expand the knowledge base in this area.

This study contributes to the research in statistics education by comparing the stochastic understanding of two groups of community college students studying elementary statistics. One set of students experienced the course in a lecture/discussion format while the other set of students had ten laboratory activities included to replace some lectures and teacher-centered instruction. The purpose of the study was to determine whether these different experiences resulted in different learning outcomes for students.

Constructivism

Recent reforms in education have been based largely on the constructivist theory of learning. Constructivists believe that students learn better through active engagement which allows them to make sense of the content they are studying in light of what they already know. "Knowledge has to be constructed (or reconstructed) by each individual learner if it is to become an integrated part of the structure of knowledge held by the individual" (Orton, 1992, p. 163). As students interact with each other and with their environment they develop deeper understanding of the subject matter. Goldin (1990) believes that

> for large numbers of students at all levels of mathematics education methods involving the statement and application of rules (i.e., methods based on a *transcriptive* model) are less successful than methods involving mathematical discovery (i.e., methods based on a *constructive* learning model). (p. 46, emphasis in original)

Constructivism has its roots in the work of Piaget and has been substantially influenced by others, including Vygotsky and von Glaserfeld.

Piaget worked with individual students as learners. He studied their cognitive abilities, specifically their ability to make sense of new information. When this new information is consistent with students' existing schema, or mental structure, they use what Piaget called "assimilation" to organize this new experience within that existing schema. However, when students' new experiences are in conflict with their existing knowledge they use a process of accommodation to restructure their understanding and make sense of the new information based on their previous experiences. Early on, assimilation and accommodation are undifferentiated in learners and act in opposing directions. That is, conflict exists between the attempt to assimilate things and the need to accommodate for them. However, "as a child's thought evolves, assimilation and accommodation are differentiated and become increasingly complementary" (Piaget, 1954, p.385).

Vygotsky's (1986) work emphasized a social component of learning, asserting that students develop understanding via their social interaction with others. He defined a pupil's zone of proximal development as "the discrepancy between a child's actual mental age and the level he reaches in solving problems with assistance" (p. 187). Further, Vygotsky claimed that the best indicator of a student's intellectual development was the ease with which he progressed from problem solving alone to problem solving with assistance.

The theory of radical constructivism was developed by von Glasersfeld through his Interdisciplinary Research on Number at the University of Georgia (Steffe and Kieren, 1994). Radical constructivists believe that "the learner does not discover an independent, preexisting world outside his or her mind" (Gadanidis, 1994, p. 94). That is, they challenge the idea that knowledge already exists in some predetermined structure and individual learners strive to rediscover that structure. As a radical constructivist, von Glasersfeld (1987) asserts that each individual interprets experiences differently, and that the interpretation is dependent on the existing conceptions of that learner. In reference to mathematics education, von Glasersfeld states that "logical or mathematical necessity does not reside in any independent world - to see it and gain satisfaction from it, one must reflect on one's own constructs and the way in which one has put them together" (p. 16).

Mathematics Education Reform

Influence of Constructivism

Constructivist ideas have affected both the learning and teaching of mathematics. Direct teaching, the traditional lecture model of education, had long been the principal mode of instruction. With the development of constructivist theories of education, teachers have been encouraged to establish

> a mathematical community – providing objects that can be used in mathematical investigation, engaging in lots of teacher-student interaction for purposes of diagnosis and guidance, encouraging student-to-student talk that focuses on mathematical issues, modeling mathematical thinking, promoting the kinds of questions and comments that help community members to challenge and defend their own constructions. (Davis, Maher, & Noddings, 1990, p. 3)

This change in the culture of the classroom has led to adjustments for both teachers and students. Typically, students have been conditioned to see the teacher as the ultimate source of knowledge, having all the answers. The mathematics student's responsibility has been to memorize and replicate facts and algorithms successfully. This model is founded on behaviorist theories of learning, influenced by the works of Thorndike and Skinner (Stiff, Johnson, & Johnson, 1993). These learning theories emphasized drill and practice along with repetitive skill development. The reforms in mathematics education have altered these traditional student and teacher roles. Students are now expected to take an active role in generating and confirming the mathematical ideas they are studying.

Constructivists recognize that students do not come to the classroom as empty vessels to be filled. Rather, they acknowledge that students' prior experiences affect their current learning situations, and they understand that these prior experiences may interfere with the learning process.

> When presented new information, we have no other option than to relate it to what we already know -- there is no blank space in our minds within which new information can be stored so as not to "contaminate" it with existing information. Learning in the classroom involves students weaving selected and interpreted teacher inputs into an existing fabric of knowledge. In this way, learning is both limited and, at the same time, made possible by prior knowledge. This constructive view of learning ... explains the frequent gap between what students report and what we, as teacher, thought we clearly communicated. (Konold, 1995, par. 15)

Students simply cannot comply with teachers' requests that they forget everything they know about a given subject prior to a new lesson on that subject.

Standards

As the behaviorist model of education is replaced by constructivist theories, many professional teachers' organizations at the K-12 level are developing guidelines for teaching and learning within this new paradigm. In addition to the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989), standards publications include *Benchmarks for Science Literacy* (American Association for the Advancement of

Science, 1993), *Expectations for Excellence: Curriculum Standards for Social Studies* (National Council for the Social Studies, 1994), and *Standards for the English Language Arts* (National Council of Teachers of English, 1996). Collectively, these documents have become known as Standards.

Reform has also been targeted at the post-secondary level, encouraged by many national reports. *A Curriculum in Flux* (Davis, 1989) was presented to the Mathematical Association of America by the Joint Subcommittee on Mathematics Curriculum at Two-Year Colleges. This document was intended to "help support and stimulate constructive change in mathematics curriculum at two-year colleges" (preface). Later, the National Research Council (1991) recommended that undergraduate mathematics be taught "in a way that engages students" (p. 45). This view of an active classroom is in striking contrast to the traditional teacher-centered format that is common in many college classrooms. In 1995, the American Mathematical Association of Two-Year Colleges (AMATYC) produced another set of recommendations specifically tailored to the first two years of college mathematics, *Crossroads in Mathematics: Standards for Introductory College Mathematics before Calculus*. As reform efforts begin to have an impact at the post-secondary level, faculty members are being encouraged to rethink the structure of their classes.

AMATYC Standards

AMATYC (1995) articulates standards for intellectual development, standards for content, and standards for pedagogy. The standards for intellectual development focus on student thinking and learning outcomes. These standards emphasize that courses at the introductory college level should provide students with a broad understanding of the nature of mathematics, including its richness and power. The seven standards for intellectual development are (a) *problem solving*, including the development of strategies and the communication of results; (b) *modeling*, choosing a model, fitting data, evaluating the appropriateness of the model, and explaining why the model is a valid representation of the physical situation; (c) *reasoning*, the development of both inductive and deductive mathematical arguments; (d) *connecting with other disciplines*, including the arts and social sciences; (e) *communicating*, the ability to listen, and to read, write, and speak mathematically; (f) *using technology*, as an aid in understanding as well as a tool for problem solving; and (g) *developing mathematical power*, experiencing mathematics in a way that builds self-confidence and perseverance.

The standards for content emphasize the problem solving aspects of mathematics. Through problem solving, students should develop an understanding of the content and be able to apply that understanding in meaningful ways. The seven standards for content are (a) *number sense*, including estimation, pattern recognition, and proportional thinking; (b) *symbolism and algebra*, using multiple representations to translate and solve problems; (c) *geometry*, the development of spatial and measurement sense; (d) *function*, understanding families of functions, the use of functions in modeling, and the behavior of functions; (e) *discrete mathematics*, including permutations, combinations, sequences, series, matrices, and linear programming; (f) *probability and statistics*, analyzing data and making inferences about real-world situations; and (g) *deductive proof*, forming and testing conjectures.

The standards for pedagogy focus on specific instructional strategies designed to engage students in mathematical activities and provide them opportunities to construct their own understanding by experiencing mathematics in context. The five standards for pedagogy are (a) *teaching with technology*, using computers and instructional media to enhance the learning experience; (b) *interactive and collaborative learning*, encouraging students to work in groups and discuss mathematics with peers; (c) *connecting with other* *experiences*, making mathematics meaningful and relevant; (d) *multiple approaches*, using various methods to solve problems and communicate results; and (e) *experiencing mathematics*, providing projects and activities which build students' confidence in mathematics and develop their ability to think independently. These standards for pedagogy were developed by the authors to be "compatible with the constructivist point of view" (AMATYC, 1995, p. 15).

Statistics Education

Influence of Constructivism

Cobb (1992) emphasizes the constructivist nature of learning in his recommendations for improving statistics instruction. In response to the "call for change" made by the Board of Governors of the Mathematical Association of America, Cobb reports on the work of the Statistics Focus Group. Three recommendations emerged from the months of e-mail discussions held by that focus group. First, they recommend an emphasis on statistical thinking, including the need for data and the importance of data collection, as well as an understanding of variability. The second recommendation calls for more data and concepts, less theory, and fewer recipes, emphasizing that "statistical concepts are best learned in the context of real data sets" (p. 7). The third focus group recommendation is to "foster active learning" (p. 8). Specifically, they suggest group problem solving and discussion, lab exercises, demonstrations based on class-generated data, written and oral presentations, and projects, either group or individual.

Recommendations for curricular reform in undergraduate statistics have called for increased student involvement, hands-on activities, and laboratories (Hogg, 1991; Hollis, 1997; Moore, 1997). Consequently, reformed curricula for elementary statistics courses have integrated a constructivist approach to teaching and learning (Aliaga & Gunderson, 1998; Rossman, 1996; Scheaffer, Gnanadesikan, Watkins, & Witmer, 1996). Mathematics laboratories have been identified as appropriate settings for providing "students with activities designed to guide them in the construction of their own understanding of mathematical concepts . . . ." (AMATYC, 1995, p. 54).

Statistics at the Community College

Mathematics requirements at community colleges vary according to program and state. Generally, mathematics credits may be earned in a number of ways, but several programs have specific requirements with elementary statistics frequently singled out. Throughout the state of North Carolina, elementary statistics is required for students in the Pre-Business Administration, Pre-Health Education, and Pre-Nursing programs. Additionally, elementary statistics is strongly encouraged for students in the Pre-Criminal Justice, Pre-Physical Education, Pre-Sociology, and Pre-Speech/Communications programs. The course may be chosen as a mathematics elective by students in other programs.

The publication of *Crossroads in Mathematics: Standards for Introductory College Mathematics before Calculus* has generated interest and enthusiasm for reforming the teaching of mathematics at the community college level. Many community college students planning to transfer to universities will complete their mathematics requirements at the community college, and as indicated previously, many programs require that a statistics course be used to meet part of the mathematics requirement. As the standards are adopted and implemented, researchers have an opportunity to determine the extent to which changing instructional practices actually improves student understanding of statistics (Garfield, 1993; Garfield and Ahlgren, 1988; Giraud, 1997).

Research in Statistics Education

Constructivist principles have been embraced by those involved with curricular reform in statistics. Similarly, the constructivist nature of learning statistics has also been recognized in the research community. Garfield (1995) asserts that students learn statistics by constructing knowledge, accepting "new ideas only when their old ideas do not work, or are shown to be inefficient for purposes they think are important" (p. 30). She also states that "students cannot learn to think critically, analyze information, communicate ideas, make arguments, tackle novel situations, unless they are permitted and encouraged to do those things over and over in many contexts" (p.30-31). She calls for research to determine which "specific small-group activities work best in helping students learn particular concepts and develop particular skills" (p. 33).

Hawkins (1997) acknowledges that many of the recent developments in statistical education "have largely been made in the absence of evidence-based understanding about the teaching/learning process" (p. 144). As past president of the International Association for Statistical Education, she considers herself among those who regret that research has not substantially driven or contributed to the reform in classroom practice. She calls for more research in this area. In light of the recent emphasis on teaching statistics in the K-12 curriculum, Shaughnessy (1992) has voiced "the need for a stepped-up, ongoing research program in the area of probability and statistics" (p. 466).

A recent review of literature on the teaching of statistics (Becker, 1996) reports that much of the published work regarding statistics education is in the form of anecdotal discussions of the experiences of those who teach statistics. Becker found that "less than 30% of the print literature reports the results of empirical studies" (p. 71). She also determined that nearly one third of the articles meeting the criteria for her review were published in the first half of the 1990s, indicating a strong current interest in the teaching and learning of statistics.

Although interest in research in statistics education has grown in the past decade, little research in this area has focused on the community college population. Some of the research involving community college elementary statistics students has involved the use of computers in teaching (Myers, 1990; Rosenbaum, 1980). Rojas (1992) investigated the use of special materials to integrate language skills in a community college probability course. More recently, Bonsangue (1994) studied the effect of collaborative learning on student achievement in elementary statistics at a community college in southern California. His work focused on students' completion of textbook problems that were directly related to the concepts presented in class.

Statement of the Problem

Researchers have begun to study the teaching and learning of elementary statistics at the community college level, but many questions remain unanswered. In particular, there is a need to determine whether the use of hands-on activities facilitates students' understanding of stochastic ideas. This study investigated community college students' learning through statistical laboratory experiences by building on the AMATYC standards for pedagogy. Specifically, the study compared community college statistics classes taught in a lecture/discussion format with those that included hands-on laboratory activities.

The AMATYC standards for intellectual development address such areas as problem solving, modeling, reasoning, and the use of technology, all of which were common to both groups of students. Similarly, the AMATYC standards for content provide general guidelines for the type of mathematics that should be taught at the community college level. It is only the AMATYC standards for pedagogy which were implemented differently for the two groups of students.

AMATYC Standards for Pedagogy

The first AMATYC standard for pedagogy is *teaching with technology*. The increased availability of technology has made it easier for students to actively collect and analyze data. Many computer software packages, including MINITAB, SAS, and SPSS, were designed to facilitate statistical investigation. Recently, graphing calculator manufacturers have incorporated statistical features into their products. This type of hand-held calculator is a computer which provides a relatively inexpensive means for effectively computing basic descriptive statistics and performing elementary hypothesis tests. When combined with portable companion units which use probes to collect temperature, voltage, motion, and light readings, among others, students gain access to convenient means for generating and analyzing data.

The introduction of hands-on activities as a means for students to collect and analyze data is consistent with the second AMATYC standard for pedagogy, *interactive and collaborative learning*. Collaborative learning provides an opportunity for students to learn through interaction with their peers. Working in small groups, students discuss strategies, solve problems, and reflect on their work. Collaborative learning is consistent with the constructivist philosophy of education and is advocated by both the NCTM (1989) and AMATYC (1995). Enjoying wide use in the K-12 setting, this type of classroom activity is gaining popularity in college classrooms (Johnson, Johnson, & Smith, 1991) and successful use of cooperative learning in undergraduate statistics courses has recently been reported (Dietz, 1993; Garfield 1993; Giraud, 1997; Keeler & Steinhorst, 1995).

The third AMATYC standard for pedagogy, *connecting with other experiences*, is addressed by the use of student investigations in a laboratory situation. These activities "provide students with experiences that connect classroom learning and real-world applications" (AMATYC, 1995, p. 25). Introductory statistics is, perhaps, one of the most natural mathematics courses for such an educational approach. As the information age continues to unfold, students increasingly will be faced with situations in which they must analyze and evaluate quantitative data. Cobb, quoted in McKenzie (1996), emphasizes that what we as statistics educators "ask of our students should be *meaningful* to them. Being meaningful has both a cognitive/intellectual component and an emotional component, and . . . both kinds of meaning often involve making connections that are new to our students" (p. 232, emphasis in original).

The penultimate AMATYC standard for pedagogy is *multiple approaches*. Students should be encouraged to solve problems of various types, and report and interpret their results numerically, graphically, orally, and in writing. Laboratory activities carried out individually and with peers provide a rich environment for students to develop these skills. The goals articulated in this standard parallel those in Cobb's (1992) third recommendation, "foster active learning."

Finally, the fifth AMATYC standard for pedagogy, *experiencing mathematics*, calls for students to engage in projects which require extensive time and effort. Extended activities provide students with this type of experience. Additionally, students experiencing statistics should be encouraged to critically review current reports and advertisements in print and broadcast media involving statistical information (Snell, 1994; Solomon, 1988). By discussing current events and the statistical arguments which are presented, students may make additional connections between the course content and their own experiences, and can begin to appreciate how stochastic ideas permeate their lives.

Purpose of the Study

The purpose of the study was two-fold. The first was to determine whether laboratory activities affect students' understanding of statistics concepts and whether students involved in the laboratory activities express their understanding differently than students enrolled in the lecture/discussion sections of the course. The second purpose was to determine whether students experiencing the course in these two different formats develop different attitudes and beliefs toward statistics.

Both quantitative and qualitative research methods were used to answer the following research questions:

1. Will students in the laboratory sections show different levels of understanding than students in the lecture/discussion sections based on course examinations, final examination items, and the Statistical Reasoning Assessment?

2. Will students in the laboratory sections write more accurate, detailed, and complete explanations on open-ended essay questions than students in the lecture/discussion sections?

3. Will students in the laboratory sections more readily see applications of statistics when compared with students in the lecture/discussion sections?

4. Will students in the laboratory sections of the course develop different attitudes and beliefs about statistics than students in the lecture/discussion sections of the course?

CHAPTER 2

REVIEW OF RELATED LITERATURE

This study was designed to compare a traditional lecture/discussion format of elementary statistics at a North Carolina community college with a course incorporating laboratory activities in the same setting. This chapter begins with a discussion of the present state of statistics education at the undergraduate level and the reform movement that is currently underway. The incorporation of laboratory investigations in this study involved many pedagogical strategies, including collaborative learning. The appropriate use of technology was encouraged during both the fall and spring semesters. Student writing was emphasized in all sections of the course and provides some data for analysis. The influence of laboratory activities on students' attitudes and beliefs about statistics was also investigated. Relevant literature from these areas is discussed.

Statistics Education

The teaching and learning of statistics at the college level has been the focus of much recent discussion "influenced by a movement to reform the teaching of the mathematical sciences in general" (Moore, 1997, p. 123). Early in this decade many statisticians and statistics educators identified areas for improvement in elementary statistics (Hogg, 1991; Snee, 1993). Hogg (1991) reports on recommendations made by 39 statisticians participating in a workshop on statistical education. This group suggests that students learn to ask questions, collect, summarize, and interpret data and "understand the limitations of statistical inference" (p.342). To that end, they further suggest an increase in teamwork, student-generated data, and projects. Snee (1993) calls for changes in both content and delivery of statistical education. He feels that real-world contextual learning with a problem-solving focus is key to content reform. Along with those content goals, he sees experiential learning, "learning by doing" through projects, labs, workshops, and group problem sessions, as the most effective instructional method.

Specific recommendations for change are also contained in a report made by a joint committee of the American Statistical Association and the Mathematical Association of America (Cobb, 1992). Their three main recommendations are to (a) emphasize statistical thinking, including the need for data and the importance of data collection, (b) use more data and concepts with less theory and fewer recipes, and (c) foster active learning, including group problem solving and discussion, lab exercises, demonstrations based on class-generated data, written and oral presentations, and projects, either group or individual.

These recommendations have been incorporated into recent curricular reform projects. Snell and Finn (1993) describe a course called Chance, based on the magazine of the same title. Incorporating current events and media clips as the basis for class discussion, students develop statistical understanding in the context of real-life examples. Group work sessions and journal-writing are components of the Chance course taught on some campuses. *Workshop Statistics* (Rossman, 1996) "is designed for courses that employ an interactive learning environment by replacing lectures with hands-on activities" (p. xv). The text focuses on conceptual understanding through the use of active learning, genuine data, and technology. *Activity-Based Statistics* (Scheaffer, Gnanadesikan, Watkins, & Witmer, 1996) also treats elementary statistics as a laboratory course rather than a traditional course by providing hands-on activities for student engagement with the content.

Garfield (1997) acknowledges that progress has been made in the areas of improved instructional materials, use of technology, and available resources for teachers of statistics, but goes on to state that despite these improvements "most statistics courses taught in institutions of higher education have changed very little" (p.138). She specifically addresses passive instruction and traditional teacher-centered lectures as one of the areas where change is needed.

Hawkins (1997) points out the limited availability of research in statistics education. She does not believe that research is guiding the reform movement as much as it should and she is concerned that current research "does not tell us about things that did not work, and therefore about what things we should avoid" (p. 145). Further, Hawkins states that "research into *why* a particular teaching approach is effective is relatively rare" (p. 145, emphasis in original).

Active Learning and Laboratories

Active learning has been encouraged by many in the statistics education community (Rossman, 1994; Scheaffer, 1994; Spurrier, Edwards, & Thombs, 1993). Richard Scheaffer (in Cobb, 1992), states that

> Statistics should be taught as a laboratory science, along the lines of physics and chemistry rather than traditional mathematics. Students must get their hands dirty with data. The laboratory must be a requirement and must contain more than just a few computers. This approach involves real data but also involves manipulative devices that include spinners, cards, bead boxes . . . . (p. 11)

Rossman (1992) asserts that much of the mystery, or apparent trickery, surrounding statistical ideas is removed when students explore and discover for themselves statistical ideas and techniques. Additionally, it is through such activity that students develop the judgment skills they need for data analysis.

Students seem to prefer activities involving group discussion and other social facets, according to Grabowski, Harkness, Birdwell, and Rosenberger (1995).

Rumsey (1998) describes two specific types of activities, those intended for students to discover concepts, and those designed for students to practice concepts. She encourages the appropriate use of both types of activities as instructional tools. Additionally, Rumsey suggests that classroom activities clearly reinforce the statistical concepts being studied. She suggests that at times the execution of the activity could obscure the underlying conceptual development rather than enhancing or reinforcing it.

Recent literature confirms that instructors are implementing some of the suggested changes, but much of that documentation is in the form of nonempirical articles (Becker, 1996). Magel (1996) implemented work-sheet based activities in a large lecture of 140 students. Often, students collected data as part of the activity. Students worked in groups to collect data, but could choose to work alone in analyzing the data. The worksheets were not graded for accuracy, but were collected and points were awarded for completion. Magel describes how many of the worksheets were used to illustrate concepts involving descriptive statistics. She gives rather detailed accounts of activities used to illustrate the central limit theorem, confidence intervals, and hypothesis testing. Although Magel acknowledges that "it is hard to judge whether more learning takes place in my classes using this approach" (p. 56), she does note that average exam scores improved and the average drop rate was lower in these sections.

Prichard (1993) suggests data collection and representation activities for college students based on analyzing the front page of a newspaper. The activity focuses on the frequency of the digits 0 through 9 appearing on the front page. Students conjecture, collect data, and display their findings. She also provides a string-tying task used to investigate probability concepts. Again, students are asked to predict the outcome and then encouraged to explore the theoretical probabilities involved in the exercise.

Somers, Dilendik, and Smolansky (1996) discuss a numerical memory exercise from which students can collect, display, and analyze data. Students are provided with a vertical list of three-digit numbers, they are given thirty seconds to memorize as many as possible, and then they have sixty seconds to write down as many of the three-digit numbers as possible. Later, students are provided another opportunity to memorize the numbers, but this time they are presented as the last three digits in a historical data. For example, 492 appeared on the first list, and 1492 (identified as the year of Columbus' expedition to San Salvador) appeared on the second list. The experiment is repeated and the two sets of data are analyzed using mean, standard deviation, and five number summary. Box plots are constructed and interpreted. The authors suggest other concepts which can be explored using this data, including correlation and regression.

Cresap (1995) discusses the use of student-generated data to illustrate the central limit theorem. He encourages students to collect data from skewed distributions such as the number of pages in books at their school library. They use a graphing calculator program to "sample" from this population and simulate the central limit theorem. Such sharing of teaching ideas makes an important contribution, but as Hawkins (1997) points out, research is needed to determine how these activities foster student understanding.

Technology

Advances in technology, including computer hardware and software, calculators, audio-visual materials, and multi-media platforms have dramatically changed the face of statistics education over recent years. Statistical software packages such as MINITAB, SAS, and SPSS have reduced tedious computations and have made it feasible to use real-world, often large or messy, data sets in the classroom. Handheld graphing calculators provide portable, relatively low-cost access to computational power, allowing "students to quickly input individual data sets, get basic descriptive statistics, and see graphical displays of their data" (Cresap, 1995, p. 357). Garfield (1995) acknowledges that "using software that allows students to visualize and interact with data appears to improve students' understanding of random phenomena and their learning of data analysis" (p. 29). Audio-visual materials, such as the series *Against All Odds,* expose students to engaging real-life, familiar, statistical situations (Mansfield, 1995).

Much emphasis in the reform of statistics education has been toward more conceptual development of the underlying ideas and away from manual calculations. Myers (1990) studied the impact of computers on community college students' understanding of two statistics concepts, random sampling and the central limit theorem. Students in one class used a computer to investigate these concepts while the other class studied the same content through a traditional lecture. Myers found that the computer users scored significantly higher on a test of concepts, but noted no significant difference between the two groups on a test of applications. A retention test given to both groups three weeks later showed no significant difference between the two classes.

Sterling and Gray (1991) investigated students' statistical understanding as a result of their interaction with simulation software. Two sections of an introductory course were used in the study; one section served as the control group and the other served as treatment group. The researchers used examination questions to measure student achievement in each group. The experimental class scored significantly higher than the control group on exam questions about concepts covered by the software.

Carson (1995) researched college students' understanding of the sampling distribution of the sample mean. Students used graphing calculators to simulate sampling from various non-normal distributions. The study focused on their construction, interpretation, and understanding of histograms resulting from the sampling activities. Carson's results were mixed; some students developed appropriate conceptions while other did not. She suggests increasing students' experiences with data analysis.

Grabowski and Harkness (1996) studied the use of expert systems by college statistics students. They conducted two studies which compared three groups of students, those who created their own expert systems, those who used an instructor-generated expert system, and those who used no expert system at all. They found that students who created their own expert systems or who used expert systems developed by the instructor had greater gains in learning than those who did not.

As technology changes and becomes more readily available, research should continue to explore the effectiveness of those changes. Hawkins (1996) encourages continued research into the use of technology in statistics education, research which identifies "a broad range of ways in which technology can assist the teaching and learning process" (p. 13).

Collaborative Learning

With more emphasis on active learning and laboratories, there is naturally an increased emphasis on group work. Many classroom activities are carried out in teams of students, either by design or by necessity due to limited resources and materials. Hogg (1991) encourages statistics educators to have students work together in teams to generate and analyze data. The need for students to be comfortable working in groups at their future workplace is often discussed as one motivator for emphasizing team projects in an introductory statistics course (Jones, 1991; Chance, 1996). Rumsey (1998) also includes "an increased respect for other viewpoints and other approaches to solving a problem" (par. 20) in her rationale for promoting cooperative learning. Active and cooperative learning techniques have gained in popularity and were recently named the number one "best technique" among the ninety-four attendees at a conference on trends in introductory applied statistics courses (Goldman, 1996). Fifty-seven percent of the college faculty in attendance indicated that they had used group work in their statistics courses.

Jones (1991) found that the use of cooperative learning strategies in his statistics course for psychology students "resulted in higher student attendance, more favorable student course evaluation, and an improved average reported attitude towards the course" (p. 5). Some changes in the course content over the period of the study made it difficult to determine changes in student achievement as a result of cooperative learning.

Keeler and Steinhorst (1994) studied university students' success with cooperative learning in elementary statistics. The study involved one section each semester for three semesters. The first semester students were taught using a traditional lecture format. The following two semesters cooperative learning techniques were incorporated into the course. Pairs of students worked together for three to five minutes to answer questions posed during a break in the lecture. These questioning breaks occurred every ten to fifteen minutes throughout the class. The researchers found that students in the cooperative classes had better course averages than students in the previous lecture course. They also found that fewer students withdrew from the cooperative sections.

Giraud (1997) also investigated the impact of cooperative learning by studying two sections of an applied statistics course in the psychology department of a state university. One section was taught in a traditional format and the other was taught using cooperative groups for problem-solving sessions. The assignments used in both classes were the same, with students in the cooperative learning class given time to work in groups during class while students in the lecture course used that time for reviewing practice problems and extra examples. Giraud found that students in the cooperative learning class displayed a greater level of achievement on the final exam than those in the traditional class. He discusses limitations of his study including the use of only one section of each type of class, the fact that one class was a morning class and the other was a late afternoon class, and that the two sections met in different classrooms, one of which was near a loud loading dock. He calls for further research in this area.

Dietz (1993) found that students in a beginning statistics course were able to construct their own sampling methods as a result of a cooperative learning activity. After the initial preparatory session, students self-selected into groups of three or four and completed a series of lab worksheets in which they generated, tested, and evaluated sampling methods. Dietz concludes by stating "Knowledge that students have constructed for themselves is understood better and remembered longer than procedures memorized from a textbook" (p. 108).

Writing to Learn Statistics

Reform efforts in statistics education have included an emphasis on writing as well as hands-on activities and cooperative learning (Iversen, 1991; Hayden, 1992; Rossman, 1992). Scheaffer (in Cobb, 1992), states,

> Students come to us with primarily an intuitive understanding of the world. It is part of our job to ferret out those intuitive processes and correct the incorrect ones. As far as I know, this can only happen by having students discuss and write about their understandings and interpretations of problems. (p. 283)

Garfield (1995) lists one of the goals of statistics courses as "learning to communicate using the statistical language, . . . drawing conclusions, and supporting conclusions by explaining the reasoning behind them" (p. 26). Goldman (1996) reports that oral and written communication were listed as the third best practice by the "techniques" subgroup of attendees at the March 1996 *Trends in Applied Introductory Statistics* conference co-sponsored by the Boston Chapter of the American Statistical Association. Interpreting and communicating results was chosen as the second best practice by those in the "topics" subgroup at the same conference (Sevin, 1996). Although writing in statistics is encouraged on many fronts and is being successfully implemented by increasing numbers of statistics educators, few research studies on the effects of writing or its impact on student understanding have been conducted.

Beins (1993) studied three levels of writing emphasis on four sections of introductory statistics for psychology majors. The first section was considered traditional-emphasis, the second and fourth sections were considered moderate-emphasis, and the third section was considered high-emphasis. He compared the four groups on three segments of the final assessment: computational, conceptual, and interpretative. The high-emphasis class scored significantly better than one of the moderate-emphasis classes on the computational segment. There was no significant difference among the three groups on the conceptual segment. He found that greater emphasis on writing during class resulted in higher scores on the interpretive portion of the assessment. That is, on interpretive items, the high-emphasis students scored better than the moderate-emphasis students who in turn scored better than the traditional-emphasis students.

Chance (1996) evaluated the use of student journals in introductory statistics. Students in one section were required to keep journals while students in the other section were not. The journals were intended to provide students with an opportunity to reflect on class activities, write chapter summaries, ask questions, and make connections to topics outside the course. Chance found that overall student achievement and course satisfaction were the same for both groups of students. However, she did find more variability among the journal writing group, concluding that better students developed deeper understanding whereas weaker students became overwhelmed by the requirement and gave up on the course.

Hayden (1990) discusses the importance of writing in introductory statistics courses as a vehicle for assessing student thought. He shares his own experiences with writing in statistics courses and offers practical suggestions for sample assignments and test questions. He does not attempt to quantitatively evaluate the success of student writing but expresses satisfaction with the program and its emphasis on meaning and understanding rather than computation.

Attitudes and Beliefs

The attitudes and beliefs about statistics which students bring to the classroom have the potential to positively or negatively impact their ability to learn and apply stochastic concepts. Additionally, attitudes and beliefs that change or are developed during a statistics course may affect the extent to which students pursue advanced coursework in statistics, or the extent to which they implement the concepts they have learned.

Gal and Ginsburg (1994) note that the limited research focusing on students' attitudes and beliefs in statistics has mainly focused on the development and use of Likert-type scales. They voice concerns regarding the development and administration of scales such as the Statistics Attitude Survey (SAS) and the Attitudes Toward Statistics (ATS) instruments. Specifically, they cite the lack of opportunity for students to supply reasons for the decisions they make as a weakness of such assessments. They identified four non-cognitive factors of student attitudes and beliefs which are of importance to statistics educators: "(a) interest or motivation for further learning, (b) self-concept or confidence regarding statistical skills, (c) willingness to think statistically in everyday situations, [and] (d) appreciation for the relevance of statistics in their personal and vocational lives" (par. 12). Subsequently, the authors report that little research has been conducted to evaluate the impact of students' attitudes and beliefs on statistics learning. There is a need for educators and researchers to extend the scope of inquiry to better capture their students' thinking and evaluation processes.

Schau, Stevens, Dauphinee, and Del Vecchio (1995) describe the development and validation of their Survey of Attitudes Toward Statistics (SATS). They carefully delineate the characteristics that they considered important in making improvements over existing instruments. These include: (a) scales which "tap the most important dimensions of attitudes toward statistics," (b) scales which can be used appropriately in most introductory statistics courses, (c) scales which are short and take little time to administer, (d) scales which have items worded both positively and negatively, and (e) scales which use student input in the development and validation. The SATS measures four dimensions of student attitudes and beliefs: affect, cognitive competence, value, and difficulty. The authors assert the need for continued research which investigates "the relationships of these attitudes to student persistence, achievement, and success in statistics" (p. 874).

Faghihi and Rakow (1995) used the SATS to compare the attitudes and beliefs of students enrolled in a self-paced introductory statistics course with students enrolled in a traditional course. Students were undergraduate and graduate students in education, and undergraduates in business and psychology. The researchers found no significant differences among the two groups of students' scores on the SATS. Additionally, no significant differences were found between men and women or between African-American and Caucasian students.

Gal, Ginsburg, and Schau (1997) discuss the importance of evaluating student attitudes and beliefs toward statistics, and they provide an overview of assessment instruments available for such evaluation. They provide three specific reasons for instructors to consider attitudes and beliefs: (a) process considerations, the influence of attitudes and beliefs on the teaching and learning of statistics; (b) outcome considerations, how attitudes and beliefs influence students after they leave the course; and (c) access considerations, how attitudes and beliefs might motivate students to continue studying statistics. The authors provide detailed information about the SATS instrument, along with an appendix which includes a copy of the post-version, scoring instructions, and potential open-ended extensions especially recommended for research purposes.

These recommendations are consistent with those made by McLeod (1992) regarding mathematics instruction. McLeod discusses "three major facets of the affective experience of mathematics students that are worthy of further study" (p. 578). He includes: (a) the beliefs that students hold about themselves as learners and about mathematics content; (b) the positive and negative emotions students experience as their study of mathematics progresses through interruptions, especially when the tasks they face are novel; and (c) the positive and negative attitudes students develop as they repeatedly experience similar mathematical situations. McLeod goes on to encourage research which combines the study of students' cognitive and affective domains. He advocates the use of both qualitative and quantitative research methods in the investigation of students' affect.

Statistics Education in Community Colleges

In 1991 approximately 69,000 community college students were enrolled in statistics and probability courses offered at 79% of community colleges nationwide (Cohan & Ignash, 1994). Many of these students will complete their college mathematics requirement at the community college prior to transferring to a four-year institution. Traditional transfer students have been joined by

> the unemployed, displaced workers, and those whose skills need upgrading; welfare recipients who are required to go to school or work as a condition of receiving further benefits; women reentering higher education after a hiatus for homemaking and child-rearing, many of whom will be seeking employment for the first time; and some who might be called *interrupted scholars* with diverse interests, objectives, and educational backgrounds. (Knoell, 1996, p.56, emphasis in original)

This student population is markedly different than the traditional student population still found at many colleges and universities.

Sevin (1995) discusses increased student work hours, changes in student attitudes and preparation, and differing student learning styles as challenges facing statistics educators, especially at community colleges. Krevisky (1994) concurs that teaching statistics to the diverse community college population is a challenging endeavor. Grosof and Sardy (1993) emphasize the importance of quantitative skills, including statistical literacy and analytical reasoning, in assuring opportunity for community college students they describe as "educationally under-served, [and] disadvantaged by conventional standards" (p.251). Research into the teaching and learning of statistics at the community college level may provide unique insight for instructors challenged by this heterogeneous population. Few studies in this area have been conducted to date (Priselac, 1995).

Bonsangue (1994) studied the effect of collaborative learning on community college students' achievement in introductory statistics. He compared two sections of the course, one taught using cooperative learning techniques for small-group problem solving and one taught using a traditional lecture format. Bonsangue measured student achievement on four common course examinations and found that the collaborative learning class did significantly better than the traditional class on the second, third, and fourth exams. There was no significant difference between the two groups on the first exam.

Rosenbaum (1980) studied community college students' use of computer programming in BASIC for elementary statistics. She was particularly interested in students' achievement and attitudes. Two curriculum topics were considered, correlation and linear regression. Rosenbaum found that students who wrote computer programs to solve problems involving correlation and linear regression did not show greater achievement in these areas. Additionally, she did not find that these students had more favorable attitudes towards statistics.

Rojas (1992) compared two randomly selected sections of a community college first course in probability. One class received special materials designed to enhance their reading, writing, and communication skills. Rojas found that students using the auxiliary materials had higher mean scores on exams than the other students. She also noted improvement in their attitudes toward writing. She suggested further research to determine whether an environment which stimulates writing increases student achievement in statistics.

Summary

Statistics education is beginning to be transformed as a result of initiatives by national and international leaders in the field. The use of hands-on activities, technology, collaborative learning, and writing provide a rich context for student learning, and have been incorporated into recent curriculum projects. Although these techniques seem to be increasingly implemented in statistics classrooms, too few research studies have focused on the roles these innovations play in enhancing student understanding. There is an alarmingly small body of research regarding community college students' experiences in elementary statistics courses. Leaders in statistics education, recognizing that research is not guiding practice, are calling for continued scholarly inquiry to determine the nature of students' understanding of stochastics.

Statement of Research Hypotheses

Based on the stated research questions, and prompted by the existing literature and the recognized need for continued study, the following null hypotheses were generated:

1. Students enrolled in the lecture/discussion format of the course and students enrolled in the format of the course including laboratory activities will show the same level of statistical understanding as measured by researcher constructed tests, final examination items, and the Statistical Reasoning Assessment.

2. Students enrolled in the lecture/discussion format of the course and students enrolled in the format of the course including laboratory activities will show the same level of statistical understanding as measured by open-ended essay questions on the final examination.

3. Students enrolled in the lecture/discussion format of the course and students enrolled in the format of the course including laboratory activities will have the same ability to see applications of statistics.

4. Students enrolled in the lecture/discussion format of the course and students enrolled in the format of the course including laboratory activities will develop the same attitudes and beliefs about statistics.

CHAPTER 3

METHODOLOGY

This study was designed to investigate the effect of laboratory activities on community college students' understanding of elementary statistics. The study was conducted over two semesters, Fall 1998 and Spring 1999, with students enrolled in four sections of elementary statistics. Each semester one class met on Monday-Wednesday-Friday mornings for 50 minutes each period and the other class met on Tuesday-Thursday mornings for 75 minutes each period. During the fall semester each class began at 9:00 AM and during the spring semester the Monday-Wednesday-Friday section met at 8:00 AM while the Tuesday-Thursday section met at 9:00 AM.

Design

The study used a quasi-experimental design with a control group and an experimental group. Both sections of the control group were taught in the fall and both sections of the treatment group were taught in the spring. This design eliminated the potential competition or perception of unfairness among students experiencing different formats of the course during the same semester.

Grabowski, Harkness, Birdwell, and Rosenberger (1995) designed a study in which one class of statistics students participated in classroom activities while a different section that same semester did not. "Students in the second class had heard of the activities being done in the earlier class and were asking why they were not being done in their class" (p. 95). The researchers subsequently modified their study design to include activities in both sections. Sterling and Gray (1991) noted the opposite phenomenon in a study they conducted involving the use of computer software in a statistics class. They claimed that the student perception of unequal treatment was actually a limitation of their study, since the software group felt they were being asked to do more work than their counterparts in the other section.

Additionally, there was a threat of experimental mortality, the loss of subjects from the study (Borg and Gall, 1989). Traditionally, substantial numbers of students at this institution withdraw from mathematics courses or receive an insufficient grade to transfer credit to a senior institution. Many of these students re-enroll in the same courses during the next semester. During the past four years, students enrolled in elementary statistics at the study site have had a success rate (completion with a C or better) of between 24% and 61%. This figure is based on the number of students enrolled at the end of the attendance documentation period who complete the course with a C or better. Students withdrawing during the add/drop period or shortly thereafter were not included in the analysis. Three different instructors taught the course over this time period, with student success rates detailed in Table 1. The researcher was not among the instructors teaching the course at this institution during this time period. The North Carolina community college system converted from a quarter system to a semester system in the Fall of 1997. Figures prior to Fall 1997 are based on an 11-week quarter and figures beginning with Fall 1997 are based on a 16-week semester. This institution did not offer elementary statistics during the winter quarters of 1995 or 1996.

Table 1 - Recent Student Success Rates in Elementary Statistics

Term	Instructor*	Success Rate
Spring 1995	Adams	40%
Fall 1995	Adams	24%
Spring 1996	Baker	61%
Fall 1996	Adams	25%
	Baker	41%
Winter 1997	Adams	28%
Spring 1997	Baker	29%
Fall 1997	Carlisle	52%
	Carlisle	51%
Spring 1998	Carlisle	50%
	Carlisle	58%
	Carlisle	44%

*pseudonyms

The researcher kept a reflective journal during the course of the study. Additionally, daily lesson plans included the order of topics, an outline of the examples used, the extent of coverage, reading assignments and practice problems assigned, and the use of technology.

Subjects

The subjects for this study were community college students in a metropolitan area of North Carolina. Students self-selected into either the Monday-Wednesday-Friday section or the Tuesday-Thursday section each semester. The researcher was the only instructor at this institution teaching statistics during this academic year, eliminating instructor preference as a factor in student enrollment decisions. During the fall semester, a third section of the course was taught via student-leased video cassettes using the series *Against All Odds*. During the spring semester, a third section of the course was an evening section taught in a three-hour block one night per week. Neither of these sections was included in the study. Thus, the four classes in the study did not comprise the entire population of students taking this course at this institution this year.

All four sections of the course involved in the study were taught using the same textbook, Brase and Brase (1995) *Understandable Statistics*, Fifth Edition. This text was selected prior to the conception of this study and the researcher was not a member of the text selection committee. While there was no attempt by the researcher to choose a text that would significantly favor one teaching approach over another, this text is widely considered to be a traditional text.

The mathematics department had been scheduled to open a laboratory facility early in the fall semester. Multiple copies of the Student Edition of MINITAB had been purchased by the department for installation in this lab. Unfortunately, this laboratory was reassigned to another department on campus and was not available as planned. As a result, the student edition of MINITAB was available for all students to use on an optional basis in the open lab, a facility that would not accommodate whole-class instruction. MINITAB demonstrations occasionally occurred in class using a portable multimedia unit, and for some topics MINITAB printouts were provided for classroom discussion and analysis.

The instructor/researcher used a TI-83 graphing calculator in class daily. Students in all sections were strongly encouraged, but not required, to purchase a TI-83. The department had anecdotally determined that some students found the cost of this technology is prohibitive. Other students already owned a different brand or model of graphing calculator, most often TI-82 or TI-85. Additionally, the department did not have classroom sets of graphing calculators available.

Student grades were based on a 1000 point total. Each student completed 3 tests (100 points each) and a cumulative final exam (200 points), and had the flexibility to complete the other 500 points based on assessment techniques including homework, writing assignments, and papers. Students had the option of choosing to double one or more test grades. Quizzes were also available as an assessment option for students during the fall semester. The spring semester students were able to choose lab reports as an assessment option instead of the quizzes given in the fall.

Parallel forms of three tests were administered to students in the fall and spring semesters. All students were assessed using the same final examination. Regardless of the scheduled length of class, students in all sections had one hour to complete the major tests and two hours to complete the final exam.

A study guide was developed and distributed to all students before the administration of the three examinations. This study guide served two major purposes. First, consistent with the NCTM *Assessment Standards* (1995) it provided students with information regarding "what they need to know [and] how they will be expected to demonstrate that knowledge . . ."(p. 17). Second, it provided students in all sections of the course with the same information about the assessment process. That is, spring semester students did not benefit from having the fall semester exams accessible from friends or classmates.

The researcher wrote both the fall and spring semester assessments simultaneously. Each examination had three parts: a short answer section, a problem section, and a short essay section. Parallel forms of each question were developed, and items were randomly assigned to either the fall assessment or the spring assessment. After the assignment of items was made, the order of the items on each part of the exam was randomly determined. Another mathematics educator verified that the study guide and both test forms covered the same material at the same level of difficulty. This educator was an adjunct faculty member at another community college in the same state and used the same statistics text.

The students in the spring semester engaged in ten data collection and analysis activities. Some activities were completed individually, some were completed in small groups, and some were whole-class activities. When groups were used they were reassigned for each activity due to the great irregularity in attendance and the high attrition rate among community college students. Descriptions of activities are included in Appendix A, beginning on page 131.

Data Collection

Cognitive Controls

At the beginning of each semester a pre-test was administered to all students. These scores, along with previous grade point averages, if available, were compared to determine whether differences in academic ability existed between the two groups. The pre-assessment items are provided in Appendix B, beginning on page 161.

Assessment

Parallel forms of three researcher constructed tests and identical items from the common final examination provided data that was used to determine the influence of the laboratory activities on student understanding. Test grades were used to determine whether overall student understanding differed throughout each semester. Specific items from the common course final examination were used to assess understanding at the end of the semester. Students were asked to complete four of six problems and three of five essay questions on the examination. The researcher was interested in overall student performance, potential differences in the question choices students might have made, and specific patterns of responses in the essay portion of the examination. Items from the three tests are included as Appendix C, beginning on page 163, and the items from the final examination are included as Appendix D, beginning on page 177.

Additionally, the Statistical Reasoning Assessment (SRA) developed by Garfield and Konold (Garfield, 1998) was used to compare student understanding. A 20-item multiple choice test, the SRA was developed and validated as part of the evaluation of an NSF-funded project, ChancePlus. "The SRA has been used . . . [with] high school and college students in a variety of statistics courses, to evaluate the effectiveness of curricular materials and approaches as well as to describe the level of students' statistical reasoning" (Garfield, 1998, p. 4). The SRA was administered during the final examination period as the third section of the final and is included as Appendix E, beginning on page 182.

Attitude Inventories

At the end of each semester, two attitude instruments were administered to students, the Survey of Attitudes Toward Statistics (SATS) and the STARC-CHANCE Abbreviated Scale (SCAS) (Gal, Ginsburg, & Schau, 1997). The SATS is a seven-point Likert-type instrument with 28 items designed to measure four different aspects of college students' attitudes toward statistics: affect, cognitive competence, value, and difficulty. The SCAS is a five-point Likert-type instrument with 10 items that reflect the outcomes of statistics education. The attitude inventories are included as Appendix F, beginning on page 195.

Interviews

Eight weeks after each semester, five students were interviewed. All ten students earned grades of B or C in the course. Originally, a student having earned a grade of A was scheduled to participate in the fall interview process. She was the final interview scheduled and withdrew at the last minute. Rescheduling with another A student was not practical at that time. Since students earning Ds or Fs were either not available or were re-enrolled in the course the following semester, only students earning Bs or Cs were interviewed during the fall. The same criterion was used for the spring. The group of 10 included men and women, black students and white students, traditional aged students and nontraditional aged students. Each audiotaped interview lasted approximately 45 minutes. The interview guide is provided in Appendix G, beginning on page 198.

Data Analysis

To insure reliability of quantitative data, an expert with a doctorate in mathematics education and extensive background in statistics graded a random sample of final examination problems from each class using the rubrics and answer keys developed by the researcher. The expert received the examinations (with names obscured) by mail and graded and returned the exams to the researcher. Scores were compared to insure that the researcher did not influence the outcome of the study through biased grading.

In addition to comparing the quantitative results of students' problem solutions, the researcher also examined the content of their essay responses to determine if patterns existed in students' answers, and if these patterns differed among students in the two versions of the course. Students' statistical reasoning evidenced in their writing was analyzed using the framework developed by Werner (1993).

Werner adapted Shaughnessy's (1992) four Types of Conceptions of Stochastics (non-statistical, naive-statistical, emergent-statistical, and pragmatic-statistical) and offers seven categories of statistical reasoning. The first is *Category 0: Non-determinable Reasoning* and is characterized by student guessing, lack of explanation, or both. Students thinking at the *Category 1: Arithmetical Reasoning* level use past understanding of number relationships to inappropriately describe statistical concepts. *Category 2: Naive Statistical Reasoning* involves the incorrect or inappropriate use of statistical terminology by students. Students who fall into *Category 3: Procedural Statistical Reasoning* correctly use formulas and procedures, but with little evidence of conceptual understanding. Students demonstrating understanding in *Category 4: Developing Statistical Reasoning* show some appropriate conceptual development behind their use of formulas and procedures. Students who make connections between concepts or show deep understanding of one or more ideas are at *Category 5: Functional Statistical Reasoning*. Finally, *Category 6: Expert Statistical Reasoning*, is a category reserved for students displaying advanced thinking based on underlying mathematical models and the interactions between probability and statistics.

Student interviews also provided data indicating their level of statistical understanding. The interviews began with an opportunity for students to reflect on the course as a whole. They were then asked to describe and expand on what they considered to be the most interesting, most important, and most challenging topics. This was followed by content questions in which students were asked to explain specific ideas and generate suggestions for collecting and analyzing data. The interview data was used to triangulate and expand on other data sources.

Summary

This study was designed to investigate the impact of hands-on laboratory activities on students' understanding of statistics. Community college students participated in two different formats of an elementary statistics course. One group of students participated in a lecture/discussion course using technology while the other group had hands-on data collection and analysis activities included.

Three tests, the course final examination problems and essays, the Statistical Reasoning Assessment, attitude inventories, and interviews provided the data for the study. Both quantitative and qualitative methodologies were used to address the research questions. Quantitative data were analyzed using SAS while qualitative data were analyzed using the constant comparative technique. This method of data analysis involves the continual comparison of qualitative data to code and categorize responses.

CHAPTER 4

RESULTS

Both quantitative and qualitative results of the study are reported in this chapter. First, the analysis of two cognitive control measures, grade point averages and pretest scores, is given. Student achievement is reported quantitatively as results of three course tests, selected final examination problems, and the Statistical Reasoning Assessment (Garfield, 1998). Final examination essay responses provided additional data used to investigate student understanding and their perceptions of the uses of statistics. These data were analyzed qualitatively using the constant comparative technique. Finally, the results of two attitude assessments, the Survey of Attitudes Toward Statistics (SATS) and the STARC-CHANCE Abbreviated Scale (SCAS), are provided. Interview data are discussed in the next chapter where they are used to support and inform results from the other data sources.

Cognitive Controls

Since students were not randomly assigned to control and treatment groups prior to the study, their previous college grade point averages (GPAs) and their scores on an instructor designed pre-test were used to determine if differences in mathematical ability existed among the students. The pre-test was an open-ended, 10-item assessment which included algebra and statistics concepts included in the K-12 curriculum. Student responses were determined to be either correct or incorrect, with integer scores from 0 to 10 possible. No partial credit was awarded.

Analyses of variance using the factors of semester (control during the fall and treatment during the spring) and day of the week (Monday-Wednesday-Friday and

Tuesday-Thursday) were conducted. Although a difference in student GPAs was not noted, there was a significant difference in the pretest scores of students enrolled on different days. Students enrolled in the Tuesday-Thursday sections both semesters had significantly higher pretest scores than students enrolled in the Monday-Wednesday-Friday sections. Therefore, pretest scores are used as a covariate when hypotheses involving cognitive measures are tested.

Descriptive statistics are summarized in Table 2 and Table 3. Because some fall semester students were enrolled in college for the first time, the number of subjects in the GPA analysis is less than that for the pretest analysis. Results of the ANOVA are provided in Table 4.

Table 2 – GPA Means and Standard Deviations

Semester	Day	n	Mean	Standard Deviation
Fall	Monday-Wednesday-Friday	21	2.86	0.71
(Control)	Tuesday-Thursday	29	2.59	0.96
Spring	Monday-Wednesday-Friday	31	2.96	0.64
(Treatment)	Tuesday-Thursday	31	2.64	0.85

Table 3 – Pretest Means and Standard Deviations

Semester	Day	n	Mean	Standard Deviation
Fall	Monday-Wednesday-Friday	33	3.70	1.42
(Control)	Tuesday-Thursday	30	4.50	2.03
Spring	Monday-Wednesday-Friday	32	4.16	2.20
(Treatment)	Tuesday-Thursday	28	4.89	2.10

Table 4 - Cognitive Control Measures ANOVA

Source	DF	SS	F value	Pr > F
GPA				
Semester	1	(0.14161657)	0.22	0.6418
Day	1	(2.27821750)	3.50	0.0641
Semester*Day	1	(0.01398051)	0.02	0.8838
Error	108	70.29968256		
Total	111	72.90999643		
Pretest				
Semester	1	(5.5599894)	1.46	0.2294
Day	1	(18.150625)	4.76	0.0310*
Semester*Day	1	(0.0337826)	0.01	0.9251
Error	119	453.3670184		
Total	122	477.040650		

Note: Values in parentheses represent Type III SS.

A total of 81 students completed the course, 39 in the fall and 42 in the spring. Of these 81, three students were absent the day the pretest was administered and were unable to schedule a retake within a few days. Thus, because the pretest score was used as a covariate for the cognitive measures, these three students were omitted from the sample for these analyses. One of these students was enrolled during the fall semester and two of these students were enrolled during the spring semester.

Course Tests

Three tests were administered each semester. A study guide was given to each group of students before each test to ensure that the spring students would not benefit by having examinations from the fall semester available. Both the fall and spring versions of the exams were written simultaneously and items were randomly assigned to one or the other version. Additionally, the order of the items on the tests was also randomly determined. Testing conditions such as tables, note cards, and time allotted were identical both semesters.

Three independent analyses of covariance (ANCOVA) tests were conducted using the total test scores as the dependent variables, and including the pretest score as the covariate. Additionally, because some students used a TI-83 calculator with internal statistics features while others did not, an extra variable for TI-83 was included in the model. Preliminary analyses indicated that neither the pretest score nor TI-83 usage showed significant interaction with the semester or day effects and these interaction terms were eliminated from the model. One student did not take the second test during the spring semester, but remained in the course and took all other tests and the final exam. Therefore, the sample size for the second test is one less than that for the other tests.

Student achievement was significantly better for spring students on the first test, but there was no significant difference in achievement noted on the second or third test. Only two laboratory activities were conducted during the beginning unit, one involving sampling variability and one involving the law of large numbers. It is unlikely that the difference in achievement on the first test can be attributed solely to the inclusion of these constructive hands-on activities. It is also plausible that students in the spring were better adjusted to school and better prepared for their first examination in January than students in the fall who took their first examination in September. Additionally, the researcher's journal indicates that five students during the fall semester did not come to the first test with the suggested formula note card prepared. Analysis of test data is provided in Table 5 and Table 6.

Table 5 – Course Test Least Squares Means and Standard Errors

	Semester	n	LSMean	Standard Error
Test 1	Fall (Control)	38	63.22	2.39
	Spring (Treatment)	40	76.74	2.26
Test 2	Fall (Control)	38	69.96	3.74
	Spring (Treatment)	39	68.75	3.56
Test 3	Fall (Control)	38	59.29	3.75
	Spring (Treatment)	40	59.52	3.55

Table 6 – Course Tests ANCOVA

Source	DF	SS	F value	Pr > F
Test 1				
Semester	1	(3277.584653)	17.47	<.0001**
Day	1	(918.621070)	4.90	0.0301*
Semester*Day	1	(38.878012)	0.21	0.6503
Pretest	1	(1148.277606)	6.12	0.0157
TI-83	1	(138.404275)	0.74	0.3933
Error	72	13508.40759		
Total	77	20795.53846		
Test 2				
Semester	1	(26.169226)	0.06	0.8118
Day	1	(3526.325730)	0.32	0.0071**
Semester*Day	1	(146.550632)	7.70	0.5734
Pretest	1	(0.745106)	0.00	0.9679
TI-83	1	(506.552349)	1.11	0.2965
Error	71	32514.18532		
Total	76	36857.53247		
Test 3				
Semester	1	(0.98779692)	0.00	0.9632
Day	1	(6.89172193)	0.00	0.9031
Semester*Day	1	(77.65307101)	0.17	0.6830
Pretest	1	(91.96382692)	0.20	0.6568
TI-83	1	(0.30934096)	0.00	0.9794
Error	72	33258.75528		
Total	77	33482.98718		

Final Examination

A common final examination containing problems, essay questions, and the Statistical Reasoning Assessment (SRA) was administered at the end of each semester. The problem section and the essay section of the final exam were constructed by the instructor/researcher and were identical for both groups of students. Students were instructed to complete four of six problems and three of five essays, but were told that if they chose to attempt more than the required number of items the scores of the best items in each section would be counted. Students were allowed to use a note card and statistical tables to complete the problem section of the final. They submitted this part of the exam along with their note cards and tables when they received the essay section. Students received the SRA upon submission of the essay responses. Two hours were allotted for the entire assessment and students paced themselves using instructor-provided guidelines with suggested time frames for each of the three components.

Final Examination Choices

Chi-Square analyses were performed to determine whether there were differences in the percentages of students each semester who had specific problems and essay items counted among the best of their attempts. Only students who had four of six problems which could be clearly identified as "best", and/or three of five essays which could be clearly identified as "best" were included in these analyses. Subsequently, students who completed less than the required number of items in each section were excluded. Similarly, students who attempted more than the required number of items, but who had identical scores on the fourth and fifth "best" problems or third and fourth "best" essays were not included in these analyses. A total of 66 students were included in the analyses for the problems (33 from the fall semester and 33 from the spring semester) and a total of 78 students were included in the analyses for the essays (38 from the fall semester and 40 from the spring semester).

The problems and essays selected for consideration were those that could reasonably have been influenced by the treatment. That is, one or more of the activities during the spring semester involved the concept(s) tested by those items. The first four problems tested concepts that were addressed in some way by one or more laboratory activities. These concepts were chi-square test for independence, correlation, binomial probabilities, and confidence intervals. Chi-square analyses for each of these indicate that the proportion of students having those problems included as one of their four best items was not significantly different from semester to semester. The results are summarized in Table 7.

Table 7 - Independent Chi-Square Analyses for Problem Selection by Students

	DF	χ^2 Value	p-value
Problem 1	1	0.6652	0.4147
Problem 2	1	0.0667	0.7962
Problem 3	1	0.7131	0.3984
Problem 4	1	2.0625	0.1510

Similarly, three of the five essay questions tested concepts that could reasonably have been influenced by the treatment. The concepts involved were hypothesis tests, least squares regression, and margin of error and confidence intervals. Chi-square analyses for each of these indicate that the proportion of students having those essays included as one of their three best items was not significantly different from semester to semester. The results are summarized in Table 8.

Table 8- Independent Chi-Square Analyses for Essay Selection by Students

	DF	χ^2 Value	p-value
Essay 2	1	0.8382	0.3599
Essay 4	1	0.0093	0.9230
Essay 5	1	0.7809	0.3769

Student Achievement

Final Exam Problem Scores

Ten final examination papers were randomly selected each semester and the problems were graded by a mathematics educator with a minor in statistics using the guidelines and rubric provided by the researcher. This individual previously taught this course at another community college in the system using the same text. Additionally, this adjunct professor teaches statistics at a local four-year liberal arts college and the local campus of the state university. Inter-rater reliability values were determined for each of the six problems, using the scores of students who had that item counted among their best four of six. These values are provided in Table 9.

Table 9 – Inter-rater Reliability Scores for Final Exam Problems

Problem	Score
Problem 1	.62
Problem 2	.99
Problem 3	.94
Problem 4	.87
Problem 5	.79
Problem 6	.63

An analysis of covariance (ANCOVA) was conducted using the total problem score as the dependent variable, and including the pretest score as the covariate. Additionally, because some students used a TI-83 calculator with internal statistics features while others did not, an extra variable for TI-83 was included in the model. Preliminary analyses indicated that neither the pretest score nor TI-83 usage showed significant interaction with the semester or day effects and these interaction terms were eliminated from the model. Table 10 shows no significant difference between the two treatment groups with respect to the total problem score.

Table 10 – Total Problem Score ANCOVA

Source	DF	SS	F value	Pr > F
Semester	1	(0.015640)	0.00	0.9956
Day	1	(148.851240)	0.29	0.5909
Semester*Day	1	(28.259750)	0.06	0.8147
Pretest	1	(1382.359145)	2.71	0.1042
TI83	1	(9.088555)	0.02	0.8942
Error	72	36761.13122		
Total	77	39005.94872		

Note: Values in parentheses represent Type III SS.

It is possible that the different groups of students performed differently on specific items although their total scores for the problem part of the exam were not significantly different. Subsequent analyses of covariance were performed on the four problems identified earlier as items that might be influenced by the different treatments. Again, only those students who had one of these items clearly among their "best" in each section were included in the analysis.

Problem 1: Chi-Square Test for Independence

The first problem was a chi-square test for independence involving a categorical variable representing three levels of cholesterol (high, borderline, and low) and a categorical variable representing five geographic regions (northeast, southeast, central, northwest, and southwest). Students were to perform a test of hypothesis to determine whether cholesterol level and geographic region of residence were independent. No computation was involved; students were provided with the calculated value of the test statistic.

During the spring semester, students collected and analyzed data to determine whether individuals' birth order and preference for playing individual or team sports were independent. Each student polled 20 adults and then each class combined their data for common analysis. The activity guided students through the development of the test statistic and the analysis of the data.

Only the interaction of pretest score and day (Monday-Wednesday-Friday vs. Tuesday-Thursday) was found to be significant in the preliminary analysis of the first problem ($p = 0.0484$); subsequently all other interaction terms involving the pretest score or the TI-83 were removed from the model. No significant differences in the performance on Problem 1 were noted between the two groups of students. Results are summarized in Table 11.

Table 11 – Problem 1: Chi-Square Test for Independence

Source	DF	SS	F value	Pr > F
Semester	1	(241.5388332)	4.12	0.0653
Day	1	(118.5663173)	2.02	0.1806
Semester*Day	1	(1.9806830)	0.03	0.8573
Pretest	1	(0.4408250)	0.01	0.9324
Pretest*Day	1	(161.2719249)	2.75	0.1232
TI83	1	(21.810943)	0.37	0.5534
Error	12	704.102731		
Total	18	1208.105263		

Note: Values in parentheses represent Type III SS.

Problem 2: Correlation

The second problem asked students to draw a scatterplot for 9 pairs of data, use a provided MINITAB printout (or their TI-83 calculators) to answer questions about the correlation coefficient and the coefficient of determination, and perform a test of hypotheses to determine whether the correlation was significant. Two laboratory activities during the spring semester involved one or more of these concepts. One activity involved the correlation between students' arm span measurements and their heights. The other activity was an ice cream taste test where students investigated the relationship between fat content and quantitative indices of vanilla ice cream quality.

None of the possible interactions between TI-83 and pretest with semester and day were significant in the preliminary analysis so these interaction terms were removed from the model. Subsequently, no significant differences in performance on Problem 2 were evident in the ANCOVA. Results are summarized in Table 12.

Table 12 – Problem 2: Correlation

Source	DF	SS	F value	Pr > F
Semester	1	(41.0899413)	0.54	0.4723
Day	1	(27.2845595)	0.36	0.5571
Semester*Day	1	(7.0207621)	0.09	0.7649
Pretest	1	(29.9815249)	0.39	0.5384
TI83	1	(35.816202)	0.47	0.5018
Error	17	1292.886531		
Total	22	1509.826087		

Note: Values in parentheses represent Type III SS.

Problem 3: Binomial Probabilities

The third final exam problem involved the computation of binomial probabilities using tables or a TI-83. The spring semester students encountered the calculation of binomial probabilities when they were introduced to the concept of hypothesis testing using a penny-spinning activity. As part of that activity they determined the probability of obtaining their individual results under the null hypothesis assumption that the probabilities of heads and tails were both .5.

A preliminary analysis showed none of the possible interactions between TI-83 and pretest with semester and day were significant so these interaction terms were removed from the model. Subsequently, no significant differences in performance on Problem 3 were evident in the ANCOVA. Results are summarized in Table 13.

Table 13 – Problem 3: Binomial Probabilities

Source	DF	SS	F value	Pr > F
Semester	1	(9.8977110)	0.17	0.6785
Day	1	(0.2508627)	0.00	0.9473
Semester*Day	1	(49.2760882)	0.87	0.3569
Pretest	1	(62.0687985)	1.09	0.3018
TI83	1	(142.643368)	2.51	0.1204
Error	43	2443.237176		
Total	48	2731.551020		

Note: Values in parentheses represent Type III SS.

Problem 4: Confidence Interval

Students were asked to compute and interpret a confidence interval for a population mean. During a spring semester activity, each class formed 10 groups and then each group constructed and interpreted a 90% confidence interval for the population proportion of blue m&ms®. The class shared results and discussed the interpretation.

None of the possible interactions between TI-83 and pretest with semester and day were significant in the preliminary analysis of Problem 4 so these interaction terms were removed from the model. No significant differences in performance on Problem 4 were evident in the ANCOVA. Results are summarized in Table 14.

Table 14 – Problem 4: Confidence Interval

Source	DF	SS	F value	Pr > F
Semester	1	(42.2510452)	0.65	0.4239
Day	1	(0.0543025)	0.00	0.9771
Semester*Day	1	(33.1496080)	0.51	0.4785
Pretest	1	(309.4708278)	4.18	0.0454
TI83	1	(4.7767043)	0.02	0.8923
Error	57	3711.773613		
Total	62	4087.079365		

Note: Values in parentheses represent Type III SS.

Summary

The quantitative analysis of the final exam problems shows no significant difference in points earned by the treatment and control groups. However, while the students' solutions to the problems appear to demonstrate the same level of achievement or understanding, the nature of their responses to the essay items are of further interest. A qualitative analysis of students' writing on the essay questions follows in a later section of this chapter.

Statistical Reasoning Assessment

The Statistical Reasoning Assessment (SRA) is a 20 item multiple choice instrument designed to measure students' correct and incorrect reasoning regarding statistics and probability. There are 8 correct reasoning skills and 8 misconceptions measured by the SRA (Garfield, 1998) and each of the 20 items contributes to at least one of the 16 specific scales. Individual items have between three and eight responses, some correct and some incorrect, which might be keyed to different correct reasoning skills or misconceptions. Some responses contribute to none of the 16 scales. Garfield's summary of the eight correct reasoning skills and eight misconceptions along with the items corresponding to each is replicated in Table 15.

Table 15 - SRA Correct Reasoning and Misconception Scales

Correct Reasoning Skills	Corresponding Items and Alternatives
1. Correctly interprets probabilities	2d, 3d
2. Understands how to select an appropriate average	1d, 4ab, 17c
3. Correctly computes probability	
a. Understands probabilities as ratios	8c
b. Uses combinatorial reasoning	13a, 18b, 19a, 20b
4. Understands independence	9e, 10df, 11e
5. Understands sampling variability	14b, 15d
6. Distinguishes between correlation and causation	16c
7. Correctly interprets two-way tables	5-1d*
8. Understands importance of large samples	6b, 12b
Misconceptions	**Corresponding Items and Alternatives**
1. Misconceptions involving averages	
a. Averages are the most common number	1a, 17e
b. Fails to take outliers into consideration when computing the mean	1c
c. Compares groups based on their averages	15bf
d. Confuses mean with median	17a
2. Outcome orientation misconception	2e, 3ab, 11abd, 12c, 13b
3. Good samples have to represent a high percentage of the population	7bc, 16ad
4. Law of small numbers	12a, 14c
5. Representative misconception	9abd, 10e, 11c
6. Correlation implies causation	16be
7. Equiprobability bias	13c, 18a, 19d, 20d
8. Groups can only be compared if they are they same size	6a

*Note: For item 5, subjects have to choose from two options before they can make further selection from four alternatives under each option.

Attempting to determine criterion-related validity, Garfield (1998) found that correlations of the SRA with student course assessments such as quizzes, exams, and final score were "all extremely low, suggesting that statistical reasoning and misconceptions are unrelated to students' performance in a first statistics course" (p. 7). Therefore, the SRA could show differences in the treatment and control groups that were not apparent in the analysis of the final examination problems.

Six students who did not complete every item in the SRA were omitted from the sample. Three of these students were from the fall semester and three were from the spring semester. In each semester, one omitted student was from one section and two omitted students were from the other section.

Four correct reasoning scales and three misconception scales measured by the SRA were determined to be connected in some way with one or more of the 10 laboratory activities. Since these scales focus on different phenomena, and successful reasoning in one area would not necessarily indicate successful reasoning in another area, univariate analyses of covariance were performed on each of the seven selected scales. Because of the limited number of scores possible for some of the scales, there is some concern that all of the assumptions for analysis of covariance are met. However, the inclusion of independent variables for day of the week and pretest score and the benefit of consistency in the analyses among the scales makes the analysis of covariance approach advantageous.

A summary of the selected SRA Scales, related laboratory activities, and statistically significant differences among students is provided in Table 16. The discussion that follows provides additional detail for each of the seven SRA scales included in the analysis.

Table 16 – Summary of Analyzed SRA Scales

SRA Scale	Associated Activities	Semester Effect p-value
C4: Understands Independence	Coin Toss	0.4932
C5: Understands Sampling Variability	Headcount Let's Go For a Spin	0.9500
C6: Distinguishes Between Correlation and Causation	Do You Measure Up? We All Scream for Ice Cream	0.3630
C8: Understands the Importance of Large Samples	Round and Round	0.0465 *
M2: Outcome Orientation	Coin Toss Let's Go For a Spin Round and Round	0.7461
M5: Representative Misconception	Coin Toss	0.8057
M6: Correlation Implies Causation	We All Scream For Ice Cream	0.7341

Correct Reasoning Scales

Correct Reasoning Scale 4: *Understands Independence*

The *Coin Toss* activity involved groups of four or five students flipping coins in sequences of four flips. The activity was designed to teach students the difference between the 16 equally likely sequences possible when a coin is flipped four times, and the discrete random variable that counts the number of heads obtained in a sequence of four flips. The concept of independence is key to students' understanding of these ideas.

The fourth correct reasoning skill measured by the SRA uses three adjacent items, 9, 10, and 11, to assess students' understanding of independence. In the first, students were asked to choose the most likely sequence of coin flips from among four choices. Next, students were asked to choose one or more reasons for the response to the previous item. The third item asked students to choose the least likely sequence of coin flips from among four choices.

The ANCOVA for the fourth correct reasoning scale did not show a significant semester effect. The pretest score was included as a cognitive covariate, but neither interaction with semester or day was significant so these interaction terms were excluded from the final analysis. Results are summarized in Table 17.

Table 17 – Correct Reasoning Scale 4: *Understands Independence*

Source	DF	SS	F value	Pr > F
Semester	1	(0.06133357)	0.47	0.4932
Day	1	(0.22108438)	1.71	0.1953
Semester*Day	1	(0.00928453)	0.07	0.7895
Pretest	1	(0.28924628)	2.24	0.1393
Error	67	8.65643175		
Total	71	9.49180000		

Note: Values in parentheses represent Type III SS.

Correct Reasoning Scale 5: *Understands Sampling Variability*

Although many activities included the ideas of sampling variability, two specific activities during the spring semester focused on this concept. *Headcount* involved individual students selecting two samples of stick-figure clusters (one of size 5 and one of size 10) by eye to be "representative" of the one hundred clusters or households. After computing and recording descriptive statistics for those samples, individual students' results were recorded for the whole class to observe and discuss. Next, students chose random samples of the same sizes using a random number table or random number generator on their TI-83s and the same procedure was repeated again. Class discussion compared students' sample statistics to the population mean and the population standard deviation, illustrating the importance of random sampling and large samples.

Later in the semester as part of the penny-spinning introduction to hypothesis testing, *Let's Go For a Spin*, students recorded the results of their ten individual penny spins on the board. Comparing the numbers of heads and tails in the sets of ten spins and discussing those outcomes illustrated the variability among their samples.

The fifth correct reasoning skill measured by the SRA involves two items, 14 and 15. The first item asked students to determine whether a hospital with 50 births or a hospital with 10 births is more likely to see 80% or more female births, or whether they are equally likely to see this outcome. The second item asked students to compare two dot plots, each involving a sample of 20 students in a sleep study. Students were to use the dot plots to choose one statement they most agreed with from among the six options provided.

The ANCOVA showed no significant semester or day effect for the fifth correct reasoning scale. Once again, the pretest was used as a covariate, but the interaction terms of the pretest with semester or day were insignificant in the preliminary analysis and so were omitted from the final analysis. Results are summarized in Table 18.

Table 18 – Correct Reasoning Scale 5: *Understands Sampling Variability*

Source	DF	SS	F value	Pr > F
Semester	1	(0.00020480)	0.00	0.9500
Day	1	(0.13342174)	2.58	0.1127
Semester*Day	1	(0.05374301)	1.04	0.3114
Pretest	1	(0.14017465)	2.71	0.1042
Error	67	3.46134051		
Total	71	3.71875000		

Note: Values in parentheses represent Type III SS.

Correct Reasoning Scale 6: *Distinguishes Between Correlation and Causation*

The activity *Do You Measure Up?* involved students computing and interpreting the correlation coefficient between their arm spans and their heights. Another activity, *We All Scream for Ice Cream*, focused on regression and correlation. Students investigated the relationships between the fat content and numerical ratings of the qualities of texture, flavor, and sweetness for different brands of vanilla ice cream.

One SRA item, 16, contributed to this correct reasoning scale. The item asks students to evaluate a research study involving the television watching habits of children and their grades in school. Students are given six statements following the description of the study, and they are asked to mark every statement of the six with which they agreed.

Results of the ANCOVA for the sixth SRA correct reasoning scale are provided in Table 19. No significant difference in the students' performance was apparent. The pretest scores are again used as covariates. The insignificant interaction terms involving pretest were omitted from the model.

Table 19 – Correct Reasoning Scale 6: *Distinguishes Between Correlation and Causation*

Source	DF	SS	F value	Pr > F
Semester	1	(0.19678805)	0.84	0.3630
Day	1	(0.01122931)	0.05	0.8275
Semester*Day	1	(0.03216874)	0.14	0.7123
Pretest	1	(0.10571507)	0.45	0.5044
Error	67	15.71852735		
Total	71	16.00000000		

Note: Values in parentheses represent Type III SS.

Correct Reasoning Scale 8: *Understands Importance of Large Samples*

Round and Round was specifically designed to teach the Law of Large numbers. Students worked in pairs recording the results of 100 spins of a spinner, completed in 10 sets of 10. They compared the results of each set of 10 spins and the cumulative results after each set of 10 spins with the theoretic outcome they would expect based on the design of the spinner. The sampling activity, coin-toss activity, and penny-spinning activity (each discussed previously) also reinforced the importance of large samples.

Two SRA items, 6 and 12, contributed to this scale. Item 6 describes a drug study with 20 patients in the treatment group and 10 patients in the control group. Students are asked to choose all reasons among the five provided that they might question the results of the study. The response that contributes to this scale is "The sample of size 30 is too small to permit drawing conclusions". Item 12 provides students with a car purchase situation. They are asked to decide whether they'd use a Consumer Reports' study of 400 vehicles of each type or the advice of three friends (one of whom had a bad experience with a car under consideration) when evaluating the purchase of that new car.

A significant semester effect is shown in the results of the ANCOVA for the eighth SRA correct reasoning scale. The spring semester (treatment) students had significantly higher scores on this correct reasoning scale, indicating a better understanding of the importance of large samples. Results are provided in Table 20. Additionally, the least squares means for each semester along with the corresponding standard errors are provided in Table 21. The least squares means are adjusted means taking the effect of the pretest covariate into account.

Table 20 – Correct Reasoning Scale 8: *Understands Importance of Large Samples*

Source	DF	SS	F value	Pr > F
Semester	1	(0.48148739)	4.12	0.0465 *
Day	1	(0.03870999)	0.33	0.5617
Semester*Day	1	(0.21974998)	1.88	0.1751
Pretest	1	(0.06121013)	0.52	0.4720
Error	67	7.83916866		
Total	71	8.81944444		

Note: Values in parentheses represent Type III SS.

Table 21 – Least Squares Means and Standard Error for Correct Reasoning Scale 8

Semester	LSMean	Standard Error
Fall (control)	0.514	0.058
Spring (treatment)	0.681	0.058

Misconception Scales

Misconception 2: *Outcome Orientation*

This misconception is illustrated by students' reluctance to take an entire sequence of events into account when evaluating probability statements. The tendency is to focus instead on single events. The laboratory activities that involved investigation of sequences of coin tosses and spins (of both spinners and pennies) could have influenced students' understanding in this area.

Five distinct items on the SRA contributed to this scale - items 2, 3, 11, 12, and 13. One of these, item 11, had three different responses that indicated this misconception. This particular item provided students with four different sequences of five coin tosses and asked them to choose the least likely sequence. A fifth choice, "All four sequences are equally unlikely," was also provided. The other four items contributing to this scale were not as closely tied to the spring semester laboratory activities.

The ANCOVA for the second misconception scale did not show a significant difference between the students' responses. Results are summarized in Table 22.

Table 22 – Misconception Scale 2: *Outcome Orientation Misconception*

Source	DF	SS	F value	Pr > F
Semester	1	(0.00364260)	0.11	0.7461
Day	1	(0.02581822)	0.75	0.3898
Semester*Day	1	(0.02188196)	0.63	0.4283
Pretest	1	(0.09034916)	2.62	0.1101
Error	67	2.30886297		
Total	71	2.49777778		

Note: Values in parentheses represent Type III SS.

Misconception 5: *Representativeness Misconception*

Students holding the representativeness misconception believe that the likelihood of a sample is influenced by how closely it mirrors the population from which it was chosen. Such students would expect a sequence of coin tosses with equal numbers of heads and tails to be more likely than a sequence with few of one outcome and many of the other. The *Coin Toss* investigation during the spring semester specifically addressed this misconception.

Three adjacent and related SRA items, 9, 10, and 11, contributed to this scale - all involved sequences of coin tosses. In the first, students were asked to choose the most likely sequence from among four choices. Next, students were asked to choose one or more reasons for the response to the previous item. The third item asked students to choose the least likely sequence from among four choices.

The ANCOVA using pretest as a covariate did not indicate a significant difference between the responses of the two groups of students. Results are displayed in Table 23.

Table 23 – Misconception Scale 5: *Representative Misconception*

Source	DF	SS	F value	Pr > F
Semester	1	(0.00244727)	0.04	0.8357
Day	1	(0.00019571)	0.00	0.9532
Semester*Day	1	(0.00214474)	0.04	0.8460
Pretest	1	(0.00102266)	0.02	0.8933
Error	67	3.78056310		
Total	71	3.78683194		

Note: Values in parentheses represent Type III SS.

Misconception 6: *Correlation Implies Causation*

Two activities involved the concept of correlation. In the first, students explored the relationship between arm span and height. Correlation was also included in the linear regression activity, a vanilla ice cream taste test. Students looked at the relationship between the fat content in grams per half-cup serving and numerical ratings of flavor, texture, and sweetness. Correlation coefficients were computed along with the least-squares linear regression equations.

One SRA item was used to assess this misconception. The item asked students to evaluate a study comparing elementary school students' television watching habits with their grades in school. (This item was also used to assess the sixth correct reasoning scale.) Students were provided with six responses and were instructed to choose all those they agreed with. Two responses contributed to this scale, so scores for individual students could be 0, 1 or 2, corresponding to the number of inappropriate responses they checked.

The ANCOVA did not show a significant difference between the responses of students who were enrolled in the lecture/discussion format course during the fall and those who were enrolled in the course including laboratory activities. However, there was a significant effect for days of the week. Students enrolled in the Tuesday-Thursday sections had significantly higher scores than students enrolled in the Monday-Wednesday-Friday sections. Table 24 provides the ANCOVA results while Table 25 reports least squares means with corresponding standard errors.

Table 24 – Misconception 6: *Correlation Implies Causation*

Source	DF	SS	F value	Pr > F
Semester	1	(0.05615781)	0.12	0.7341
Day	1	(2.13901745)	4.43	0.0391*
Semester*Day	1	(1.52724078)	3.16	0.0799
Pretest	1	(1.24093567)	2.57	0.1136
Error	67	32.34845826		
Total	71	36.61111111		

Note: Values in parentheses represent Type III SS.

Table 25 – Least Squares Means and Standard Error for Misconception 6

Day	LSMean	Standard Error
MWF	0.482	0.109
TT	0.847	0.130

Summary

Four correct reasoning scales and three misconception scales were related to one or more laboratory activities included in the spring semester treatment. Each scale was examined using independent univariate analyses of covariance with the pretest score being used as the covariate in each case. A significant treatment effect was detected for the eighth SRA correct reasoning scale, *Understands the Importance of Large Samples*. A significant day effect was noted for the sixth misconception scale, *Correlation Implies Causation*. No other significant differences among the selected SRA items were found.

Qualitative Analysis of Essays

Three essay questions from the final examination were analyzed qualitatively. These questions included content that related to some of the laboratory activities conducted during the spring semester. Concepts involved in these essays were hypothesis testing, linear regression, and confidence intervals and margin of error. Students were allowed to attempt more than three essays, understanding that the best three of five would be counted towards the exam grade. Only student responses that were deemed one of the three best responses were analyzed. Student writing was examined to determine patterns, common characteristics, and differences in their responses. Classification according to Werner's (1995) framework, which was discussed on page 39, is included along with tables summarizing the responses for each essay.

Essay 2: Hypothesis Testing

Students were asked to explain hypothesis testing to a friend who had registered for statistics. They were instructed to include the ideas of null and alternative hypotheses, α, one-tailed and two-tailed tests, decisions, and conclusions in their explanations. In addition to students' general discussion of hypothesis tests, explanations of these five concepts were investigated separately. Thirty-four students from the fall control group and 40 students from the spring experimental group attempted this item. All of the 74 students had this item count as one of their three best essays.

Six percent of the fall semester students and 8% of the spring semester students who answered this question did not attempt to explain the general concept of hypothesis testing but immediately began discussing the individual ideas provided as prompts in the question. These non-responses might be classified in Werner's *Category 0: Non-Determinable Reasoning*, which includes "lack of explanation when asked".

Students who responded with an explanation of hypothesis testing generally used one sentence to summarize their thoughts. These single sentence explanations included the notion of hypothesis testing as a procedure for rejecting or not rejecting hypotheses, determining the truth of a statement or proving something was true, as a tool for decision making or determining relationships, or as a way to figure the chance of something happening.

One pattern in student responses was what Werner would refer to as "statistical dump". Students providing this type of explanation use various statistical terms and ideas haphazardly, in a nonsensical fashion. Eight percent of the fall semester students had explanations for hypothesis testing that could be categorized in this way, while twenty percent of spring responses were considered "statistical dumps". An extreme example from the spring semester is, "A hypothesis test is a test to see if a random sample population of a group will be reject or failed to reject a region." Werner includes such responses in *Category 2: Naïve Statistical Reasoning.*

Other responses included those claiming that hypothesis tests were a way to "prove" something or find out if something was "true" or "correct". Twenty-eight percent of fall semester students made such claims, while 12% of spring semester students used this type of terminology. For example, one fall semester student said, "A hypothesis test is testing to see if what you believe to be true is actually true or not." A spring semester student offered the following: "Hypothesis tests are experiments used to prove a theory."

Many students described hypothesis testing as a process for rejecting or failing to reject (some used the word "accept") hypotheses. Nineteen percent of the fall students offered this type of explanation, while 10% of spring students' responses were categorized this way. One fall semester student wrote that a hypothesis test " . . . is the procedure whereby we decide to accept or reject a hypothesis." Similarly, a spring semester student wrote, "Hypothesis tests are used to determine if you can reject or accept a hypothesis." Werner's *Category 3: Procedural Statistical Reasoning*, would include such explanations.

The fourth response type that differed in frequency among the fall and spring students was that of hypothesis testing as a way to determine if change has occurred. Ten percent of fall students included this idea, while seventeen percent of spring students captured this concept. A fall semester student said "You would usually use this kind of test to see if one thing was better than another or if you are going to have higher numbers if you do something a new way." A spring semester student stated that "Hypothesis tests are used in experimental studies. They compare studies to determine if something has changed or will change." Answers of this type would be characteristic of Werner's *Category 4: Developing Statistical Reasoning*. Student's reasoning is mostly correct, but the depth or originality needed for *Category 5* reasoning is absent.

Students were asked to include the concepts of null and alternative hypotheses in their essays. Fall semester students were more likely to correctly differentiate between the null and the alternative hypotheses. Two common responses from this control group described the null hypothesis as the "status quo" or the hypothesis indicating no change, while the alternative hypothesis contained a statement of change. Students responding with the phrase "status quo" were relying on memories of a classroom explanation taken directly from their notes, which is evidence of reasoning in Werner's *Category 2: Naïve*

Statistical Reasoning. Approximately 30% of control group responses fell in one of these classifications, while 15% of the experimental group responded in this way. Spring semester students were more likely to claim that the null hypothesis is the "thing that you believed was true". Eighteen percent of this experimental group responded in such a fashion – twice the proportion as the fall.

The other common response from the spring semester students that differed from the fall semester was that the null and alternative hypotheses were "opposites" without offering further explanation. Most of the responses were limited to examples using one mean or one proportion. Additionally, two spring semester students included in their answers the concept of testing for relationships. One student explained that "The null hypothesis says that there is no relationship or that the original data has stayed the same." None of the fall semester students referred to relationships among variables in their explanations of hypothesis tests.

Students were also asked to include an explanation of α in their responses. Twelve percent of fall semester students and 20% of spring semester students did not attempt an explanation. Common themes in student answers were the translation of the symbol α as "level of significance" without further comment (provided by 12% each semester), connection of α with determining the rejection region or critical value, and an association of α with the precision or accuracy of results or a way of knowing "how sure" you are. Students providing a translation from the symbolic α to the phrase "level of significance" show reliance on memory as indicated by Werner's *Category 2: Naïve Statistical Reasoning.* Some fall semester students (9%) demonstrated deeper understanding by connecting α to the probability of making a Type I error, and then going on to explain what a Type I error involved. This concept was not included as a prompt in the examination question and demonstrates *Category 5: Functional Statistical*

Reasoning. No spring semester students extended their explanation of α to include ideas relating to Type I error. However, one spring student offered a more specific explanation of the relationship between α and the rejection region. He was the only student to state that "alpha is the amount of area in the rejection region." Other students made broad, less sophisticated statements such as "α will help in determining your rejection region" or "you use α to help you find your critical value."

Although some students in both versions of the course seemed to confuse α with other statistical concepts and/or provide erroneous explanations of α, this was more prevalent during the spring. Fifteen percent of fall students and 23% of spring students had such responses. One fall student wrote "This symbol α stands for your confidence interval." Some spring semester students labeled α as sigma, the confidence level, the degrees of freedom, and one claimed it was a value "which your data must exceed or your statement could be wrong." Werner would characterize these students as reasoning in *Category 2: Naïve Statistical Reasoning*, for using statistical terminology incorrectly, incoherently or inappropriately.

Students' explanations of one-tailed tests and two-tailed tests did not vary much from semester to semester. Common responses included discussion or sketches of the rejection region(s), and association with the symbols >, <, or $\neq$. Twenty-six percent of fall students and 17% of spring students mentioned this concept without providing any explanation. Examples of this kind of statement include "The diagrams can be one or two tails depending on the type of question being answered" and "You must decide on whether it is a one-tailed or two-tailed test" which were provided by spring students. A fall semester student offered "They can be one-tailed test or two-tailed test. These you will find out when you take the class." These students provided *Category 0: Non-determinable Reasoning* responses.

A few responses were in *Category 1: Arithmetic Reasoning*. Here, students inappropriately use previous number relationships in their explanations. One fall semester student, 3% of the control group, and four spring students, 10% of the treatment group, had responses in this category. These students claimed that one-tailed tests were used when one set of data was being analyzed and two-tailed tests were used to compare two things.

Approximately 33% of students each semester provided responses that included mention of increase or decrease in values vs. change in an unknown manner. For example, a spring semester student said, "If you want to see if something is better or worse then you use a one-tailed test. If you just want to see if it has changed then you use a two-tailed test." These responses were at *Category 3: Procedural Statistical Reasoning* and *Category 4: Developing Statistical Reasoning.*

One fall semester student demonstrated thinking in *Category 5: Functional Statistical Reasoning*. After explaining the difference between one-tailed and two-tailed tests, he added, "This is important because with a 2-tailed test your rejection regions are higher which means we must have a higher calculation with our new data in order to reject the null hypothesis." Although some aspects of this statement could be clearer, he is making connections among different concepts, an indicator of *Category 5: Functional Statistical Reasoning.*

Finally, students were asked to discuss the ideas of decisions and conclusions in their explanations. The expectation was that they would describe the decision as either to reject the null hypothesis or fail to reject the null hypothesis, and then provide a conclusion in plain language using the context of the problem. Approximately 40% of students each semester provided such responses. Twenty-six percent of the fall students and 10% of the spring students failed to address these ideas in their essays. Otherwise, student responses did not differ greatly from semester to semester. One exception would be that 18% of fall students appropriately discussed the decision-making process without mentioning a conclusion while 30% of spring students responded in this fashion.

Two fall semester students had unique responses. One said, "decisions are what size α to use and whether to use a 1 tailed test or 2 tailed test." This statement incorporates decisions that a researcher would make when designing a study. Another fall student included the idea that "this decision can change based on the confidence level you are looking for or the percentage of time you are willing to be wrong." While she shows some confusion regarding level of significance and Type I error, she does demonstrate some understanding of how those ideas relate to the decision to reject or fail to reject the null hypothesis.

Table 26 - Werner's Framework Applied to Essay #2

		Category 0 Non-determinable Reasoning	Category 1 Arithmetical Reasoning	Category 2 Naïve Statistical Reasoning	Category 3 Procedural Statistical Reasoning	Category 4 Developing Statistical Reasoning	Category 5 Functional Statistical Reasoning	Category 6 Expert Statistical Reasoning
What are hypothesis tests?	Fall	6%	0%	20%	47%	27%		
	Spring	8%	0%	43%	22%	27%		
Explanation of H_0 and H_1	Fall	8%	0%	42%	50%			
	Spring	10%	0%	70%	20%			
Explanation of α	Fall	9%	0%	27%	29%	26%	9%	
	Spring	10%	0%	43%	36%	12%	5%	
One Tailed/ Two Tailed	Fall	26%	3%	34%	15%	18%	3%	
	Spring	17%	10%	37%	24%	12%	0%	
Decision/ Conclusion	Fall	32%	0%	6%	18%	44%		
	Spring	15%	0%	7%	35%	43%		

Essay 4: Linear Regression and Correlation

Students were asked to describe the least-squares regression line, explain the mathematical meaning of "least squares," tell why researchers use regression analysis, and relate the correlation coefficient to the least-squares regression line. Student responses to each of the four prompts will be discussed separately. Few students chose to answer this item, only seven in the fall and eight in the spring. Six fall semester students and seven spring semester students had their responses to this item count as one of their three best essays.

When students were asked to describe the least-squares regression line, one student, 17% of the fall group, wrote "The least squares regression helps you after you made your graph from your regression problem." Although this student makes a connection between the regression line and the scatterplot, she does not make any meaningful statement and therefore is reasoning at Werner's *Category 2: Naïve Statistical Reasoning*. With this exception, the responses from the control group in the fall were mostly algebraic, reasoning at Werner's *Category 3: Procedural Statistical Reasoning*, because of their reliance on formulas. These students included the equation $y = ax + b$ and/or stated that the line passes through $(\bar{x}, \bar{y})$. Eighty-three percent of the fall responses were of this nature.

Only one of the seven spring semester students, 14%, gave a *Category 3: Procedural Statistical Reasoning* algebraic response similar to those given by most of the fall students. The students in the experimental spring group gave responses that were largely graphical rather than algebraic. Four of the seven, or 57%, described the least-squares regression line in terms of the scatterplot. For example, one student wrote "The least-squares regression line is a line that passes through a graph in the most precise manner possible, with the most equal amount of data on both sides." This type of reasoning is characteristic of Werner's *Category 4: Developing Statistical Reasoning* due to the mostly correct conceptual nature of the response. Two students, or 29% of those responding, wrote that the regression line provides information about relationships among the data. One said that it "shows the general tendency of a group of data." These responses also show characteristics of *Category 4: Developing Statistical Reasoning* because they indicate understanding without showing depth.

Next, students were asked to address the mathematical meaning of "least-squares." Fifty percent of the fall students did not attempt to answer. The other 50% gave answers that did not capture the mathematical meaning behind "least squares." Their responses were more conceptual and focused on regression in a general sense, not the least-squares criteria specifically. For example, one student offered, "Mathematically, it is a measuring utensil of how close the data is related or how much of an impact the x value has on the y value." These statements would fall in *Category 2: Naïve Statistical Reasoning*.

All of the spring semester students attempted to explain the mathematical meaning of "least squares." One student gave a *Category 1: Arithmetical Reasoning* response claiming that least squares means the line "intersects the least data possible while dividing it." She is using her numerical understanding of "least" to describe this idea. Three students, 43% of the group, gave an explanation of correlation. Two students, 29%, provided answers that captured the concept well with mostly-correct reasoning, evidence of thinking in *Category 4: Developing Statistical Reasoning*. Their answers included "the points it doesn't go through are the least amount away" and "the best line so that the distances squared are the minimal." Neither of these students explicitly mentioned that the distances measured are vertical.

Responses of the fall and spring students to the prompt "Why do researchers use regression analysis?" differed in two ways. Fifty percent of fall students and 29% of spring students gave appropriate responses that mentioned predicting future events or determining relationships among variables. These responses were at the *Category 4: Developing Statistical Reasoning* level. Another 29% of the spring students offered responses that showed some confusion with correlation. For example, one suggested that "this line is used to show the correlation or the noncorrelation between two separate or independent items." These students demonstrated reasoning in *Category 2: Naïve Statistical Reasoning.* Approximately 30% of each group did not offer an answer.

When students were asked to describe the relationship between regression and correlation, 50% of the fall control group students and 43% of the spring experimental group indicated that the correlation coefficient tells how close the data lies to the regression line. These statements fall in *Category 4: Developing Statistical Reasoning.* One student in each group, 17% of the fall students and 14% of the spring students, said that the correlation coefficient showed whether the regression line had a positive slope or a negative slope. These responses, less specific than those at *Category 4: Developing Statistical Reasoning*, would be classified in *Category 3: Procedural Statistical Reasoning.* Forty-three percent of spring students gave answers that did not appropriately describe this relationship. These students showed some confusion or lack of understanding. Their answers included statements such as "The regression line shows how much correlation there is between two sets of data" and "The correlation coefficient relates to the regression line because they are both necessary to do this computation." These are *Category 2: Naïve Statistical Reasoning* responses due to the incorrect or this category. He said that "r is the range of the least squares line and r^2 is the value of the least squares line."

Table 27 - Werner's Framework Applied to Essay #4

		Category 0 Non-determinable Reasoning	Category 1 Arithmetical Reasoning	Category 2 Naïve Statistical Reasoning	Category 3 Procedural Statistical Reasoning	Category 4 Developing Statistical Reasoning	Category 5 Functional Statistical Reasoning	Category 6 Expert Statistical Reasoning
Description of the least-squares regression line	Fall Spring	0% 0%	0% 0%	17% 0%	83% 14%	0% 86%		
Meaning of "least-squares"	Fall Spring	50% 43%	0% 14%	50% 0%	0% 0%	0% 29%		
Reason for regression analysis	Fall Spring	33% 29%	0% 0%	0% 29%	17% 14%	50% 29%		
Relationship of correlation coefficient to least-squares regression line	Fall Spring	0% 0%	0% 0%	17% 43%	33% 14%	50% 43%		

Essay 5: Confidence Intervals and Margin of Error

Students were provided with a clipping from USA TODAY which summarized results of two political polls. They were asked to explain margin of error, interpret the margin of error statements, and explain the connection between margin of error and confidence intervals. Twenty-eight fall semester students and 31 spring semester students answered this item. All of the students who responded had this essay among their three best essays.

First, students were asked to explain margin of error. Only one spring semester student neglected to attempt an explanation. Fifty-seven percent of fall control group students and 35% of spring treatment students gave answers that indicated an understanding of margin of error providing "give or take," "slide," or "padding" around an estimate. These responses fall in Werner's *Category 4: Developing Statistical Reasoning*. Few students specifically identified sampling error as the source of this uncertainty. Four percent of fall students and 13% of spring students included a reference to sampling error in their responses. These answers are indicative of *Category 5: Functional Statistical Reasoning* as students showed more depth in their responses.

Other students' explanations are classified as *Category 2: Naïve Statistical Reasoning* responses for a variety of reasons. Werner's *Category 2: Naïve Statistical Reasoning* is characterized by incorrect, inappropriate, or incoherent use of statistical terminology, reasoning with little or no meaning, the use of concrete examples from the text or from class, memorization, use of trial and error, or statistical dumps, "explanations which are densely impregnated with statistical terms used in a nonsensical manner" (Werner, 1995, p. 142). Fourteen percent of fall semester students and 19% of spring semester students gave explanations that included references to non-sampling error including instrumentation error, dishonest respondents, and those undecided or with no opinion. Another 11% of fall students gave responses in this category that included "the area between the difference between information" and "how much error is allowed." An additional 29% of spring students gave such explanations including "compare two things and try to find the value of it" and "margin that is calculated to base the statistical study on."

Students were asked to use an example from the news clipping to illustrate their understanding of margin of error. Student responses were grouped in four broad categories: those making no attempt, those providing inappropriate examples or explanations, those providing an example from the article without any explanation or connection to their previous statements about margin of error, and those providing appropriate examples with explanations or connections to previous remarks. The 39% of fall control group students and 16% of spring experimental students making no attempt to provide an example are classified as *Category 0: Non-Determinable Reasoning.*

Students giving inappropriate responses are reasoning in *Category 2: Naïve Statistical Reasoning* because of the incorrect, incoherent, or inappropriate use of statistical terminology. Eighteen percent of fall students and 32% of spring students made such statements. One fall semester student noted that 49% of respondents would have voted democratic, 45% would have voted republican and 6% were undecided. She claimed that the gap between 49% and 45% and the gap between 45% and 6% represented the margins of error. Another student from the control group reported a result from the poll and then said, "Well, if the whole population was asked the question, the percentage could waive by a percentage or two." The stated margin of error for this poll was plus or minus three percentage points. Examples from the spring experimental group included a student who referred to the same example previously given and said, "For all adults we are 97% sure that all adults have these ratios of views on the election."

Another spring student stated "The margin of error for the poll of likely voters is +/- 5%. That means that about 5% of the data is possibly wrong."

Eighteen percent of fall students and 29% of spring students provided appropriate examples from the clipping but did not attempt to explain the statements. These responses are included in *Category 3: Procedural Statistical Reasoning* because although they were able to identify an appropriate margin of error statement from the clipping, their lack of explanation provides no evidence of conceptual understanding.

Approximately 25% of each group gave appropriate examples from the clipping along with mostly appropriate explanations. These statements are classified in Werner's *Category 4: Developing Statistical Reasoning* since the reasoning is mostly correct and some understanding is demonstrated. One such response from a spring student was "For example, there might not really be 49% of voters that will vote for the democratic candidate. The margin of error is +/-3 points so the percentage is between 46-51%." Although this student communicates understanding, she doesn't include a sense of uncertainty in her explanation preventing her response from being classified as *Category 5: Functional Statistical Reasoning*.

Next, students were asked to make a connection between margin of error and confidence intervals. Twenty-nine percent of fall control group students and 16% of spring experimental students did not attempt to make this connection. Approximately 35% of students each semester offered incorrect responses characteristic of Werner's *Category 2: Naïve Statistical Reasoning*.

Included in those figures are students who showed evidence of the most common misconception among both sets of students - the idea that a margin of error and a confidence interval were complements. Twenty-five percent of fall students and 16% of spring students gave responses that indicated this misunderstanding. For example, a fall semester student offered, "Confidence interval of say 99% means that you are 99% sure and leaving 1% for a margin of error (room for being wrong)." Another fall student suggested that "Confidence intervals and margin of error are connected in that if you choose a 95% CI then your margin of error – chance [you're] willing to take of being wrong is 5%. The two have to equal to 100%." Spring semester students displayed similar thinking. One claimed "Margin of error is just the complement to a confidence interval. In the first poll the margin of error was +/- 3 percentage points so the confidence interval was 97%." Another spring student said "In confidence intervals the margin of error is 1 - the confidence interval."

Eighteen percent of fall students and 13% of spring students gave responses considered characteristic of *Category 4: Developing Statistical Reasoning.* These students made comments about confidence intervals that were not inaccurate, yet did not fully communicate understanding of the relationship between the two ideas. A fall semester student said "Margin of error is connected to confidence intervals both of them give you a range for mistakes." Similarly, a spring student suggested that "Margin of error is connected to confidence intervals by being able to say either it falls within the margin of error or it falls outside the margin of error."

The remaining students, 21% of the fall control group and 35% of the spring experimental group, gave explanations in *Category 5: Functional Statistical Reasoning.* An example from the fall group is "Confidence intervals are related to margin of error because both values relate to how accurate the data values are. If a large confidence interval (95-98%) and a small margin of error (1-3%) are both presented the reader can be confident that the information is accurate." A spring semester student said that margin of error was "sort of like constructing a confidence interval where you discover a zone around a point estimate, but in this case you're predicting a percentage, give or take (+,-) a few percentage points."

Finally, students were asked why the writer included two different margin of error statements. Twenty-five percent of the fall students and 16% of the spring students offered no attempt at an explanation. The majority of students each semester responded appropriately, citing the two different populations, all adults and likely voters, and the two different sample sizes. These students comprise 57% of the fall group and 65% of the spring group. The remaining students, approximately 20% of each group, provided inappropriate explanations characteristic of Werner's *Category 2: Naïve Statistical Reasoning*. An example of such a response from the fall semester is "The writer used two different margin of errors because he or she had two different pieces of data that could be wrong some of the time." A spring student wrote "Perhaps the writer included two statements about margin of error to illustrate how small the change was and therefore it would have little effect on the margin."

Table 28 - Werner's Framework Applied to Essay #5

		Category 0 Non-determinable Reasoning	Category 1 Arithmetical Reasoning	Category 2 Naïve Statistical Reasoning	Category 3 Procedural Statistical Reasoning	Category 4 Developing Statistical Reasoning	Category 5 Functional Statistical Reasoning	Category 6 Expert Statistical Reasoning
Description of margin of error	Fall	0%	0%	39%	0%	57%	4%	
	Spring	3%	0%	49%	0%	35%	13%	
Margin of error illustration	Fall	39%	0%	18%	18%	25%		
	Spring	16%	0%	32%	29%	23%		
Margin of error connection with confidence intervals	Fall	29%	0%	32%	0%	18%	21%	
	Spring	16%	0%	35%	0%	13%	35%	
Why were two margin of error statements included?	Fall	25%	0%	18%	0%	57%		
	Spring	16%	0%	19%	0%	65%		

Student Perceptions of the Uses of Statistics

Thirty-five fall semester control group students and thirty-seven spring semester experimental group students chose to complete the first final exam essay question which focused on the meaning of, and uses for, statistics. Responses to this essay item were analyzed to determine whether students in the two different versions of the course ended the semester with different perceptions of the usefulness of statistics in their everyday lives and their chosen field of study.

Students' examples of the uses of statistics in everyday life were classified into one of four different categories: did not attempt to address the question, gave a weak response, gave an example including the use of descriptive statistics, or gave an example including the use of inferential statistics. (Note that students who included an example of inferential statistics in their response and are counted in this category might also have included uses considered descriptive in nature.) Weak responses were characterized by very general language and lack of specific examples. For example, "I see statistics in my life now in the newspaper, on the television, and at work" would be considered a weak response.

The percentages of student responses in each category are provided in Table 26. Approximately 80% of students each semester provided reasonable uses of statistics in everyday life. About 60% of fall control group students and 40% of spring experimental groups students provided examples which were descriptive in nature, including charts, graphs, sports statistics, and course grades. One fall semester student offered, "I see statistics being used everyday in the business world. When it comes to stocks being purchased to hearing of the stats on the news such as murder." Similarly, a response from a spring student in this category was "Statistics is used everyday in the news and media showing polls or war situations and economic statis [sic] reports."

Another 20% of fall control group students and 38% of spring experimental group gave responses that focused on inferential ideas such as decision making and prediction. A fall semester student gave this response:

> With statistics you find out a lot about better deals and scores and what the people want. You can use it in companies to find out how productive you are, what changes should be made and using past data, how should you plan for the future. Statistics is used in everything around. [A nearby town's] water shortage is one example. They use data from average water use by [the nearby town], how much of a supply we have left and how much rain is expected to determine how long [the nearby town] will have water.

A spring semester international student, whose native language was not English, said "Statistics include random sampling and other sampling, all different ways. Wherever you go you hear about statistics in magazines or conducting a poll. They statistical [sic] measure / are mostly conducted with info from the census population, different ways to conduct surveys. These are done to study the change and possibilities of one thing relating to the other."

Table 29 – Distribution of Student Perceptions of the Use of Statistics in Everyday Life

	Control (Fall) n = 35	Experimental (Spring) n = 37
Did not attempt to answer	6%	8%
Weak	14%	14%
Descriptive	60%	40%
Inferential	20%	38%

Students' examples of the uses of statistics in their chosen career field were classified in one of five different categories. These categories included the four in the previous analysis, but added one to capture the few students who acknowledged the question but could not give an example of the use of statistics in their field. These responses were considered different than those who did not respond to the question at all.

The percentages of student responses in each category are provided in Table 27. One apparent difference in the distribution of responses is noted in the categories "weak response" and "descriptive statistics". Approximately 10% fewer students in the spring experimental classes gave weak responses, and approximately 10% more students in the same group gave responses that focused on descriptive statistics. The one fall semester student who could not apply statistics to her field of study was a nursing major; the two such spring semester students were English majors.

Table 30 – Distribution of Student Perceptions of the Use of Statistics in Their Careers

	Control (Fall) n = 35	Experimental (Spring) n = 37
Did not attempt to answer	6%	8%
Didn't know	3%	5%
Weak	20%	11%
Descriptive	40%	49%
Inferential	31%	27%

Attitude Assessments

Two attitude assessments were administered at the end of each semester. Students were asked to complete the instruments anonymously, and completed surveys were coded as control (fall) or treatment (spring). The researcher could not identify whether students were part of the Monday-Wednesday-Friday class or the Tuesday-

Thursday class because some students who could not attend their regular section chose to come to the other class that week.

Survey of Attitudes Towards Statistics (SATS)

The 28-item SATS instrument measures four subscales: affect, cognitive competence, value, and difficulty. Six of the 28 items contribute to the affect scale, six others contribute to the cognitive competence scale, nine items contribute to the value scale, and the remaining seven items contribute to the difficulty scale. Seventy-three students completed the SATS, thirty-six in the fall and thirty-seven in the spring.

The response scale for each item is a 7-point Likert scale with 1 corresponding to *strongly disagree*, 4 corresponding to *neither agree nor disagree*, and 7 corresponding to *strongly agree*. In determining the scale scores, responses for negatively worded items are reversed so that high scores represent more positive attitudes for each scale.

Two students each omitted one of the 28 items. One fall semester student omitted one of the six items contributing to the cognitive competence scale and one spring semester student omitted one of the nine items contributing to the value scale. Because these scales are computed as the sum of the responses for items in each category, the mean of the remaining items for that scale was used for the missing values.

F-tests for equality of variances were non-significant ($p > .05$) for each of the four variables, affect, cognitive competence, value, and difficulty. Independent, pooled t-tests were conducted to detect any significant differences in the responses of fall and spring students. No significant differences were found. Results are summarized in Table 31.

Table 31 – Mean Scale Scores for Survey of Attitudes Towards Statistics (SATS)

	Control (Fall) n = 36	Experimental (Spring) n = 37	t-value (df = 71)	p-value
Affect	25.306	25.703	-0.25	0.8045
	(6.948)	(6.708)		
Cognitive Competence	28.061	28.662	-0.42	0.6764
	(5.783)	(6.440)		
Value	44.403	45.311	-0.40	0.6905
	(9.216)	(10.150)		
Difficulty	22.056	23.730	-1.01	0.3136
	(7.243)	(6.850)		

Note: Standard deviations are given in parentheses.

STARC-CHANCE Abbreviated Scale (SCAS)

The STARC-CHANCE Abbreviated Scale is composed of 10 items to be interpreted independently, not as components of two or more scales. The response scale for each item is a 5-point Likert scale with 1 corresponding to *strongly disagree*, 3 corresponding to *neither agree nor disagree*, and 5 corresponding to *strongly agree*. Thirty-six fall semester control group students and 35 spring semester experimental students completed the assessment. Although 37 spring students completed the SATS instrument described in the previous section, one of those students did not complete all items on the SCAS and one student did not answer any of the SCAS items. These two students were omitted from the analysis.

Simultaneous t-tests were conducted employing the Bonferroni procedure to adjust for the increased probability of making a Type I error. For the 10 simultaneous t-tests the threshold for statistical significance becomes $\alpha/10$. Using this adjustment, no significant differences were found among students' responses to the 10 SCAS items.

Results are summarized in Table 32. Each Bonferroni adjusted p-value is the raw p-value multiplied by 10, the number of tests. Values exceeding 1 are reported as 1.

Table 32 – Mean Scale Scores for STARC-CHANCE Abbreviated Scale (SCAS)

	Control (Fall) n = 36	Experimental (Spring) n = 35	raw p-value	Bonferroni
I often use statistical information in forming my opinions or making my decisions.	3.00 (0.956)	3.086 (1.011)	0.7146	1.000
To be an intelligent consumer, it is necessary to know something about statistics.	3.917 (0.967)	3.800 (1.079)	0.6328	1.000
Because it is easy to lie with statistics, I don't trust them at all.	2.833 (0.971)	2.486 (1.068)	0.1555	1.000
Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract.	3.528 (1.028)	3.457 (1.039)	0.7742	1.000
Given the chance, I would like to learn more about probability and statistics.	3.167 (1.231)	3.857 (0.879)	0.0084	0.0840
You must be good at mathematics to understand basic statistical concepts.	3.236 (1.045)	3.171 (1.175)	0.8070	1.000
When buying a car, asking a few friends about problems they have had with their cars is preferable to consulting an owner satisfaction survey in a consumer magazine.	2.819 (1.070)	2.886 (1.255)	0.8113	1.000
Statements about probability (such as what the odds are of winning a lottery) seem very clear to me.	3.556 (0.909)	3.686 (1.022)	0.5723	1.000
I can understand almost all of the statistical terms that I encounter in newspapers or on television.	3.500 (0.971)	3.286 (0.987)	0.3597	1.000
I could easily explain how an opinion poll works.	3.556 (0.969)	3.514 (0.981)	0.8590	1.000

Note: Standard deviations are given in parentheses.

CHAPTER 5
DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

This study investigated the influence of laboratory activities on community college students' understanding of elementary statistics. Both quantitative and qualitative methodologies were used to examine students' knowledge and communication of statistical ideas. Additionally, students' attitudes and their ability to see applications of statistics in everyday life and in their future careers were studied. This chapter provides discussion and conclusions in the form of answers to the four research questions posed. Data from ten student interviews is used throughout to triangulate results from the final exam, the Statistical Reasoning Assessment, and the two attitude inventories. Additionally, limitations of the study and recommendations for future research are provided.

Student Understanding

Question 1: Will students in the laboratory sections show different levels of understanding than students in the lecture/discussion sections?

The spring semester treatment classes showed significantly greater understanding on the first of the three course tests. There were two laboratory activities completed during this time period, one involving sampling variability and one involving the law of large numbers. Because few of the items on the first test were directly related to these activities it is possible, but unlikely, that this difference in achievement could be attributed to the laboratories. Rather, it is reasonable that students enrolled in the spring semester had already acclimated to the academic environment as a result of being enrolled during the fall.

The spring group also showed significantly better understanding of one concept measured by the Statistical Reasoning Assessment – the importance of large samples. This result is discussed in greater detail in the section that follows. No statistically significant differences in student understanding were detected in the final examination problems. There was a significant difference in understanding between students in the Monday-Wednesday-Friday sections and the Tuesday-Thursday sections with regard to correlation not implying causation. This might be a result of the ice cream taste activity being completed in one full 75 minute period with the Tuesday-Thursday class rather than parts of two consecutive 50 minute periods with the Monday-Wednesday-Friday class. It is also possible that due to the large number of tests conducted in this analysis that this result could be considered a Type I error.

The Importance of Large Samples

As measured by the Statistical Reasoning Assessment, students in the treatment group showed significantly greater understanding of the importance of large samples than students in the control group. Many of the activities these students participated in during the semester illustrated sampling variability and the tendency of larger samples to more accurately reflect the population. This continual reinforcement in the context of learning various statistical concepts may explain their greater understanding of this idea.

The following item was used on the SRA to test for the misconception "law of small numbers". The voluntary explanation of one spring semester student illustrates the greater understanding this group had with regard to the importance of large samples.

Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

______ *a. Hospital A (with 50 births a day)*

______ *b. Hospital B (with 10 births a day)*

______ *c. The two hospitals are equally likely to record such an event.*

The appropriate response is b. The Statistical Reasoning Assessment was not administered via computer score sheet; students were instructed to mark their answers directly on the question sheets. One spring semester treatment student added the following comment underneath his correct response to item 14, "The law of large numbers would make Hospital A less likely – it is possible, but more possible for Hospital B". This item had the appropriate answer keyed to the correct reasoning skill of "understands sampling variability" and one incorrect answer keyed to the misconception "law of small numbers". This student was voluntarily justifying his correct response by making a connection between sampling variability and the larger number of births at Hospital A. No other students offered explanations of their SRA responses.

During an interview, a spring semester student offered this thought about sample size and the laboratory activities: "When I was taking a test I would remember – one thing that came to mind was the spinner, when we had the spinner and four sections, and the Law of Large Numbers. During the test I could go back and it made sense to me because all that hands-on sort of gets it into your brain. Reading in the book I probably never would have gotten that concept".

Other concepts included in the Statistical Reasoning Assessment were related to the treatment activities, but the activities involved were more isolated in nature. For example, although one activity early in the semester was specifically designed to teach independence concepts, at the end of the semester students in the treatment group showed the same level of understanding in this area as their counterparts in the control group. This suggests that isolated activities are not sufficient for students to develop and retain understanding. Instructors should consider sequences of activities that revisit key concepts over a long period of time to facilitate understanding.

Student Writing

Question 2: Will students in the laboratory sections write more accurate, detailed, and complete explanations on open-ended essay questions than students in the lecture/discussion sections?

<u>Hypothesis Testing</u>

One of the essay choices on the final exam asked students to explain the concept of hypothesis testing to a friend who was registering for the course and saw it listed as one of the topics. Spring semester students seemed to have less focus in their explanations of hypothesis testing, 20% providing a response in Werner's *Category 2: Naïve Statistical Reasoning*, due to the inappropriate, jumbled use of statistical terminology. This is compared with 8% of fall semester students providing this type of explanation.

All of the *Category 5: Functional Statistical Reasoning* statements made in response to this item were made by students in the fall semester. These statements focused on concepts connected to α, the size of the rejection region(s) when comparing one-tailed and two-tailed tests, and the likelihood of rejecting the null hypothesis.

Another interesting difference in the responses of the two groups of students involves the percentage of students who claimed that null hypothesis was the thing that you "believed was true." Twice the proportion of students in the spring experimental group, 18%, made this statement, while only 9% of the fall students did. It seems these students are confusing the null hypothesis with the research hypothesis, usually the claim that the researcher wants to make.

Perhaps spring semester students had this misunderstanding in greater numbers than the fall students because of the penny spinning activity that was used to introduce concepts of hypothesis testing. Student's intuition prompted them to say that the probability of heads resulting when a penny was spun rather than flipped was . 5. This statement, $p = .5$, was used as the null hypothesis. When the students collected their data it was clear that this null hypothesis should be rejected in favor of the alternative, $p \neq .5$. The design of this activity might have caused many of them to confuse the assumption of the null hypothesis with the research goal of the alternative hypothesis.

When students were interviewed they were asked to describe hypothesis tests and were prompted by the researcher to expand their explanations if necessary. The following interview dialog with a student from the spring semester illustrates the perception discussed above.

Researcher: What are hypothesis tests?

Student S1: They're tests of what you think is going to happen – one is the null hypothesis. Oh man (pause) um (pause) I don't remember that one. But the null hypothesis is what you think is gonna happen and the alternative hypothesis is the opposite of the null hypothesis.

Researcher: And so, what do you do with those?

Student S1: You test, um, to see if you're right or if you're wrong and you either reject the null hypothesis or fail to reject the null hypothesis.

Researcher: And what does that mean?

Student S1: If you reject the null hypothesis that means the opposite of what you thought was gonna happen happened and if you accept the null hypothesis that means that what you thought was gonna happen happened.

As noted earlier, 9% of fall students showed this misconception in their final exam essay along with 18% of spring students. One of the fall interviewees shared his understanding of hypothesis tests as follows:

Researcher: What is a hypothesis test?

Student F1: You have a null and a (pause, and trails off) The null hypothesis is the one you're trying to prove correct and the uh, um,

Researcher: Alternative?

Student F1: The alternative hypothesis is the one you're going to accept if the null is false or if there's not enough evidence to prove that it's correct.

Researcher: So, why don't we say "false"?

Student F1: It could still be true, but you just have a lack of evidence to prove it true – you can't say it is false because of the amount of information you're given.

Compare his remarks with another fall student who offered a unique example to illustrate her points.

Researcher: What is hypothesis testing?

Student F3: Um (pause) I thought it was something about accuracy. Being able to determine how accurate something was – like um (pause) No, wait. I think there was something like we did (pause) OK let's say about sales or something. Like if a logo improved sales. Yup. So I would tell, um, it's kind of like a way to determine if A affected B. Is that clear?

Researcher: If you want to tell me more, that would be helpful.

Student F3: OK – if you wanted to figure out if accidents were caused by a curve or something, if you would, like, compare how many accidents happened without the curve on just a straight highway and then compare to how many happened with the curve to see if the curve made any difference.

Researcher: How would you see if that data showed a difference?

Student F3: Um – seemed like it was something like if it failed (pause) Oh, I know. If the number fell within a certain percentage of something then it would be said that yes, this did affect the number of accidents that happened.

Researcher: Can you tell me about the null and alternative hypotheses?

Student F3: OK, the null hypothesis is what you start with, I think, and the alternative is what you're trying to (pause) what you're trying to determine. If it may have been good.

Linear Regression and Correlation

Based on the written essay responses, the spring semester treatment students had a better understanding of the "least squares" concept than their fall counterparts. None of the fall students explained this sufficiently, and 50% did not even attempt an explanation. In contrast, 29% of the spring respondents gave answers that demonstrated understanding. Yet, they neglected to include the idea that the squared distances to be minimized were vertical. All of the spring semester students who chose this essay attempted to explain the meaning of "least squares."

While 30% of students each semester chose not to attempt an explanation of the use of regression in research, 50% of fall students and 29% of spring students gave appropriate responses which included the discussion of prediction or determining relationships among variables. However, 29% of spring students expressed some confusion between correlation concepts and regression concepts.

The ideas of correlation and regression were among the last topics discussed both semesters. When students were asked to identify the most challenging topic during the interview, 6 of the 10 students (3 each semester) said that the topics covered at the end of the semester were most challenging. One fall student said, " . . . because at the end I wasn't paying much attention" and the other said "I didn't apply myself much at the end. It was mostly me not applying myself." MINITAB printouts and the TI-83 were used extensively for this material and that student also said he wasn't "much of a computer person." The third fall student did not offer an explanation for her assertion that these were the most challenging concepts.

One of the three spring semester students who said the most challenging topics were those at the end of the semester had a similar response, saying " . . . probably the last, just because I never felt like I really got it, I was just kind of floating through those last weeks." The other two students described the process as lengthy and complicated. One said, "The two I don't remember (pause) the two very last things we did, right before the final exam, um, the y calculations with the ice cream. Yeah, and the graphs and the plots we had to do with that. Those were hard because you had (pause) You couldn't just take one equation and do it. You had to solve this one and then this one and then you go here – that was really difficult."

The other student shared similar thoughts. "Probably, the last three weeks. The (pause) I think that was the correlation stuff and the (pause) I know I kept getting lost when ever we would do the (pause) like the graphs on the TI-83 – You would have the little dots and then you would have another little set of dots and the linear correlation. I kept getting that confused. That was hard for me." When she was prompted by the interviewer to explain her understanding of correlation she replied

> For me, correlation is kinda how things relate, but at the same time, explaining it specifically, what it was and the numbers and stuff (pause) But, see, that's what I said, I wasn't doing my part. I wasn't studying and doing my homework like I should. The first half of the semester I was doing stuff and I would take the time and sit down and do it. I was going to the library and doing the homework and whenever I did that and I came to class, then I knew what was going on.

Many of these students, both fall control group and spring treatment group, admitted that they were not applying themselves as they should have during the last weeks of the semester. This tendency to get lazier at the end of the semester would have a greater detrimental impact on understanding in a class with laboratory activities designed for students to construct understanding than in traditional settings that are more teacher-centered.

Students were asked in the interview to describe how they would determine if there was a relationship between the number of times a student was absent from class and his or her grade in the course. The researcher anticipated that students would suggest plotting the data and computing the correlation coefficient. Three of the five fall semester students suggested correlation, one suggested a "time trial" to compare several classes over time, and one didn't know. Of the five spring semester students, none suggested correlation. Three suggested a hypothesis test to compare the grades of those who attended regularly and those who didn't, one suggested a bar graph, and one didn't know.

Each of the fall semester students who suggested correlation as an appropriate technique to determine whether there was a relationship between these variables was asked if causation could be determined from correlation. One of the three said yes, one was noncommittal saying "I'd probably recommend it to them" and the other indicated that correlation does not indicate causation. Student F2 believed that correlation indicated causation.

Researcher: How can you determine whether there's a relationship between the number of times a student comes to class and his or her grade in the course?

Student F2: Like a correlation.

Researcher: What kind of information would that correlation give you?

Student F2: Graph attendance and certain grades. The more straight the line was the more correlated. If they're all scattered over there might not be much correlation. Which way the line's going shows if it's positive or negative.

Researcher: So, what does that mean? If it was positive, what would that tell you?

Student F2: Positive means that it affects it, I think. The higher the coefficient is, the more closely they're related.

Researcher: Can you say that increased attendance would cause increased grades?

Student F2: Yeah, you could say that. Yeah.

Students both semesters showed through their essay responses and their interviews that the ideas of correlation and regression were among the most confusing they encountered. Spring semester treatment students had a greater understanding of the relationship between the observed data and the regression line and provided responses that were more graphically oriented than the fall control students who relied more heavily on algebraic interpretations.

Spring semester treatment students seemed to have more difficulty separating the ideas of correlation and regression, perhaps because the lab that was used to teach regression ideas incorporated the concepts of correlation coefficient and coefficient of determination. Students did not fully understand that correlation is a measure of association but does not imply causation, while regression depends on the identification of one variable as explanatory and the other as response.

Confidence Intervals and Margin of Error

A greater percentage of spring semester students than fall semester students gave higher level responses to prompts for this item according to Werner's (1995) framework. This is particularly evident when *Category 5: Functional Statistical Reasoning* responses are examined.

Student essay responses on the final exam showed that while only 4% of fall semester students identified sampling error as the source of uncertainty in margin of error, 13% of the spring semester students made this statement. These students were classified in *Category 5: Functional Statistical Reasoning* for not only correctly describing margin of error, but for providing supplemental information showing greater understanding.

Additionally, when asked to explain the relationship between margin of error and confidence intervals, 21% of the fall control group and 35% of the spring experimental group provided responses in *Category 5: Functional Statistical Reasoning*, showing connections between these two ideas. This connection was communicated by a spring semester student in the following interview excerpt.

Researcher: What is a confidence interval?

Student S3: How sure you are about the data you got. Gives you leeway to be wrong.

Researcher: Can you give me an example?

Student S3: It gives you a margin of error on either side.

Another spring semester student showed confusion and misunderstanding regarding confidence intervals during her interview. In an effort to see how effective the lab designed for this concept was in teaching the ideas, the researcher prompted the student with questions about the laboratory activity.

Researcher: What is a confidence interval?

Student S4: I remember doing 'em. I remember we did one – I'm not sure I did it right, but you know (pause) we had a set price – that was the thing, we had this price and we had to find the confidence interval (pause) I don't know.

Researcher: We did a lab involving m&ms®, does that help you remember? Were you there?

Student S4: Yes, I was. We were trying to see how many blue m&ms® were in the bag. And what we did was take the number of m&ms®, put how many blue ones we had in our little group set over the number (trails off)

Researcher: And what did that give us?

Student S4: Confidence?

Researcher: The sample proportion.

Student S4: The sample proportion. And then you take that and you plug it into a formula, right?

Researcher: And then you have two numbers, the endpoints of the confidence interval. What does that tell you?

Student S4: The number of m&ms® in the bag is somewhere between that number and that number.

Researcher: And what does the word "confidence" have to do with the whole thing?

Student S4: We're pretty confident – we're pretty sure that it falls in between there and there.

This student does not understand that a confidence interval gives an estimate of the population parameter based on a sample statistic, in this case the sample proportion. She does not realize that she can count the exact number of blue m&ms® in that particular bag and constructing a confidence interval in an attempt to determine this is inappropriate. She appears to see the bag of m&ms® as the population being considered (rather than all m&ms® produced by the company) and her group's little sample as a way to predict the number of m&ms® in that bag.

Applications of Statistics

Question 3: Will students in the laboratory sections more readily see applications of statistics when compared with students in the lecture/discussion sections?

Students were asked to discuss uses of statistics they encounter in their daily lives and uses of statistics they anticipate in their future careers. Approximately equal proportions of students each semester, 80% of fall students and 78% of spring students, gave acceptable responses for the use of statistics in everyday life. The fall control group responses were 60% descriptive in nature and 20% inferential in nature, while the spring semester treatment group had approximately 40% descriptive examples and 38% inferential examples.

When asked to provide examples of the uses of statistics in their future careers, 71% of fall students and 76% of spring students provided acceptable responses. In both groups, more students generated descriptive uses than inferential uses. During the fall 40% of students gave examples that were descriptive in nature while 31% gave inferential uses. Spring semester results were 49% descriptive and 27% inferential.

Students enrolled in the spring semester laboratory version of the course were able to see more sophisticated inferential uses of statistics in everyday life than students in the fall semester group. It is not evident that they were able to see those same kinds of applications as readily in their own chosen careers.

During the interviews students were asked, "What kinds of questions can statistics help you answer?" in an effort to have them generate applications of statistics in daily life and in their future work. Students' answers from both semesters were generally short and vague. Of the five fall semester students who were interviewed, two responded with descriptive examples, including making a graph to organize data. The other three responded with inferential examples focused on decision making and comparing.

Three of the spring semester students gave descriptive responses such as knowing how much money you spent and how you spent it, consumer preferences for products, and the organization of information. Two students gave inferential responses. One suggested predicting course grades based on test performance. The other discussed the interpretation of charts and research in her field, drug and alcohol counseling. She also included some descriptive uses such as collecting and reporting demographic information about clients. This student was the only one of the ten interviewed who gave specific examples of how statistics could be used in her future career.

There does not seem to be much difference in the ways the two groups of students view the use of statistics in their daily lives or their future careers. Students are more likely to give examples that focus on the reporting of descriptive statistics, rather than the use of inferential techniques. Additionally, both groups of students seem more comfortable discussing the applications of statistics to their daily lives than imagining what kinds of on-the-job applications they might encounter.

Attitudes and Beliefs

Question 4: Will students in the laboratory sections of the course develop different attitudes and beliefs about statistics than students in the lecture/discussion sections of the course?

Two attitude assessments were administered at the end of each semester, the Survey of Attitudes Towards Statistics (SATS) and the STARC-CHANCE Abbreviated Scale (SCAS). The SATS measured four scales: affect, cognitive competence, value, and difficulty. There were no statistically significant differences between the two groups on these four scales or the ten items assessed by the SCAS.

Conclusions

Quantitative student achievement data in the form of three course examinations, four final exam problems, and seven Statistical Reasoning Assessment scales showed only two areas of statistically different results when the fall control group and the spring experimental group were compared. Given this number of statistical tests, it would not be surprising to find one significant result even if no difference was present.

Spring semester students did significantly better ($p < .0001$) than fall semester students on the first test given during the course. Only two lab activities had been completed during this time period, one on sampling variability and one on the Law of Large Numbers. It is also possible that the fall students were not yet acclimated to school and were less prepared for the exam than their spring counterparts. Evidence to support this conjecture is found in the researcher's journal where five fall semester students came to the first exam without the suggested note card of formulas.

The other statistically significant result involved the Statistical Reasoning Assessment scale *Understands the Importance of Large Samples*. Many of the activities used throughout the spring semester involved individual students or groups of students collecting data and then either pooling it with the rest of the class, or comparing to other groups. This continual reinforcement that different samples from the same population provide different sample statistics and that larger samples more accurately reflect the population would explain the spring students' increased understanding of this idea.

It was surprising that none of the other cognitive measures indicated a statistically different level of understanding between the treatment and control groups. Some activities were specifically designed to teach concepts such as independence, confidence interval interpretation, and hypothesis testing. Yet, students in the treatment group did not show increased understanding of these ideas.

Two attitude inventories, one containing four scales and the other containing ten stand-alone items, were administered. No statistically significant differences were determined among the four scales on the Survey of Attitudes Towards Statistics (affect, cognitive competence, value, and difficulty) or the ten included on the STARC-Chance Abbreviated Scale.

Qualitative data in the form of student essay responses and interviews support the findings of the quantitative data. In many cases, students from the spring treatment group showed a greater tendency to have "statistical dumps" in their writing, including many ideas and terms stung together in a haphazard way. They had difficulty organizing the concepts taught through constructive activities and often showed confusion when explaining related ideas such as confidence intervals and margin of error, or correlation and regression. Spring semester students also showed less understanding and retention in their interviews, responding with "I don't know", "I'm not sure", and "I don't remember" more often than students interviewed after the fall semester.

According to Hawkins (1997), among the priorities for research in statistics education should be studies that "tell us about things that did not work, and therefore things we should avoid" (p. 145). Injecting ten hands-on constructive activities in a course that was otherwise traditional in scope and content was insufficient to develop the desired improvement in student understanding. The data suggest that repeating key ideas throughout many activities, for example the importance of large samples, provides students with ample opportunity to assimilate those concepts. However, disjoint laboratories that have no connections or common threads may actually impede students' understanding because they have difficulty organizing the content and relating it to previous experiences.

Additionally, this research supports the conjecture by Rumsey (1998) that hands-on activities could interfere with students' conceptual development. This point is illustrated in the confidence interval discussion between the researcher and student S4 included earlier. The results of this study are also consistent with those of Konold (1995) who describes the following three research findings with respect to student understanding of probability and statistics:

> (1) students come into our courses with some strongly-held yet basically incorrect intuitions;
>
> (2) these intuitions prove extremely difficult to alter;
>
> (3) altering them is complicated by the fact that a student can hold multiple and often contradictory beliefs about a particular situation. (par. 2)

Limitations of the Study

As the course instructor, the researcher was not in a position to investigate her own role in the classroom and the impact that her words and deeds had on students' understanding. While a thorough set of lesson plans and reflective notes was constructed by the teacher/researcher during the year, it is not possible to determine what, if any, unseen influence or bias might have impacted the study.

By design, the activities were largely disjoint, with little, if any, connection from one to another. While this was intentional because of the potential problems introduced by frequent absenteeism at the community college, and the difficulty of "making up" these activities, a set of labs that build on each other and reinforce ideas previously studied would prove more beneficial to students. This is evidenced by the spring semester students' significantly better understanding of the importance of large samples, an idea they visited over and over again throughout the set of activities.

Students in the laboratory sections of the course had greater confusion regarding names of concepts and the differentiation between related topics, for example correlation and regression. Students in the traditional version of the class had concepts organized during lectures and had major points emphasized by the instructor. Students in the experimental class had to make many of these distinctions themselves and were responsible for drawing conclusions through the activities. At times, these conclusions were inaccurate and students did not always clarify these misconceptions.

During the interviews, students in both versions of the course expressed their tendency to "coast" through the last few weeks of the semester. Because much more of the responsibility for learning was placed on the students during the spring treatment semester, the impact of this phenomenon may have been exaggerated. In fact, during the interview one rather conscientious student admitted "Yeah, and then at the end, when we had to go ask the people questions, I didn't do that because I don't like to go up to people."

Although the instructor/researcher had no input into the text selection for this course, the text used, *Understandable Statistics*, 5th Edition (Brase and Brase, 1995) was a traditional text with traditional examples and exercises. Including ten hands-on activities in an otherwise traditional course did not provide students with enough practice making conjectures, working together, asking "what if" questions, expressing these ideas clearly in writing, etc. It is natural to inquire whether a reformed curriculum used throughout the entire semester, one such as *Workshop Statistics* (Rossman, 1996) or *Interactive Statistics* (Aliaga & Gunderson, 1998), would facilitate deeper understanding and allow students to develop the skills they need to be effective learners in a student centered classroom. Alternatively, activities that might be teacher-led guided discovery rather than completely constructive might provide a balance between the goal of having students uncover statistical ideas and relationships on their own and their need for a post-organizer to help them differentiate between related concepts, and establish structure in the content.

Technology was not required for either for the fall control group or the spring treatment group. Students were strongly encouraged to purchase and use a TI-83 calculator, and MINITAB was available in the open lab for out-of-class assignments. Although the instructor used a TI-83 daily, the regular use of technology by all students would have enhanced their ability to focus on concepts rather than computations.

Interviews were conducted eight weeks after the end of each semester. The researcher was the only instructor teaching this course at this institution this year and wanted to avoid the possibility of interviewing students who might subsequently be enrolled in this course again, or enrolled in another course taught by this instructor. Students were not selected to be interviewed until after the add/drop period in the subsequent semester. Scheduling and conducting the interviews at that point meant that the last students were interviewed about eight weeks after the course final examination. If interviews were conducted closer to the end of the semester, the quality of the interview data might have been richer and might have discriminated better between the understanding of the two groups of students.

The sample size for both the fall control group and the spring treatment group is relatively small. Two sections of students were included each semester, but a total of only 39 students completed the course in the fall semester classes and a total of 42 students completed the course in the two spring semester classes. This high rate of attrition is common in community college mathematics classes at this institution, throughout the state of North Carolina, and throughout the country.

Recommendations for Further Study

The results of this study indicate that further research is needed to investigate the role of hands-on laboratory activities in developing students' understanding of elementary statistics. The following recommendations are made:

1. Similar research should be conducted using a completely reformed curriculum where students are immersed in data collection and analysis on a more regular basis.

2. Similar research should be conducted by an investigator who is not the course instructor.

3. Many sections taught by more than one professor should be included, and if possible, more than one site institution should participate. In addition to providing an increased sample size, and a more varied source of data, the role of the instructor can also be studied.

4. Student interviews should be conducted in a sequence throughout the semester, perhaps once a month. The interviews might include specific questions about the recent laboratory activities to ascertain student understanding and possible misconceptions.

5. Pairs of students or groups of students working together should be audiotaped or videotaped during the laboratory activities so researchers could record and analyze their thought processes, the kinds of questions they encounter, and the avenue they choose to resolve those questions.

6. Populations of students other than community college students should be studied to determine the influence of hands-on laboratory activities on a broader group of undergraduate students' conceptual understanding.

7. Later activities should explicitly reinforce previously studied concepts where appropriate.

REFERENCES

Aliaga, M., & Gunderson, B. (1998). *Interactive statistics* (Preliminary Edition). Upper Saddlebrook, NJ: Prentice Hall.

American Mathematical Association of Two-Year Colleges. (1995). *Crossroads in mathematics: Standards for introductory college mathematics before calculus.* Memphis, TN: Author.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

Becker, B. J. (1996). A look at the literature (and other resources) on teaching statistics. *Journal of Educational and Behavioral Statistics, 21*(1), 71-90.

Beins, B. C. (1993). Writing assignments in statistics classes encourage students to learn interpretation. *Teaching of Psychology, 20*(3), 161-164.

Bonsangue, M. V. (1994). Symbiotic effects of collaboration, verbalization, and cognition in elementary statistics. In J. J. Kaput & E. Dubinsky (Eds.), *Research issues in undergraduate mathematics learning: Preliminary analyses and results* (pp. 109-117). Washington, DC: Mathematical Association of America.

Borg, W. R., & Gall, M. D. (1989). *Educational research.* (5th ed.). White Plains, NY: Longman.

Brase, C. H., & Brase C. P. (1995). *Understandable statistics,* (5th ed.). Lexington, MA: D. C. Heath and Company.

Carson, V. M. (1995). Data analysis to explore the sampling distribution of the sample mean. (Doctoral dissertation, Georgia State University, 1995). *Dissertation Abstracts International, 57-02A,* 598.

Chance, B. L. (1996). Experiences with authentic assessment techniques in an undergraduate introductory statistics course. *American Statistical Association 1996 Proceedings of the Section on Statistical Education* (pp. 36-44). Alexandria, VA: American Statistical Association.

Cobb, G. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action* (pp. 3-43). Washington, DC: Mathematical Association of America.

Cohen, A. M., & Ignash, J. M. (1994). An overview of the total credit curriculum. In A. M. Cohen (Ed.), *Relating curriculum and transfer* (pp. 13-30). San Francisco: Jossey-Bass Publishers.

Cresap, D. (1995). Using student-generated data and a graphing calculator to illustrate the central limit theorem. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 355-357). Alexandria, VA: American Statistical Association.

Davis, R. B., Maher, C. A., & Noddings, N. (1990). Constructivist views on the teaching and learning of mathematics. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 1-3). Reston, VA: National Council of Teachers of Mathematics.

Davis, R. M. (Ed.). (1989). *A curriculum in flux: Mathematics at two-year colleges.* Washington, DC: Mathematical Association of America.

Dietz, E. J. (1993). A cooperative learning activity on methods of selecting a sample. *The American Statistician, 47*(2), 104-108.

Faghihi, F., & Rakow, E. A. (1995, November). *The relationship of instructional methods with student responses to the survey of attitudes toward statistics.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Biloxi, MS.

Gadanidis, G. (1994). Deconstructing constructivism. *Mathematics Teacher, 87*(2), 91-97.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. [54 paragraphs]. *Journal of Statistics Education* [On-line serial], *2*(2). http://www.amstat.org/publications/jse/v2n2/gal.html

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal and J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). Amsterdam, Netherlands: IOS Press.

Garfield, J. (1993). Teaching statistics using small-group cooperative learning. [40 paragraphs]. *Journal of Statistics Education* [On-line serial], *1*(1) . http://www.amstat.org/publications/jse/v1n1/Garfield.html

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*(1), 25-34.

Garfield, J. (1997). Discussion. In D. S. Moore, New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-165.

Garfield, J. (1998). Challenges in assessing statistical reasoning. In *Innovative assessment strategies for improving the teaching and learning of statistics.* Symposium conducted at the meeting of the American Educational Research Association, San Diego, CA.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Giraud, G. (1997). Cooperative learning and statistics education [39 paragraphs]. *Journal of Statistics Education* [On-line serial], *5*(3). http://www.amstat.org/publications/jse/v5n3/giraud.html

Goldin, G. A. (1990). Epistemology, constructivism, and discovery learning in mathematics. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (pp. 31-47). Reston, VA: National Council of Teachers of Mathematics.

Goldman, R. N. (1996). Trends in introductory applied statistics courses: Techniques. *American Statistical Association 1996 Proceedings of the Section on Statistical Education* (pp. 223-228). Alexandria, VA: American Statistical Association.

Grabowski, B. L., & Harkness, W. L. (1996). Enhancing statistics education with expert systems: More than an advisory system. [39 paragraphs]. *Journal of Statistics Education* [On-line serial], *4*(3). http://www.amstat.org/publications/jse/v4n3/grabowski.html

Grabowski, B. L., Harkness, W. L., Birdwell, D., & Rosenberger, J. (1995). Small group activities in larger classes: A controlled study. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 95-99). Alexandria, VA: American Statistical Association.

Grosof, M. S., & Sardy, H. (1993). Statistical literacy for whom: The case of the two-year colleges. *American Statistical Association 1993 Proceedings of the Section on Statistical Education* (pp. 247-253). Alexandria, VA: American Statistical Association.

Hawkins, A. (1996). Myth-conceptions! In J. B. Garfield and G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (pp. 1-14). Voorburg, The Netherlands: International Statistical Institute.

Hawkins, A. (1997). Discussion. In D. S. Moore, New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-165.

Hayden, R. W. (1990). Using writing to improve student learning of statistics. In A. Sterrett (Ed.), *Using writing to teach mathematics* (pp. 125-126). Washington, DC: Mathematical Association of America.

Hogg, R. V. (1991). Statistical education: Improvements are badly needed. *American Statistician, 45*(4), 342-343.

Hollis, P. J. (1997). Ideas for improving statistics education. *International Journal of Mathematical Education in Science and Technology, 28*(4), 569-573.

Iversen, G. R. (1991) Writing papers in a statistics course. *American Statistical Association 1991 Proceedings of the Section on Statistical Education* (pp. 29-32). Alexandria, VA: American Statistical Association.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Active learning: Cooperation in the college classroom.* Edina, MN: Interaction Book Company.

Jones, L. V. (1991). *Using cooperative learning to teach statistics.* (Research Report 91-2). Chapel Hill, NC: The L. L. Thurston Psychometric Laboratory, University of North Carolina.

Keeler, C. M., & Steinhorst, R. K. (1994). Cooperative learning in statistics. *Teaching Statistics 16*(3), 81-84.

Keeler, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. [22 paragraphs]. *Journal of Statistics Education* [On-line serial], *3*(2). http://www.amstat.org/publications/jse/v3n2/keeler.html

Knoell, D. M. (1996). Moving toward collaboration in transfer and articulation. In T. Rifkin (Ed.), *Transfer and articulation: Improving policies to meet new needs.* (pp. 55-64). San Francisco: Jossey-Bass Publishers.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. [24 paragraphs]. *Journal of Statistics Education* [On-line serial], *3*(1). http://www.amstat.org/publications/jse/v3n1/konold.html

Krevisky, S. (1994). You can get there from here: An analysis of student projects in a community college statistics class. *Proceedings of the Fourth International Conference on Teaching Statistics* (p. 27). Morocco: National Institute of Statistics and Applied Economics.

Liu, H. C. (1998). A cross-cultural study of sex differences in statistical reasoning for college students in Taiwan and the United States. Unpublished doctoral dissertation, University of Minnesota.

Magel, R. C. (1996). Increasing student participation in large introductory statistics classes. *American Statistician, 50*(1), 51-56.

Mansfield, E. R. (1995). A time-subject index for *Against all odds: Inside statistics. American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 341-345). Alexandria, VA: American Statistical Association.

McKenzie, J. D. (1996). Trends in introductory applied statistics courses: Technology. *American Statistical Association 1996 Proceedings of the Section on Statistical Education* (pp. 229-232). Alexandria, VA: American Statistical Association.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: Macmillan.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-165.

Myers, K. N. (1990). An exploratory study of the effectiveness of computer graphics and simulations in a computer-student interactive environment in illustrating random sampling and the central limit theorem (Doctoral dissertation, The Florida State University, 1990). *Dissertation Abstracts International, 51-02A*, 441.

National Council for the Social Studies. (1994). *Expectations of excellence: Curriculum standards for social studies.* Washington, DC: Author.

National Council of Teachers of English. (1996). *Standards for the English language arts*. Urbana, IL: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Research Council. (1991). *Moving beyond myths: Revitalizing undergraduate mathematics*. Washington, DC: National Academy Press.

Orton, A. (1992). *Learning mathematics: Issues, theory, and classroom practice*. (2nd ed.). New York: Cassell.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books, Inc.

Priselac, N. J. (1995). Distance learning and introductory statistics. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 262-267). Alexandria, VA: American Statistical Association.

Prichard, G. R. (1995). The NCTM standards and community colleges: Opportunities and challenges. *Community College Review, 23*(1), 23-32.

Prichard, M. K. (1993). Applying the standards to the college classroom: Ideas and obstacles. *Mathematics Teacher, 86*(9), 744-747.

Rojas, M. E. (1992). Enhancing the learning of probability through developing students' skills in reading and writing (Doctoral dissertation, Columbia University Teachers College, 1992). *Dissertation Abstracts International, 53-05A*, 1437.

Rosenbaum, R. S. (1980). A study to determine the effect on achievement and course attitude when community college students write and execute computer programs for selected topics in elementary statistics (Doctoral dissertation, New York University, 1980). *Dissertation Abstracts International, 41-12A*, 5013.

Rossman, A. J. (1992). Introductory statistics: The "workshop" approach. *American Statistical Association 1992 Proceedings of the Section on Statistical Education* (pp. 352-357). Alexandria, VA: American Statistical Association.

Rossman, A. J. (1994). Learning statistics through self-discovery. *Proceedings of the Fourth International Conference on Teaching Statistics* (p. 30). Morocco: National Institute of Statistics and Applied Economics.

Rossman, A. J. (1996). *Workshop statistics: Discovery with data.* New York: Springer-Verlag.

Rumsey, D. J. (1998). A cooperative teaching approach to introductory statistics. [77 paragraphs]. *Journal of Statistics Education* [On-line serial], *6*(1). http://www.amstat.org/publications/jse/v6n1/rumsey.html

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement, 55*(5), 868-875.

Scheaffer, R. L. (1994). Activity-based statistics: Engaging students in the learning process. *Proceedings of the Fourth International Conference on Teaching Statistics* (p. 27). Morocco: National Institute of Statistics and Applied Economics.

Scheaffer, R., Gnanadesikan, M., Watkins, A., & Witmer, J. (1996) *Activity-based statistics.* New York: Springer-Verlag.

Sevin, A. D. (1995). Some tips for helping students in introductory statistics classes carry out successful data analysis projects. *American Statistical Association 1995 Proceedings of the Section on Statistical Education* (pp. 159-164). Alexandria, VA: American Statistical Association.

Sevin, A. D. (1996). Trends in introductory applied statistics courses: Topics. *American Statistical Association 1996 Proceedings of the Section on Statistical Education* (pp. 217-221). Alexandria, VA: American Statistical Association.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.

Snee, R. D. (1993). What's missing in statistical education? *American Statistician, 47*(2), 149-154.

Snell, J. L., & Finn, J. (1993). *A course called "Chance"* [On-line]. Available: http://www.geom.umn.edu/docs/snell/chance/course/sf/sf.html

Snell, J. L. (1994). Chance: A course about current *Chance* news. In Committee on Applied and Theoretical Statistics, National Research Council (Ed.), *Modern interdisciplinary university statistics education: Proceedings of a symposium.* (pp. 86-90). Washington, DC: National Academy Press.

Solomon, P. R. (1988). Science and television commercials: Adding relevance to the research methodology course. In M. E. Ware and C. L. Brewer (Eds.), *Handbook for teaching statistics and research methods* (pp. 146-150). Hillsdale, NJ: Lawrence Erlbaum Associates.

Somers, K., Dilendik, J., & Smolansky, B. (1996) Class activities with student-generated data. *Mathematics Teacher, 89*(2), 105-107.

Spurrier, J. D., Edwards, D., & Thombs, L. A. (1993). Elementary statistics laboratory. *American Statistical Association 1993 Proceedings of the Section on Statistical Education* (pp. 23-28). Alexandria, VA: American Statistical Association.

Steffe, L. P., & Kieren, T. (1994). Radical constructivism and mathematics education. *Journal for Research in Mathematics Education, 25*(6), 711-733.

Sterling, J., & Gray, M. W. (1991). The effect of simulation software on students' attitudes and understanding in introductory statistics. *Journal of Computers in Mathematics and Science Teaching, 10*(4), 51-56.

Stiff, L. V., Johnson, J. L., & Johnson, M. R. (1993). Cognitive issues in mathematics education. In P. Wilson (Ed.), *Research ideas for the classroom: High school mathematics* (pp. 3-20). New York, NY: Macmillan Publishing Company.

Tannenbaum, P. & Arnold, R. (1995). *Excursions in modern mathematics* (2nd Ed.). Upper Saddlebrook, NJ: Prentice Hall.

von Glasersfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 3-18). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: The MIT Press.

Werner, J. S. (1995). A study of the effects on statistical reasoning, skill and conceptual development, and attitude of students in an elementary statistics class which is taught in a contextually and technologically rich environment. Unpublished doctoral dissertation, The Pennsylvania State University.

APPENDICES

APPENDIX A
Descriptions of Laboratory Activities

Activity #1: *Head Count*

This activity was based on activities in *Interactive Stat*istics (Aliaga & Gunderson, 1998) and *Activity-Based Statistics* (Scheaffer et al., 1996). It was designed to show students the importance of random sampling and the influence of increasing sample size. Students were shown a transparency with 100 groups of stick figures and they were instructed to guess the average number of stick figures in each group, or the average number of "people per home". After recording their guesses they were asked to select a representative set of 5 "homes" and determine the mean number of stick figures per home. They repeated this for a representative set of 10 homes.

Next, students used a random number table or the random number generator on their calculators to select random samples of 5 homes and random samples of 10 homes, computing the mean number of occupants in each case. Then, the class pooled their results for each sample and computed the mean and standard deviation for each (guesses, representative samples of 5 and 10, random samples of 5 and 10). Results were discussed and students were asked to make conjectures and draw conclusions.

Name ______________________________

Headcount

You want to estimate the average number of people per home in a small town of 100 households. The instructor will display a transparency of the 100 family units on the overhead projector. In the short time available, make your best guess as to the average number of people per home.

MY GUESS : __________

v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v

Now, you will receive your own copy of this graphical display. Choose 5 homes that seem to represent the group and determine the average number of people per home.

'House Numbers' of the *Representative Sample of 5*:

____ ____ ____ ____ ____

Mean number of people living in those homes: ________

v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v

Choose a sample of 10 homes which seem representative of the entire group of 100.

'House Numbers' of the *Representative Sample of 10:*

____ ____ ____ ____ ____ ____ ____ ____ ____ ____

Mean number of people living in those homes: ________

v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v

Now, use the random number table in the back of your textbook to determine a random sample of 5 homes from the 100 in town.

'House Numbers' of the *Random Sample of 5:*

____ ____ ____ ____ ____

Mean number of people living in those homes: ________

Write a sentence that compares this value with the guess you made at the beginning of class and with your representative sample of 5.

Now, choose a random sample of 10 homes from the 100 in town.

'House Numbers' of the *Random Sample of 10:*

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

Mean number of people living in those homes: __________

Write a sentence that compares this value with the guess you made at the beginning of class and with your representative sample of 10.

v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v v

As a class, use the TI-83 to compute each of the following:

Mean of class guesses:

Standard deviation of class guesses:

Mean of class averages
for representative sample of size 5:

Standard deviation of class averages
for representative sample of size 5:

Mean of class averages
for representative sample of size 10:

Standard deviation of class averages
for representative sample of size 10:

Mean of class averages
for randomsample of size 5:

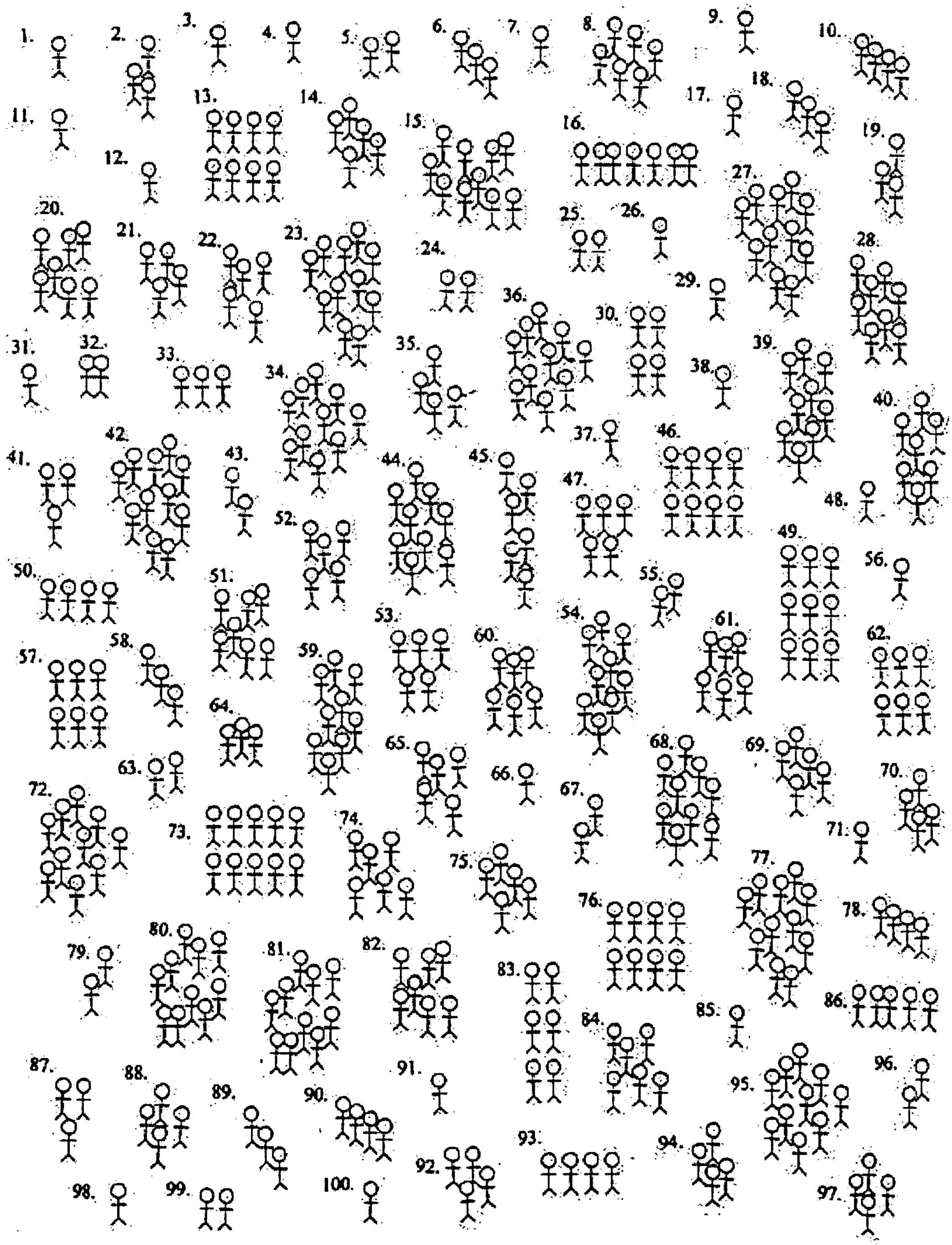
Standard deviation of class averages
for random sample of size 5:

Mean of class averages
for random sample of size 10:

Standard deviation of class averages
for random sample of size 10:

Compare these values with the mean and standard deviation for the entire population of 100 homes, provided by the instructor. Which sample(s) provided the best estimate for the mean number of people per home in this town? Why?

What conclusions can you draw about the importance of random sampling?


1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43.
44.
45.
46.
47.
48.
49.
50.
51.
52.
53.
54.
55.
56.
57.
58.
59.
60.
61.
62.
63.
64.
65.
66.
67.
68.
69.
70.
71.
72.
73.
74.
75.
76.
77.
78.
79.
80.
81.
82.
83.
84.
85.
86.
87.
88.
89.
90.
91.
92.
93.
94.
95.
96.
97.
98.
99.
100.

Activity #2: *'Round and Round*

This activity was designed to teach the Law of Large Numbers constructively. Students completed this activity in pairs with each pair sharing a spinner chosen from among many color and number combinations available. They were asked to describe the spinner they chose, select an outcome considered "success", and determine the theoretical probability of that outcome occurring.

Students were instructed to spin the spinner 100 times, in 10 sets of 10 spins. They were provided with a table to keep track of the results for each set of ten spins, as well as their cumulative results. Later, students were asked to compare the results of each set of ten spins to the theoretical probability they determined earlier. Then they were asked to compare the empirical results with the theoretical expectation as the total number of spins increased.

Name(s) ____

'Round and 'Round

You are familiar with a fair coin having "heads" on one side and "tails" on the other. Because the two possible outcomes are equally likely, we say that the probability that a single flip will result in "heads" is 1/2. We use the phrase *theoretical probability* to describe probability determined by such mathematical means. On the other hand, when probability is determined based on the frequency of repeated observations we use the term *empirical probability*. How are theoretical and empirical probabilities related? The results of this activity should help you answer that question.

For this activity, you will use a spinner with various colors and/or numbers on it. Draw a sketch of your spinner, and/or write a brief description of it in the space below.

If an experiment consists of spinning the spinner once, and if we assume that the spinner lands in one sector (not on a line), write the sample space for the experiment:

Choose an event (for example, *the spinner lands on brown*) and write the corresponding probability statement based on the theoretical probability for that event.

For example, you might write **P(brown) = 5/6**.

What is the decimal representation for this probability?

Review the data collection sheet below. Spin the spinner 10 times, keeping track of the number of successes in the first row of the table. Then, spin 10 more times, recording the number of successes resulting from those ten spins, as well as the cumulative data from the entire set of twenty spins. Then complete rows 3-10 of the table in a similar fashion.

	# of successes for this set of 10 spins	proportion of successes for this set of ten spins	cumulative number of successes	proportion of successes based on all spins	decimal estimation of successes based on all spins
first set of 10 spins		/ 10		/ 10	
second set of 10 spins		/ 10		/ 20	
third set of 10 spins		/ 10		/ 30	
fourth set of 10 spins		/ 10		/ 40	
fifth set of 10 spins		/ 10		/ 50	
sixth set of 10 spins		/ 10		/ 60	
seventh set of 10 spins		/ 10		/ 70	
eighth set of 10 spins		/ 10		/ 80	
ninth set of 10 spins		/ 10		/ 90	
tenth set of 10 spins		/ 10		/100	

How does the actual proportion of successes for **each set** of ten spins compare to the theoretical probability you stated on the previous page?

Now consider the far column of the table. How does the empirical probability compare to the theoretical probability as the **total number of trials** increases?
(This phenomenon is referred to as the *Law of Large Numbers.*)

Activity # 3: *Coin Toss*

This activity was completed by groups of 4 or 5 students working together. It was designed to teach the concepts of sample space and discrete random variables. Students generated the sample spaces of sequences of 2, 3, and 4 coin tosses. They were instructed to complete 25 sequences of 4 tosses at home and record the results. During the next class period, group members pooled their results and investigated the data, attempting to determine whether any of the sequences seemed more likely to occur than any other.

The next part of the activity involved the discrete random variable that counts the number of heads tossed in a sequence of four tosses. Students worked together to construct the probability distribution. They later reflected on the difference between the sixteen equally likely outcomes possible when a coin is tossed four times and the discrete random variable that counts the number of heads obtained in a sequence of four tosses.

Names ________________________________ (Write your name first.)

____

____

____

____

Coin Toss

You will work in groups of four or five to complete this activity. Discuss the answers to the following questions with the members of your group, and write responses that reflect the consensus of the group.

If you toss a coin once, how many outcomes are possible?

List the possible outcomes (or the *sample space*):

If you toss a coin twice, how many outcomes are possible?

List the possible outcomes (or the *sample space*):

If you toss a coin three times, how many outcomes are possible?

List the possible outcomes (or the *sample space*):

If you toss a coin four times, how many outcomes are possible?

List the possible outcomes (or the *sample space*):

Let's investigate what happens when a coin is tossed four times successively. Do you think that any of the outcomes listed in your sample space above are more likely to occur than any of the others? Explain your response completely.

Each member of your group should toss a coin four times, keeping track of the sequences of heads and tails. Repeat the tossing for a total of 25 sequences per person. (Flip four times then mark the appropriate sequence row. Flip four more times then mark the appropriate sequencc row. Repeat for a total of 25 groups of 4, or 100 total flips per person.)

Individual results:

Outcomes	Tally

Pool your results together and count the number of times your group obtained the sequences listed in your sample space on the previous page.

Make a group tally using the table below. Your tally total should be 100 or 125 (25 times the number of people in your group).

Outcomes	Tally

Does it appear that any of the outcomes is more likely than any other outcome?

Is this consistent with your prediction on the previous page? Explain.

Now, suppose we define a random variable, x, which counts the number of heads appearing when a coin is tossed four times. Complete the probability distribution table below, using the outcomes listed in the left column of the table on the previous page to determine the probability of each value of x:

x					
P(x)					

Is the probability the same for each value of x?

Is this consistent with your response regarding the likelihood of the sequences of heads and tails above? Explain.

Write a paragraph which compares and contrasts the two ideas you've looked at today – the possible outcomes resulting from four flips of a coin and the possible values for the random variable which counts the number of heads for each set of four flips. Include the concept of "equally likely" in your explanation.

Activity #4: *Lucky Numbers*

This was a teacher-led whole-class activity designed to illustrate the Central Limit Theorem. Each student was provided with a small toy lottery machine containing balls marked with integers 1-39. When a button was pushed the balls hopped in the machine and randomly fell into a cylinder at the side of the toy. The instructor led a discussion of the random variable taking on the value of the first ball to drop in the cylinder.

The class drew a histogram of the uniform random variable, and determined the expected value and the standard deviation. Then, each student used the toy to collect two samples of size 30 from this distribution. Students computed the mean for each sample of size thirty. (We called these their "personal" $\bar{x}$ values.)

As a class we investigated the distribution of these sample means. Using the TI-83 with viewscreen capability we computed the mean and standard deviation and compared these to the mean and standard deviation of the original uniform random variable. We also constructed a histogram and compared its mound shape to the flat, rectangular shape of the uniform histogram.

A discussion of the Central Limit Theorem followed and students were asked to conjecture what would happen to the mean and standard deviation if each sample contained 50 values or 15 values rather than 30.

Activity #5: *Where are all the blue m&ms®?*

This activity was designed to teach the concepts involved in constructing and interpreting confidence intervals for proportions. The class was divided into 10 groups of about 3 students each. Each group was provided with a sample of m&ms® scooped from a large 3-pound bag. (Each sample contained approximately 100 m&ms®.) Each group had at least one member with a TI-83 calculator to facilitate the computation of the confidence interval.

Students were asked to verify that their sample met the necessary criteria for approximating the binomial distribution with the normal distribution. If the sample was not sufficient they were provided with additional m&ms®. Each group of students was asked to construct and interpret a 90% confidence interval for the true population proportion of blue m&ms® using their group sample. These confidence intervals were posted on the board and compared. The instructor then provided the M&M/Mars Company data giving the population proportion of blue m&ms® (10%) and students were asked to reflect on the results.

Name(s) ____ Group # _______

Where are all the blue m&ms®?

In 1995 Mars, Inc. replaced tan m&ms® with blue...the public's choice. How does this public relations campaign impact the typical bag of m&ms®? A confidence interval for the proportion of blue m&ms® might shed some light on this situation.

When np > 5 and n(1-p) > 5 the distribution of the sample proportion is approximately normal with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

For samples meeting these criteria, a $c\%$ confidence interval for the population proportion is given by

$$\hat{p} - z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

We will use this formula (or alternatively the TI-83) to determine a 90% confidence interval for p.

m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m m

The class will break into 10 groups and each group will determine a 90% confidence interval for the population proportion of blue m&ms®.

Determine the total number of m&ms® in your group sample. TOTAL : _____________

Now determine how many of those are blue. BLUE : _____________

Use this information to calculate $\hat{p}$, the sample proportion.

Does your sample data meet the necessary criteria for approximating the binomial distribution with the normal distribution? Explain.

Use the data from your group to construct a 90% confidence interval for the proportion of blue m&ms®.

Write a sentence explaining what that confidence interval represents.

Members of the class constructed 10 different confidence intervals. How many of these confidence intervals would you expect to contain the true population proportion? Explain.

Write your confidence interval on the board in the appropriate place. Then, copy the confidence intervals of the other groups so you have a complete list of all ten confidence intervals.

Group # 1 ______________________________

Group # 2 ______________________________

Group # 3 ______________________________

Group #4 ______________________________

Group # 5 ______________________________

Group # 6 ______________________________

Group # 7 ______________________________

Group # 8 ______________________________

Group # 9 ______________________________

Group # 10 ______________________________

Based on the intervals above, can you predict the true proportion of blue m&ms®? Explain.

At this point, bring your completed lab sheet to the instructor who will provide you with information from M&M/Mars which discloses the true proportion of blue m&ms® in production.

Is the class experiment consistent with the claim made by Mars? Explain.

Are the results of this activity surprising to you based on your own experience with m&ms® candies?

Activity #6: *Let's Go for a Spin!*

This teacher-led activity introduced students to the basic concepts involved in hypothesis testing and was developed to give students an intuitive feel for "rejecting the null hypothesis". It was adapted from *Instructor Resources for Activity-Based Statistics* (Schaeffer, et al., 1998) where it is reported that the probability of a 1962 penny landing heads up when spun rather than flipped is 10% (p. 248).

The day of the activity the researcher handed each student a penny stating that she'd learned from experience that not everyone carried change to class. (Students were **not** told that they all had 1962 pennies.) As expected, students predicted that the probability of getting heads when the penny was spun would be .5. This was used as the value for the null hypothesis, $p = .5$ The alternative $p \neq .5$ was also written on the board and discussed. Next, students were asked to spin their pennies 10 times and record the results.

As students began spinning their pennies many of them made statements to others such as "I must be spinning wrong" or "I must have a bad penny" since the outcomes were generally not what they expected. Class results were compared and the instructor led a discussion of the outcomes. A few students had results close to 50% heads, but most had many fewer heads than tails. Students computed the probability of getting the results they did under the assumption of the null hypothesis $p = .5$. This was used in a later class as the concept of p-values was introduced. Type I and Type II errors were discussed.

Activity #7: *Pepsi® or Coke®?*

This taste-test activity was designed to guide students through the hypothesis testing process. Pairs of students worked together to set up the null and alternative hypotheses and to select a level of significance for the test.

Then students were provided with two cola samples labeled A and B. They attempted to identify the samples as Pepsi® or Coke® and then submitted their guesses to the instructor who tallied the class results and provided students with the $\hat{p}$ value. Students used this information to complete the hypothesis test.

Name(s) ____

Pepsi® or Coke®?

Can people can tell the difference between Pepsi® and Coke®? In this experiment we will use a taste test to compare the proportion of students who correctly identify Pepsi® and Coke® with the proportion we'd expect to correctly identify the two brands based on guessing.

If the class members truly cannot identify which cola sample was Pepsi® and which was Coke®, then we'd expect the proportion of students guessing correctly would be ______ .

Write in sentences the null and alternative hypotheses for this test.

H_0:

H_1:

Now write each of the hypotheses using statistical notation.

$H_0: p =$

H_1:

At what level of significance will you test this claim? $\alpha =$ __________

Each class member will receive two small cups of cola, one labeled Brand A and the other labeled Brand B. Silently, attempt to identify one sample as Pepsi® and one sample as Coke®. Record your decision on the paper ballot provided and submit it to the instructor. When all ballots have been collected the instructor will announce the correct brands and the class will determine the proportion of those correctly identifying the two beverages.

When np > 5 and n(1-p) > 5 the distribution of the sample proportion is approximately normal with mean np and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

Is the class sample size sufficiently large? Explain.

The test statistic for this situation is:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Use the space below to test the hypothesis you asserted on page 1.

What conclusion can you draw? Write a sentence that communicates the decision in everyday language.

Activity #8: *Do you Measure Up?*

This activity was modeled after an activity in *Workshop Statistics: Discovery with Data and the Graphing Calculator* (Rossman & Von Oehsen, 1997). Students worked in pairs to measure their arm spans and heights using pre-marked masking tape on classroom walls. Three stations were set up for each measurement. Students were assigned a number and were instructed to record their arm spans and heights on a class overhead transparency so all students could easily access all values.

Students constructed and described scatterplots. Then they computed and interpreted the correlation coefficient for this set of data. Follow-up questions asked students to generate examples of other pairs of body measurements that might be similarly correlated and explain an example of two variables that might be negatively correlated.

Name(s) ________________________________ Number ______

Do you measure up?

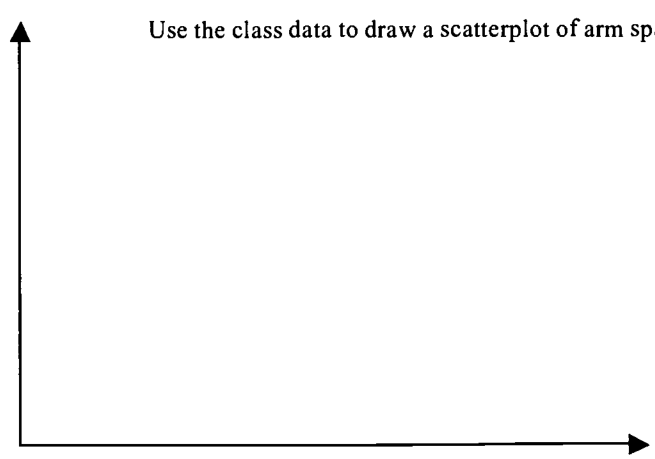
Is there a relationship between a person's arm span and height?

Use the masking tape "meter sticks" on the classroom walls to determine each of these body measurements. Then, enter your data in the table below according to your student number at the top of the page.

Student Number	Arm Span in centimeters	Height in centimeters
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		

Now enter your personal data on the class table (overhead transparency) on the line corresponding to your student number. Copy the entire table of student data on this page above.

Use the class data to draw a scatterplot of arm span vs. height.



Describe the scatterplot in words. Does there appear to be a relationship between arm span and height? Explain.

It is possible to describe the linear relationship between two variables mathematically. The *correlation coefficient, r*, is a number which indicates the strength of the linear relationship. The r value will always be between -1 and 1. If the r value is -1 there is a perfect negative relationship between the two variables . . . as one increases the other decreases. If the r value is 1 then there is a perfect positive relationship between the two variables . . . as one variable increases the other increases. The closer the r value is to -1 or 1, the stronger the relationship between the two variables. An r value of zero indicates no linear relationship between the two variables (although there could be some non-linear relationship).

Use a TI-82, TI-83 or MINITAB to determine the correlation coefficient for this data.
{If you have a TI-83, be sure to check that your "DiagnosticsOn" option is activated by pushing 2nd 0 (for CATALOG), then scrolling down to "DiagnosticsOn" and pushing ENTER, ENTER.}

Based on this value, how strong is the linear relationship between arm span and height?

Does this surprise you? Why or why not?

What other pair of body measurements might have a similar correlation coefficient?

Can you think of a pair of "real life" variables that might have a negative correlation coefficient? Identify the two variables and explain why you think the linear relationship between them might be negative.

Activity #9: *We All Scream for Ice Cream*

Inspired by a chocolate-chip cookie taste testing activity at a local mathematics teachers' conference, this activity involved a taste test with six different brands of vanilla ice cream. Individual students tasted six samples labeled A-F in random order and rated each sample on a 5-point scale for the quality of vanilla flavor, texture, and sweetness.

Class means for each of the quality categories for each brand were computed. The instructor then provided brand names and fat content per half-cup serving for each brand. Students constructed three scatterplots, and determined the equations of the least-squares regression line for each quality.

The least-squares regression lines were used to answer questions and make predictions. Students were asked to select the taste quality that was best predicted by the fat content of the ice cream. They were also asked to interpret the coefficient of determination in the context of this activity.

Name(s) ____

"...we all scream for Ice Cream!"

Does the fat content of vanilla ice cream indicate its quality? In this experiment we will measure consumer satisfaction by considering three components of ice cream quality. We will then attempt to determine if the fat content of the ice cream is an indicator of its quality.

Six brands of vanilla ice cream will be taste tested. Small samples of each will be labeled A, B, C, D, E, and F. You will be provided with a set of samples, a data collection sheet, and, if you wish, a glass of water to "cleanse your palette" after each taste. You will begin by randomly determining your tasting order sequence . Follow your tasting sequence and mark your data collection sheet appropriately. After you have tasted all the samples in the sequence provided, you may return to taste previous samples again, if desired.

Your Randomly Selected Tasting Sequence: ______________________________

Data collection

Outstanding (5) ***Very Good (4)*** ***Good (3)*** ***Fair (2)*** ***Poor(1)***

Brand	Vanilla Flavor	Texture	Sweetness
A	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1
B	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1
C	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1
D	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1
E	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1
F	5 4 3 2 1	5 4 3 2 1	5 4 3 2 1

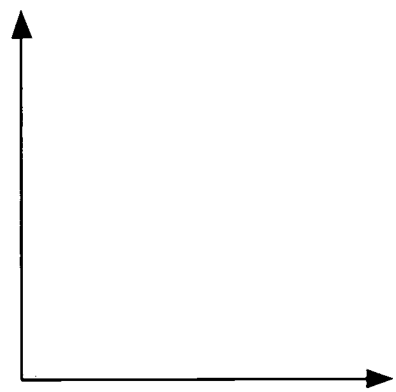
The class will average their ratings for each brand of ice cream. The fat content per half cup serving for each brand along with the average rating for each brand in each category will form sets of 6 ordered pairs.

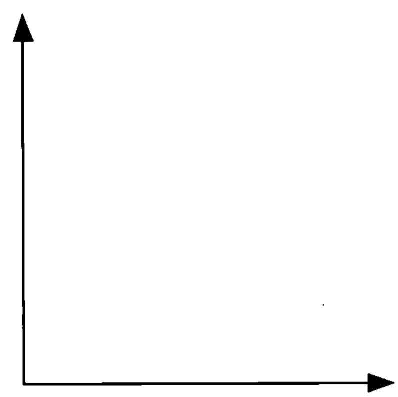
Record the fat content for each of half cup serving of ice cream and the average class rating for each criteria.

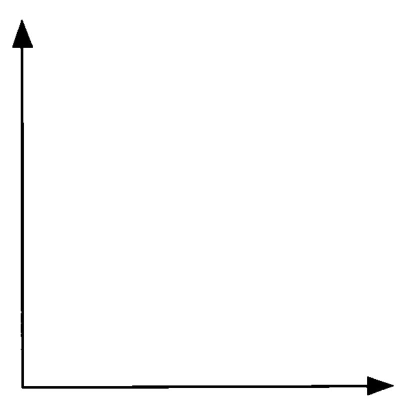
Brand	Fat Content (per half cup serving)	Class Average Vanilla Flavor Rating	Class Average Texture Rating	Class Average Sweetness Rating
A	______	______	______	______
B	______	______	______	______
C	______	______	______	______
D	______	______	______	______
E	______	______	______	______
F	______	______	______	______

Using a graphing calculator (or MINITAB) enter the fat content in L1 (or C1) and the class rating average for that brand in each of the three categories in L2, L3, and L4 (or C2, C3, and C4).

Draw and label three scatterplots, one for each set of (fat content, rating) ordered pairs.







The square of the correlation coefficient, r^2, is called the coefficient of determination. The coefficient of determination indicates the percentage of variability in the dependent variable (quality indicator) that is explained by the variability in the independent variable (fat content).

The *least squares regression line* is the line that best fits the observed data. The equation is written $\hat{y} = a + bx$.

For each of the three quality indicators, determine a linear regression model and record the corresponding correlation coefficients and coefficients of determination. {If you have a TI-83, be sure to check that your "DiagnosticsOn" option is activated by pushing 2nd CATALOG, then scrolling down to "DiagnosticsOn" and pushing ENTER, ENTER.}

Vanilla Flavor : ______________________ $r =$ _______ $r^2 =$ ________

Texture: ______________________ $r =$ _______ $r^2 =$ ________

Sweetness: ______________________ $r =$ _______ $r^2 =$ ________

Which of these has the strongest linear correlation? Briefly discuss the meaning of this measure in the context of this experiment.

What is the predicted Vanilla Flavor rating of ice cream with 4.5g of fat per serving? How did you determine this?

If an ice cream had a predicated texture rating of 4, what would expect as the fat content per serving? How did you determine this?

For wh ich of these criteria is fat content the best predictor of quality? How do you know?

How can the r^2 values be interpreted in the context of this experiment?

Activity #10: *Play Ball!*

Students collected data for this activity outside of class, asking twenty adults their preference for playing individual sports, teams sports, both, or neither and then asking their birth order (oldest, youngest, middle, only child). They recorded this information in a grid provided by the instructor.

At the beginning of the following class period the students shared their data and the pooled data was used for the analysis. [Note: In each class every cell had at least 5 entries so this potential problem did not have to be addressed.] Students worked through the activity doing some pieces on their own, working with a partner at times, and participating in a whole-class discussion other times.

Name ___________________________________

Play Ball!

Do you think that preference for individual or team sports is related to an individual's birth order? The Chi-Square test for independence will help us answer this question.

Data Collection: Before the next class session, ask 20 adults whether they prefer to ***play*** individual sports, team sports, both kinds of sports, or neither. (We're not interested in what kind of sports they like to watch.) After the person answers in one of the four possible ways, ask where they fall in family birth order: first, last, somewhere in the middle, or only child. Keep your tally of responses in the chart below. Ask 20 "independent" people – no siblings or softball teammates! Also, be sure that the people you ask haven't already participated by answering these questions for one of your classmates.

	Oldest Child	Youngest Child	Somewhere in the Middle	Only Child
Prefers Individual Sports (swimming, cycling, tennis, golf, etc.)				
Prefers Team Sports (softball, football, soccer, etc.)				
Likes to play both kinds of sports				
Doesn't like to participate in sports at all				

Data Analysis:

For this test, the null hypothesis is that the two characteristics are independent, and the alternative is that the two characteristics are dependent. Use the context of this activity to write the null and alternative hypotheses in words.

H_o:

H_1:

We will compare the number of observations in each cell with the number of expected responses in each cell if in fact these two characteristics are independent. In the table below, enter the class totals for observations in each of the 16 cells containing possible responses. Compute the row and column totals as well.

	Oldest Child	**Youngest Child**	**Somewhere in the Middle**	**Only Child**	***Row Total***
Likes to Play Individual Sports					
Likes to Play Team Sports					
Likes to play both kinds of sports					
Doesn't like to play sports at all					
Column Total					

Given the data above, what is the probability that a randomly selected person (in any birth order position) likes to play individual sports?

What is the probability that a randomly selected person (with any sports preference) is an oldest child?

Based on our knowledge of probability, if the two variables were truly independent we'd expect that P(individual sports *and* oldest child) would equal P(individual sports) *times* P(oldest child).

Use your results to compute the probability that a randomly selected person would prefer to play individual sports **and** would be an oldest child.

The number of people we'd expect to be in the first cell {individual sports and oldest child} is equal to the probability of any one person being in that cell multiplied by the total number of people surveyed.

Using this idea, determine the expected number of people in the first cell. ___________

Now write a simplified formula for the expected number in each cell using the row total, the column total, and the grand total.

Expected cell entry =

Compute the expected number of respondents in each cell.

Cell 1:	Cell 2:	Cell 3:	Cell 4:
Cell 5:	Cell 6:	Cell 7:	Cell 8:
Cell 9:	Cell 10:	Cell 11:	Cell 12:
Cell 13:	Cell 14:	Cell 15:	Cell 16:

The test statistic is $\chi^2 = \sum \frac{(O-E)^2}{E}$. If the characteristics are truly independent, then the observed values will be close to the expected values and the χ^2 statistic will be relatively small. Compute the test statistic.

We will compare this computed value with the χ^2 critical value having (r-1)(c-1) degrees of freedom, where r is the number of rows in our table and c is the number of columns.

Draw a diagram of the Chi-Square distribution below, showing the rejection region and the indicating the critical value. Use $\alpha = .05$.

Will you reject H_0 or fail to reject H_0? Why?

Write a sentence that communicates the meaning of your decision using the specific context of the problem.

APPENDIX B
Preassessment Items

Preassessment Items

1. Evaluate the expression $ab^2 - \sqrt{bc}$ if $a = 3$, $b = -2$, and $c = -8$.

2. Ten chips marked with the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 are mixed in a bag. What is the probability that a randomly selected chip is marked with a multiple of three?

3. Solve for p: $3p + 2r = 15$.

4. Determine the mean of the following set of values: 12, 19, 8, 4, 10, 7.

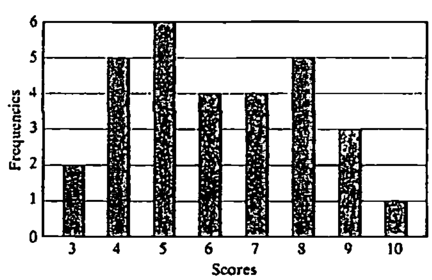
5. Sketch the graph of any line having a slope of -1.

6. What is meant by the term *median*?

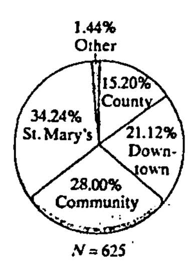
7. A seed company wants to include 200 seeds per packet and wants 32% to be zinnia seeds. If 25 zinnia seeds have already been included in a packet, how many more need to be included?

8. Write the equation of the line passing through the points (2, -4) and (-1, 8).

9. The following bar graph shows the frequencies of various scores received by students in Math A on a 10-point pop quiz. How many students took the quiz? *



10. The pie chart shown below gives the percentage of babies born at each of the four hospitals in the city of Smithtown in the last year. How many babies were born at Downtown Hospital? *



*From Tannenbaum, P. & Arnold, R. (1995). *Excursions in Modern Mathematics* (2nd Ed).

APPENDIX C
Course Test Items

Test 1 Items, Fall 1998

Part A: Short answer. Write the letter of the best choice in each blank. [3 points each]

1A. _______ Isabel scored 78 on a physics test, placing her at the 85th percentile. After returning the papers, the instructor realized that she had mistakenly added an extra 10 points to each score and she reduced every grade by 10 points. Isabel's new score remained at the 85th percentile.

A) True B) False

2A. _______ The mean of a set of values is considered a measure resistant to outliers.

A) True B) False

3A. _______ Le Thu knows that the probability her son Alex will forget to take out the trash on trash night is .1 . The probability he remembers to take out the trash is .9 .

A) True B) False

4A. _______ The agriculture department developed a new variety of tomato plant. For one particular set of conditions, eight plants were grown to maturity and the yield over a two week period was recorded. The mean number of tomatoes was 15 per plant. Which of the following is true?

A) The most typical yield was 15 tomatoes per plant.
B) There were a total of 120 tomatoes.
C) Half the plants yielded more than 15 tomatoes and half the plants yielded fewer than 15 tomatoes.
D) All of the above.

5A. _______ Three fair coins are tossed. There are eight equally likely outcomes:

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

A) True B) False

6A. _______ The stem and leaf display shows ages of students in a recent section of a psychology course. (3|2 means 32 years old.)

1 | 8 8 9 9 9
2 | 0 1 2 2 5 8 9
3 | 0 0 1 3 4
4 | 1 3
5 | 6 9

What is the median age? (Fill in the blank.)

7A. _______ A Girl Scout council wants to determine the effectiveness of their leader training program. Each of the leaders in that council completed and returned a survey. This is an example of

A) sampling B) census
C) simulation D) experimentation

8A. _______ Which of the following is true regarding the mean and the mode of a data set (if a unique mode exists)?

A) The mean is always greater than the mode.
B) The mean is always less than the mode.
C) The mean is always equal to the mode.
D) The mean can be less than, greater than, or equal to the mode.

9A. _______ Which level of measurement best describes data listing the ages of infants in months?

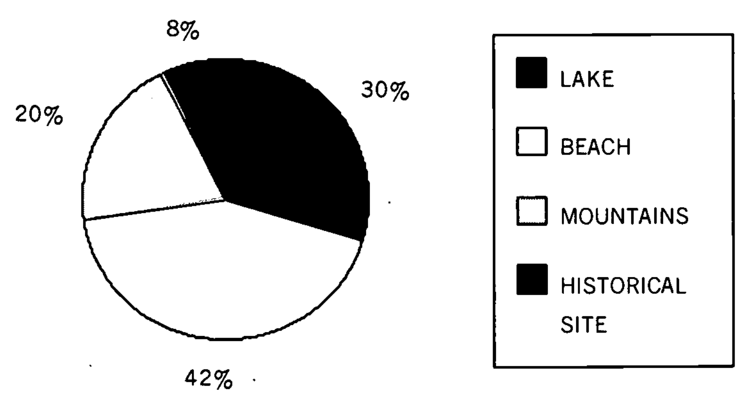
A) nominal B) ordinal C) interval D) ratio

10A. _______ To determine customer satisfaction with their long-distance telephone service, a research firm surveyed 100 customers in every area code. This is an example of systematic sampling.

A) True B) False

Part B: Free Response. **Choose 4 of the following 5 problems.** [15 points each]

1B. The following circle graph displays the favorite vacation location of 160 members of a private pilots' association. Construct and label a bar graph to display the number of members preferring each location.



2B. Five identically shaped cards are numbered 1 thorough 5, one digit per card. A six-sided fair die is marked with the letters A, E, H, O, M, U. One card is selected at random and the die is tossed.

What is the sample space for this experiment?

What is the probability of selecting an odd number and a vowel?

What is the probability of selecting an odd number or a vowel?

3B. An audit of 10 households determined that the number of years the homeowners had occupied that residence were as follows:

1	11	7	4	8	2	3	10	6	8

Determine the range and the sample standard deviation.

Explain why one measure might be considered better than the other.

4B. Thirteen adults were asked to keep track of the number of days they ate breakfast during the month of June. The reported the following data:

22	27	16	30	19	14	21	20	21	4	12	17	9

Give the five number summary and draw a box-and-whisker plot.

Write a sentence or two explaining how the box-and-whisker plot visually displays the variation of the data. Be specific.

5B. The management at a local mall wanted to estimate the amount of time people spent shopping there during the period from Thanksgiving to New Year's Day. Twenty-five shoppers reported the following numbers of hours:

5	8	2	11	6	12	10	3	3	7	18	14	
7	4	12	3	10	7	10	7	4	13	16	8	7

Create and label a frequency histogram using 5 classes. Use class boundaries as horizontal axis labels.

What statement can you make about the data using the information displayed in the histogram?

Part C: Short essay. [5 points each]

1C. What is the Law of Large Numbers? Explain this concept using at least one example.

2C. Write 3-4 sentences connecting some of the concepts you've learned in this course with each other. How do some of the concepts you've studied relate to your world outside the classroom?

Test 2 Items, Fall 1998

Part A. Choose the best answer from among the choices provided. [4 points each]

A1. _______ If the following table represents a probability distribution, determine the value of **a**.

x	3	4	5	6	7
P(x)	.05	a	.25	.45	.15

A) a = 0 B) a = .1 C) a = .15

D) a = .2 E) impossible to determine

A2. _______ A binomial distribution with n = 10 and p = .5 has a probability histogram which is

A) skewed right B) skewed left C) symmetric

A3. _______ The area under the normal curve between $\mu - 2\sigma$ and $\mu + 2\sigma$ is closest to

A) 34% B) 68% C) 95% D) 99.7 %

A4. ______ Which of the following is a characteristic of all binomial experiments?

A) fixed number of trials
B) dependent trials
C) P(success) = P(failure)
D) none of the above

A5. _____ Which of the following is **not** characteristic of a normal distribution?

A) symmetric
B) bell shaped curve
C) the median is equal to the mode
D) the total area under the curve is 1
E) none of the above

Part B. Answer each question completely, showing sufficient work to justify full credit.

B1. The following paragraph describes a situation that does not meet the conditions of a binomial experiment. Why does it fail to be binomial experiment? *[8 points]*

A researcher is studying the gender of children born at a certain hospital. He assumes that the probability of a male birth is .5 and the probability of a female birth is .5. The records of 100 newborns were examined, including 5 sets of twins and a set of triplets. What is the probability that 40 of the newborn children are boys?

B2. Suppose it is commonly known that 90% of all high school graduates attend their senior prom. The current senior class at Jones High School has 150 students.

a) Determine and interpret the mean of this probability distribution. *[8 points]*

b) Determine and interpret the standard deviation of this probability distribution. *[8 pts]*

B3. Michael is a supervisor for workers at a electronics manufacturing company. He has a consistent record of 85% attendance in his department on any given day. Michael has 20 workers in his department.

a) Determine the probability that exactly 17 of his workers show up tomorrow. *[5 points]*

b) Determine the probability that at least 19 of his workers show up tomorrow. *[7 points]*

B4. The height of a new variety of corn is normally distributed with a mean of 9 feet and a standard deviation of .3 feet.

a) What is the probability that a randomly selected corn stalk would measure between 8.5 feet and 9 feet? *[5 points]*

b) What is the probability that a randomly selected corn stalk would measure more than 9.7 feet? *[7 points]*

c) What is the probability that a randomly selected corn stalk would measure between 8.4 and 8.9 feet? *[7 points]*

B5. Cool-off is a company which manufactures air conditioning units. The life of their deluxe model is normally distributed with a mean of 15 years and a standard deviation of 2 years. They will replace any machine that fails during their guarantee period. How long should the guarantee run if they want to replace no more than 10% of the air conditioners? *[5 pts]*

Part C. Answer two of the following three questions. ***[10 points each]***

C1. What is a random variable? What is the difference between a discrete random variable and a continuous random variable? Give one example of each.

C2. What is the purpose of standardizing normal random variables? How is the standardizing accomplished?

C3. Compare and contrast the characteristics of probability histograms for discrete probability distributions with the area under the curve for continuous probability distributions.

Test 3 Items, Fall 1998

Part A. Mark each statement TRUE or FALSE. (4 points each)

A1. _________ If you want to be more sure that your confidence interval contains the true population mean, you increase your confidence level and your confidence interval gets wider.

A2. _________ The Central Limit Theorem states that the mean of the distribution of sample means, $\bar{x}$, is equal to the mean of the original distribution.

A3. _________ Statistical inference involves making a conjecture about a sample based on a similar sample.

A4. _________ If a 90% confidence interval for the population mean is $8.3 < \mu < 9.7$, then we can say that if many samples were taken from this population and similar confidence intervals were constructed, 90% of them will contain the true population mean.

A5. _________If a computer program such as MINITAB generates a p-value of .06 for a test of hypothesis at the $\alpha = .05$ level of significance, the researcher will reject the null hypothesis.

PART B. (15 points each)

Answer each question, using the procedure(s) from class. For hypothesis tests, include the null and alternative hypotheses, the level of significance, a sketch of the rejection region and the corresponding critical value, a decision and a verbal interpretation of that decision using the context of the problem. For confidence intervals, write the formula for the appropriate confidence interval first, then substitute all appropriate values and give a numerical answer as well as a verbal interpretation of the interval using the context of the problem.

B1. A national organization holds a science contest each year. Some science professors wonder if coaching students before the test would increase their scores. Eight randomly selected students are tested with different versions of the contest material before and after the coaching. The students' scores on the 20 point test are given both before and after the coaching. Do the data suggest that coaching increases students' scores? Test using $\alpha = .05$.

Student	1	2	3	4	5	6	7	8
Before	8	10	7	6	12	10	8	2
After	10	10	6	7	10	13	10	5

B2. The local merchant association has learned that the national average for holiday gift expenditures is $\mu = \$300$ per household with a standard deviation of $\sigma = \$80$. They randomly sample 40 shoppers in their county and find that the average holiday spending among this group is \$330 per household. Does this indicate that shoppers in this county have spending habits different than the national average? Test at the 5% level of significance.

B3. A random sample of 150 drivers showed that 37 regularly talked on a cellular telephone while driving. Determine a 90% confidence interval for the true population proportion of drivers who regularly talk on cellular phones.

B4. Freshness First, a grocery chain, knows that the average amount spent per week on groceries at their stores is $\mu = \$120$ with a standard deviation of $\sigma = \$25$. If a random sample of 48 shoppers is surveyed, what is the probability that the mean of their grocery bills is less than \$110?

PART C: Answer 2 of the following 3 questions, concisely but completely. (10 points each)

C1. What is the Central Limit Theorem? Why is it so important to the field of inferential statistics?

C2. What is meant by "a 99% confidence interval for the population mean?"

C3. What is a Type I error? What is Type II error? Which is generally considered more serious? Explain.

Test 1 Items, Spring 1999

Part A: Short Answer. [3 points each]

Write the letter of the best choice in each blank.

1A. ________ A family has three children. There are four equally likely outcomes: 3 girls, 2 girls and a boy, 2 boys and a girl, 3 boys.

A) True B) False

2A. ________ Which of the following is true regarding the median and the mode of a data set (if a unique mode exists)?

A) The median is always greater than the mode.
B) The median is always less than the mode.
C) The median is always equal to the mode.
D) The median can be less than, greater than, or equal to the mode.

3A. __________ The stem and leaf display shows scores on a recent test. (8|7 means 87%)

5 | 2 8
6 | 5 7 7 9
7 | 0 0 1 5 6
8 | 3 4 4 4 9
9 | 1 5 8 8 9

What is the median score? (Fill in the blank.)

4A. _____ Twenty people suffering from colds are asked to use a zinc lozenge to treat their symptoms. Twenty others are given a hard candy instead. Later, each group was asked to report on the severity of their symptoms. This is an example of

A) sampling B) census C) simulation D) experimentation

5A. ______ Mary knows from experience that her probability of getting stopped at a certain red light on Capitol Blvd. is .7. The probability she will not get stopped at that light is .3.

A) True B) False

6A. ______ Last Friday, Your Local Bank measured the length of time every fifth person coming into the bank waited for a teller. This is an example of systematic sampling.

A) True B) False

7A. _____ One recent day the airport authority recorded the number of flights departing each hour during the 12 hour period from 8:00 am - 8:00 PM. The mean number of flights was 7 per hour. Which of the following is true?

A) There were a total of 84 departing flights.
B) During half the hours, more than 7 flights departed and during half the hours less than 7 flights departed.
C) The typical number of departures was 7 per hour.
D) All of the above.

8A. _____ Elizabeth scored 83 on a chemistry test, placing her at the 90th percentile. If three points were added to each score, her new score would be at the 93rd percentile.

A) True B) False

9A. _____ The median of a set of scores is considered a measure resistant to outliers.

A) True B) False

10A. _____ Which level of measurement best describes data listing the heights of people standing in line at an amusement park.

A) Nominal B) Ordinal C) Interval D) Ratio

Part B: Free Response. Choose 4 of the following 5 problems. [15 points each]

1B. The president of Whispering Pines homeowners association wanted to know how much time residents spent at the neighborhood swimming pool. A random sample of 25 residents reported the following hours per week:

10	2	23	19	6	9	6	4	3
0	15	13	7	8	6	5	8	9
11	6	10	9	6	0	20		

Create and label a histogram using 5 classes. Use class boundaries as horizontal axis labels.

What statement can you make about the data using the information displayed in the histogram?

2B. A sample of 10 families were asked how often each month they ate dinner at a restaurant. The responses were:

2	7	5	4	10	3	2	2	4	5

Determine the range and the sample standard deviation.

Explain why one measure might be considered better than the other.

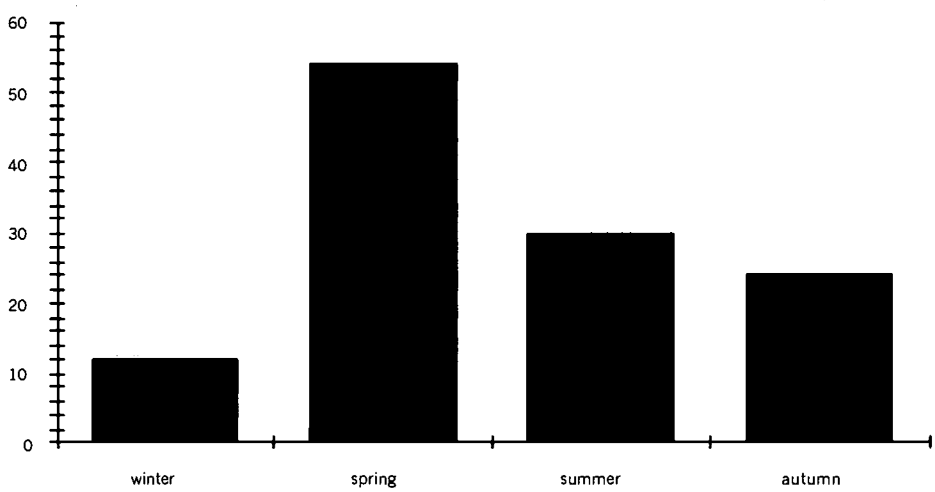
3B. On a personality assessment test, one group of questions relate to confidence. A random sample of 13 scores on the confidence portion are:

5	20	22	27	30	18	13
17	28	12	15	16	9	

Give the five number summary and draw a box-and-whisker plot.

Write a sentence or two explaining how the box-and-whisker plot visually displays the variation of the data. Be specific.

4B. The following bar graph shows the number of members of a square dance club reporting their favorite season. Construct and label a circle graph to display this information.



5B. Six balls are in an urn, one each colored red, yellow, blue, green, white and purple. A four-sided (tetrahedral) die is marked with the numerals 1, 2, 3, 4. The die is tossed and one ball is selected at random.

What is the sample space for this experiment?

What is the probability of getting an even number and a color whose name ends in E?

What is the probability of getting an even number or a color whose name ends in E?

<u>Part C</u>: Short essay. [5 points each]

1C. What is the Law of Large Numbers? Explain this concept using at least one example.

2C. Write 3-4 sentences connecting some of the concepts you've learned in this course with each other. How do some of the concepts you've studied relate to your world outside the classroom?

Test 2 Items, Spring 1999

Part A. Choose the best answer from among the choices provided. [4 points each]

A1. _____ Which of the following values of p corresponds to a binomial distribution which is skewed right when n = 9?

A) $p = .1$ B) $p = .5$ C) $p = .75$ D) $p = .95$ E) none of these

A2. _____ Approximately 68% of the area under a normal curve lies between

A) $\mu - 2\sigma$ and $\mu + 2\sigma$ B) $\mu - 2\sigma$ and $\mu + \sigma$

C) $\mu - \sigma$ and $\mu + \sigma$ D) $\mu - \sigma$ and $\mu + 2\sigma$

A3. _____ If the following table represents a probability distribution, determine the value of **a**.

x	2	3	4	5	6
P(x)	.25	.05	.3	a	.4

A) a = 0 B) a = .1 C) a = .15

D) a = .2 E) impossible to determine

A4. _____ Which of the following is **not** characteristic of a normal distribution?

A) symmetric
B) bell shaped curve
C) the mean is equal to the median
D) the total area under the curve is 1
E) none of the above

A5. _____ Which of the following is **not** a characteristic of all binomial experiments?

A) fixed number of trials
B) independent trials
C) P(success) = P(failure)
D) none of the above

Part B. Answer each question completely, showing sufficient work to justify full credit.

B1. Debbie works at a blood bank. She knows from experience that 25% of the people with appointments don't show up. She has 12 appointments scheduled for Thursday.

a) Determine the probability that exactly 2 of the people will not show up. *[5 points]*

b) Determine the probability that more than 4 of the people will not show up. *[7 points]*

B2. The weight of a certain breed of dog is normally distributed with a mean of 35 pounds and a standard deviation of 4.2 pounds.

a) What is the probability that a randomly selected dog of this breed would weigh between 30 and 35 pounds? *[5 points]*

b) What is the probability that a randomly selected dog of this breed would weigh more than 40.5 pounds? *[7 points]*

c) What is the probability that a randomly selected dog of this breed would weigh between 32 and 36 pounds? *[7 points]*

B3. The following paragraph describes a situation that does not meet the conditions of a binomial experiment. Why does it fail to be binomial experiment? *[8 points]*

A local bakery advertises three award winning cheesecake variations. The baker knows from taste tests that 45% of his customers prefer Strawberry Surprise, 40% of his customers prefer Double Chocolate and 15% of his customers prefer Pumpkin Swirl. If ten customers were asked their preference, what is the probability that they'd all recommend the same variety?

B4. The local Buffalo club has a scholarship contest each year. They administer a test and award scholarships to the top 5% of test-takers. This year, the test scores were normally distributed with a mean of 38 points and a standard deviation of 3 points. What is the minimum score required to recieve a scholarship? *[5 points]*

B5. A major medical plan has data showing that 85% of its enrollees participate in the optional dental coverage. The plan has enrolled 180 new members this month.

a) Determine and interpret the mean of this probability distribution. *[8 points]*

b) Determine and interpret the standard deviation of this probability distribution. *[8 pts]*

Part C. Answer two of the following three questions. ***[10 points each]***

C1. What is a random variable? What is the difference between a discrete random variable and a continuous random variable? Give one example of each.

C2. What is the purpose of standardizing normal random variables? How is the standardizing accomplished?

C3. Compare and contrast the characteristics of probability histograms for discrete probability distributions with the area under the curve for continuous probability distributions.

Test 3 Items, Spring 1999

Part A. Mark each statement TRUE or FALSE. (4 points each)

A1. _________ As the confidence level decreases, the width of the confidence interval also decreases.

A2. _________ The use of sample data to make a claim about the population is called "statistical inference".

A3. _________A p-value of .04 would cause a researcher to reject the null hypothesis at the $\alpha = .05$ level of significance.

A4. _________ If $.45 < p < .53$ represents a 99% confidence interval the population proportion, then we can say that the true population proportion will be between .45 and .53 about 99% of the time.

A5. _________ One result of the Central Limit Theorem is that the standard deviation of the distribution of sample means, $\bar{x}$, is smaller that the standard deviation of the original distribution of x.

PART B. (15 points each)

Answer each question, using the procedure(s) from class. For hypothesis tests, include the null and alternative hypotheses, the level of significance, a sketch of the rejection region and the corresponding critical value, a decision and a verbal interpretation of that decision using the context of the problem. For confidence intervals, write the formula for the appropriate confidence interval first, then substitute all appropriate values and give a numerical answer as well as a verbal interpretation of the interval using the context of the problem.

B1. Yourtown hosts a holiday light display each December. They took a random sample of 45 vehicles this past year and determined that the sample mean wait time was 12 minutes and the sample standard deviation was 3 minutes. Find a 99% confidence interval for the true population mean waiting time.

B2. A pediatrician knows from years of experience that the mean time required to cure an ear infection is $\mu = 9$ days with a standard deviation $\sigma = 2.4$ days. If a random sample of 42 children with ear infections is selected, what is the probability that the average time needed for the infections to clear is greater than 9.5 days?

B3. In 1990, 85% of dogs in a western state had been inoculated against rabies. This year, a random sample of 1300 dogs in that state indicated that 1085 were inoculated against rabies. Does this data indicate that the actual proportion of dogs inoculated against rabies in this state has decreased? Test at the 5% level of significance.

B4. The local school board keeps records of the amount and type of playground equipment at elementary schools in their district. Traditionally, playgrounds in this area have had an average of 11 pieces of equipment. A random sample of 19 elementary schools in this district this year has shown a sample mean of 9 pieces of equipment, with a sample standard deviation of 3.1 pieces. Does this indicate that the mean number of pieces of playground equipment at elementary schools in this district has changed? Test at the .05 level of significance.

PART C: Answer 2 of the following 3 questions, concisely but completely. (10 points each)

C1. What is the Central Limit Theorem? Why is it so important to the field of inferential statistics?
C2. Compare and contrast point estimates with interval estimates. Which is generally preferred? Why?
C3. Under what conditions can the normal distribution approximate the binomial distribution? Explain.

APPENDIX D
Final Examination Problems and Essays

Final Examination Problems

1. A test was conducted to determine if cholesterol level and geographic region of residence were independent. The three categories of cholesterol level were high, borderline, and low. The five geographic regions were northeast, southeast, central, northwest, and southwest. The test was conducted at the .05 level of significance and the calculated value of the test statistic was 13.76.

a) What are the null and alternative hypotheses for this test?

b) What is the formula for the appropriate test statistic?

c) Sketch a graph of this distribution, including the rejection region and the critical value.

d) What decision is made based on the calculated value of the test statistic? What can you say about the independence of cholesterol level and geographic location of residence?

2. Nine women reported the average number of hours they exercise per week and their weight as follows:

hours of exercise	6	0	1.5	5	4	1	2	5	6
weight	134	166	144	132	139	160	148	120	109

a) Draw a scatterplot for this data. Label your axes appropriately.

A researcher used MINITAB to perform a regression analysis. Part of the results are printed below.

MTB > Regress c2 1 c1.

The regression equation is
C2 = 163 - 7.09 C1

Predictor	Coef	Stdev	t-ratio	p
Constant	163.130	5.281	30.89	0.000
C1	-7.088	1.315	-5.39	0.000

s = 8.508 R-sq = 80.6% R-sq(adj) = 77.8%

b) What is the correlation coefficient for this data? What does it mean?

c) What is the coefficient of determination? What does it mean?

d) Use $\alpha = .05$ to test the claim that there is a nonzero correlation between weekly hours of exercise and weight.

3. An exam for a prestigious scholarship is prepared so that only 15% of all high school seniors qualify.

a) If 12 high school seniors take the exam, what is the probability that 3 or more of them will qualify for the scholarship?

b) If 9 high school seniors take the exam, what is the probability that exactly 7 of them will NOT qualify for the scholarship?

4. A random sample of 12 fast food employees in Millertown showed they earned an average of \$6.50 per hour, with a standard deviation of \$.50 per hour.

a) Construct a 95% confidence interval for the true mean hourly wage of all fast-food employees in this town.

b) Write a sentence explaining the meaning of this confidence interval.

5. Researchers at a leading car manufacturer are testing the latest model of a popular sedan. The previous model was known to drive an average of 29 miles per gallon of gasoline, with a standard deviation of 3.2 miles per gallon. A random sample of 37 new cars of this type drove an average of 30.3 miles per gallon under similar conditions as the previous year. Does this indicate that the new model gets better gas mileage than the old model? Perform a test of hypothesis at the .01 level of significance.

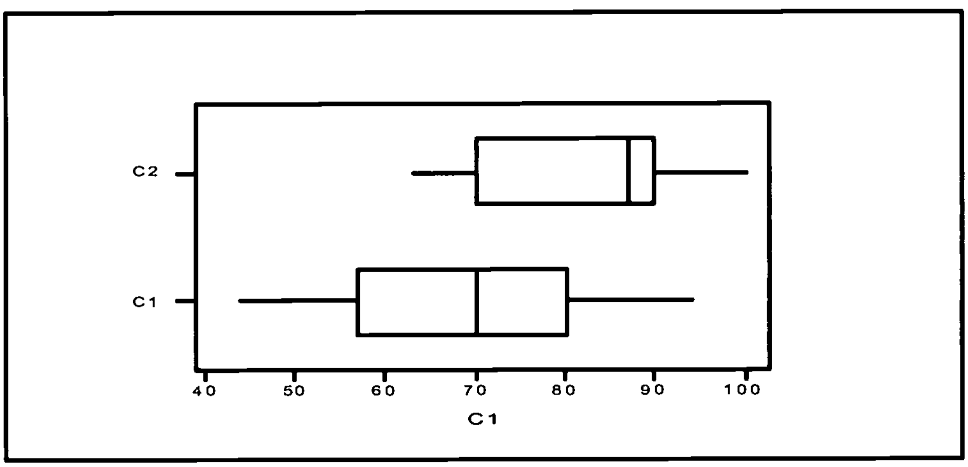
a) Give the null and alternative hypotheses for this test.

b) Sketch a graph of the rejection region and label the critical value.

c) Calculate the test statistic.

d) Make a decision and give a conclusion based on the context of the problem.

6. The box-and-whisker plots represent grades for two classes on a Statistics exam. Use this information to answer the questions below.



a) Which class did better on the exam? Why?

b) What does the grade of 70 represent for each class?

c) Estimate the median grade for Class 2. What does this value represent?

d) Compare the "box" part of each plot. What similarities and differences do you see? What does this tell you about the grades in each case?

Final Examination Essay Questions

1. Discuss in your own words the meaning of "statistics" (include both descriptive and inferential statistics). How do you see statistics used in your everyday life? How do you envision statistics being used in your future career?

2. A friend is registering for statistics next semester and saw "hypothesis tests" listed in the catalog description. Write a response to her question, "What are hypothesis tests?" Include the following ideas: null and alternative hypotheses, α, one-tailed and two-tailed tests, decisions, and conclusions.

3. Describe as completely as possible the characteristics and importance of the normal distribution. How does the normal distribution compare with the *Student's t* distribution?

4. Describe the least-squares regression line. Mathematically, what does "least-squares" mean? Why do researchers use regression analysis? How does the correlation coefficient for a set of data relate to the least-squares regression line?

5. The accompanying news clipping contains two statements regarding margin of error. What is margin of error? Use an example from this clipping to illustrate your ideas. How is margin of error connected to confidence intervals? Why were two different margin of error statements included by the writer?

USA TODAY · MONDAY, NOVEMBER 2, 1998 · 3A

Slight shift before elections

The last full week before the elections Tuesday showed some slight shift of public opinion in favor of Democrats, although the change was within the margin of error.

If the elections for Congress were being held today, which party's candidate would you vote for in your congressional district?

	Likely voters
Democratic candidate	49%
Republican candidate	45%
Undecided/Other	6%

USA TODAY CNN GALLUP POLL

Compared with previous elections, are you more enthusiastic about voting than usual or less enthusiastic?

All adults	Republican/Lean Republican	Democrat/Lean Democratic
More	39%	39%
Less	46%	44%
Same	14%	15%
No opinion	1%	2%

Do you think Congress should or should not impeach Bill Clinton and remove him from office?

	All adults	Likely voters
Yes, should impeach	27%	29%
No, should not	67%	66%
No opinion	6%	5%

Source: All adults, USA Today/CNN/Gallup Poll of 2,084 adults Oct. 29 to Nov. 1. Margin of error: +/-3 percentage points.
Likely voters, USA Today/CNN/Gallup Poll of 1,105 adults Oct. 29 to Nov. 1. Margin of error: +/-5 percentage points.

APPENDIX E
Statistical Reasoning Assessment

Statistical Reasoning Assessment Items

1. A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

The students want to determine as accurately as they can the actual weight of this object. Of the following methods, which would you recommend they use?

_____ a. Use the most common number, which is 6.2.

_____ b. Use 6.15 since it is the most accurate weighing.

_____ c. Add up the 9 numbers and divide by 9.

_____ d. Throw out the 15.3, add up the other 8 numbers and divide by 8.

2. 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?

_____ a. Don't use the medication on your skin – there is a good chance of developing a rash.

_____ b. For application to the skin, apply only 15% of the recommended dose.

_____ c. If a rash develops, it will probably involve only 15% of the skin.

_____ d. About 15 of 100 people who use this medication develop a rash.

_____ e. There is hardly a chance of getting a rash using this medication.

3. The Springfield Meteorological Center wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster has reported a 70% chance of rain. They compared these forecasts to the records of whether or not it actually rained on those particular days.

The forecast of 70% chance of rain can be considered <u>very</u> accurate if it rained on:

_____ a. 95%-100% of those days

_____ b. 85%-94% of those days

_____ c. 75%-84% of those days

_____ d. 65%-74% of those days

_____ e. 55%-64% of those days

4. A teacher wants to change the seating arrangement in her class in the hope that it will increase the number of comments her students make. She first decides to see how many comments students make with the current seating arrangement. A record of the number of comments made by her 8 students during one class period is shown below.

Student Initials	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Number of Comments	0	5	2	22	3	2	1	2

She wants to summarize this data by computing the typical number of comments made that day. Of the following methods, which would you recommend she use?

_____ a. Use the most common number, which is 2.

_____ b. Add up the 8 numbers and divide by 8.

_____ c. Throw out the 22, add up the other 7 numbers and divide by 7.

_____ d. Throw out the 0, add up the other 7 numbers and divide by 7.

5. A new medication is being tested to determine its effectiveness in the treatment of eczema, an inflammatory condition of the skin. Thirty patients with eczema were selected to participate in the study. The patients were randomly divided into two groups. Twenty patients in an experimental group received the medication, while ten patients in a control group received no medication. The results after two months are shown below.

Experimental Group (Medication)		**Control Group (No Medication)**	
Improved	8	Improved	2
No Improvement	12	No Improvement	8

Based on this data, I think the medication was:

_____ 1. somewhat effective

_____ 2. basically ineffective

If you chose option 1, select the one explanation below that best describes your reasoning.

_____ a. 40% of the people (8/20) in the experimental group improved.

_____ b. 8 people improved in the experimental group while only 2 people improved in the control group.

_____ c. In the experimental group, the number of people who improved is only 4 less than the number who didn't improve (12-8), while in the control group the difference is 6 (8-2).

_____ d. 40% of the patients in the experimental group improved (8/20) while only 20% improved in the control group (2/10).

If you chose option 2, select the one explanation below that best describes your reasoning.

_____ a. In the control group, 2 people improved even without the medication.

_____ b. In the experimental group more people didn't get better than did (12 vs. 8).

_____ c. The difference between the numbers who improved and didn't is about the same in each group (4 vs. 6).

_____ d. In the experimental group, only 40% of the patients improved (8/20).

6. Listed below are several possible reasons one might question the results of the experiment described above. Place a check by <u>every</u> reason you agree with.

_____ a. It's not legitimate to compare the two groups because there are different numbers of patients in each group.

_____ b. The sample of 30 is too small to permit drawing conclusions.

_____ c. The patients should not have been randomly put into groups, because the most severe cases may have just by chance ended up in one of the groups.

_____ d. I'm not given enough information about how doctors decided whether or not patients improved. Doctors may have been biased in their judgments.

_____ e. I don't agree with any of these statements.

7. A marketing research company was asked to determine how much money teenagers (ages 13-19) spend on recorded music (cassette tapes, CDs, and records). The company randomly selected 80 malls located around the country. A field researcher stood in a central location in the mall and asked passers-by who appeared to be the appropriate age to fill out a questionnaire. A total of 2,050 questionnaires were completed by teenagers. On the basis of this survey, the research company reported that the average teenager in this country spends $155 each year on recorded music.

Listed below are several statements concerning this survey. Place a check by every statement that you agree with.

_____ a. The average is based on teenagers' estimates of what they spend and therefore could be quite different from what teenagers actually spend.

_____ b. They should have done the survey at more than 80 malls if they wanted an average based on teenagers throughout the country.

_____ c. The sample of 2,050 teenagers is too small to permit drawing conclusions about the entire country.

_____ d. They should have asked teenagers coming out of music stores.

_____ e. The average could be a poor estimate of the spending of all teenagers given that teenagers were not randomly chosen to fill out the questionnaire.

_____ f. The average could be a poor estimate of the spending of all teenagers given that that only teenagers in malls were sampled.

_____ g. Calculating an average in this case is inappropriate since there is a lot of variation in how much teenagers spend.

_____ h. I don't agree with any of these statements.

8. Two containers, labeled A and B are filled with red and blue marbles in the following quantities:

Conatiner	Red	Blue
A	6	4
B	60	40

Each container is shaken vigorously. After choosing one of the containers, you will reach in, and without looking, draw out a marble. If the marble is blue, you win $50. Which container gives you the best chance of drawing a blue marble?

_____ a. Container A (with 6 red and 4 blue)

_____ b. Container B (with 60 red and 40 blue)

_____ c. Equal chances from each container

9. Which of the following sequences is most likely to result from flipping a fair coin 5 times?

_____ a. H H H T T

_____ b. T H H T H

_____ c. T H T T T

_____ d. H T H T H

_____ e. All four sequences are equally likely

10. Select one or more explanations for the answer you gave for the item above.

_____ a. Since the coin is fair, you ought to get roughly equal numbers of heads and tails.

_____ b. Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.

_____ c. Any of the sequences could occur.

_____ d. If you repeatedly flipped a coin five times, each of these sequences would occur about as often as any other sequence.

_____ e. If you get a couple of heads in a row, the probability of tails on the next flip increases.

_____ f. Every sequence of five flips has exactly the same probability of occurring.

11. Listed below are the same sequences of Hs and Ts that were listed in item 9. Which of the sequences is <u>least</u> likely to result from flipping a fair coin 5 times?

_____ a. H H H T T

_____ b. T H H T H

_____ c. T H T T T

_____ d. H T H T H

_____ e. All four sequences are equally likely

12. The Caldwells want to buy a new car, and they have narrowed their choices to a Buick or an Oldsmobile. They first consulted an issue of Consumer Reports, which compared rates of repairs for various cars. Records of repairs done on 400 cars of each type showed somewhat fewer mechanical problems with the Buick than with the Oldsmobile.

The Caldwells then talked to three friends, two Oldsmobile owners and one former Buick owner. Both Oldsmobile owners reported having a few mechanical problems, but nothing major. The Buick owner, however, exploded when asked how he liked his car:

> First, the fuel injection went out - $250 bucks. Next, I started having trouble with the rear end and had to replace it. I finally decided to sell it after the transmission went. I'd never buy another Buick.

The Caldwells want to buy the car that is less likely to require major repair work. Given what they currently know, which car would you recommend that they buy?

_____ a. I would recommend that they buy the Oldsmobile, primarily because of all the trouble their friend had with his Buick. Since they haven't heard similar horror stories about the Oldsmobile, they should go with it.

_____ b. I would recommend that they buy the Buick in spite of their friend's bad experience. This is just one case, while the information reported in Consumer Reports is based on many cases. And according to that data, the Buick is somewhat less likely to require repairs.

_____ c. I would tell them that it didn't matter which car they bought. Even though one of the models might be more likely than the other to require repairs, they could still, just by chance, get stuck with a particular car that would need a lot of repairs. They may as well toss a coin to decide.

13. Five faces of a fair die are painted black, and one face is painted white. The die is rolled six times. Which of the following is more likely?

_____ a. Black side up on five of the rolls; white side up on the other roll

_____ b. Black side up on all six rolls

_____ c. **a** and **b** are equally likely

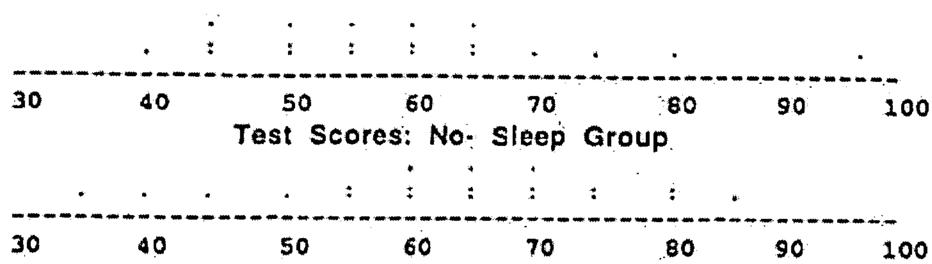
14. Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births per day. On a particular day, which hospital is more likely to record 80% or more female births?

_____ a. Hospital A (with 50 births a day)

_____ b. Hospital B (with 10 births per day)

_____ c. The two hospitals are equally likely to record such an event.

15. Forty college students participated in a study of the effect of sleep on test scores. Twenty of the students volunteered to stay up all night studying the night before the test (no-sleep group). The other 20 students (the control group) went to bed by 11:00 PM on the evening before the test. The test scores for each group are shown in the graphs below. Each dot on the graph represents a particular student's score. For example, the two dots above the 80 in the bottom graph indicate that two students in the sleep group scored 80 on the test.



Test Scores: Sleep Group

Examine the two graphs carefully. Then choose from the 6 possible conclusions listed below the one you <u>most</u> agree with.

_____ a. The no-sleep group did better because none of these students scored below 40 and the highest score achieved was by a student in this group.

_____ b. The no-sleep group did better because its average appears to be a little higher than the average of the sleep group.

_____ c. There is no difference between the two groups because there is considerable overlap in the scores of the two groups.

_____ d. There is no difference between the two groups because the difference between their averages is small compared to the amount of variation in the scores.

_____ e. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

_____ f. The sleep group did better because its average appears to be a little higher than the average of the no-sleep group.

16. For one month, 500 elementary students kept a daily record of the hours they spent watching television. The average number of hours per week spent watching television was 28. The researchers conducting the study also obtained report cards for each of the students. They found that the students who did well in school spent less time watching television than those students who did poorly.

Listed below are several possible statements concerning the results of this research. Place a check by every statement that you agree with.

_____ a. The sample of 500 is too small to permit drawing conclusions.

_____ b. If a student decreased the amount of time spent watching television, his or her performance in school would improve.

_____ c. Even though students who did well watched less television, this doesn't necessarily mean that watching television hurts school performance.

_____ d. One month is not a long enough period of time to estimate how many hours the students really spend watching television.

_____ e. The research demonstrates that watching television causes poorer performance in school.

_____ f. I don't agree with any of these statements.

17. The school committee of a small town wanted to determine the average number of children per household in their town. They divided the total number of children in the town by 50, the total number of households. Which of the following <u>must</u> be true if the average children per household is 2.2?

_____ a. Half the households in the town have more than 2 children.

_____ b. More households in the town have 3 children than have 2 children.

_____ c. There are a total of 110 children in the town.

_____ d. There are 2.2 children in the town for every adult.

_____ e. The most common number of children in a household is 2.

_____ f. None of the above.

18. When two dice are simultaneously thrown, it is possible that one of the following two results occurs:

Result 1: A 5 and a 6 are obtained.
Result 2: A 5 is obtained twice.

Select the response you agree with the most:

_____ a. The chance of obtaining each of these results is equal.

_____ b. There is more chance of obtaining result 1.

_____ c. There is more chance of obtaining result 2.

_____ d. It is impossible to give an answer. (Please explain why.)

19. When three dice are simultaneously thrown, which of the following results is MOST LIKELY to be obtained?

_____ a. Result 1: a 5, a 3, and a 6

_____ b. Result 2: a 5 three times

_____ c. Result 3: a five twice and a three

_____ d. All three results are equally likely

20. When three dice are simultaneously thrown, which of the following results is LEAST LIKELY to be obtained?

_____ a. Result 1: a 5, a 3, and a 6

_____ b. Result 2: a 5 three times

_____ c. Result 3: a five twice and a three

_____ d. All three results are equally unlikely

APPENDIX F
Attitude Inventories

Survey of Attitudes Toward Statistics (SATS)

DIRECTIONS: The questions below are designed to identify your attitudes about statistics. The item scale has 7 possible responses ranging from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). Please read each question. From the 7 point scale, carefully mark the one response that most clearly represents your agreement with that statement. Use the entire 7 point scale to indicate your degree of agreement or disagreement with the items. Try not to think too deeply about each response. Record your answer and move quickly to the next item.

	strongly disagree		neither agree nor disagree			strongly agree	
1. I like statistics.	1	2	3	4	5	6	7
2. I feel insecure when I have to do statistics problems.	1	2	3	4	5	6	7
3. I have trouble understanding statistics because of how I think.	1	2	3	4	5	6	7
4. Statistics formulas are easy to understand.	1	2	3	4	5	6	7
5. Statistics is worthless.	1	2	3	4	5	6	7
6. Statistics is a complicated subject.	1	2	3	4	5	6	7
7. Statistics should be a required part of my professional training.	1	2	3	4	5	6	7
8. Statistical skills will make me more employable.	1	2	3	4	5	6	7
9. I have no idea what's going on in statistics.	1	2	3	4	5	6	7
10. Statistics is not useful to the typical professional.	1	2	3	4	5	6	7
11. I get frustrated going over statistics tests in class.	1	2	3	4	5	6	7
12. Statistical thinking is not applicable in my life outside my job.	1	2	3	4	5	6	7
13. I use statistics in my everyday life.	1	2	3	4	5	6	7
14. I am under stress during statistics class.	1	2	3	4	5	6	7
15. I enjoy taking statistics courses.	1	2	3	4	5	6	7
16. Statistics conclusions are rarely presented in everyday life.	1	2	3	4	5	6	7
17. Statistics is a subject quickly learned by most people.	1	2	3	4	5	6	7
18. Learning statistics requires a great deal of discipline.	1	2	3	4	5	6	7
19. I will have no application for statistics in my profession.	1	2	3	4	5	6	7
20. I make a lot of math errors in statistics.	1	2	3	4	5	6	7
21. I am scared by statistics.	1	2	3	4	5	6	7
22. Statistics involves massive computations.	1	2	3	4	5	6	7
23. I can learn statistics.	1	2	3	4	5	6	7
24. I understand statistics equations.	1	2	3	4	5	6	7
25. Statistics is irrelevant in my life.	1	2	3	4	5	6	7
26. Statistics is highly technical.	1	2	3	4	5	6	7
27. I find it difficult to understand statistics concepts.	1	2	3	4	5	6	7
28. Most people have to learn a new way of thinking to do statistics.	1	2	3	4	5	6	7

SCAS Instrument

Scale:	1 Strongly Disagree	2 Disagree	3 Neither Agree, nor Disagree	4 Agree	5 Strongly Agree

1.	I often use statistical information in forming my opinions or making decisions.	1	2	3	4	5
2.	To be an intelligent consumer, it is necessary to know something about statistics.	1	2	3	4	5
3.	Because it is easy to lie with statistics, I don't trust them at all.	1	2	3	4	5
4.	Understanding probability and statistics is becoming increasingly important in our society, and may become as essential as being able to add and subtract.	1	2	3	4	5
5.	Given the chance, I would like to learn more about probability and statistics.	1	2	3	4	5
6.	You must be good at mathematics to understand basic statistical concepts.	1	2	3	4	5
7.	When buying a new car, asking a few friends about problems they have had with their cars is preferable to consulting an owner satisfaction survey in a consumer magazine.	1	2	3	4	5
8.	Statements about probability (such as the odds of winning a lottery) seem very clear to me.	1	2	3	4	5
9.	I can understand almost all of the statistical terms that I encounter in newspapers or on television.	1	2	3	4	5
10.	I could easily explain how an opinion poll works.	1	2	3	4	5

APPENDIX G
Interview Guide

Interview Protocol

What were your overall thoughts about the course?

Were your expectations met? Explain.

Would you recommend the course to other students?

What do you think was the most interesting topic we discussed? Why?
(Ask the student to explain any statistical concepts involved in the response.)

What do you think was the most important topic we discussed? Why?
(Ask the student to explain any statistical concepts involved in the response.)

What do you think was the most challenging topic we discussed? Why?
(Ask the student to explain any statistical concepts involved in the response.)

What kinds of questions can statistics help us answer?

What is the distinction between descriptive statistics and inferential statistics?

Why do you think we emphasized random sampling so much during the course?

What are hypothesis tests?

(Significance Level? Type I and Type II error?)

("Reject H_o" / "Fail to Reject H_o"?)

How would you use statistics to determine whether there is a relationship between student attendance and grades? Does better attendance cause better grades?

What is a confidence interval?

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Data Collection and Analysis: Examining Community College Students' Understanding of Elementary Statistics through Laboratory Activities	
Author(s): Jane Ann Brandsma	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy. ✓	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Jane Ann Brandsma	Printed Name/Position/Title: Jane Ann Brandsma	
Organization/Address: Greensboro College 815 W Market Street, Greensboro NC 27401	Telephone: 336-272-7102	FAX:
	E-Mail Address: jbrandsma@gborocollege.edu	Date: 1/16/02